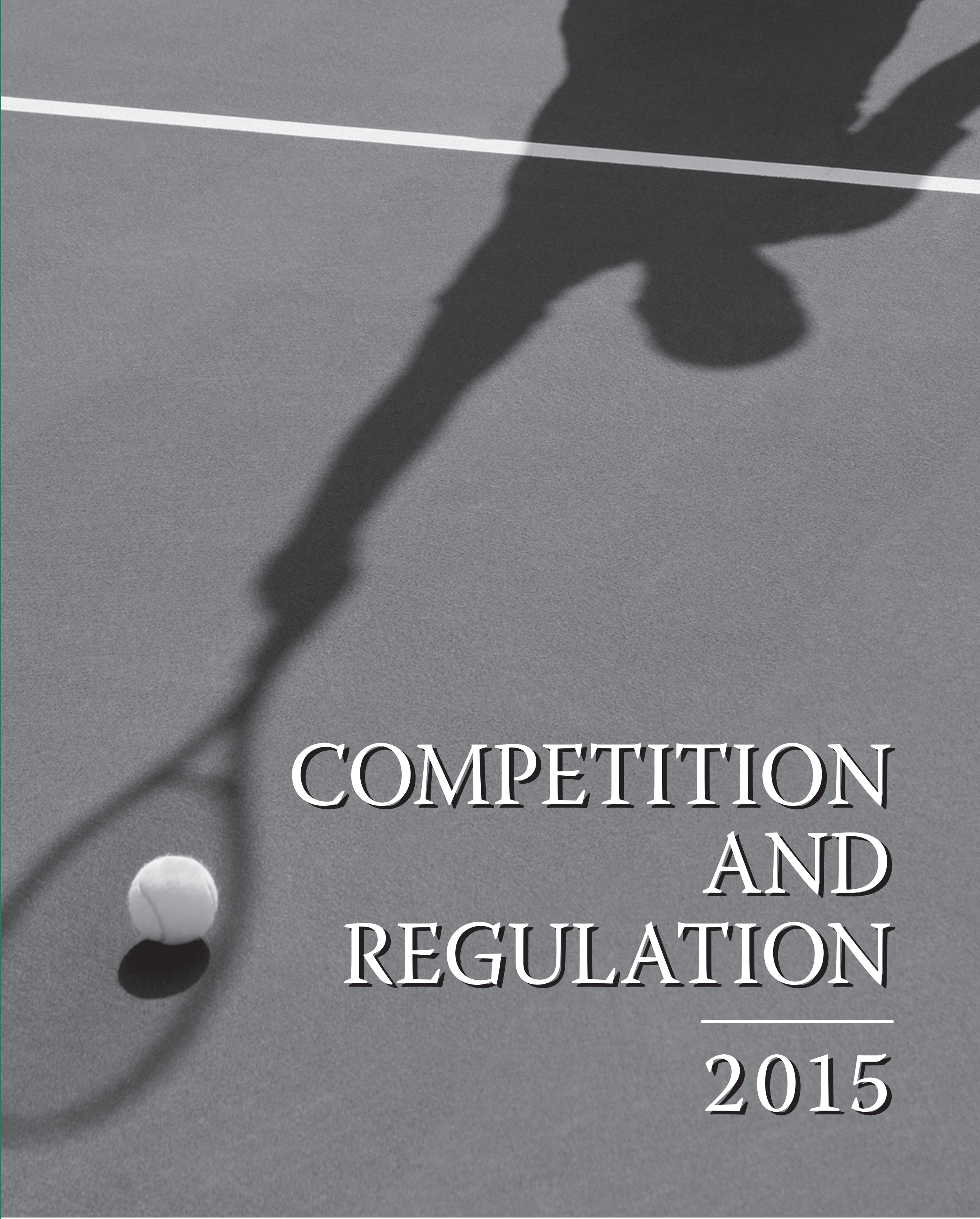


COMPETITION AND REGULATION

2015

INSTITUTE OF ECONOMICS • CERS • HUNGARIAN ACADEMY OF SCIENCES

COMPETITION AND REGULATION

• 2015 •

COMPETITION AND REGULATION

COMPETITION AND REGULATION

• 2015 •

Editors

PÁL VALENTINY • FERENC LÁSZLÓ KISS •
KRISZTINA ANTAL-POMÁZI • CSONGOR ISTVÁN NAGY

INSTITUTE OF ECONOMICS
CENTRE FOR ECONOMIC AND REGIONAL STUDIES,
HUNGARIAN ACADEMY OF SCIENCES
Budapest • 2015

The publication of this volume has been financially supported by
the *Hungarian Competition Authority*.

Editors
Pál Valentiny • Ferenc László Kiss • Krisztina Antal-Pomázi • Csongor István Nagy

ISSN 1789-9702

Published by:
Institute of Economics
Centre for Economic and Regional Studies, Hungarian Academy of Sciences
1112 Budapest, Budaörsi út 45.
Online available at http://econ.core.hu/kiadvany/vesz.html

Publisher: Károly Fazekas
Copy editor: Anna Patkós
Design, page layout: Anikó Környei
Cover image: Schiz-Art, iStock
Printing: Mondat Kft.
Managing director: László Nagy

CONTENTS

PREFACE

The Institute of Economic Sciences at the Hungarian Academy of Sciences launched a new series of publications entitled *"Verseny és szabályozás"* [Competition and Regulation] in 2007. Eight annual volumes have been published so far, all in Hungarian. The current volume is the first one in English, and it contains 12 selected translations from the crop of the first seven years. It offers the reader a glimpse into the current state of research in its chosen field in Hungary.

As the title suggests, the main objective of the publications was to open up a much-needed new forum for home-grown Hungarian research on the legal and economic issues of the regulation of imperfectly competitive markets. The published studies covered a very broad range of topics. Some articles of general theoretical and methodological nature dealt with the background in the law and economics of regulated markets. Others investigated current legal, economic and policy issues and cases. Others again dealt with regulation and the regulators themselves. The functions, methods, analytical tools, the institutions and the impact of regulation were discussed in those articles. Special attention was paid to regulation by the European Union, and also to recently de-monopolized key industries such as communications, energy, media, the postal sector or water and sewage. More than a half of the articles of the first seven publications dealt with the problems of key industries. The publications were designed to provide a meeting place for economists and lawyers to work together on the economic background of legal problems and the legal solutions to economic problems. They also had an educational function. In an introductory manner and by relying on timely surveys of recent developments in the analysis of imperfect markets and regulation, articles suitable for educational use have been regularly published.

Over the years, the series has become a major undertaking. Its 49 contributing researchers (40 economists and 9 lawyers) appeared in it a total of 100 times as authors or co-authors of 84 articles. Twenty of them became recurring contributors, authoring or co-authoring at least two articles each. A steadily growing interdisciplinary circle of dedicated researchers has formed around the publications. Interactions among the authors increased over time. Significant lawyer-economist cooperation demonstrated itself by the large number of contributing lawyers and articles about legal issues (9 lawyer authors produced 18 such articles), and by the emergence of articles co-authored by economists and lawyers.

Five of the 12 articles selected for publication in English in this tome deal with broad economic and legal issues of regulation, while the remaining 7 discuss the state and specific problems of key industries in Hungary and, in some cases, in the surrounding region.

The first article by *A. Kovács* examines the practice of judicial reviews of regulatory decisions, and finds it harmful if information which is pertinent to, or even crucially important for, the reviewed case is declared a commercial secret in the regulatory proceedings, and thus remains unavailable for the reviewer. The author recommends limitations to the right to protect commercial secrets. Reaching back to Law & Economics, it is suggested that protection may be appropriate only when it improves resource allocation in the economy. Otherwise commercial secrets should be made public to varying degrees, depending on their nature.

The second article by *B. Muraközy* and *P. Valentiny* takes a look at two of the various possible alternatives to straightforward government regulation: self-regulation and co-regulation (self+governmental). The authors explain how these regimes work, what variants may exist, how their history evolved, and how the Anglo-Saxon, French (Napoleonic), German, Scandinavian and the so-called socialist legal systems provide incentives and counterincentives for the development of self-regulation. An extensive survey of studies and models of self- and co-regulation is provided.

The third article by *G. Csorba* is a survey of empirical studies of Hungarian market structures. In this field of research, often referred to as Empirical Industrial Organisation, a fairly large number of studies have been produced in Hungary in recent years. The author identifies numerous generic uses in competition policy and regulation of the results of market structure studies. Among them market definitions, evaluations of market power, and the estimated consequences of various forms of market behaviour are emphasised as particularly important ones. An outline of the history of empirical market structure studies is followed by descriptions of the surveyed studies, which are organised into three groups according to the nature of the data they utilise. Studies of the first group use price data and estimate the relationship between market concentration and prices or the impact of structural and behavioural changes on prices. In the second group we find price and volume data on products and/or services sold in the target markets of the studies. The most basic findings of these studies are estimates of demand characteristics, particularly those of own price and cross price elasticities. Some variants of the classic demand function are mentioned. The third group contains models using elaborate structural information on the supply side of the market and estimating the properties of demand for factor inputs.

The fourth article by *L. Halpern* and *B. Muraközy* tackles one of the most popular topics: the relationship between competition and the research and development (R&D) activities of competing firms. Initially the authors draw attention to some problems, such as the erroneous measurement of R&D by its inputs rather than its outputs. Errors caused by not distinguishing between endogenous and exogenous R&D are also mentioned. Based on their empirical results, the authors accept the inverted U-shaped relationship between competition and R&D at the firm level as well as at the industry level, meaning that innovative activity is lower in firms that operate either in highly concentrated or highly competitive industries than in firms in moderately competitive sectors of the economy. Interpretations are provided. There is some inconsistency in the results. The presence and intensity of competition were measured by different metrics and only some of them seemed to have had an impact on the innovative investments of firms. The empirical results are accompanied by an account of difficulties in modelling the relationship. Simultaneity (*i.e.,* competition causing R&D and R&D causing competition at the same time) is singled out as an important problem.

The fifth article by *F.L. Kiss* revisits the important but lately somewhat ignored field of productivity analysis. The author argues that the socially responsible regulator must study and understand what makes productivity grow in the regulated suppliers of imperfect markets. Following a brief conceptual clarification and overview of the method of measurement, the author focuses his attention on two topics: 1. intra-firm and inter-firm comparisons of productivity levels and changes (gains), and 2. decomposition of firm-level productivity gains into components caused by increased production volumes, cost-saving technological changes, pricing and the characteristics of corporate demand for inputs. Having accomplished these analytical tasks, the author then combines the two and causally decomposes the compared measures of firm-level productivity. An empirical study rounds out the presentation. Productivity comparisons and decompositions as well as an attempt at econometric forecasting of factor inputs are carried out for two real-life regulated companies. The resulting deep insight into the efficiency of the production processes of the two firms is demonstrated.

The sixth article by *C.I. Nagy* examines the influence of the European Union on price regulation by its member states. Three areas are identified in which such influences may exist: competition law, liberalisation law, and internal market (free movement) law. A thorough investigation of EU jurisdiction and the relevant legal cases revealed no document that would deal with this issue. The only positive finding of the author is a declaration in one legal case, that price regulation by the state does not constitute subsidy by the state, because it does not involve subsidy by the state budget. Thus price regulation by member states is neither forbidden nor explicitly hindered by the EU. However, the author notes that limitations and restriction may exist in connection with the application of other principles and provisions. It is forbidden to hinder the free movement of goods and services within the EU by any means, including price regulation. There are cases mentioned in the article in which setting maximum or minimum prices or some upper boundary such as unit cost would violate internal market law. Regulated prices may also act as obstacles to market liberalisation, free access to markets. The natural gas and electricity markets are examined in some detail. Price regulation is also one of the possible means of artificially creating economic advantages and disadvantages among market players.

The seventh article by *I. Major* and *K.M. Kiss* deals with the regulation of access prices for firms with interconnected networks under imperfect and asymmetric information. Its point of departure is a long series of studies that address the issue of interconnection, access prices, and termination charges in particular, under the assumption that the regulator has perfect information about the true costs of providing inter-firm network access services. Interconnection is revisited under the assumption of imperfect regulatory information and informational asymmetry between the regulator and the regulated firms. Comparing incentive regulation (whereby the regulator offers the regulated firm an incentive-based contract menu) with regulation by cost-based pricing, it is concluded that under imperfect information cost-based pricing may give perverse incentives to regulated firms not to improve the efficiency of interconnection, and cost-based pricing of call termination ultimately rewards the less efficient types of regulated firms. In contrast, incentive regulation produces no perverse incentives and allows the efficient firm to earn higher profits. Various aspects of incentive regulation are discussed. It is concluded that incentive regulation works with smaller social welfare loss than cost-based pricing or bottom-up cost accounting. Principal-agent models of price regulation are more "knowledge intensive" but less time consuming than cost accounting. Most importantly, a regulatory mechanism that takes into account the existence of asymmetric information between the regulator and the regulated firm induces cooperation between the contracting parties, while cost-based pricing induces cost manipulations by the companies and inevitably brings about conflicts with the regulator.

The eighth article by *L. Paizs* is about the electricity balancing market. The term "balancing" refers to the correction of very-short-term market disequilibria that regularly result from various unforeseen fluctuations in the volume of demand for electricity. Deviations between demand and supply require efficient real-time corrections to the volume of energy supply. These are crucially important for network safety and the efficiency of the entire market. Liberalisation has brought with itself market-based institutional arrangements and processes of balancing. These and their applications to the Hungarian electricity industry are described in the article. Balancing takes place through positive and negative corrections to volumes, and involves prices and penalties. In the article it is investigated how the stakeholders' behaviour is affected by the properties of these prices and penalties. The main conclusions are: Suppliers have strong incentives to keep their portfolio balanced. Asymmetry in the penalties makes suppliers inclined to under-contract. The structure of purchase and settlement prices is such that it motivates the public utility wholesaler to nominate more than their expected load.

For years, the Regional Centre for Energy Policy Research at Corvinus University in Budapest has been modelling European regional electricity markets. The ninth article by *A. Kiss* presents a regional simulation model of wholesale electricity markets, comprising of Hungary and six neighbouring countries. It is emphasized during the presentation of the structure and workings of the model that this market is characterised by very high concentration. Few large electricity generators, having high degrees of market power, dominate the wholesale market and drive up prices. Mark-ups due to market power are estimated to have amounted to 25-40 percent of actual wholesale prices. An examination of congestions and shortages indicates that prices could be lowered by tighter market integration resulting in improved flows of electricity and reduced market power. Simulation results, however, show that market integration alone is not nearly sufficient to eliminate the dominance by large power stations, or to realize the potential welfare gains of competition.

The tenth article by *P. Kaderják et al.* presents the elaborate Danube Region Gas Market Model of the Regional Centre for Energy Policy Research. This model depicts the combined gas infrastructure of 15 countries of Central and Southern Europe in a consistent and unified manner. Following a detailed description of the model, simulations are carried out in order to determine various economic consequences of a series of important existing or planned or proposed gas infrastructure investments. The estimated effects include price effects (among them spill-overs to third countries), benefit-cost analyses, impacts on social welfare, improvements in European market integration and the security of gas supplies in the region.

As the use of mobile telephony started spreading beyond the wildest initial dreams of market analysts, worldwide attention became urgently focussed on the issue of the relationship (substitutability and/or complementarity) between fixed and mobile telephony. This relationship was of enormous importance for the future size and structure of telecommunications markets. Results from an empirical study were reported in the eleventh article by *B. Édes et al.* A Hungarian survey of the access and usage preferences of 1000 people was used to analyse consumers' responses to price changes and the effects of lifestyle and demographic characteristics on demand. The own-price elasticities implicit the responses were low (-0,3) for mobile access and high (-1,4) for fixed access, suggesting that mobile access was much less readily substitutable by fixed access than fixed access was by mobile access. However, results with respect to cross-elasticities were somewhat inconsistent and required deeper analysis and evaluation by the authors, whose ultimate conclusion was that in Hungary mobile access was a substitute for fixed access. The rate of usage substitution was significantly lower in the short run (*i.e.,* in the presence of existing subscriptions) than in longer-term access decisions.

The last study by *K.M. Kiss* is of monographic nature, as it introduces the reader to a single key market, the Hungarian postal sector. The article consists of three parts. The first one provides detailed descriptive information on the sector's recent history and current structure. The second part deals with the introduction of competition into, and regulation of, the postal sector. Market and regulatory developments are placed into the broader environment of the European postal reforms. The author mentions some negative phenomena, discusses important legal-regulatory cases of anti-competitive behaviour, and also points to areas in which some progress has been made in Hungary. The third part completes the description by introducing performance indicators and using them to measure the sector's performance.

The editors

• *András György Kovács* •

ECONOMIC ANALYSIS OF THE RIGHT TO PROTECT COMMERCIAL SECRETS

The fundamental statement of this article is that in the regulated markets, the basic hindrance to the effective judicial review of the regulator's decisions derives from the fact that the factual basis for the regulator's decision is usually considered to be a commercial secret. For this reason, a new relationship must be established between the right to the protection of commercial secrets and regulatory intervention, by placing limitations on the right to the protection of commercial secrets. This article uses the accepted apparatus of law and economics to assess whether the goals of the right to the protection of commercial secrets are acceptable. It aims at finding an answer to the questions when the protection of commercial secrets enhances the proper allocation of resources and when it does not. The article puts a special emphasis on the economic effects of the information which constitutes a commercial secret of future behaviour. The conclusion of the article is that the legislator should consider making the commercial secrets used in regulatory procedures partially or fully public.

INTRODUCTION

The legal institution of commercial secrets is an inherent element of a market economy. We consider it natural that the constitution – along with the freedom to conduct business and the freedom of competition – protects the privacy rights of businesses, thus we presume that the protection of the secrets of businesses constitute a fundamental right. Consequently, commercial secrets – if not formally, but substantially – are constitutional rights.[1] Nonetheless, it is far from obvious to associate legal persons and corporations with Article VI Paragraph 1 and 2 of the Fundamental Law of Hungary, containing the provisions on the right to private and family life, home, communications and good reputation, and the right to the protection of personal data. The origin of these rights is to be found in the relationship between the state and its citizens, derived from the protection of the separated private sphere, and its extension only seems necessary in a world ruled by modern market economy institutions (businesses and mostly legal persons). This is a global phenomenon and the "global law" can be traced back to the interpretation of the

[1] The practice of the Constitutional Court is mostly related to the constitutional questions of special types of secrets, but many decisions clearly prove this statement. See for example the 24/1998. (VI. 9) and 61/B/2005. decisions of the Constitutional Court.

Hungarian constitution as well.[2] For instance, the European Court of Human Rights (ECtHR) held that Article 8 of the European Convention on Human Rights[3] about the right to respect for private and family life also applies to business entities, thus the private and family life of businesses is also entitled to protection in democratic, market economy, rule of law states.[4]

This extended interpretation, however, has its own price. On the one hand, businesses have to precisely keep track of all the data related to their economic activities (rules of accounting), and make them at least partially public (business registry, balance sheet). The reason for this is that the state has to be aware of all the relevant economic data of a company (mainly for tax purposes), moreover, some information from the balance sheet and the annual report has to be disclosed to competitors and other market participants (e.g. creditors) in order to ensure safe business relations and safe transactions.

There are some business relations (for example between a bank and its client) that make it necessary to disclose private data and commercial secrets. Due to these special business relations and the state's need for information, specialized sectoral secrets (tax-secret, bank-secret, insurance-secret, etc.) have been separated from the legal institution of commercial secrets. Beyond the state's want for information, the requirement of transparency in government functions also calls for limitations on the right to protect commercial secrets. In a modern market economy, where the state is the largest investor, the transparency of the functioning of the state is not only a question of democracy (and the possible violation of democratic principles), but also a concern for competition policy.[5] The economic relations between a business and the state are generally seen from the perspective of the business as a commercial secret, however, the state views data as being of public interest. Commercial secrets enable not market-oriented, irrational state decisions – mainly through the dangers of corruption – which deteriorates market economy efficiency and distorts competition.

---

[2] The definition of commercial secrets as defined in the civil code (Act V of 2013) is based on the fundamental international norm, the TRIPS Agreement (*Trade Related Aspects of Intellectual Property Rights)* Article 39. 1. and 2. (Act IX of 1998. 1. *C)* supplement). (See: *Bobrovszky* [2006] p. 1385, which refers to the fundamental international agreement of the protection of industrial property, the 1883 Paris Convention; *Nagy* [2008] p. 555) The TRIPS Agreement is a cornerstone of the world trade system established in Marrakesh (together with the GATT and GATS), which was published in Hungary by Act IX of 1998. Section 2 of Article 39 of the Agreement defines commercial secrets by referring to Article 10bis of the Paris Convention, which latter was published in Hungary by the 18th Regulation with the force of a statute in 1970.

[3] Act XXXI of 1993 on the publication of the Convention for the Protection of Human Rights and Fundamental Freedoms of Rome 4.XI.1950 and of the supplementing eight Protocols.

[4] Case of Societé Colas Est and others *versus* France (Application no. 37971/97). European Court of Human Right, Strasbourg, 16 April 2002.

[5] The regulation of public procurement – as part of the Community's competition policy – is based on community directives. (See: the Preamble of the 89/665/EGK Directive of the European Council of 21 December 1989. or Section 4 of the Preamble of the 2004/18/EK Directive of the European Parliament and Council of 31 March 2004., etc.)

This phenomenon was acknowledged by the former Civil Code of Hungary (Ptk.), when the commercial secret definition envisioned by the fundamental international norm, the TRIPS Agreement (Trade Related Aspects of Intellectual Property Rights), was incorporated and supplemented with additional provisions.[6] These made it clear that the business relations between the state and the private entities, in relation to public procurement, state (and EU) aid or other financial relations connected to the state budget, are public information, and could not be classified as a commercial secret. As a result of this trend, the Hungarian state could not have any commercial secrets from 22 May 2009, based on Paragraph 1 of the XXXI Act of 2009. However, the new Civil Code of Hungary that entered into force on 15 March 2014 has seemingly reversed this trend, diminishing the results of the early years of the 2000s in substantive law, by returning the commercial secret definition applicable before 2003, which does not contain the limitations on state functions.[7] So long as the Hungarian legal development did not abandoned the general standards,[8] commercial secrets have gradually become increasingly subordinated to pubic interest in the economic relations between state and business. This can be detected even today, when the right to the protection of commercial secrets is in conflict with other constitutional right. The fact that the protection of a commercial secret as a fundamental right can only be justified by an extended interpretation of the constitutional text which results in a situation, in which when it is in conflict with another constitutional right – especially with one of the fundamental norms defining our socio-economic system – the protection of commercial secrets will turn out to be the weaker and can be restricted. This is also the cost of an extended interpretation. For instance the fundamental right to legal remedies, laid down by Article XXVIII paragraph 7 of the Fundamental Law, is supposed to be a stronger right than the right to the protection of commercial secrets, thus courts – based on the necessity and proportionality test – must provide access to data, information or documents classified as commercial secrets. Also this is generally true for administrative procedures.[9]

These clashes lead to a number of practical problems. On the one hand, businesses often tend to classify their submissions in administrative and judicial proceedings

[6] Act IV of 1959 (old Civil Code Section 2 Paragraph 81). The restrictive provisions were added by the Paragraph 16 of the Act XXIV of 2003, which entered into force on 9 June 2003 (old Civil Code Section 3-4 Paragraph 81)

[7] Act V of 2013 (Civil Code) Section 1 Paragraph 2:47

[8] We will see later that even before the FIDESZ government with the two-third constitutional majority from 2010 and since, the situation was not clear, because the development of substantive law has gradually restricted the right to the protection of commercial secrets, but in the procedural legislation the lobby power of the opposite side has appeared as restrictions strengthening the protection of commercial secrets appeared, which made the protection of commercial secrets powerful even against constitutional rights. See the next footnote and the conclusion of the study.

[9] This was undoubtedly true before the amendment of Act III of 1952 on the Civil Procedure Code on 1 January 2009. We will discuss the current procedural rules at the end of the study.

as commercial secrets, however, when they are asked to specify which exact data they are referring to as a commercial secret, hesitation is prevalent.

A detailed – yet at this point failing – regulation would be needed for the access to commercial secrets in judicial proceedings, which requires a lot of administration in the course of judicial review. This is especially true in the cases of judicial review of regulatory authorities' decisions. A good example for this is the legality review of price regulation decisions related to dominant market position in the field of info-communications, where the determination of cost-based price is based on the use of fundamentally important commercial secrets. The resistance to making these data available is nicely shown by the fact that in the field of info-communication even the regulatory authority is refusing – contrary to the law – to publish the preparatory documents for its market regulatory decisions, apart from the draft decision. Nonetheless, so far this approach has not hindered the judicial review, since the administrative authority is forbidden from making such documents public that were classified as commercial secrets by the interested parties.[10] It is, however, also doubtful that the cost-calculation method used by the authority to assess an effective service [bottom-up long-run incremental cost model (BU-LRIC)] is published in such detailed fashion as it is required by the statute.[11]

Effective and substantial judicial review is, however, unimaginable without access to the most fundamental commercial secrets. [12] If for example the judicial proceeding is about whether the cost-model used to determine the cost-based price was appropriate, the plaintiff company affected by the price regulation is allowed to access the fundamentally important commercial secrets of other service providers, since without such access the appropriateness of the cost-model could not be assessed. This alone – without considering the outcome of the case – provides a competitive advantage to the plaintiff company, which could unfairly distort competition, as opposed to the regulatory objectives. [13]

[10] Point b) Section 1 Paragraph 36 of Act C of 2003 on electronic info-communication (Eht.). The interested parties classify basically all existing data as commercial secrets.

[11] See: Section 4 Paragraph 108 of the Eht. This conclusion is based on the experiences of the judicial proceedings of ex 16, then 7 markets. (voice transmission call termination wholesale service in specific mobil radio-telephone network wholesale markets).

[12] The regulation of electronic info-communications is based on community directives, and community law obliges nation courts to effectively enforce community law in the judicial cases. This is the *principle of effectiveness.* (*Steiner–Woods* [2000] p. 441–443.). Before the entry into force of the Lisbon Treating on 1 December 2009, this principle could be deducted from Article 10 (before Article 5) of the Treaty on the European Communities. After the Lisbon Treaty it is based on the second sentence of Section 2 Article 19 of the Treaty on the European Union. In the field of electronic info-communications, Section 1 Article 4 of the 2002/21/EK Directive of the European Parliament and Council of 7 March 2002 mandates the member states that effective judicial remedies against the decisions of the national regulatory authorities must be provided.

[13] The fundamental hypothesis of this article is that the existence of effective judicial remedies is an essential legal and economic-efficiency element of an effective regulatory regime. This is, however,

By providing a general analysis to the legal institution of commercial secrets, this article aims to show that there is a theoretical possibility to make fundamentally important commercial secrets public based on the regulatory interests. This question should be worth exploring from practical aspects as well, however, since the author is not an economist, it is outside the scope of the article. Thus in the general economic analysis of the right to the protection of commercial secrets, we will not be providing a detailed description to those questions that do not relate tightly to the issues mentioned above, even though they might be essential and important elements of the economic analysis of the right to the protection of commercial secrets. The protection of commercial secrets plays a crucial role in vertical relations, such as between employer and employee, the analysis of which is mostly needed for understanding the justification for the legal institution of commercial secrets. Due to reasons of space, however, we will only make some brief remarks in this regard.

Lastly, it is important to note that this analysis is building on the current Hungarian legal environment, thus the conclusions are adapted to the Hungarian situation, consequently, it describes a special case of the economic analysis of the right to the protection of commercial secrets, the generalisation of which might need some corrections.

THE LEGAL AND ECONOMIC CONSTRUCTION OF COMMERCIAL SECRETS

The legal definition of commercial secret

"Business secrets shall comprise all of the facts, information, conclusions or data pertaining to economic activities that, if published or released to or used by unauthorized persons, are likely to imperil the rightful financial, economic or market interest of the owner of such secrets – other than the State of Hungary –, provided the owner has taken all of the necessary steps to keep such information confidential." (Section 2 of Paragraph 81 of the former Civil Code).[14] This was the general definition of commercial secrets, applicable to all fields of law, based on the old Civil Code

not at all evident, so it is worth analysing. Similarly a further hypothesis of the study: a substantial judicial review includes the economic overview of the authority's discretionary power based on economic considerations, which could also be debated.
See: *Tóth* [2006], *Kovács* [2006], *Koppányi* [2006–2007] EU law, nonetheless, requires the effective judicial review. (See: previous footnote.)

[14] Section 1 of Paragraph 4 of the act on competition refers back to this definition. (Act LVII of 1996 on the prohibition of unfair trading practices and unfair competition, Competition Act) The Criminal Code uses the same definition for commercial secrets, see the reasoning for Paragraph 18 of Act XCI of 2005.

up until 15 March 2014.[15] Section 1 of Paragraph 2:47 of the new Civil Code contains the new definition, which from a functional perspective is not fundamentally different. Presumably, similar conclusions could be drawn from the new definition of commercial secrets as well, nonetheless, since it is in force only since 15 March 2014, no relevant case-law and practice has evolved around it. Thus we will use the old definition in the article to show what the general definition of commercial secrets could be, which could also be applicable in any legal system.[16] The definition shall be approached in three ways. The first is the subject of the commercial secret, the second is the relevant conduct that could result in an injury of interests, and finally is the required conduct of the person entitled to the secret (the formal element of the definition) to make the commercial secret concretely identifiable with an external interference.

The subject of a commercial secret is the information. The definition of information is, however, an exceedingly wide category.

- According to some the world is nothing else then matter, energy and information. Nevertheless, others think that it is a *fact* that *Sz. L.* is a member of *B.* law firm, while it is a *circumstance* that he has an armchair in the left corner of his office, and the way *Sz. L.* usually sits in this chair, his individual body position is some sort of a *solution*. And it is just a mere *data* that *Sz. L.* writes 15-page longer claims than the average length of others' claims. If we can acknowledge a connection (even if there is or is not) between these facts, circumstances, solutions and data that is an information. Given that it is due to *Sz. L.*'s individual way of sitting in front of the computer that he is able to stare at his monitor 20 % more each day than the others (which can be verified by the average of time spent by the other lawyers in front of the computer), and thus he is able to write 15-page longer claims, then *Sz. L.* can evidently give a competitive advantage to his employer. This is an important commercial secret, because if it was made public, then either others would copy his special way of sitting, or *Sz. L.* would have to be paid more in order for him to be able to refuse the different daily job offers.

[15] In the course of the analysis we will use this statutory definition, even though the current statutory definition of commercial secrets is partially different, and this definition could be analysed separately in each legal system *Nagy* [2008] (p. 554). For example five major theories may be distinguished for the justification of the regulation of commercial secrets. The Hungarian dogmatic approach is based on the personality and its protection, as we have already mentioned it in the introduction. However, the study must refer to the so-called contractual theory, the fiduciary theory (United Kingdom), and the misappropriation theory (United States), since these theories has significantly influenced the international legal literature of the economic analysis of commercial secrets.

[16] Making this decision we took into account that the Hungarian version of this article was closed on 31 December 2009, and in this English version we aim only at signalising for the reader the changes that have occurred since, but we were unable to completely rewrite some parts based on the new regulatory regime/reaching the same conclusions, since this collection contains the original studies, not new analyses.

The terms fact, solution, data and circumstance are thus seen as elements of the information.[17] The statutory text expands the definition of commercial secret to all valuable sub-information that in themselves do not contain information, that could not be the subject of a commercial secret, because the threat of an injury in interests is only present with the acquisition of the information itself. The law considers the questions of evidence. It is extremely difficult – often impossible – to prove the realisation of the causal link between two facts that constitute two sub-information, which means that even the acquisition of the two sub-information that are individually invaluable can violate commercial secrets, if putting them together they can threaten financial, economic or market interests.

Making an information public or letting it be used by unauthorized persons can only violate the legitimate financial, economic or market interests of a business, if that information provides some kind of competitive advantage, including the level of command over resources.[18] An information kept in secret, thus, is nothing more in economic sense, then a competitive advantage. Then it is not surprising that beyond the general norms of the Civil and the Criminal Code, competition law is the one that protects commercial secrets with a separate provision (Competition Act, Article 4).

The person entitled to the commercial secret has to perform all necessary measures in order to keep the information in secret.[19] This element of the definition has a role in making the commercial secret, the legally protected information identifiable for third persons, including the law enforcement authorities. This shows that there is an information kept in secret, thus the owner of the information realised its value.

This element of the definition contains other important substantive criteria for the economic analysis, namely that it is the owner of the secret solely that can decide whether the information is valuable or not. Thus the commercial secret has no normative content.

[17] By using the results of the formalistic information theory, we could have a more exact starting point for the analysis, so the results could be better generalised. It should be noted that in the data protection regulation the definition of data is wider, while the category of information is narrower. This approach is due to the special word-set of the data protection regulation, which is distinct from the legal vocabulary. Section 1 Paragraph 2:27 of the new Civil Code uses fact, information and other data, or a compilation thereof.

[18] The “rightful” part could be separately analysed. From the standpoint of our study this is only relevant, because the reference to “rightful” strengthens the hypothesis that the advantage cannot come from outside the normal functioning of a market economy, thus it can only refer to advantages gained from the economy, so only legitimate competitive advantages. For example the real information behind a commercial that states content unfairly influencing consumer choices cannot be the subject of commercial secrets. However, the information behind a commercial with valid content can be, so it is often only an authority that can assess the validity of commercial statements, but not the consumer.

[19] Section 1 Paragraph 2:47 of the new Civil Code states the condition goes as: the entitled person is not liable for protecting the secret, so their conduct related to the protection was what can be generally expected in the given situation. There is no substantive difference between the two solutions from the perspective of our analysis.

- This, however, does not mean that it would constitute a commercial secret if the owner of the secret would classify information as commercial secret that do not violate or threaten economic or market interests, or already public, let alone of public interest. Obviously this element of definition is important in a criminal proceeding regarding the violation of economic secrets or in a judicial case of commercial secret violation or of access to public interest data. First and foremost judicial cases regarding access to public interest data are when this element of the definition plays an important role. The cases that are relevant to this article, so when the subject of the judicial or administrative procedure is not the classification of a commercial secret, then this element of the definition is disregarded in the realm of the classification of commercial secret, and it only has a role in the disclosure of commercial secrets relating to assessment of the necessity and proportionality of the reasons for disclosure. This means that so long as in a case about the classification of a commercial secret does not decide on the character of a certain information, law influences the regulated subjects' conduct as if any information classified as a commercial secret by the owner of the information would in fact be a commercial secret. This determines both the procedure of the regulatory authority and the judicial review thereof.

Without presuming the economic theory related consequences of information society's impact on modern market economy, it should be noted that precisely the competitive advantages gained from information are the greatest in modern market economies, because the core competences which cannot be copied by others, are the ones that can ensure a long-lasting competitive advantage. Such core competences derive from institutional culture, institutional knowledge that are specific to the institutional structure, and are the collection of such institutional practices and knowledge that might only be partially known or stay hidden even from the management, because the procession and evaluation of this enormous amount of information is almost impossible. Due to this later fact, businesses attempt to classify as commercial secrets all information related to their economic functioning, and it is due to this that they have difficulties in giving reasons for such classification in an official – administrative – procedure. Nonetheless, for an economist it is clear that a rational business company is the sole authentic decision-maker in the question which information is providing its competitive advantage, thus which information is worth spending money on classifying and keeping as a commercial secret.

In summary: *from an economic perspective a commercial secret is all the company's information kept as a secret that is able to provide a competitive advantage against the competitor companies.* This is exactly the Anglo-Saxon definition of commercial secrets, which deeply influenced the TRIPS Agreement.

The *Restatement of Torts* (1939) for example says that a commercial secret is any information which is used in one's business, and which gives him an opportunity to obtain an advantage over competitors who do not know or use it.[20] Under US law the definition goes as follows: A commercial secret is any information that can be used in the operation of a business or other enterprise and that is sufficiently valuable and secret to afford an actual or potential economic advantage over others.[21]

Information in microeconomic models

As a starting point, we should remember what presumptions the classical – static – economics model (competitors' model) establishes: there are a large number of smaller buyers and sellers, with competing homogeneous products, the capital goods also come in homogeneous units, none of the market participants are able to change to price alone (everyone is a price-taker), market entry is free, prices and goods can move without any limitations, market participants possess all relevant information to make an informed decision (even the consumer knows all the possible alternatives). There are no mechanisms in place to win over buyers, such as reducing prices, increasing the quality of goods, or using advertisements, also there is no personal relationship between buyers and sellers. In such a market the long-term profit is zero, both the buyers and the sellers act as a *homo oeconomicus* (who can make optimal decisions) and there are no transaction costs.

Thus in the classical analysis the existence of information is a crucial starting point in numerous regards (advertisement, winning over buyers, even acting as a *homo oeconomicus* assumes it). If being perfectly informed is such an important starting condition, then we should rightly presume that the existence of a legal institution like the protection of commercial secrets is against the competition, consequently, competition law should *per se* prohibit it. Controversially, the situation is that competition law does not only prohibit commercial secrets, but even protects them.

The obvious model-nature of the starting conditions of a competitors' market is even apparent – contrary to public opinion – in the classical microeconomic studies. Economics views asymmetric information as one of the main reasons for market failures. If asymmetric information causes market failures in the functioning of market economy, then the existence of the legal institution of commercial secrets, which protects secret information, still seems unjustifiable, and thus the existence of asymmetric information shall be removed through legal measures.

[20] Restatement (First) of Torts Sec 757, Comment b (1939), http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=standards&p_id= 10103 (25 February 2010).

[21] Section 39 of the Restatement (Third) of Unfair Competition, Business Torts & Unfair Competition: A Practitioner's Handbook, American Bar Association, 1996, ISBN: 1-57073-294-9168.

Contrary to this viewpoint, there are other economic models describing a competitors' market. For example the paradigm of the new Austrian school[22] – based on the more realistic presumption of limited rationality (*Simon* [1982]) – sets the unavoidable imperfection of human knowledge as a starting condition, and its focus is not the determination and theory of an equilibrium price, rather the market as a mechanism for spreading information. Its perspective is fundamentally different from the classical theory, since it considers the differentiation of products to be an immanent element of competition. The market is in motion not because of the buyer and the seller (producer), but only because of the intermediary merchant, the profit-oriented company. While the buyer and the seller are simply price-takers, the competition of entrepreneurs makes the profit disappear, because the difference between production prices and retail prices are always levelled. It must be noted that in this theory information has a completely different role as in the classic competitors' market model. Here *information is the driver of competition*, and in this regard this theory stands on entirely different grounds as the classic competitors' market model.

The existence of asymmetric information belongs to competition, without it we could not talk about competition. By this the legal institution of commercial secrets could be nicely explained. The legal institution of commercial secrets protects the intermediary entrepreneurs, who can – by disseminating information – influence the prices and who are the cornerstones of market economy and competition. In this context, however, the Pareto-optimality of market competition comes into question. The less profit those market participants who are able to influence prices can make, meaning that the less the price of information is, the more efficient a market competition in the allocation of resources is.

The modern theories of institutional economics, such as the property rights theory, the principal-agent theory, or the theory of transactional costs, may further differentiate our views on market economy as the dominant economic-regulatory mechanism. These theories influenced other disciplines, including organisational studies, or some areas of law, thus creating the school of *law and economics* that holds the economic analysis of legal institutions as its core subject of inquiry.[23]

Market exchange is just one form of economic processes, and distributions. The reproduction of goods in a company happens through administrative channels, instead of market regulatory mechanism, consequently, it cannot be stated that market exchange, the far from uncontroversial price-system is alone or even dominantly

[22] The new Austrian school was founded in the 1960s and 1970s by Ludwig von Mises, its most influential representative was Friedrich von Hayek, nowadays its lead figure is I. M. Kirzner (In this topic we are relying on the monography of *Mátyás* [2003] and the study of *Mátyás* [2004])

[23] Law and economics is the subject of major legal researches also in Hungary, especially in the field of civil law. (*Vékás* [1998], *Sajó* [1984]). The international literature of the subject is enormous. See on the different viewpoints: *Burrows–Veljanovski* [1981], *Cooter–Ulen* [2005], Kelman [1987], *Polinsky* [1989], *Posner* [1996].

responsible for economic regulation, even in a so-called market economy (*Coase* [1990]). The importance of Coase's work lays in his acknowledgement that the basis for well-defined property rights and a functioning system of market exchange – in other words the basis for the prevalence of market economy – is a more or less uncontroversial price-system. Without well-defined property rights, it is unavoidable that one of the reasons for market failures is to be the existence of negative (and positive) external effects (*Coase* [1990]). It follows Coase's basic argument that optimal solutions can arise, only if the value of transactional costs were to be zero. Consequently, the existence of transactional costs is the reason for the losses in efficiency. In cases where transactional costs create barriers to market solutions, law has to intervene. So if commercial secrets increase transactional costs, then it results in the loss of efficiency and can be a barrier to market mechanisms. Thus law has to intervene against commercial secrets, as against one element of transactional costs.

The principal-agent theory is also influential on our views about market economy, which describes processes beyond the well-defined property rights.

According to the members of the new institutional school (*Williamson* [1981]), who considered Coase as their forerunner, in the case of special capital goods, transactional costs are exceptionally high. This fact has, however, become dominant in the extremities of international division of labour, and become the obvious reason for the rise of bilateral monopolies. Two other models of microeconomics (the Azariadis–Baily–Gordon-model and the Okun-model) – based on the contract theory – explain the long-term contractual relations between seller and buyer, which relationship increasingly resembles the long-standing relations between employer and employee, by the high-priced nature of information. This explains not only the permanency of wages, but also of prices, thus imposing limitations on the functioning of classical market mechanisms. Based on commercial contracts, Williamson concluded that in the long-term commercial relations for specific capital goods the contractual partners develop so-called relational contracts and – due to the high transactional (mainly exchange) costs – they are often interested in collective profit-maximisation.

The literature of negotiation game theory is expansive, and includes a large number of meticulously executed experiments. One of the main results of the experiments was the realisation: the more definite the rights of the bargaining partners are, the more they tend to co-operate, while the less transparent their legal relations are, the smaller the chance is that they reach an agreement. According to Robert Cooter and Thomas Ulen, the negotiations become complex and burdensome, when private information is needed for the decision. Private information hinders negotiations, because mostly they must be made public in order to reach the rational conditions of coordination. In general: a negotiation is costly, if a lot of private information needs to become common for a deal (*Cooter–Ulen* [2004] p. 93). This makes it clear that commercial secrets between seller and buyer increase transactional costs.

Lastly, it is worth referring to the empirical study showing that within a company the most effective tool to increasing competition between employees is to withhold information, especially in relation to employees who have accomplished performance-based successes within the organisation (*Hámori et al.* [2007]). However, here the causality chain is reversed. According to company leaders, one of the most crucial detrimental effects of in-house competition is the hindrance to the flow of information, which encourages the avoidance of too intense in-house competition. This fact shows that an intense competition leads to limitations on the flow of information, consequently, too intense competition is avoidable. Is it possible that the legal institution of commercial secrets was brought to life by the too intense competition? Would this mean that the legal institution of commercial secrets legitimises a detrimental effect?

We can conclude so far that information plays an increasingly important role in economic theories, and could be the basis of a new theory. The heightened attention is, however, understandable, since the "informational boom", the evolution of information technologies created a new – information – society, in which the functioning of institutions and market participants, and thus the functioning of the market economy, is fundamentally transforming. By the 21st century, information has become a key resource, while due to globalisation, market competition is ever increasing. There is almost no production factor left, including natural treasures, cheaper manufacturing technologies, qualified or cheaper workforce, which is out of the reach of a multinational company. Consequently, complex information-centred competition strategies have arisen, and the acquisition of unmatchable competitive advantages, the achievement of long-term competitive advantages has become a core competence.

All of this has the consequence that the problem of asymmetric information appears in a more complicated, complex form. Although it would follow – especially in the markets ruled by multinational companies – that the legal institution of commercial secrets, as a means of competition between companies, has been integrated into the protection of privacy and private secrets, nonetheless, the general purpose of that right – along with the rules on data protection – is to ensure information freedom rights, and to eliminate asymmetric information between companies and consumers for the sake of private individuals, and for the loss of companies (*Vikman* [2006] p. 23).[24]

As a consequence of this evolution, the literature on law and economics does not consider information as an external condition anymore, rather as a good with its own market. In the following sections, we will examine this theory and its plausibility.

[24] However, this development is a dichotomy. While under the prior practice of the data protection ombudsman the commercial data handled by authorities was considered public data, and for example a decision stating the violation of a statute was a clearly public interest data, but under the newer practice the commercial secrets handled by authorities do not considered public data. (*Majtényi* [2006] p. 428, *Jóri–Bártfai* [2005] p. 159–164).

THE MARKET OF INFORMATION

Robert Cooter and Thomas Ulen apply in their already mentioned work the property rights theory to information as well (*Cooter–Ulen* [2004] pp. 120–122). Here two difficulties are identified regarding the property rights of information and the creation of a market for information. Information has two distinct features that differentiate transactions of information from the transactions of regular private goods. The first of these features its *non-excludability*, while the other is its *authenticity*.

Its feature of *non-excludability*, makes information resemble to common goods. Information is difficult to create, however, it is usually easy to transfer. Information is sold by its creator for a fraction of its value. The use of information is free of competition, because – as opposed to other goods – the use of information does not reduces its quantity and its gains for others. "The use of information is thus *free of competition*" (*Cooter–Ulen* [2004] pp. 120–121). It is difficult to exclude others from the use of information, due to this the free rider problem exists. Consequently, similarly to public goods, the market is unable to produce the sufficient amount of information. Since the private sector on market grounds offers less than the optimal amount of information, in unregulated markets there is a lack of ideas, knowledge and most importantly creations embodying thereof.

It follows from the public good nature of information that it is either ensured by the state,[25] or in the realms of contracts the protection of commercial secrets creates the regulated market for commercial secrets, or supplementing the protection of commercial secrets intellectual property rights are also regulated. It seems from this that the legal institution of commercial secrets could be sufficiently justified.

The question of *authenticity* is usually mentioned in relation with contracts. This is based on the negotiation theory in the US literature, which provides a perfect terrain for game theory analysis.

The aim of the contracts system that can be created based on game theory is to transform games with non-efficient solutions to games with efficient outcomes. The enforceable contract transforms a game with a non-cooperative outcome to cooperative. The further aim of contracts is to promote the efficient publication of information within contractual relations. Situations with asymmetric information could be managed with this, leading to the redistribution of welfare rather than the extension of welfare, thus barely relating to our topic.

The two aims stand in a means-ends relationship. The efficient distribution of information enables cooperative outcomes. The problem of authenticity stems from the fact that the buyer is unable to assess the value of information before receiving it. It is a common problem that the information has to be revealed before the buyer in

[25] *Cooter–Ulen* [2005] refer to the system of charity donations (p. 133), which, however, is equivalent with indirect state financing.

order to determine the value of it, but then what is the reason to pay for the known information? To understand the problem, we will shortly describe how contract law can contribute to the efficient disclosure and transmission of information.

In economics information is public, if in a negotiation process both parties are aware of it, while it is private, if only one party knows it, while the other does not. The stimulator of a transaction is private information. The transmission of information and the trade of goods enables one to take over control of knowledge and resources. Due to the fact that private information lies at the heart of transactions, law usually treats contracts based on asymmetric information as being enforceable. Nonetheless, efficiency requires that the merger of the control over knowledge and resources would be of the lowest cost, respectively to the costs of information transmission and of the trade of goods. Consequently, a contract is not legally enforceable in cases of omission of guidance, fraud, or bilateral misconception (in this case there is not even a bargain), however, it is enforceable in the case of a unilateral misconception.[26] By this, law attempts to promote efficiency through benefiting the pursuit of information and the merger of control of knowledge and goods. There is a possibility, nevertheless, that information was acquired by chance, thus without the costs of pursuit, and so the unilateral mistake of the other party does not lead to a boost in efficiency.

a) For this reason, the literature classifies information based on its effects on economic efficiency. According to their nature, there are information that *enhance welfare* (productive information) and that *redistribute welfare* (redistributive information). Productive information are for example discoveries, inventions, etc. Contrary to this, redistributive information provide such an advantage to its holder, which can be used in a negotiation in order to redistribute welfare according to the holder's interests. For instance, if someone acquires the information before others where new rail-roads will be built by the state, it gives him a great advantage on the real estate market. Investments made for the acquisition of redistributive information may seem on the one hand like a luxury, but on the other hand it encourages those who do not wish to suffer welfare losses to be better informed so as to carry out defensive investments. The investments with a defensive aim are, however, only created obstacles to redistribution, but do not create new value.

b) Additionally, information can be labelled according to the *method of acquisition*. Information can be acquired in an active manner, namely by investing resources into the recovery of information, or by chance, accidentally.

[26] Under Hungarian law, a contract may only be challenged based on unilateral misconception, if the clearly false information was provided by a legal counsel advising both parties, and the misconception was regarding an essential question.

From an efficiency standpoint, there is only one combination of the nature and manner of acquisition of the information that clearly justifies the enforceability of a contract. *This is the productive information, which was a result of an intentional investment.*

Most information, however, is in practice both productive and redistributive at the same time, thus mixed information. Most information also seems to be mixed from the aspect of being acquired through investment or accidentally. It can be asked for example whether the information acquired from the market situation itself – not intentionally, but as some kind of positive externalities – was gained accidentally or in an active way. If for instance we pursue a legal education, and later as a lawyer handle a lot of real estate contracts, we might accidentally acquire the information about where the next rail-road will be built. Conducting any kind of economic activity, we can come across a number of accidental experiences, which can be acquired by anyone pursuing the same economic activity, but in order to start such an activity we need a large amount of information and knowledge. Are these information the fruits of the investment into knowledge or the results of chance? It is a further difficult question, whether the accumulation of huge corporations' institutional knowledge is the result of intentional investments or accidents, the latter of which is a statistical necessity.

The literature distinguishes between three principles of economics: "*1.* A contract has to be enforced, if some productive information was not at the disposal of all parties, especially if that information was a result of the investment of one party. *2.* Most contracts should be enforced, in which a mixed information (both productive and redistributive) was not at the disposal of both parties at the time of signing the contract. *3.* A contract shall be annulled, if the party holding the information has not increased, only redistributed welfare, or the information was acquired accidentally. " (*Cooter–Ulen* [2004] p. 283)

c) The third possible way to classify information according to its nature may be connected to the questions of authenticity: whether there is an *obligation to give guidance* or not. The obligation to provide guidance triggers the definition of security information. Security information refers to a knowledge that helps people to avoid damages. Naturally, law requires the parties to share with each other all security information they possess. Law often requires the seller to be aware of such information in an explicit manner.

There is another side to the problem of authenticity, which is independent of the lack or ambiguity of information and may be understood from the game theory analysis of such single transactions that are of great value. Single transactions of great value can often be described as the results of manoeuvres, of using unfair, but not illegal techniques against the other party. In these cases, the parties making offers to each other mostly disregard the losses caused by the breach of the promises. Separate studies deal with the issue of the appropriate amount of damages that deteriorate from the breach of contracts, but at the same time do not cause a loss in efficiency, thus the disproportionate amount of damages do not discourage from contracting (*Cooter–Ulen* [2004] pp. 205–206).

Contemplating this, the risks (ambiguities) are not only caused by the incomplete information known about the other side, which could only be eliminated through contract law with considerable transactional costs, but also they are caused by the fact that there are only a known and a definite number of games. If we are building a long-term, lasting relationship, of which duration is unknown, then we are facing an infinitely repeated game, in which cooperation is more likely than competition. Logic is that simple. It is well know that at any round of the repeated game, in which the principal (first player) invests money, the agent (second player) gains immediate profit by the expropriation. The principal may strike back by not investing anything in the following rounds, as a result of which the return of the agent will be zero. So long as the agent is unaware of which round is the last, and may assume that there are an infinite number of further rounds, expropriation is not a winning strategy, because he can expect more profit from the next rounds than from a once-only expropriation. As a result of this, long-term business relations are far more efficient than the single-time relations.

It can be observed in the economy that the intermediary commercial activities are attempted to be covered by exclusive distribution contracts, through which the advantage given to the agent ensures the continuous and long-lasting relationship. This on the other hand is advantageous to the principal.

Infinite games contribute to the enhancement of information authenticity, the improvement of business trust. A number of risk factors may be eliminated through this, nonetheless, the costs of exchange increase. Apart from the market of goods that can be acquired through single-time transactions of law value, the markets of all other goods and services are built on business confidence, which prerequisites, however, long-term relations and contracts, leading to the permanency of prices. Consequently, the distortion of market competition is not the result of the mere existence of asymmetric information, but of a game theory proved situation that derives from the lack of information regarding future action.

It can be noted that for the analysts of law and economics, asymmetric information brings market processes, market exchange and competition into motion. Similar conclusions have been reached by the members of the new Austrian school as well, reserving that they do not differentiate between productive and redistributive information, rather consider both as a source of profit, thus the driver of market competition.

Law and economics use the theory of transactional costs to show that transactional costs are the cost of the disclosure of asymmetric information. The bargaining process is about nothing else but the costs of negotiations and other expenses of signing a contract. Contract law attempts to reduce these costs. The other cornerstone of law and economics holds: in an efficient market economy, the most resources belonging to the one who pays the most for them, because he values them the most. We can thus conclude that the one who values resources the most, is the one with the better information.[27]

If *ceteris paribus* the person (company) possess the same amount of resources, the individual who will be more efficient is the one who is able to utilise those resources more effectively, consequently attaining a higher level of production. This derives only from advantages in information. In general terms, this means that there is no competition without asymmetric information.

As a consequence the question is not only whether competition is the most efficient allocator of resources, but in a dynamic perspective also whether competition is putting technical innovation into motion. An innovative enterprise in the Schumpeterian-sense is the one that induces market competition. The *existence of asymmetric information* is an essential condition of *technical innovation*. This asymmetric information situation is efficient, only if it involves new information – yet completely unknown and created not by chance, but through investment. *Consequently, the new information is without doubt a productive information.*

However, the static competitors' model that considers a perfect informational situation as a baseline, is not in contravention with the information-market approach. Namely, the new information could mean a new product and thus a new market, which is the basis of product diversification. There are claims that product diversification reduces the intensity of competition due to incomparability. If we take the approach of the new Austrian school, which considers product diversification not as feature of monopolies, but rather as a natural by-product of market competition, then we can conclude that the above mentioned critical view is only true, if there is no new information involved in the product diversification. The fact that in a certain market the intensification of competition can be sensed when new information is used, only means that the new product is a close substitute of the previous one. If we talk about new markets in these situations, then only a correction mechanism dependent on interchangeability relations starts on the previously Pareto-optimal competition market. If the two products are completely interchangeable, then the previous market disappears. If they are only partially interchangeable, then due to the reduction in the demanded amount, the market for the older product is necessarily curtailed. If the new information is disclosed with others and can be utilised, then soon the equilibrium price will be reached once more.

This, however, also shows that it is not new and non-productive information which sustains competition, but in fact hinder the emergence of a Pareto-optimal situation. Thus the product diversification leads to a loss of efficiency, only if it is

[27] The one, who is badly informed and thus pays larger sums, will not be the holder of resources for long, because shortly will go bankrupt…

based on redistributive information. It follows that legal instruments may contribute to the strengthening of market competition, only if they are aiming at creating a market for productive information. Since the completely new information are definitely and undoubtedly productive, the legal system couples these up with property entitlements. This is the function of intellectual property rights. This is the basis for the recognition of intellectual property rights, patent rights, and trade mark rights[28]. The innovations and know-how also create such property rights, but these are special or borderline cases, because the former prevails in an employer-employee relationship, while the subject of the latter is difficult to define, thus in both cases the productive effect is harder to prove.

If we simplify the concept of know-how it can be considered as a special subsection of a commercial secret, which deals with information that is partially or fully protected with property rights, and in the given market situation it is more reasonable for the company to treat them as commercial secrets. Know-how is a special commercial secret in way that it is the most productive. As it is explicitly mentioned in the Civil Code, [29] this analysis disregards know-how, so all the conclusions of this study is limited to non-know-how commercial secrets. The reason behind this, is that know-how would require a separate analysis, which could lead to different conclusions in a number of questions.

The main feature of a market is that it enables the appropriation of information for a limited time – and sometimes with limitations, thus temporary monopolies can arise. The temporary nature of such monopolies compels their utilization, which could have major effect on other markets as well. The time limitation on the monopoly should be construed in a way to allow the emergence of a new market. The regulation should allow for monopoly rights so that after the emergence of a new market they should enable the evolution of a competitive market. This question could be analysed concretely, and it is the subject of the discipline of law and economics. With regard to trade marks the situation differs, because the time-frame of the protection is determined by the duration of the actual utilization.

The protection of trade marks is productive, because market value is only attached to these rights, if they are indicating a quality above the market average (or at least they are perceived by consumers as such). Since quality has a productive effect by definition, moreover, it reduces the consumer's need for information, strengthening business confidence, and thus eliminating problems regarding the authenticity of information about the product, it can be considered as a border-line means of productive product diversification.

[28] This is only limitedly true for trade mark, because the trade mark provides new information for the consumer in the producer-consumer relation. The role of trade mark is more important in the reduction of search costs and in the fight against "market for lemons". The legal institution of trade mark would also need an economic analysis, just as now this article is providing it for commercial secrets.

[29] Section 2 Paragraph 2:47 of the Civil Code.

It should be emphasised that our analysis of the legal institution of commercial secrets focuses on information other than the above mentioned, and although there might be interesting overlaps with know-how (for example it can be kept as a secret as well), but it is not the subject of this study, just as patents and intellectual properties.

We could think based on this that the legal institution of commercial secret protects information, which kept as a secret results in wastefulness, and loss of efficiency, thus the legal institution is not efficient in an economic sense. Especially so, as the subject of commercial secrets covers all economy-related partial information, thus its scope is seemingly endlessly expandable – except for the statutorily, explicitly defined, concrete information. Companies try to classify all information regarding their management and functioning as commercial secrets, which is only limited by the costs of the necessary actions to protect commercial secrets. We can thus assume that *most of this information is not productive, other legal institutions remove most of the productive information from the range of commercial secrets. Consequently, we consider the legal institution of commercial secrets as consisting of mixed and distributive information.* [30] Such an interpretation of commercial secrets raises a number of basic questions in the literature, which although can be answered based on our analysis, are not part of this article (See on these issues: *Cooter–Ulen* [2007]).

First of all, it must be observed that commercial secrets as information have no legally created market. The Anglo-Saxon legal theory has come up with two justifications for the necessity of the protection of commercial secrets. *1.* Based on the property rights theory a commercial secret is a property, which inspires its owners for innovation. [See the US Supreme Court decision in Ruckelshaus v. Monsanto Co., 467 U.S. 986, 1001-1004 (1984)] Due to this, some consider a commercial secret to be an intellectual property right. *2.* Under the other theory, the protection of commercial secrets derives from the law of damages (and contracts). If someone transfers someone else's commercial secret without permission, he breaks a contract, and violates the due diligence requirement towards the entitled person, and owner. This due diligence requirement can be justified with the preservation of commercial morals and fair competition.

Since a commercial secret as an information has no market, the conclusions of the property rights theory may be debated. First, it is not clear how innovation is promoted by commercial secrets. Second, commercial secrets are distinguishable from intellectual property rights, since the protection is attached not to ownership, but to possession.

[30] This assumption might be debated based on a more detailed analysis. After finishing the study, anyone could correct the results based on a wider interpretation. It should be noted that security information could not be classified as commercial secret, the reasons for which we will deal with in the next chapters.

According to *Bobrovszky* [2006] "the cohesion, the common ground of intellectual property… lays not in exclusive rights, those are only the core of it, but rather in its subject being different goods with intellectual values, and the two levels of protection… are separated from a private law standpoint": 1. One level being a *de facto* possession-like based on the protection of commercial secrets, 2. And the other being a *de jure* ownership-like based on patent and other exclusive rights (*Bobrovszky* [2006] 1388.p.).

Nevertheless, the *de facto* character of the protection of commercial secrets differentiates it from the protection of "other" forms of intellectual property, so defining it as a right is difficult. It follows that even the elimination of the legal institution of commercial secrets would not mean that the commercial secrets would not exist. It would be a mistake to assume that the economic analysis of the right to the protection of commercial secrets means that the scientific analysis that if there are no commercial secrets, then there are no redistributive or mixed information in market competition, which appear as asymmetric information. Studying this would only be reasonable in connection to such a legal institution that requires all redistributive and mixed information to be published and not kept as a secret. Such a regulation would cause "infinite" social costs, since the number of this information is practically endless, for this reason there exists no such regulation.

If we consider that commercial secrets exist without the legal institution of commercial secrets, then the function of the legal institution of commercial secrets should be searched elsewhere. This statement is proven by the fact that *the acquisition of market information by deducting the competitor's commercial secret from the competitor's product is an approved practice. So the existence of a commercial secret is a factual matter, not a legal one.*

If we take into account that information in itself (as opposed to a patent as a right) cannot be considered as an object, thus it cannot be the subject of a property right,[31] then it is easy to realise that the right to the protection of commercial secrets is not to be associated with property rights, but rather it is a product of contract law.

This view is underlined by the analysts of law and economics, who always refer to examples taken from contract law, when dealing with problems of the protection of commercial secrets – especially from the employer-employee contractual relations (*Cooter–Ulen* [2008] p. 134). They also state as a general opinion that the weaknesses of the legal regulation of the protection of commercial secrets undermine the efficiency of the system as a whole.

[31] Section 1 of Paragraph 5:14 of the Civil Code holds that the subject of ownership may be all things of a tangible nature which are capable of appropriation. Section 2 states that the definition of things also include money and securities, including natural resources that can be utilized as capital goods. While Section 3 refers to the special rules regarding animals. It is, however, difficult to interpret information as a "natural resource".

If for instance *A* as an investor signs a non-disclosure agreement with his employee *B*, who then discloses *A*'s secret to *C*, and *C* did not know or could not have known about the breach of contract, then without a contractual relationship between *A* and *C, A* cannot bring a claim against *C*. Moreover, the disclosed information that is known by the given industrial sector may be utilised by anyone without compensation, if everyone is aware of the fact the information was made public with the violation of the non-disclosure agreement (ibid).

This example shows one of the main functions of the legal institution of commercial secrets. If non-disclosure could be obliged only by means of contract law – namely to keep information and prevent its transmission (so an obligation to endure) – then *A* can typically bring a case only against a (former) employee based on the contract, but then *C* could not even be held responsible, even if he knew (or should have known) that it was regarding a commercial secret. Not even if it had been *C*, who convinced *B* of breaking the contract. For one thing, it is quite probable that *B* as a regular (former) employee without the necessary funds would not be able to reimburse the damages caused by the breach of the secret. This is apparent, because *B* can calculate the loss from the breach of contract based on the multiplication of the probability of being caught and the damages caused. Since the subject of the commercial secret is unreasonably wide, and not well-defined, moreover, its utilization necessarily happens in secret and its public results only appear indirectly, then the multiplier of the probability of being caught is below 1, close to 0. It follows that an employee can be easily tempted, since chances are low that the breach of contract will be exposed, thus the compensation given by *C* for the breach of contract would not cover the caused damages. The legal institution of commercial secrets eliminates the too extensive – and thus not efficient – risk of breach of contract. It ensures that *C* can have a claim brought against them, if he knew (or should have known) that it is regarding a commercial secret, also if *C* abetted *B* to disclose the secret. If the legal institution of commercial secrets did not exist in civil law – without being of a contractual character – it would be as if criminal law only punished the thief, but not the dealer of stolen goods.

It should be noted that the legal institution of commercial secret creates a legal relation between *A* and *C* by a unilateral declaration. Consequently, *C* is not allowed to lawfully use a document labelled as *commercial secret*, even if he acquired it by coincidence and legally. (Such a document can be used only for what was permitted by the entitled person. If no permission was given, then the document cannot be used at all. It follows that even without special regulations, employees of public authorities and courts are obliged to keep the commercial secret. In the course of a judicial proceeding, the person entitled to the commercial secret discloses the commercial secret voluntarily – generally in a civil case for example because he wishes to use it for the case, so the other party has to make a statement of non-disclosure.)

It is, however, still a question whether innovation and market competition is enhanced by the acknowledgement and enforcement of such contracts between *A* and *B*, or by the legal relationship between *A* and *C* based on the existence of the protection of commercial secrets.

Cooter–Ulen [2007] (p. 134) has brought attention to the limited efficiency of the protection of commercial secrets. Empirical studies conducted in the Silicon-valley showed that employees working there often switch workplaces and in such cases they bring with them most of the commercial secrets of their prior employer. In more cases, employees do not even notice when they are breaking the contract, because the laws governing commercial secrets are in violation with the business norms of Silicon-valley. It is well-known that the real places of innovation in Silicon-valley are the pubs, where employees of similar status but coming from different companies spend their spare time.

The case of the Silicon-valley is a nice example of how the weaknesses of the protection of commercial secrets may be the driver of innovation, since the world's most successful IT companies are in Silicon-valley, which proves that regulating commercial secrets as a contractual matter hinders competition, so does the legal institution of commercial secrets.

If the legal institution of commercial secrets is an obstacle to innovation, what might be the reason for sustaining the legal institution of commercial secrets from the perspective of economic efficiency? It is shown by our earlier example that the legal institution of commercial secrets has important functions within the company, in employer-employee relations.

Section 2 Paragraph 4 of the Competition Act emphasises that "an unfair access to trade secrets shall also mean where access to such trade secrets has been obtained without the consent of the data proprietor through a party in a business relationship - including the provision of information, negotiations and making proposals prior to making a deal, where no contract is signed subsequently in consequence - or in a confidential relationship with such person - such as a contract of employment or any similar relationship, or membership at the time of, or prior to, gaining access to the secrets." Under b) and c) Subsections of Section 3 *'confidential relationship'* shall, in particular, mean employment relationship, other work-related contractual relationship and membership; while '*business relationship*' shall comprise the provision of information, negotiations and making proposals prior to making a deal, whether or not a contract is subsequently signed in consequence.

This definition of commercial secrets fulfils an important function even in other business relationships, such as the buyer-seller (company – principal – consumer). Consequently, in the following sub-chapter we will shortly summarise our conclusions on the role of commercial secrets in vertical relationships.

THE ROLE OF COMMERCIAL SECRETS IN VERTICAL RELATIONS

The main research area of asymmetric information even within the buyer-seller relation is the consumer markets, retail markets (*Carlton–Perloff* [2000]). In these cases usually commercial secrets are not – or only indirectly – responsible for the existence of asymmetric information situations. For one thing, the sellers are obliged to provide consumers with all safety information, so these cannot constitute a commercial secret. Safety information helps to eliminate the problem of limited information regarding the quality of the product, and the company is highly motivated to make all the positive quality characteristics public. Since the Competition Act prohibits – in the course of advertisement and consumer information – the concealment of information regarding the essential features of a product, thus none of this information can constitute a commercial secret.

Nonetheless, the legal institution of commercial secrets has a direct effect on how informed the consumer is, because if information as a whole or part constitute a commercial secret, then the validity of the facts and data behind the consumer information cannot be controlled. For example a credit-line contract of a bank and a retail company behind the interest-free, "costless" credit offered by the retail company might be a commercial secret, and the credibility of the provided information can only be checked through administrative procedures. Moreover, the protection of commercial secrets as a legal institution do not even play a role in these cases, as the holders of the commercial secret – the employees of the retail company or the bank – are not at all interested in the disclosure of the commercial secret. Consequently, this information would be kept as a secret, even if the legal institution of commercial secrets did not exist.

Since the sole interest of the consumer is to acquire all relevant information regarding the price and quality of the product, which is also required by other legal provisions, the company has no obligation to provide information about either the other features of the product or the market opportunities related (e.g. where the product is on sale), so the legal institution of commercial secrets has no influence over these market relations.

Regarding the quality and the price of a product, the company is not allowed to refer to commercial secrets against the consumers. This information in the company-consumer relation is protected by neither contract law, nor by the legal institution of commercial secrets. The fact that the consumer can acquire such information anyway, the information loses its commercial secret characteristic, since it can be freely transferred (regardless of whether the company would like to withhold it from the competitors).[32]

[32] An example for this is the case of double price-discrimination between new and existing customers, when the existing customer does not terminate the contract only because the service provider – when realising the determinate intention – offers the discounts given to new customers, although

When neither retail or consumer markets are involved, but rather it is regarding the market of production factors in a broader sense, including the distributional, wholesale markets, when typically companies close deals with companies, then the information channels for prices and quality might differ. These questions have a wide literature, mostly in marketing.[33] Commercial secrets play a crucial role in the negotiation of the parties, and not only regarding contractual terms, but also the prices. Contrary to consumer markets, in the market of production factors the list prices and price reductions of the delivery contract between the parties constitute a commercial secret, thus the buyer is obliged to keep it. This only limits the buyer in using certain physiological techniques in the bargaining process (negotiations with other companies regarding prices cannot be referred to), but in general the company is not restrained in making an informed decision. At the same time, the legal institution of commercial secrets protects this information from the competitors on the seller's side.

It is, however, often not efficient that the seller[34] provides a greater price reduction to only one of its buyers, without being able to double-check the information from the buyer's competitors. Moreover, information regarding price as a commercial secret can only be redistributive information, thus it is not efficient in an economic sense either. Handling price information as a commercial secret on the market of production factors thus only leads to loss of efficiency.

The role of commercial secrets (and its legal institution) in these situations is limited to horizontal relations, so the efficiency of the regulation is dependent upon whether the legal institution of commercial secrets can be considered efficient in the relations of the competitors.

The employer-employee relation, in connection with the principal-agent theory in the literature on asymmetric information, has been in the spotlight of economic analysis as well (e.g. Spence [1973]). We have already showed that one of the main features of the legal institution of commercial secrets is related to the relationship of employer and employee. This comes from the fact that in the modern market economy governed by organizations, a large number of people represent

under its official communication it is not allowed to. This option is a commercial secret towards the competitors, but the customer can get hold of the information. This information then could be shared with other customers, or even competitors.

33 Microeconomic studies do not really deal with this. The reason for this is that microeconomic models assume that the companies making rational and optimal decisions are also acting as well-informed and expert participants in the market of production factors. Otherwise they would stay behind in the competition. Alternatively these supplier markets in the system of long-term contracts are prevalent, which follows from the already mentioned views of game theory and from the problems of the validity of information.

34 The literature of economics uses the term supplier for the seller of factors markets based on accounting jargon. We kept the term seller in this case, because supplier as a legal terms means something else.

the certain knowledge and information, which provides the company's competitive advantage, thus without the protection of commercial secrets this competitive advantage would be in all likelihood lost. For reasons of space, we do not have the opportunity to give a detailed account of this question from the aspects of economic analysis, so we will only discuss shortly the relevant assertions of an otherwise detailed analysis.

In the course of an employment, a commercial secret is best protected by the common/mutual interests of the employer and the employee, for this reason neither the contract law provisions for commercial secrets, nor the legal institution of commercial secrets can be justified from an efficiency standpoint by the conflict of interests between them, which is the alleged policy justification for regulating commercial secrets. However, as the conflict of interests can most efficiently be resolved by property rights on the side of the employer, thus without the regulation of commercial secrets inefficient situations may arise. The reason for this is that even the owner decides on "selling" a commercial secret based on the amount of the foreseeable profit. If the commercial secret belongs to more owners (and it is so in the case of property rights on the part of the employees), then the marginal cost of a single owner will be lower than the marginal benefit thereof of another company. In a situation like this a deal is struck even if the company selling the commercial secret could benefit more from the utilization of it than the other company. This outcome is not efficient.

Those employees are especially valuable for the company who make strategical decisions, and determine the company's business plan, goals, and specific actions for the future.

Information regarding future market behaviour is such a special information that there is a marginal benefit for a competitor company – if they exist – is always greater. The reason for this is the following. Let's assume that *A* company acquires the commercial secret of *B* about their future market conduct. This creates an asymmetric information for *A* company, because *B* company is not aware of *A's* future market actions. Let's assume that *B* company also acquires the commercial secret of *A* about their future market conduct. In this case *B* company has an advantage based on asymmetric information. Since one company's behaviour is modified by the information about the alleged actions of the other company, it is always that company with the competitive advantage based on asymmetric information, which acquired the other's commercial secret last. Let's assume that *A* and *B* companies acquire each other's commercial secrets at the same time. For this situation, game theory can give a description on when the returns are the highest. The outcome depends on whether they are aware of the fact that the other acquired their secret, and whether they know both sides are aware that the other side knows this. If both *A* and *B* knows all the facts, then they are in exactly the same situation regarding competitive advantages, as if they had never known each other's commercial secrets. Because the new steps that are based on the acquired information will not be known by either of them. If either company has more knowledge, for example if *A* knows about the simultaneous acquisition of commercial secrets, but *B* does not, then *A* has an informational advantage.

All this means that *the protection of commercial secrets creates a symmetrical informational situation in relation to information on future market behaviour, and without the protection of commercial secrets far more asymmetrical informational situations would arise!*

If such information (business plans, future price increases, expected innovations, patents, advertising campaigns, etc.) could be transferred by ex-employees, then the cost of keeping such information undisclosed would follow its benefits and advantages for the competitor. In such a situation the wage of the employees handling such information would be disproportionately high. The protection of commercial secrets is not efficient even at this point, which is proven by the fact that managers possessing such information have a higher income as compared to their performance, responsibility, etc.[35]

All these questions, however, belong not to the relations of employee and employer, but rather to the relations of competitors.[36]

COMMERCIAL SECRET IN COMPETITIVE RELATIONS

So far we have asserted that the legal institution of commercial secrets only increases situations with asymmetric information, with regards to mixed and purely distributive information. However, we have also shown that in case of some information, for example regarding future market behaviour, the number of situations with asymmetric information would increase without the legal institution of commercial secrets.[37] As a general observation, we can realise that since commercial secrets exist even without the legal institution of commercial secrets, then the commercial secret itself is responsible for asymmetric informational situations, and the legal institution of commercial secrets only worsens this by making the acquisition of commercial secrets more costly. For example, if *A* competitor company discovers

[35] We should note that this is not only due to the informational power of some people, this is only a factor. The control over resources for example can be just as important, which could justify the high level of management salaries.

[36] For reasons of space we are not dealing with the non-compete agreements after the termination of an employment relation, the treatment of which is similar, because just as for the existing employment relations they ensure the protection of secrets on a contractual basis.

[37] It would only “increase”, because even without the legal institution of commercial secrets commercial secrets would exist, so generally even without the legal institution of commercial secrets two companies would not know each other’s future market steps.

B company's cost-structure for a certain product, then this creates an asymmetric informational situation, which could only be balanced by *B* getting acquainted with *A's* cost-structure for the same product.[38] In order to preclude an asymmetric informational situation from arising, not only the legal institution of commercial secrets shall be abolished, but also the data underlying the commercial secret shall be mandated to be made public. Nevertheless, as we have shown, this would mean "infinite costs", thus such an intervention would not be efficient.

It has also become clear that the real function of the legal institution of commercial secrets appears not between relations of company and consumer or company and employee, but rather it influenced informational situations between competitors. [39] Consequently, the law and economics analysis of the right to the protection of commercial secrets must be carried out for horizontal relations of market competitors. This, however, brings us to the problem that it matters what kind of commercial secret we are dealing with.

So far we have discussed that the commercial secret is regarding as mixed and redistributive information, and the protection of these cannot be justified with any kind of argument for economic efficiency, except for their advancement of symmetrical informational situations. However, we have also seen, that *making public the information about market actors' future behaviour cannot be symmetrical, thus concealing these is required if asymmetrical informational situations are to be avoided.*

For the reasons above, first we will differentiate between the main categories of information constituting commercial secrets, and with these categories taken into consideration we will – assuming different market structures – analyse what effects the elimination of the legal institution of commercial secrets would have. According to this, we differentiate between *information regarding the future (behaviour) and factual information*, and within this latter part also between price- and cost-information. Naturally, a number of other types of information might exist, but since the category of commercial secrets is open logically, a conclusive and closed system of categories cannot be created. As for the market structures, the role of commercial secrets and role of the legal institution of commercial secrets will be examined on competitive markets, on oligopolies, on monopolies and on monopolistic market situations, as well as on a special case of regulated markets (price regulation).

---

[38] Naturally, it is another question that based on the market structure this could lead to various competition situations. For example in case of a duopoly, the intensity of competition could decrease.

[39] This statement is only true with the limitation that we disregard the seller-buyer relations on the market of workforce regarding employer and employee. However, even this proves that on the market of workforce the legal institution of commercial secrets is rather harmful in the seller-buyer relationship, and it pushes the average of wages down from the market balance to the detriment of the seller-employee (the buyer side is better-informed).

Categories of information

Information regarding the future is created by companies based on factual information. Factual information is not only information about the prices and costs, but also on quality, on selling conditions, or on clients and suppliers, etc. The availability of prices as factual information is critically important for consumers, but also crucial for market competitors. So it is in the interest of the company that the least people know what they are selling to whom and for what price.

- The ignorance of consumers regarding prices is valuable for companies. *Carlton–Perloff* [2000] describes the case of Ronald Kahlow, who attempted to take notes in a Best-Buy shop of the prices of different television sets. The shop took action against him and asked him that "for the sake of competition, please, do not take notes of the prices. It upsets the other costumers". The court held that taking note of prices is not against the law, so Kahlow was innocent.

It makes one wonder why a Best-Buy shop would go through all the trouble just to refrain a single consumer from making an informed choice. It is more likely that Best-Buy mistook Mr. Kahlow for someone coming from the competition. What argument can Best-Buy make against noting down the prices? Only something relating to commercial secrets. As we have mentioned the prices listed in a show room cannot be the subject of a commercial secret, while prices determined in closed negotiations (mainly in the market of production factors), prices used in trade are strictly under the protection of commercial secrets.

If the single most important indicator of an efficiently functioning market economy is a prices system, then how is it possible that the concealment of information about prices is protected by law? This is also a crucial question, since buyers are only interested in keeping the prices of a long-term frame-agreement in secret, if there exists some kind of anti-competitive alliance of interests between seller and buyer.[40]

Internet-based price comparison pages are able to make consumers better informed and also enable the comparison of different products. The success-story of mandatory motor vehicle liability insurance can serve as an interesting example, when the period for changing contracts was reduced to one month. (This statutory provision was abolished in 2010.) The reduction resulted not only in a price competition between service providers, but also for this one month (November) real competition had evolved. By reducing the competitive market for one month, information had become more concentrated and transparent, and the costs of transferring

[40] This is for example when a major public undertaking stands on the buyer side, which is not only profit oriented, but political connections might also play a role in the decision-making.

had been significantly reduced (for example the certification of reward and penalty were handled between insurance companies, the whole process was conductible via internet, etc.). The fact that the prices were not constantly changing, made the price information on the market more traceable, resulting in well-informed consumers. It was less feasible to use higher prices than the competitive price, moreover, companies had to adopt a more focused and deliberate price strategy, also by analysing the competitors' prices. [41]

Companies can gain a dominant position on the relevant market, if consumers are unaware of the prices, but also if they are uninformed about the quality. Limited information may lead to a monopoly price even on a market, where otherwise competition would dominate.

Since the other side of the price-competition is the competition in quality (product diversification), regarding which information is more complex, thus the analysis of price information might be coupled with the questions of – here not discussed – standardised contracts.

Information related to the cost-structure of a certain product can tell us not only where and what kind of competitive advantage does one company have regarding the use of production factors, but also it can be assumed whether the relevant market is competitive. The cost-structure also shows how much the capital-cost of the certain product is for the company. The reports and the balance sheets of a company are only appropriate to assess the cost-structure, if the company is a single-product company. However, in case of multi-product companies all these constitute commercial secrets.[42] In the intermediary commerce, for example, it is understandable that no one would like to reveal to the competitors what it sells for, how much and to whom, because the existence of intermediary commerce is based on this information constituting commercial secrets.

Lastly, it should not be forgotten that price- and cost-information could be information regarding the future, for instance, if they are concerning the future prices of a company. Competition authorities consider it a cartel, if companies inform their competition about their future prices, because it enables them to coordinate their behaviour (concerted practice, collusion). The most difficult question related to information regarding the future is to decide whether this may be mixed information, or only redistributive by effect. It may be entirely possible that none of the above mentioned categories is appropriate for information on the future market conduct.

[41] The LXII Act of 2009 eliminated this system of contracting

[42] It should be added that precisely determining the item-cost of one product of a multi-product company raises serious problems in methodology, so it is not only about keeping the cost-structure as a commercial secret, but often the commercially valuable information are not even known by the company that is producing the given product.

Commercial secrets on the competitive market

To begin with, we will use a game of logic to represent the real reasons behind the existence of commercial secrets. In the static model of the competitors' market the notion of commercial secrets is unknown, thus dissolving the assumptions of this model, we will attempt to introduce this term.

On a competitive market every actor is a price-taker, consequently, price is public data by definition. If a specific price (which is not the market price) is classified as a commercial secret by the company, then it would influence the equilibrium price. This, nonetheless, can only happen, if the consumer is under-informed, which creates some monopolistic power. The buyers are well-informed about the price on a competitive market, it is not possible to reach a price that is different from the market price, so there is nothing to be kept as a secret. Alternatively, if someone succeeds in selling at a higher price, then it has to be the result of product diversification. If a company finds out that a different company was able to sell something (somewhere, sometime, to someone, etc.) at a higher price, then it will attempt to acquire this segment of the market. Since on a competitive market there is no transactional cost, and the products are homogeneous, it cannot occur that prices are handled as a commercial secret.

The situation is comparable with costs and cost-structure. In theory, the costs of companies on a competitive market cannot be different, thus concealing the cost-structure would not create a competitive advantage.[43]

Although it is a rare case when new and cheaper technology is used to create the same product as the prior ones, but it is not unimaginable (this happened for example in the case of industrialisation of agricultural production). If the technology is indeed new, then it is not protected by commercial secret, but by the legal institutions of intellectual property law. If it is a solution constituting a commercial secret that is causing the reduction in the volume of production and at the same time in marginal cost (increasing economies of scale),[44] then a natural monopoly is created.[45] This example thus does not belong to the questions of competitive

[43] Naturally, in reality the cost-structures are different even in the competitive market, because there is never a long-standing perfect balance situation, thus in the course of a competition law analysis in practice the characteristic of a competitive market is the price dispersion due to the distinct cost-structure, which is a result of a number of objective circumstances (for example some level of market dominance that is always present in practice). The harmonised raise of prices – due to the differences in cost-structure – always raise the suspicion of cartel.

[44] It is natural to not refer to the case, when with increasing marginal costs, the marginal cost still remains the average cost even with a production size covering the whole market, because then the conclusions of the previous footnote are relevant.

[45] This conclusion is only true in case of mono-product companies. For more products *Evans–Heckman* [1983] showed that it can be economies of scale even without cost-subadditivity (natural monopoly) (See *Kiss* [2009] p. 93). All this, however, does not affect the validity of the statement.

markets. If by increasing the volume of production, marginal cost is increasing after reaching the ideal size of production (decreasing economies of scale), then the acquisition of a market share is not depending on whether the company treated the solution as a commercial secret. However, if a company conceals that it is conducting investments for the implementation of the newer technology – that is know to others – and later unexpectedly appears on the market with lower prices, then time can be highly relevant to the increase in market share, and so can also the treatment of the investment as a commercial secret. In this case the information treated as a commercial secret is productive, since it enables the company to reach the necessary production volume for the optimal production size. If others become aware of the company's intention of making an investment in technology, then – for the sake of staying on the market – they will also start investing, which could easily lead to a situation in which the advantages of the optimal production size could not be exhausted, and the industry will be characterised by oversupply and surplus capacities.

Nevertheless, it can be observed that this case is an example not for the concealment of costs as factual information, but for keeping the future investment plans of the company a secret, so the only consequence that could be drawn from it would be that the knowledge of future market behaviour is important even in competitive markets. Information of costs as facts, however, are generally known market information.

- If someone is considering giving up his well-paid profession, in order to live on truffle cultivation, he can look up all the websites that roughly show the returns of the investment in truffle cultivation. These would show that the cultivation of fruits or potato would bring at least the same returns as truffle cultivation, but due to the differences in technology, with different cash-flow. If someone has been cultivating truffles for years, he has such experiences that could give him a competitive advantage. Presumably, he would not be keen on sharing the knowledge gained through hard work with anyone, but the inherent characteristic of such information is that they result from combined experiences, so they cannot be easily transferred. The theory of easy transferability and impossible appropriation of information does not prevail for these kind of professional knowledge, so there is no need to protect these as commercial secrets.

At first, we could assume that the competitive model of the classical economics based on all information does not even require knowledge about the individual future market behaviour. The need for information is always connected to the specific situation, and the price movement carries this market information, based on which we may determine our future market behaviour. It follows, however, that information on other companies' specific future market behaviour is not market information. Since the competitive model presupposes that none of the market actors can influence the price, then the cost-benefit arising from one market actor's investment cannot be of a volume that can in itself influence market price, even if the company itself is able to sell at a lower price.

As a consequence, even if a company does not conceal its cost-reducing investments, it cannot happen that for this reason others start cost-reducing investments. In fact, the number of companies on a competitive market is so large that market actors usually conceive technological changes in the industry as a market incident. It follows that at times of technological change competitive markets may be highly unstable, because a lot of companies going through technological change at the same time results in huge oversupply.[46]

All this means that the future market behaviour – as opposed to our earlier stand – is in fact not a relevant market information, in order to collect all relevant market information, it is sufficient to follow past market occurrences.

As a result, the legal institution of commercial secrets plays absolutely no role in competitive markets. It is true though that the legal institution of commercial secrets is not beneficial, but it is also not detrimental in these markets, since commercial secrets have an insignificant effect on market competition.

Commercial secrets in oligopolies and monopolies

In an oligopoly the price can be influenced by a single market actor. The mechanism of this influence is disparate based on type of the oligopoly and the market situation. For example in an oligopoly with one dominant actor, it is most commonly the dominant actor who dictates the price, if it raises, the others will follow a bit later. As is evidenced , by the market of production factors vertical restrictions are common, such as treating prices as commercial secrets, which is due to the fact that in these market – even on the side of the buyers – the market situation is often oligopolistic (or oligopsonic). The concealment of list prices and price reductions from competitors in the case of long term contracts of huge volume enables the less effective functioning of the market price-mechanism. This also effects the stability of economic relations, and mainly – based on the conclusions of game theory – the more effective functioning of cooperation.

The transparency of prices and cost-structure would make the competitors' future market behaviour more predictable. The more transparent the functioning of an oligopoly, the more it can be expected that the competitors' reactions will be predictable. *On completely transparent markets, companies are able to concert their*

[46] It is a common situation in agriculture, and not only at times of technology change. If the high cost of strawberries induce a change in one year, then as a consequence everyone will operate with losses on the strawberry market in the following years.

practices even without agreements, which is considered parallel conduct in absence of intentional information sharing.

If companies share their future intentions, prices with each other, it leads to collusion (concerted practice), for this reason in oligopolies the law prohibits the sharing of any information regarding the future conduct, so these have to be kept secret.

If the legal institution of commercial secrets did not exist, and one company was more dominant than the others in the market, then this dominant company could invest more resources into the acquisition of commercial secrets, which would create asymmetrical informational situations to its advantage, leading to a competitive advantage against the others.

The real problem, nonetheless, is that the acquisition of commercial secrets is often a result of chance or a series of coincidences, and in absence of the legal protection of commercial secrets the companies of an oligopoly would gain competitive advantages in an unpredictable way. If *A* company hears that *B* company's competitiveness highly depends on its suppliers, and the market has high barriers to entry, then it can easily occupy the capacities of the supplier without the fear that *B* company will shortly be able to find a new supplier.

For this reason, in all the oligopolies, where one company has some kind of advantage against the others, the lack of protection of commercial secrets would accelerate the processes leading to a monopolistic market situation.

It is also clear that the protection of the commercial secrets of a monopoly company is a factor increasing barriers to entry. If these commercial secrets were not protected by law, then the entry costs will be lower, reducing monopoly prices to a certain level at which other companies could not be able to enter the market.[47]

Without the legal institution of commercial secrets, the competitive advantage of the monopolistic company with a competitive margin will increase the most. Without the legal institution of commercial secrets, oligopolies would be more transparent, which would enhance the chance of deviation from the equilibrium price, since it would be easier to predict future market behaviour.

In summary: the legal institution of commercial secrets is *inevitable* in case of oligopolies, which can counterbalance the absence of the starting conditions of a competitive market (barrier to entry, price-taking, etc.). The legal institution of

[47] This is the common case, when in a geographically well-defined market the dominant companies define – according to wording of the Supreme Court – such "imaginary prices", which are just under the price increased by entry (transfer, local knowledge, etc.) of the product that is outside the geographic market but same or substitute. If one element of the barrier to entry decreases, then it means that the company outside the geographic market becomes competitive, and can enter the geographically defined monopoly market, unless the monopoly company decreases the price. If the monopoly still remains after the decrease of entry costs, then this reduced price will still be over the competitive price. For term "imaginary price" see the Magyar Autóklub contra GVH case of the Supreme Court (Kf.II.39048/2002/13.) concerning the judicial review of the competition authority's Vj.152/2000/51. decision.

commercial secrets prevents the unfavourable market situation from developing into a worse situation, and helps reaching a *status quo*, which is characterised by the limited competition of some market actors. *The absence of the legal institution of commercial secrets in such markets would only be beneficial for companies with the most dominant market position.*

If there is a dominant company on a market, then the complete lack of the legal institution of commercial secrets would help him. *However, if law did not protect the commercial secrets of only the dominant companies, then it would cause a significant competitive disadvantage to the dominant company. In this situation, the acquisition of the commercial secret by other companies would redistribute welfare for the weaker companies, as well as decreasing the costs of market entry.*

PRICE REGULATION AND COMMERCIAL SECRETS – CONCLUSIONS

What kind of conclusions could be drawn in the light of the foregoing for a market with price regulation?

As an example, we chose the already mentioned area of electronic info-communication.

In the case when the state regulates the prices of the service providers with significant market power in electronic info-communication law, it subjects the economically dominant companies to an asymmetric regulation to the favour of other companies. Such a price regulation is often coupled with transparency, publication, accounting separation standards, without which the service providers with significant market power would handle a number of information as commercial secrets. The existence of commercial secrets is so significant in these procedures that the legal dogmatic of the regulatory procedures fundamentally differs from a traditional administrative procedure. These distinctions derive from the fact that a large number of commercial secrets are managed by the procedure (See also *Kovács* [2008*a*], 2012]).

In these regulated markets, if the cost-data on which the price regulation is based was not protected by the legal institution of commercial secrets, then for instance, in a court proceeding the service providers with significant market power would not challenge the cost-model behind the price regulation, because their cost-data could easily become public as opposed to the others not challenging the regulation.

Contrary to this, *if law mandated the publication of all cost-data* that served as a basis for the price regulation – namely that the national info-communications authority would publish its draft decision with all the evidence in full length – then *this would burden all service providers with significant market powe*r, while *also redistributing welfare among the companies without significant market power.*

In this case companies would not see any risk relating to commercial secrets in the judicial remedies, while the judicial proceedings and the information shared between parties would accelerate, making the legal remedy more effective.

In these instances the welfare redistribution is not to be criticised, because it reduces significant market power, which intensifies competition, thus creating the effect of welfare increase for consumers. In theory, there is a possibility that even less efficient service providers can enter the market,[48] but after the regulatory peak this problem is corrected by the market, if the significant market power and the related additional obligations disappear in time.

It would be especially advantageous, if the market could better control the reliability of information provided by service providers. At present authorities are not equipped with any kind of reliable control-mechanism to verify the validity of the data provided.[49] The national statute enables the authority to impose fines for providing false or misleading information, but the question is rather how the submission of false information can be detected.

Since companies with significant market power have some assumptions regarding their competitors – mainly based on their own market experience – they often have an estimate of the costs and other features of the competitors. If the provided data is public, then the market actors themselves are able to check the validity of the competitors' data, and signal if they have doubts about the reliability of the data, since in this game situation, it is in their best interest. If someone is submitting real data, then it is in his fundamental interest that others would do so.

It must be also seen that if cost-data serves as a basis for price regulation, then treating a set of cost-data related to the administrative price as a commercial secret could result in limited competitive advantage. This question is, however, more complex in the case of multi-product companies, since the cost-model calculated for one product might contain the cost-data not only for the price regulated wholesale product (service), but also for the freely priced product competing on the retail market. The data acquired in such way may be made quite transparent by an oligopoly retail market structure, leading to increasing prices even without concerted conduct.

Moreover, in an oligopoly the growing transparency might in itself reduce the intensity of competition, stimulating the emergence of parallel conduct, which could result in the consequences of intentionally concerted conduct, in a way that the *ex post* legal remedies of competition law could not be applied.

[48] Since as a result of the redistribution of welfare the new entries or the smaller actors gain advantage, it can happen in case of a sufficiently big advantage that they can operate in a competitive manner even if compared to the incumbent it is less efficient.

[49] Problems stemming from asymmetrical information based on the principle-agent theory also arise in the relation of regulatory authority and regulated service provider. These questions have a vast literature in regulation-economics. (See: *Kiss* [2007] p. 63, in detail: *Lafont–Tirole* [1991]).

It may be suggested that in the course of procedures on significant market power, there are instances when the market as a whole provides information classified as commercial secrets for the regulator, in order to decide who can be considered a service provider with significant market power. Additionally, the commercial secrets of service providers without significant market power become available. Since in the procedures on significant market power far less information is made public on service providers without significant market power (as they still constitute commercial secrets), thus the asymmetric informational advantage deriving from the asymmetric regulation is still on the side of service providers without significant market power. Additionally, treating smaller providers' data as a commercial secret is a manageable problem even in the legal remedy proceedings, because – due to the large number of smaller market actors – there are acceptable technical options to recover these in an anonymous way. Obviously, it makes no sense to create anonymous versions of the data of the large service providers with significant market power for the judicial procedure, because this data shows that it could be easily connected to the provider.

The above mentioned proposal, consequently, is not about diminishing the legal institution of commercial secrets, but rather about the classification of the data needed for the regulation of significant market power as public data. Making these data available for the public could eliminate the asymmetric informational situation, which results from the neglect of the legal institution of commercial secrets, between those initiating judicial review and those who do not.

Nevertheless, since the publication of such data could result in various consequences depending upon the oligopoly market situations on the adjacent and interconnected markets, and the enhancement of transparency on markets with few actors increases the threat of "legal collusion". The advantages and disadvantages of making data public shall be considered based on the detailed and precise analysis of types of data, in order to efficiently assess the set of data that could be made public. In case of the various marginal cost-based cost-models[50] that are currently used in the regulatory practice – based on the requirements of the European regime – the results could differ from for example the regulatory price determined through the optimal Ramsey-margin that uses the price-flexibility of demand. Our regulatory recommendation is worth considering in the former case, so in the current practice the publication of most information used for the regulatory process is viable.

When designing a regulatory regime, especially, if the law-maker intends to create a functioning system of legal remedies, it is necessary to change the system in this direction. Also EU law mandates all member states must ensure effective and substantial legal remedies.[51]

[50] Different variations of Fully distributed costs (FDC), Long-Run Average Incremental Cost (LRIC), etc.
[51] See footnote 12.

In light of EU law and the constitutional criteria mentioned in the introduction, the rules under the Code on Civil Procedure (Act III of 1952) are entirely unexplainable. The rule[52] amending section 2 of paragraph 119 of the Code on Civil Procedure and the related section 3 of paragraph 192, which entered into force on 1st January 2009, made it dependent on the statement of the person entitled to the commercial secret whether the commercial secret can be used in the judicial proceeding.

This solution questions our basic hypothesis that the right to legal remedies as a constitutional right is stronger than the right to the protection of commercial secrets. Moreover, in the recent – procedural – legislation it is apparent that there is a tendency of placing the right to the protection of commercial secrets ahead of the right to legal remedies, contrary to constitutional arguments.

This legislative tendency is questionable not only from a constitutional, but also from an EU law standpoint, since in the future, the judicial trials on price regulation cannot be conducted, because the data serving as the basis for the decisions will constitute commercial secrets. The procedures may become increasingly complex and slow in a technical sense, especially in cases, when defining the relevant market is based on the data constituting commercial secrets of hundreds of service providers. This is because in these cases hundreds of notifications must be sent out – with the signalisation of the specific character of the commercial secret. Moreover, in relation to market definition and market analysis, even one service provider withholding consent could be enough to block substantial review, since the judge has no margin of discretion, and the basic data of market definition and market analysis can only be assessed based on the submissions of all market participants. If there is only one service provider, who is withholding consent, then the differences between the aggregated data of submissions and revealable data submissions precisely show the data of the exact company, who prohibited the revelation of its commercial secrets, so the commercial secret will remain concealed even when the company – actively or by not submitting a statement – consented to the recognisability of its commercial secrets.

The new regulatory regime, however, not only eliminated implicitly the substantial judicial review of cases regarding regulated markets, but it can also make reaching a judicial decision impossible in a large number of other cases. These procedural rules result in a situation, where if an authority has used a commercial secret, but the entitled person does not consent to the revelation of the specific commercial secret, then the data cannot be used as evidence. The only loophole in the regulation is, when the plaintiff withholds consent of using his own commercial secrets as evidence, since under section 1 paragraph 164 of the Code on Civil Procedure in most cases the plaintiff is obliged to prove his case, so then he could not succeed,

[52] Paragraph 10 and Section 1 of Paragraph 32 of the XXX Act of 2008 amended the cited provisions of the Civil Procedure Act.

which is clearly not in his interest. In all the cases, however, when the person entitled to the secret is a co-defendant/co-plaintiff, the opposite party or a third person outside the procedure, it is impossible to finish the procedure, if consent is withheld. Since in a judicial proceeding the defendant, administrative authority is an equal party to the plaintiff, the commercial secrets used by the authority without consent must be excluded from evidence. Consequently, the administrative decision must be vacated based on the lack of substantial evidence.[53]

Although the judge and his assistant may access this evidence in theory, but in practice they will not, if the given evidence cannot be used in the case, because such evidence should not influence the judicial decision.

So far in the judicial practice, a similar problem only arose in case of classified data. Then the right solution[54] seemed to be that the judge can request access from the person entitled to the secret (mostly from the National Security Authority, NSA), but the parties[55] are not allowed to access the data during the trial, and the judgement – just like the administrative decision – cannot contain a substantial reasoning. This procedure, however, could not be considered as effective and substantial judicial review. The only legal guarantee against the authority available to the client is that the final decision in the case was not made by a public servant of the secret service, but an independent judge had – at least a formalistic – oversight over the "decision" of the secret service. In case of a clear abuse of discretionary powers, there is the theoretical possibility to reverse the administrative decision – although without providing a reasoning. But even for this, the consent of the NSA was required, which was usually given after consultations between the leadership of the courts and the NSA. The National Security Authority as an important organisation of a democratic, rule-of-law state was aware of the criteria of the rule-of-law, and only upheld a theoretical option for refusing consent.

This kind of self-restrained behaviour is, however, not to be expected from a business company, moreover, they explicitly have – even a constitutional – right to prohibit the use of their commercial secrets.

The current rules on commercial secrets intensely interfere with the functioning of one of the pillars of rule of law, the judiciary, by not even providing a formal control like in the case of classified information, and by allowing for the exclusion of the use of commercial secrets in a trial, if the consent of the entitled person is withheld.

All this is not only makes it impossible to conduct judicial review over regulated markets, but also effects other judicial proceedings of administrative law, including

[53] At least this would follow based on the rules of formal logic. The practice would in all likelihood try to come up with a more elaborate solution.

[54] We cannot say that the judicial practice is coherent, due to the serious difficulties of handling the problem.

[55] To be precise, the plaintiff, because the defendant administrative authority is usually aware of the used state secrets.

for example competition law cases. Cases of cartel and abuse of dominant position are built on a large amounts of commercial secrets. In these cases – even if unintended by the parties – the effective judicial review might be eliminated. Additionally, it effects also civil law cases, where the decision depends on the commercial secret.

The legislative intention behind hindering the judicial review in large economic and administrative cases related to business companies is unknown, along with the question that whose interest, or lobbying resulted in the rules, which are clearly and vigorously violate the rule of law, the principle of legal certainty, and the right to effective legal remedies. The legislative report of Act XXX of 2008, which introduced the amendments, is silent about the motivations and substantive reasons behind the amendment. Consequently, it is clear that creating a nuanced regulatory regime for the commercial secrets of regulated markets, is only possible after the abolishment of these obstacles.

REFERENCES

Bobrovszky, J. [2006]: Rejtélyek és fortélyok. Hozzászólás az üzleti titok és a know-how kérdésköréhez a polgári törvénykönyv reformja kapcsán [Mysteries and Ploys. Contribution to the questions of commercial secrets and know-how in light of the reform of the civil code]. Magyar Tudomány, 11, pp. 1385–1395.

Carlton, D. W.–Perloff, J. M. [2000]: Modern Industrial Organisation. Third Edition, Pearson/Addison Wesley.

Coase, R. H. [1990]: The firm, the market, and the law. University of Chicago Press, Chicago–London.

Cooter, R.–Ulen, T. [2004]: Law and economics, 4th edition. Pearson/Addison Wesley.

Cooter, R.–Ulen, T. [2008]: Law and economics, 5th edition. Pearson/Addison Wesley, Boston.

Evans, D. S–Heckman, J. J. [1983]: Multiproduct Cost Function Estimates and Natural Monopoly Tests for the Bell System. In: *Evans, D. S.* (ed.): Breaking Up Bell: Essays on Industrial Organization and Regulation. A CERA Research Study. North-Holland, New York–Amszterdam–Oxford, pp. 253–282.

Hámori, B.–Szabó, K.–Hurta, H.–Tóth, L. [2007]: A vállalaton belüli verseny- A munkahelyi rivalizálás empirikus vizsgálata. [Competition within the company – An empirical analysis of workplace rivalry]. Vezetéstudomány, 11, pp.2–16.

Harmathy, A.–Sajó, A. (ed.) [1984]: A jog gazdasági elemzése. Válogatott tanulmányok, Gazdasági-jogi tanulmányok, 2. [Law and Economics. Selected Studies, Economic-legal Studies, Part 2]. Közgazdasági és Jogi Könyvkiadó, Budapest.

Jager, M. F. [1995]: Legal Expertise on Commercial secret Law. September http://www.theta.com/copyright/jager.htm.

Jóri, A.–Bártfai, Zs. [2005]: Vitás kérdések az adatvédelmi törvény értelmezése körül. [Controversial questions regarding the interpretation of the data protection act]. Infokommunikáció és Jog, 9, pp. 159–164.

Kiss, F. L. [2008]: Bevezetés a szabályozás gazdaságtanába [Introduction to the economics of regulation]. In: *Valentiny, Pál–Kiss, F. L.* (eds.): Verseny és szabályozás [Competition and Regulation, 2007]. MTA KTI, Budapest.

Kiss, F. L. [2009]: Belső gazdaságosság a termelési folyamatban [Internal economy of the production process]. In: *Valentiny, P.–Kiss, F. László* (eds.): Verseny és szabályozás [Competition and Regulation, 2008]. MTA KTI, Budapest.

Kitch, E. W. [1980]: The Law and Economics of Rights in Valuable Information. Journal of Legal Studies, 9, pp. 683–723.

Koppányi, Sz. [2006–2007]: A hatékony jogorvoslat terjedelme, lehetőségei és korlátai az elektronikus hírközlésben, különös tekintettel a piac szabályozásával kapcsolatos eljárásokra, 1–2 [The scope, the possibilities and limitations of the effective legal remedy in electronic info-communications, with a special focus on procedures of market regulation. Part 1–2]. Infokommunikáció és Jog, 16, (2006) pp. 199–202. and 1, (2007) pp. 19–27.

Kovács, A. Gy. [2006*a*]: A bírói döntések karakterisztikája komplex problémák esetén, az ítéleti indokolás meggyőző erejének egyes kérdései [The characteristics of judicial decisions in complex problems, questions of the persuasive power of the reasoning of judgements]. Állam és Jogtudomány, 2, pp. 269–289.

Kovács, A. Gy. [2006*b*]: Nemcsak a bíróságokon múlik… Hozzászólás Tóth András: Az elektronikus hírközlésre vonatkozó európai szabályozási keret 2006. évi felülvizsgálata: a tagállami bíróságokon múlik a szabályozási keret hatékony működése? című cikkhez. [It not only depends on courts… Contribution to András Tóth's article on the 2006 revision of the European regulatory framework of electronic info-communications: does the effectivity of the regulatory framework depends on the national courts?]. Infokommunikáció és Jog, vol. 13. Infokommunikáció és Jog, 14, pp. 128–132.

Kovács, A. Gy. [2007]: A szabályozó hatóságok eljárásjogának modellje a határozatok jogereje és végrehajthatósága tükrében [The procedural model of regulatory authorities in light of the force of law effect and enforceability]. (Manuscript)

Kovács, A. Gy. [2008*a*]: The procedural problems of SMP-regulation in the Hungarian judicial practice. The conflicts of community law and national law in electronic infocommunications law. Infokommunikáció és Jog, 2, pp. 71–77.

Kovács, A. Gy. [2008*b*]: A piacszabályozás, mint önálló jogterül [Market regulation as an independent legal field]. Európai Tükör, 7–8, pp. 12–32.

Kovács, A. Gy. [2012]: Piacszabályozás és jogorvoslat [Market regulation and legal remedy]. HVG-Orac, Budapest.

Lafont, J-J.–Tirole, J. [1991]: The Politics of Government Decision-Making: A Theory of Regulatory Capture. Quarterly Journal of Economics, Vol. 106, No. 4, pp. 1089–1127.

Lafont, J-J.–Tirole, J. [1991]: The Politics of Government Decision-Making: A Theory of Regulatory Capture. Quarterly Journal of Economics, Vol. 106, No. 4, pp. 1089–1127.

Landes, W. M.–Posner, R. A. [2003]: The Economy Structure of Intellectual Property Law, President and Fellows of Harvard College, Harvard University Press, Cambridge, Mass.

Majtényi, L. [2006]: Az információs szabadságok [The information liberties]. Complex, Budapest.

Mátyás, A. [2003]: A modern közgazdaságtan története [The History of modern economics]. Aula kiadó, Budapest.

Mátyás, A. [2004]: General analysis of the new Austrian school. Közgazdasági Szemle. Vol. 51, No. 10, pp. 936–947.

Nagy, Cs. I. [2008]: A magyar versenyjog üzletititok-szabályának néhány értelmezési kérdéséről [Some questions of interpretation of the commercial secret rules of Hungarian competition law]. Jogtudományi Közlöny, 11, pp. 553–561.

Posner. R. A. [1996]: Law and economics. In: Law and legal theory in England and America, Clarendon Press/Oxford University Press.

Sajó, András [1984]: Közgazdaságtani vizsgálódások a jogról [Economic inquiries about law]. In: *Harmathy–Sajó* [1984] pp. 5–43.

Schumpeter, J. A. [1934]: The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and Business Cycle. Harvard University Press. Cambridge MA.

Simon, H. A. [1982]: Models of Bounded Rationality. Behavioral Economics and Business Organization. MIT Press Cambridge, MA.

Spence, M. [1973]: Job market signaling, Quarterly Journal of Economics Vol. 87, No. 3, pp. 355–374.

Steiner, J.–Woods, L. [2000]: Textbook on EC law. 7th Ed., Blackstone Press, London.

Tóth, A. [2006]: Az elektronikus hírközlésre vonatkozó európai szabályozási keret 2006. évi felülvizsgálata: a tagállami bíróságokon múlik a szabályozási keret hatékony működése? [The 2006 revision of the European regulatory framework of electronic info-communications: does the effectivity of the regulatory framework depends on the national courts?]. Infokommunikáció és Jog, 13, pp. 101–105.

Vékás, L. [1998]: A magánjog gazdasági elemzése [Economic analysis of civil law. Állam és Jogtudomány]. 1–2, pp. 3–19.

Vikman, L. [2006]: Az üzleti célú információmenedzsment adatfeldolgozó eszközeiről [On the data processing devices of business information management]. Infokommunikáció és Jog, 14, pp. 8–30.

Williamson, O. E. [1981]: The modern Corporation: Origins, Evolution and Attributes. Journal of Economics Literature. Vol. 19, No. 4, pp. 1537–1568.

Williamson, O. E. [1984]: Transaction-Cost Economics: The Governance of Contractual Relations. The Journal of Law and Economics Vol. 22, No. 2, pp. 233–261

• *Balázs Muraközy–Pál Valentiny* •

ALTERNATIVES TO STATE REGULATION: SELF- AND CO-REGULATION*

In this study a brief overview is provided of the alternatives to direct governmental regulation of imperfectly competitive markets and of the evolution of the use of self-regulation in the past decades. We take into account the arguments in favour of alternative regulatory forms and compare these with their possible shortcomings. We show how the divergent features of different legal origins influence the framework of alternative regulation, including that of self-regulation. Because of the diversity of markets affected – at present or perhaps in the future – by self-regulation (from food industry through environment and lawyer services to internet, media and network services), we provide a detailed review of the literature dealing with the theoretical models of self-regulation, and attempt to categorise the various types of regulations according to their actors, origin of licences as well as type and degree of regulations.

INTRODUCTION

The past decade witnessed an upsurge of interest in the alternatives of governmental regulation of markets. This can be explained by a withering faith in the omnipotence of the modern regulating state that was established in the second half of the 20th century, intention to improve the quality of regulation, need for better governance, reduction of administrative burdens, and new solutions generated by regulatory failures. In relation to the 2008 crisis, the analyses mention the deficiencies of previous regulations and the need for strengthening governmental regulation. Some of the more in-depth studies call attention to the fact that the coexistence of various modes of regulations dates back to a longer period, their relative weight changed a number of times even in the past century, and alternative regulatory measures often complement each other (*Bartle–Vass* [2005]). This is the approach we chose as well, since we believe that the activity of market actors – especially in modern economies – is regulated by differing degrees of state influence, and activities of self-regulation, co-regulation and joint regulation can be placed along this line as alternative, in some cases, supplementary solutions to direct governmental regulation.

* This paper is a shortened, edited version of an earlier one prepared with the support of a grant GVH VKK AL/1206/2011. We are grateful to *István Ilyés* for his excellent research assistance. We also acknowledge support from the grant of OTKA No. 81235.

QUESTIONS OF DEFINITION AND OPTIONS OF ALTERNATIVES

Between the two extremes of governmental regulation and no regulation, there are several options for shaping and influencing the behaviour of market actors. Among these, the most explored themes in the literature are opportunities in information provision, the operation of certain market mechanisms, self-regulation and co-regulation (*BRTF* [2005], *Bartle–Vass* [2005], *Hepburn* [2009]). Those wanting to intervene because of negative developments in the market conditions (market failures), must consider if any type of intervention delivers greater benefits than the costs of market failures. If it does not, then there is no intervention. An extreme point among alternative choices is governmental regulation,[1] when legal means are used to create a regulatory framework and an organisation responsible for observing and enforcing compliance. As a middle ground, the desired aim is achievable with the help of certain market mechanisms (such as tax and support incentives); or perhaps information and education campaigns can be launched with the use of possible certificates, labels and emblems; the self-regulation of market actors can be trusted; or self-regulation can be developed by government incentives (co-regulation).

Most reports and studies dealing with the topic divide state interventions according to their intensity, with some variation in the number categories defined. However, questions pertaining to regulations may also be analysed from other perspectives, for instance, according to the market structure of regulated areas, formal or informal modes of regulation, root causes, objectives, or the measures of regulations. The focus of analysis may of course also differ depending on whether the purpose is to describe an existing regulatory condition, or to change it.

Alternatives beyond existing governmental regulations however show that different divisions and typologies do not neatly correspond to regulations in the real world. The diversity of regulatory processes and differences in real life scenarios demand diverse regulatory solutions, the majority of which are some combination of the versions described in typologies (*Bartle–Vass* [2005], *Coglianese–Mendelson* [2010]).

In the following, we restrict our analysis to self-regulation and joint regulation.[2] partly because even this area contains ample varieties, partly because these regulatory forms – which incorporate elements (for example, certain market mechanisms, information provision agreements) of other kinds of regulation – is the most prevalent. At the same time, it must be borne in mind that this restriction also distances our analysis from the traditional approach to discussing market and sectoral regula-

[1] Governmental regulation may include: legislation, governmental implementation of public policy, general competition regulation and sectorial regulation as well. The categorisation of rule-makers and regulations as well as the role of legal and technological rule-making are analysed in detail in the studies of Ferenc Kiss (*Kiss–Major–Valentiny* [2000], *Kiss* [2008]).

[2] Co-regulation might be called as meta-regulation (*Coglianese–Mendelson* [2010]), situated between governmental and self-regulation. It alloys of the features of both.

tions, as it inevitably addresses regulation on the supply side of the market, which is primarily driven by the need for compliance. For this reason, self-regulation was left outside the regulatory literature of classic economic theoretical framework for a long time, and only surfaced on the horizon of regulatory analyses in the past decades.

Coglianese–Mendelson [2010] worked out a useful analytical method for the separation of basic characteristics of regulation. Their method takes into account four factors of regulations: the regulated, the regulator, the regulation as a command and the consequences of regulation (command).

1. The regulated is usually a business firm but it can also include individuals, government organisations, or non-profit organisations. One of the main features of the target is that it bears the consequences of non-compliance.
2. In this relationship, the *regulator* creates the rules and enforces compliance. Traditionally, the state is regarded as the regulator but as we will see this is holds only in some cases. It is not true if, for example, regulation is devoid of government regulation, or it is an activity remote from governmental interest, or there are signs of independency from the government. In reality, the modern state exhibits at least a "passive interest" in self-regulation (*Bartle–Vass* [2005]).
3. In the *regulatory process,* commands encourage or discourage certain forms of behaviour by the regulated (target) entities. Regulation can specify not only the goals but also the means to achieve them, for example, when they direct the regulated activity into the desired direction by standards, or they can prescribe performance targets.
4. Regulatory commands can have negative and positive *consequences*. Fines and sanctions can be expected for non-compliance, and subsidies or perhaps, exemptions from restrictions for compliance. However beyond a certain magnitude of consequences the direction of negative and positive effects may no longer make sense. A massive subsidy given to firms that comply, for example, can be equal to a very serious penalty to firms that fail to comply.

This theoretical framework can also be applied to self-regulation and co-regulation. Self-regulation means regulatory conditions, whereby the regulated entity gives commands for itself and bears the consequences. Thus, in this situation the regulator and the target are in a close relationship with each other. In contrast, in co-regulation the main role is played by an external regulator, and only the remaining process phases may concur with those experienced in self-regulation. The term, compelled regulation is, therefore, also often used for this type, indicating that the regulation was initiated by an external regulator.

The origin of self-regulation is typically associated with the regulatory processes of primitive societies, where belonging to or being excluded from a group make certain behaviour more or less desirable. In these circumstances modern theories examine, among others, the free-rider problem, network effects, or for instance, the question of credibility (*Ogus–Carbonara* [2011]). If conditions are given, and there is an opportunity for self-regulation, a number of advantages can be mentioned in comparison with government regulation. As a result of proximity to the regulated, the accumulated experiences and professional knowledge can be used more efficiently, self-regulation can respond to changes more rapidly and flexibly, putting less burden (cost) on the state and the target, and finally, the markets also work more efficiently due to a higher degree of commitment and loyalty on the part of the target. Naturally, all these advantages can only be enjoyed if public interest can – beside the private one – prevail during self-regulation, anti-market endeavours can be prevented, and efficiency is strengthened by transparency and accountability.

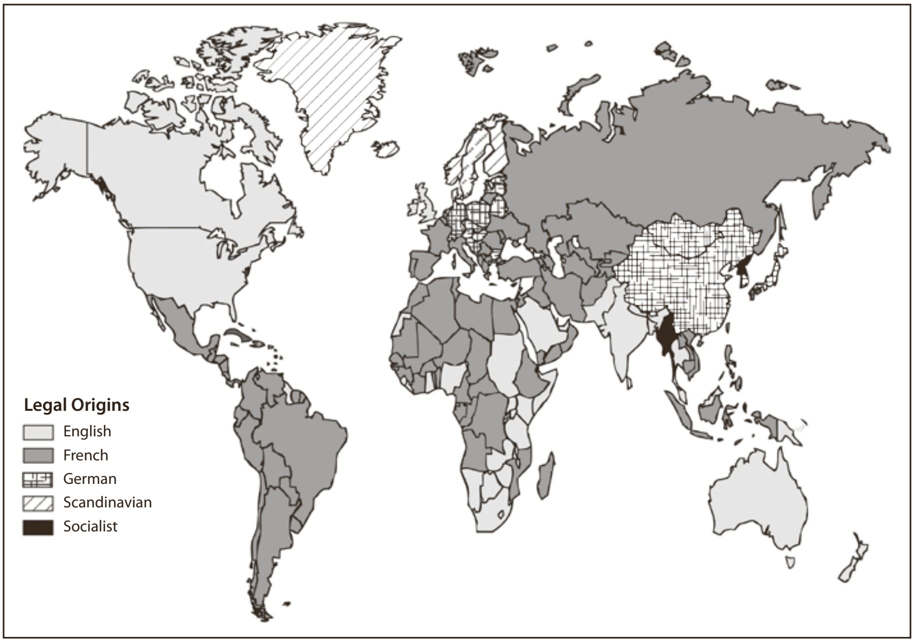
Self-regulation or co-regulation most often take place when the collection of information indispensable for regulation can be solved by them. This situation may arise, among others, in fast changing sectors or in highly complex regulatory scenarios. There are many cases when an external, governmental regulator does not even recognise the existence of the problem awaiting regulation, or if it does, then cannot see the full scale and expected effects of regulation. Regulators should be aware of the weight of the problem that awaits solution, the damages associated with unsolved problems, and the likelihood of damages. The difficulty of judging these issues may tip the balance in favour of self-regulation when it comes to choosing between alternative forms of regulations. After all, what matters in practice is whether the entities in self-regulation (and co-regulation) succeed in deciding in favour of the common social interests over the individual interests (*Coglianese–Mendelson* [2010]).

REGULATORY ALTERNATIVES AND LEGAL ORIGINS

The global map of regulation has significantly changed in the past two decades. Even though the regulation of competitive markets is not a new phenomenon, comprehensive legislations that regulate competition are a relatively recent development. Some literary sources estimate that there are about a hundred countries that adopted such legislation. In one available sample of seventy countries it became apparent that 60-70 per cent of the countries adopted the first modern competition laws in the past two decades. According to analyses of the relationship between legal traditions and competition rules, the differences in legal traditions are reflected in the institutional and procedural systems of the application of competition rules (*Lee* [2005]). Summarising the effects of legal origins, *La Porta et al.* [2008] also find that the differences in rule making and regulations are to a significant extent determined by legal origins. Previous colonial empires played a crucial role in the spread of different legal origins. There are five larger legal origins differentiated in the 152 examined countries: the Anglo-Saxon legal origin based on legal precedence (*common law*), 42 countries are listed here, the continental legal origin (*civil law*) including the French (84 countries), German (19 countries), and Scandinavian (5 countries) sub-system; and finally the socialist legal origin (2 countries). *Figure 1* shows the influence of each legal origin in the world.

It is worth comparing this map to the annual report of the World Bank that takes into account the most important factors of business environment (administrative burden, constraints, costs, legal certainty, predictability) and ranks countries according to the broader regulatory environment of doing business (http://www.doingbusiness.org/rankings). According to the June 2011 survey, more than half (11) of the first 20 countries most conducive for doing business belonged to the Anglo-Saxon legal system, five into the Scandinavian, and four into the German.

This picture can be further elaborated if we look at the history of public service regulations. In the past century, three countries played the most important role in the creation of sectoral regulation: the United States, Canada, and the United Kingdom. The former two because of their more than century-old regulatory traditions, the latter because of the new regulatory structure created in the 1980s



Source: La Porta et al. [2008] p. 289.

FIGURE 1 • **The influence of legal origins** that diverged from the American model and followed a European organisational framework for public services. The two regulatory frameworks also proved seminal for each other, many elements were transferred between the two, and this process was later enriched by the experiences of other countries as well. There were two other Anglo-Saxon system countries that have become front-runners in radical recreation of regulations, developing new methods and incentives, and reducing over-regulation: Australia and New Zealand. The accomplishments and failures of these countries also often feature in the literature of regulatory theory and practice, but noteworthy solutions were used in South-Africa (Anglo-Saxon), Malaysia (Anglo-Saxon) Korea (German) and Chile (French). In Europe mostly the Scandinavian countries, the Netherlands (French) and occasionally Spain (French) followed the increasingly prevalent British regulatory innovations.

While law making originating from the Continental law characterised, primarily the area of public services (the prevalence of state monopoly, centralised law making and regulation), the Anglo-Saxon legal order – where case laws are characterised by higher uncertainty – provided more opportunity for the creation of decentralised regulatory forms.

Soon, besides government regulations, other regulatory solutions also appeared in the Anglo-Saxon countries, and these served as examples for other nations. At the same time, in the use of government regulation and self-regulation, one can observe a period of varying intensity even in the Anglo-Saxon countries. There were fluctuations between both the increase and decrease of demand for regulation as well as the two regulatory forms. In the United States the progressive period is considered to be the development of the governmental regulation (the period between 1890 and 1920), while during *New Deal* new forms of regulations had been developed (*Ogus–Carbonara* [2011]). In the United States the use of self-regulatory systems has by now become a standard practice. The Federal Trade Commission of the United States recently prepared a report on the self-regulatory systems in the alcohol industry (2008) online behaviour advertising (2009), and marketing food to children (2008) (http://www.ftc.gov/reports/index.shtm).

The fluctuations were typical in Great Britain as well. In the 19th century, after the industrial revolution, a number of forms of self-regulations were established, but by the second half of the 20th century their further applicability had been questioned, especially in periods when corporate bankruptcies increased because of business management problems, and in order to ensure compliance with laws, more effective deterrents were needed than before. Despite all this, today, the 21st century is considered the renaissance of self-regulation in Great Britain.

Using the evaluation of *Bartle–Vass* [2005], it is worth further exploring the changes in self-regulation and co-regulation in the British system. The areas that, even today, exhibit various forms of self-regulation have strengthened in the 19th century: manufacturing industry, various trades (doctors, lawyers, engineers, auditors) and financial affairs. A series of laws provided opportunity for self-regulation [*Factory Act* (1833), *Medical Act* (1858), *Companies Act* (1862)] that only laid down the general framework of regulation and essentially relied on cooperation, agreement and supervision by the regulated entities themselves. Thus, in this case one can talk about co-regulation that constituted a deeply rooted and fundamental element of British regulation until the second half of the 20th century. For this period – despite the emergence of opening markets in numerous sectors, and a general trend of dismantling unnecessary regulations (deregulation) – there were also several signs of strengthening government regulation. The privatisation of public services in the 1980s and 1990s created a new regulatory environment and government regulatory system, but in the financial regulatory environment, the former self-regulatory system was also tightened [*Financial services and Markets Act* (2000)] which resulted in the incorporation of nine, previously partially self-regulatory bodies into a single government regulatory organisation. In some professions the extent of self-regulation was reduced: the profession of auditors was re-regulated in 1990, then in 2002, that of lawyers in 1990, and in both cases the former autonomy of sectors was reduced. In the field of education and health, in this period, incentives of certain market mechanism were increasingly used, but especially, for this reason the role of controls and government regulators also increased.

Among the reasons that decreased self-regulation, *Bartle–Vass* [2005] highlight the decline of trust, adaptability and the strengthening of risk-avers behaviour, which, coupled with cases of business scandals and abuses of dominant positions, compelled successive governments to introduce stronger regulations. The rearrangement between the individual types of regulations, however, was not unidirectional. Due to globalisation, contradictory processes were also under way in trans-national regulations. In the case of trans-national activities, new regulatory forms were initiated by business actors, most of which took self-regulatory or joint-regulatory forms. One of the examples is the internet, where self-regulation by the actors of the industry was later supplemented by state actors as well. The classical areas of government regulation, such as the regulation of public services, have also gradually transformed. Many believed that established regulatory mechanisms were too rigid and there was over-regulation. Even those who did not share this opinion, had increasingly admitted that there were more and more areas and submarkets in these sectors, which could be opened up to market mechanisms. To this end, *ex ante* type sectoral regulations were limited and the use of analytical tools in competition rules were adopted as well. In some cases, there were attempts to develop a regulatory framework based on self-regulation, but, for example, in the case of connection fees, the attempt of Oftel, a British telecommunication regulator, had proved to be unsuccessful.

By the second half of the 1990s, demand increased for rethinking different forms of regulations as well as developing better, more efficient regulations, which sometimes led to various conferences and the proliferation of different organisations, institutions and committees responsible for regulating regulations. Driven by a desire to improve conditions, and later, as part of its agenda, the OECD organised a number of conferences, held roundtables and prepared recommendations in the theme of regulatory reforms (*OECD* [1997*a*], [2001], [2004], [2009*a*], [2010*a*], [2012]). In Great Britain, an advisory committee (*Better Regulation Task Force, BRTF*) set up by the government in 1997 prepared recommendations for the improvement of regulation, which prescribed the reduction of direct government interventions and a more frequent use of self-regulation as an option to consider again (*BRTF* [1999], [2000], [2003], [2005]).

Between 2006 and 2008, the advisory body was renamed *Better Regulation Commission* and its powers were increased. Concerns for improvement were also enacted in legislation affecting regulatory authorities. The Energy Act (2004) obliged authorities to follow the principles of better regulation and implement good practices. Pursuant to the Communication Act (2003), regulatory authorities, besides the previously mentioned obligations, had to take into account the expected burden of regulations, and where possible, were obliged to promote self-regulation. In certain markets, the act also recommended the use of codes of conducts adopted in self-regulatory frameworks. Also, touching on also the operation of Office of Fair Trading (OFT), the Enterprise Act (2002) emphasised the importance of the prevalence of the codes of conduct, which was later clearly interpreted by the OFT as a broader applicability of self-regulation.

In 2005, the British government established the Better Regulation Executive (BRE) that coordinates the government's activities in regulatory affairs. The office is currently under the Department for Business, Innovations and Skills (BIS) and its task is to evaluate the regulatory plans of the government in two respects. First, it must be examined whether a regulatory alternative exists that could replace and bring the same result as the regulation intended by the government. If it does not exist, the introduction of new regulation can only be endorsed if it reduces red tape created by existing regulations or implies deregulation (http://www.bis.gov.uk/policies/bre/principles-of-regulation). This is also helped by the principle that every time a regulation is approved an existing one needs to be cancelled, and in the case of new regulations, their planned end data or termination must also be indicated (*BRE* [2011*a*], [2011*b*]). A recent innovation is that in order to facilitate better selections from classical regulatory opportunities (the use of market mechanisms, information and education campaigns, self- and joint regulation), the findings of behavioural economics must be taken into consideration (*Dolan et al.* [2010], *OFT* [2010]).

As well as the British government, the Australian one also made great efforts to promote a more widespread use of self-regulation. The minister responsible for consumer protection and the regulation of financial markets established an advisory body called Taskforce on Industry Self-regulation (TIS) in 1999. A report published by the advisory body in 2000 examined the cases when self-regulation could become the most and least efficient regulatory solution (*TIS* [2000]). A Consumer Competition Act enacted in 2010 dealt with the form of self-regulation set out in the codes of conduct. The Australian Competition and Consumer Protection Commission (ACCC) developed self-regulatory guidelines to facilitate compliance with the act by professional and occupational organisations as well as companies operating in consumer markets (*ACCC* [2011]). Not even the regulated markets were left out from the new wave of self-regulation. A case in point is the Australian Communications and Media Authority (ACMA). A report developed by this authority summarised the optimal conditions of efficient self-regulation and co-regulation (*ACMA* [2010]).

The development of better regulatory systems was supported by the systematic use of Regulatory Impact Analyses (RIA). A more substantial review of the quality of regulations also enabled increasingly more in-depth analyses of regulations in terms of transparency, consultation mechanisms, institutional solutions, monitoring and progress checks (*Valentiny* [2008]). This process that had started in the United States in the 1970s, spread to a number of countries by the 1980s, and afterwards the use of regulatory impact analysis was articulated by the OECD recommendations and incorporated into the practice of the European Union as well (*OECD* [1997*b*]). A recurring aspect of impact analyses is the consideration of regulatory alternatives. According to a survey by the OECD published in 2008, the consideration of regulatory alternatives was obligatory in all of the 31 examined member states, but there was no obligation to do so in writing in nine countries, at least one written analysis was required in 14 countries, and at least two in eight countries (*OECD* [2009*b*]). Data pertaining to individual countries is provided in *Table 1*.

Regular quality control of regulations and impact assessments were introduced in the European Union at end of the 1990s. Many consider the disclosure of the Mandelken report as a decisive moment in this process (*Mandelkern Group* [2001]). The report summarised the most important principles of good regulation and emphasised the importance of impact assessment and weighting of alternative instruments. Following the Mandelken report, annual reports analysing the practice of better regulation (http://ec.europa.eu/governance/better_regulation/reports_en.htm) and the use of impact assessments (http://ec.europa.eu/governance/impact/key_docs/key_docs_en.htm) were established in the European Union. The Commission thus decided that in this way it could speed up the process. As a result of improved regulation, it was expected that, on one hand – to use a new term – smart regulation would be realised in the whole policy cycle, from the design of pieces of legislation to implementation, enforcement, evaluation and revision, and on the other hand, the most affected parties would have a key role in the process (*EC* [2010]).

TABLE 1 • **Alternative regulations in OECD countries**

Country	Assessment of potentially feasible alternative instruments: Regulators are required to assess alternative instruments before adopting new regulation	It is required to be provided in a written form	This written assessment is required to include more than one alternative
Australia	always	Y	Y
Austria	always	Y	N
Belgium	occasionally	N	–
Canada	always	Y	N
Czech Republic	always	Y	N
Denmark	always	N	–
Finland	always	Y	Y
France	occasionally	–	–
Germany	always	Y	N
Greece	occasionally	Y	N
Holland	always	Y	N
Hungary	occasionally	Y	Y
Iceland	always	Y	N
Ireland	always	Y	N
Italy	occasionally	–	–
Japan	always	Y	Y
Korea	always	Y	N
Luxembourg	occasionally	N	N
Mexico	occasionally	Y	Y
New Zealand	always	Y	N
Norway	always	N	–
Poland	always	Y	N
Portugal	occasionally	N	–
Slovakia	occasionally	N	–
Spain	occasionally	N	–
Sweden	always	Y	N
Switzerland	always	Y	Y
Turkey	occasionally	Y	Y
United Kingdom	always	Y	N
United States	always	Y	Y
European Union	always	Y	N

Source: OECD [2009*b*] p. 106.

TABLE 1 • **Alternative regulations in OECD countries** (continued)

Country	Guidance on using alternative policy instruments						
		Topics addressed					
	Has been issued	Performance based regulation	Process (or management) based regulation	Co-regulation	Economic instruments	The use of quasi regulatory guidelines	Voluntary approaches
Australia	Y	Y	Y	Y	Y	Y	Y
Austria	N	–	–	–	–	–	–
Belgium	Y	Y	N	Y	Y	N	Y
Canada	Y	Y	N	N	Y	N	Y
Czech Republic	Y	N	N	Y	Y	N	Y
Denmark	Y	Y	Y	Y	Y	Y	Y
Finland	Y	Y	Y	Y	Y	Y	Y
France	Y	N	N	Y	Y	N	Y
Germany	Y	Y	Y	Y	Y	Y	Y
Greece	N	–	–	–	–	–	–
Holland	Y	Y	Y	Y	Y	Y	Y
Hungary	N	–	–	–	–	–	–
Iceland	N	–	–	–	–	–	–
Ireland	Y	Y	N	Y	Y	Y	Y
Italy	Y	N	N	Y	Y	Y	Y
Japan	Y	N	N	Y	Y	N	Y
Korea	Y	Y	Y	Y	Y	Y	Y
Luxembourg	N	–	–	–	–	–	–
Mexico	N	–	–	–	–	–	–
New Zealand	Y	Y	Y	Y	Y	Y	Y
Norway	N	–	–	–	–	–	–
Poland	Y	Y	Y	Y	Y	Y	Y
Portugal	N	–	–	–	–	–	–
Slovakia	N	–	–	–	–	–	–
Spain	N	–	–	–	–	–	–
Sweden	Y	N	N	N	Y	Y	Y
Switzerland	Y	N	Y	N	Y	Y	Y
Turkey	Y	N	N	Y	Y	Y	Y
United Kingdom	Y	Y	Y	Y	Y	Y	Y
United States	Y	Y	Y	Y	Y	Y	Y
European Union	Y	N	N	Y	Y	N	Y

Source: OECD [2009*b*] p. 106.

The search for direct alternative instruments was from the outset included in the implementation of better regulation. Earlier, the widespread use of self-regulation and co-regulation only formed an integral part in the decentralised "soft law" framework of Anglo-Saxon legal systems. Other countries, concerned about the potential erosion of government, put restraint to such initiatives. However, a White Paper published in 2001 as well as subsequent sectoral recommendations laid down the generally accepted principles of self- and co-regulation. The final push was given by an Inter-institutional Agreement on better law-making (2003) between the three institutions of the European Union (Parliament, Council and Commission). The joint resolution stipulated that in cases where the Maastricht Treaty did not require the use of legal instruments, there was also an opportunity for the use of alternative regulatory instruments. The resolution also identified cases where this could not be done, namely, the cases, where there were fundamental rights and important political questions at stake, or where laws had to be used uniformly in all member states.

This document was the first attempt to define in the framework of the European Union what was meant by self- and co-regulation. Self-regulation is perceived as an opportunity by economic actors, social partners as well as non-governmental organisations and associations to develop directives (primarily codes of conduct or sectoral agreements) for and among themselves (*Interinstitutional Agreement …* [2003] Article 22). In co-regulation, the legislative power defines an objective and empowers the above actors (economic actors, social partners, non-governmental organisations and associations) to reach that by their own means (*Ibid.* Article 22). The resolution articulated the conditions for the use of regulatory instruments, and the control of implementation of the agreement.

In most member states, self-regulation, of course, has already been at place for a long time in certain occupations and in the form of technical standards. But a more extensive, European-level self-regulation has been taking place only since the beginning of the 1990s (for example, advertising agencies, legal counsellors, restaurants, travel agencies, internet providers, hairdressers, real estate agents, etc). These were primarily concerned with training, recognition of qualifications, rules of rule-making, and hence, facilitated the flow of activities in the given professions among the member states. In the past years, self- and co-regulation have been extended by the inclusion of consumers. The number of European-level self- and co-regulations has been increased by the improvement of product information quality, the development of security-enhancing purchase conditions (payment, service provisions, maintenance, handling of complaints, etc.) and new challenges posed by electronic commerce (*EESC* [2005]).

In order to increase the scope of alternative regulatory forms and share best practices, an independent database was established, linked to the monitoring system of the internal market, which would support the analysis of self- and co-regulation.

In the evaluation in 2009, there were 108 operating and 17 completed regulatory initiatives in the database. 83 were related to the European Commission and 25 to the European Economic and Social Committee. Among these, there were 47 (5 terminated) self-regulations and 61 (12 terminated) co-regulations. The European Commission examined 78 regulations (including 17 closed cases). In 2011, the list was extended by four new regulations (http://www.eesc.europa.eu/?i=portal.en.self-and-co-regulation-enter-the-database). The database features the legal background that led to the creation of the regulation, the objective, methods of monitoring and sanctions, geographical coverage, type of financing and the Commission's opinion on the regulation.

The summaries of the database analysis provide a glimpse into self- and co-regulations in some of the more important fields of activity of the European Union (*Hoogen–Nowak* [2009]). Most regulatory initiatives arose in connection with the European Union's internal market, this was followed by regulations pertaining to enterprises and industries, and thirdly, by the energy and transport sector (*Table 2*).

TABLE 2 • **Self- and co-regulation by sectors**

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	1990–2007
Agriculture			1																1
Employment, social affairs								1			4	1	1	3		1			11
Energy, transport								3		3			3	2		2			13
Enterprise, industry	2						1		1	1		3	1	1	4	2		1	17
Environment									1	2	2			1	1				7
Fisheries and maritime affairs										1									1
Health, consumer protection												2	2	1	1	3	1		10
Information society, media									1			1		2	1		2		7
Internal market, services		1	1			1			1		2	3	4	4	6	4	4	2	33
Public administration																1	1		2
Research											1					2			3
Total	2	1	2	0	0	1	1	4	4	7	9	10	11	14	13	15	8	3	105

Source: Hoogen–Nowak [2009] p. 151.

EXAMPLES OF SELF- AND CO-REGULATION

Empirical case studies

Self-regulation often occurs when the regulated entities are "threatened" by upcoming government regulation. In these cases, those involved in self-regulation usually choose self-regulation as a preventative action. By accepting quality parameters for products and services in the framework of self-regulation, it might be possible to achieve more favourable outcomes for the industry than under the conditions of possibly stricter government regulations. But these tactics can backfire, as it is often the self-regulatory steps that draw the government's attention to an area that is not or not adequately regulated. In other cases, self-regulation is developed or changed in reaction to a shock effect. As mentioned by *Coglianese–Mendelson* [2010], among the classical examples is a relatively unsuccessful self-regulatory attempt following the Indian chemical disaster in Bhopal, and a more successful one in reaction to the nuclear accident on the Three Mile Island in the United States. Both cases have been studied extensively.

In reaction to the former case, the Chemical Manufacturers Association in the United States launched a programme called *Responsible Care* in 1988, which was shortly followed by Australia and Great Britain as well. In the framework of the programme, participants made a commitment that they would develop codes of conduct pertaining to environmental, health and safety measures. These commitments were developed by companies one by one regarding their activities and they defined how the stipulated objectives should be achieved. However, the association did not disclose to the public if its members fulfilled their commitments and no company was excluded for non-compliance or poor compliance from the association. The flow of information between the members was poor, board members could learn the names of non-complying companies only since 1996, and internal ranking lists on the compliance of companies have been created only since 2000. Participants' compliance have only been verified by a third party since 2007. Research dealing with the programme could identify few positive aspects and claimed that there was more paper work than impact on the environment. There was a study which found that companies not participating in the programme could more significantly reduce their toxic emissions than those who participated in the programme (*Coglianese–Mendelson* [2010] pp. 154–155).

The other example confirms though that self-regulation can be successful. Before the nuclear disaster on the Three Mile Island there was no need in the nuclear industry to develop safety plans on a sectoral scale. After the disaster, a report prepared by the Kemeny Comission recommended the revision of sectoral standards, the regular collection of information and the preparation of independent, third-party evaluations. According to some, the institute (*Institute of Nuclear Power Operations, INPO*) established by the leaders of the nuclear industry prevented the federal ownership of nuclear power plants. The INPO conducts regular reviews in power plants based on a list of recommendations compiled from 417 reports that have been prepared up to now. The two-week review is carried out by 20 persons and upon the completion of their task they prepare operative recommendations. During reviews a ranking is created comprising all INPO members, which stimulates sectoral actors to comply with standards to the fullest extent. The review material is, however, confidential, not even the members can have access to them. Studies examining the operation of the INPO consider the organisation as a good practice for self-regulation and add that it may be a further success factor if the self-regulatory organisation operates in a sector that is made up of closely cooperating members that are few in numbers and relatively homogeneous (*Coglianese–Mendelson* [2010] pp. 155–156.).

In the empirical analyses summarised by *Ogus–Carbonara* [2011], the advertising industry is frequently mentioned as a good example of self-regulation. It was quickly realised in this industry that the credibility of services can be increased by an emphasis on professional responsibility. In spite of this, their self-regulatory organisations only took a strong line against non-complying organisations, when the threat of governmental regulation increased. Similar conclusion were drawn about the operation of commodity exchanges.

For a long time, the cyber space (online communication systems) was considered to be a typical example of self-regulation. In the 1990s this area was characterised by self-regulation and in the past decade co-regulation has increasingly become prominent, primarily in those fields where regulatory principles needed to be coordinated with existing governmental regulatory organisations (for example, in the case of offensive content). The security of cyber space is, in general, considered to be an area where, due to the free-rider problem, market solutions are less effective. There is a need for some sort of regulation, but the global nature of the network makes it difficult to develop any feasible arrangement. There is a need for a joint application of self-regulation and international cooperation.

Empirical studies of self-regulation in various occupations have found rent-seeking opportunities in a wide range of professions such as opticians, dry cleaners, lawyers or dentists. Standards developed by self-regulatory professional organisations – for example tariffs or advertising restrictions or rules of professional ethics – often protected the interests of regulators rather than those of consumers, and prevented the use of cost reduction measures. According to *Kleiner's* [2006] calculations, in professions subject to licencing the social costs of maintenance of licencing significantly exceed expected benefits. He believes that the introduction of professional certificates demonstrating compliance would create higher competition and lower barriers to entry than licencing.

Typical areas of co-regulation are financial services, management of hazardous materials, food safety, or for example, pollution. Co-regulation aimed at the reduction of toxic waste in certain states of the United States provide adequate empirical data for the analyses of co-regulation (*Coglianese–Mendelson* [2010] pp. 157–158.). First, it was the state of Massachusetts that imposed a law to limit waste pollution by 50 per cent. In order to meet this requirement, the state compelled respective companies to prepare a waste reduction plan, but only planning was obligatory, the content could be decided by the company and the implementation was not verified. The commitment merely consisted of a reduction plan that needed to be prepared in every other year. Nevertheless, waste reduction (between 1988 and 2007) was higher – 90 per cent – than required by the law. A survey conducted three years after the law had taken force found that 81 per cent of responding companies realised at least some, but some companies did so with all measures that they had planned. 67 per cent also perceived a cost reduction effect by the introduced measures, and 86 per cent declared that they would intend to carry on with the planning practice, even if the state did not prescribe it any longer. At the same time, studies of the effects of legislation pointed out that in the given period, waste emissions were reduced by 81 per cent overall in the United States. The practice of Massachusetts, that is, the prescription of planning was followed by 13 member states. In these states together the degree of reduction was 30 per cent higher than in other states. The benefits of this type of regulation however, gradually decreased, and were only significant in the first six years. Overall, the above mentioned methods of co-regulation were considered successful, but it is assumed that in the long run, they would not remain effective.

Self-regulatory systems, however, are not able to adequately attend to their tasks in all cases. The following two examples illuminate inadequacies in self-regulation and a need for more direct state control that evolved in the areas of audits and credit rating.

The audit market

Audits were characterised by self-regulation for a long time. The professional association of audits developed the rules and standards of auditing, and enforced compliance. Self-regulation was also justified by the complexity of professional knowledge. Auditors detected it much easier if among those performing similar tasks, one of the parties went wrong in the client-contractor relationship system, or an auditor got excessively under the influence of or was potentially mislead by a client. In the case of clients with diverse activities, innovative and often changing portfolios, those with daily contact with companies were really in an advantageous position. For a while, it was considered as an advantage of self-regulation that the costs of regulation were borne by the regulated entities. As a result of regulation, the reliability of the profession increases, the service becomes more valuable, and the costs of regulation are absorbed and feature as a price increase factor.

After some time, however, financing was put in a different perspective. The financing of a self-regulating organisation within the trade questioned the independence of the regulator. In the United States, the behaviour of auditors massively undermined the trust in self-regulation when during a dispute on the independence of the self-regulatory organisation, the auditing companies considered the reduction of funding (*Pritchard–Puri* [2006]).

Cartel formation is usually also considered among the dangers of self-regulation, and many see their concerns justified as the number of largest auditing firms decreased to four (*Valentiny* [2012]). Self-regulation can also encourage the increase of professional standards to an extremely high level, which can lead, on the one hand, to the exclusion of certain companies, and on the other hand, to the artificial stimulation of demand for services. The consistency of use of sanctions, in some cases, may raise doubts, but penalties can have negative effects for the whole trade. In the regulation of audits the mode of obtaining information is not resolved: while member organisations cooperate in this respect, the cooperation of the most important party, that of the client, is usually not possible due to a conflict of interest.

In the North-American continent two parallel and in many respects different audit regulations have evolved. In the United States the supervision of audits was under the stock exchange supervisory authority (Securities and Exchange Commission), but in practice it was professional associations that were in charge of carrying out the task. In Canada the laws did not directly affect audits. The corporation law entrusted self-regulatory bodies with the development of professional standards and rules of independence. The committees of professional organisation developed the rules and the mechanisms of checks and accountability. The two countries differently reacted to the corporate crisis affecting the audits (for example, Enron, Worldcom) as well as the 2008 crises. In the United States the functioning of self-regulatory body called Financial Accounting Standards Board (FASB) has received a lot of criticism. They reacted slowly to the demand of enforcing technologically driven changes in audits, while at the same time, their rules enabled certain companies to disclose false revenue and profit data, thereby artificially increasing their share prices and credit ratings.

A United States law on the reporting and responsibility of companies and audits, which is named after the claimants as the Sarbane–Oxley Act [Corporate and Auditing Accountability and Responsibility Act (2002)], was aimed at tightening checks. A new organisation called Public Company Accounting Oversight Board (PCAOB) was established to supervise audits. The act stipulated the separation of audits and consultancy, the obligation to fully disclose risks, the exchange of auditors, and a significant improvement of internal supervision (*Romano* [2004], *Zhang* [2005]).

The assessments of PCAOB became regular and the activities of the four large auditors in 2010 were summarised in four reports. The mistakes made during audits were revealed and recommendations were put forward for their correction. Irregularities were found in 26 cases out of 57 at Deloitte, in 15 cases out of 62 at Ernst

& Young, in 12 cases of out of 54 at KPMG, in 28 cases out of 76 at PwC (*PCAOB* [2011*a*], [2011*b*], [2011*c*], [2011*d*]). The supervision prepared a separate report on the activities of auditors during the crisis, it identified areas with specifically many problems and where evaluation standards developed by the supervision were not adequately used – for example, fair value accounting, income taxes, stocks, calculation of revenues, accounting off-balance-sheet items, devaluation of goodwill, etc. However, in relation to audits as a whole, the reports did not identify serious problems that correlated with the crisis (*PCAOB* [2010]).

Canadian self-regulation reacted sensitively to the series of company failures. The professional association established an independent body for the supervision of auditors in 2002. Members of the body consisted of famous personalities of the business life and the representatives of regulatory organisations. The body oversees the standard and rule development process and keeps contact with the public. The Canadian Public Accountability Board (CPAB), a regulatory organisation established under the Enterprise Act in 2003, has dealt with the regulation of auditors of publically listed companies. Apart from the professional auditing association there are two other self-regulating audit organisations operating in Canada. After 2004, the legislation of certain provinces made it possible that following adequate qualification these organisations could also audit the publically listed companies, and thereby contribute to the dilution of high concentration. In line with changes to the enterprise act, the regulation of financial reports and auditing committees had also changed, as they made steps to increase their independence. The federal government also declared the applicability of criminal code in relation to the failures of corporate management. The Canadian reforms are less drastic than the ones in the United States, the CPAB is not directly under a government body, as opposed to the PCAOB, which is overseen by the US Securities and Exchange Comission (SEC). The Canadian oversight – in contrast to the one operating in the United States – does not compile standards, and the traits of self-regulation are still very strong in the Canadian regulation.

Studies and recommendations have been made in Europe as well to analyse and resolve the problems of audit market. A Green paper published by the European Commission in October 2010 summarised the lessons of the crisis and proposed solutions. In certain cases, the proposed solutions follow those in the Sarbane–Oxley Act, in other cases they are more radical. The primary aim of the recommendations of the European Commission is to strengthen the independence of the auditors and to "diversify" the auditing market. An important part of the recommendations is the strengthening of supervision of auditors on the national and European level.

The internal rules of audit firms are also changing. Regarding property relations, the rule which required that partners must constitute more than half of the owners is cancelled. In line with the basic principles of audit procedures the draft decree emphasises professional scepticism in a separate chapter as basic rule of conduct. In order to create a single market for compulsory audits, a European passport is introduced for the audit profession. In order to strengthen audit oversight, the investigatory rights of national supervisors are increased and their independence from professional organisations is required everywhere. The Commission recommends that the coordination of supervision matters should be undertaken by the European Securities Markets Authority (ESMA) (*Staff Working Paper* [2011] pp. 256–259.).

Credit rating market

Large and centuries-old credit rating agencies started their activities with collecting and selling information and statistics about business actors. Analysing, rating and classification practices have evolved from these activities. Credit rating agencies – similarly to auditors and insurance companies – occupy a special status in the institutional system of business: their findings can be used by regulatory, control institutions, or even by courts, which render credit rating agencies part of a regulatory process. There was a time when credit rating agencies were described as ideal cases of self-regulation, since their products were clearly visible and their ratings were well-measurable. For this reason, good reputation is an extremely important element of their functioning and mistakes or anti-competitive behaviour can cause a lot of harm to their reputation. Their activities can virtually not be overtaken by other economic actors, the tasks requiring vast information and lots of experience cannot even be assumed – among others, due to lack of impartiality – by the state (*Sen* [2011], *Mulligan* [2009]).

The extension of the credit rating market was generated by various laws that related to banks, insurance companies, pension funds in the 1930s, and this circle had increasingly widened by the 1970s, as institutes under state control also became increasingly reliant on the services of credit rating agencies (*White* [2012]). Basically, until 1975, the only instrument for the regulation of credit rating agencies was the adoption of a handbook that contained rating principles. The American Security and Exchange Commission (SEC) decided in that year to establish a new category – the Nationally Recognized Statistical Rating Organization (NRSRO) – for the companies that provide information, and among these it immediately acknowledged the three biggest credit rating agencies. The control was, in fact, informal and relied chiefly on feedback from the market rather than standards.

In the following 25 years only four other agencies became recognised organisations, but due to mergers and bankruptcies, only three remained again by 2000. The exact conditions for inclusion in the recognised club were not disclosed by the SEC. The performance of credit agencies during the 2008 crisis was considered worse than that of auditors. In the United States the control rights of SEC were significantly strengthened by the Dodd–Frank Act intended to improve control over the financial system (2010), and in June 2012, an independent body, the Office of Credit Ratings, was established to oversee credit rating agencies. Incompatibility rules were taken much more seriously for credit rating agencies than before: those working on the ratings have been banned from participation in the selling of ratings, and several aspects of the functions of credit rating agencies have been regulated. It must be publically disclosed what kind of conditions were used in the ratings, whether the rating was a paid or non-paid one. Furthermore, historical data of earlier ratings must also be disclosed for the evaluation of rating's accuracy, and if an employee of a credit rating agency becomes later the employee of a rated company, the employee's rating activity must subsequently be checked and evaluated. (*Dodd–Frank* [2010] SEC.931–939H).

The European regulation also tried to keep pace with the problems that arose during the crisis, and create a framework for regulation by formulating directives and decrees. The European regulation on credit rating agencies, which has been in force since December 2010 (*EPC* [2009]) was amended in May 2011, after the establishment of the European Securities and Markets Authority (ESMA). In effect, the ESMA was entrusted with the oversight of the market and the details of implementing supervision have also been developed (*EPC* [2011]). Conditions for registering credit rating agencies, rules of business conduct, quality assurance, and incompatibility were also defined. The changes of rating methodology, after they are discussed with issuers and investors, must be submitted to the ESMA, which supervises implementation according to the principles developed in March 2012.

The cases of self- and co-regulation presented up to this point demonstrate that the development of this type of regulation depends on the particular legal system and economic environment as well as many other market and professional conditions. The theoretical models attempted to provide a typology for these conditions, weigh their respective prevalence, and examine their effect. In the next section we review the economic models dealing with self-regulation.

THEORETICAL RESULTS IN RELATION TO SELF-REGULATION

The most important goal of economic regulation is to correct market failures. The most frequently cited market failure is deadweight loss caused by market power.[3] Market power and the resultant excessive pricing is usually addressed in the framework of classical regulation, since the "self-regulation" of actors would at best lead to the development of cartels and thus, to more significant deadweight loss. Monopoly power does not only result in high prices: if the monopoly decides about multiple

[3] Deadweight loss may arise if a firm prices above marginal cost, hence some consumers with a higher marginal utility than marginal cost will not buy the good. This means that some socially optimal exchanges do not take place.

factors simultaneously (for example, about the price and the quality) then – depending on the elasticity of the demand function with respect to quality – it may find optimal to choose a too high or too low level of quality (*Spence* [1975]). But since this also follows from market power, classical regulation in this dimension cannot be substituted with self-regulation even under complete information.

Self-regulation can therefore be applied successfully only in those areas where the coordination of corporate decisions is in the interest of both the firms and society. A simple example for this is standardization. Fundamentally, however, the literature focuses on information asymmetries between producers and consumers. This usually arises in the cases of search goods, experience goods and credence goods (*Scarpa* [1999]). For search goods, consumers can only determine the quality of goods after paying some search costs, hence search will be higher than optimal. In the case of experience goods, consumers can get to know the quality of the product only after purchasing it, while in the case of credence goods, not even after that; this last group consists of medical, certain financial and legal services.

Market failure related to asymmetric information can take two forms: adverse selection and moral hazard; which is to say, that on the one hand, market failure may originate from consumers' inability to observe the type of individual service providers, and on the other hand, it can also stem from inability to observe how much effort service providers put into improving service quality. *Akerlof's* [1970] model demonstrates that as a result of adverse selection, the better producers are crowded out of the market, and, the allocation of products among consumers will not be optimal either. This approach is later followed by *Leland* [1979] and *Shaked–Sutton* [1981]. Nonetheless, in the more recent literature, authors primary focus on moral hazard rather than adverse selection.

Reducing the problems stemming from quality-related asymmetric information – in contrast with excessively high prices – may be in the common interest of all stakeholders, therefore, industry-wide self-regulation in these cases may represent a viable alternative to classical regulation. In case of a very strong adverse selection – when a market cannot even operate – it is clear that it is advantageous for both service providers and consumers if industrial self-regulation can control quality and thus restore the operation of the market. In the case of a more moderate degree of adverse selection, it can be similarly argued that self-regulation preventing the crowding out of high quality service providers is beneficial both for consumers and producers (*Leland* [1979]).

Information had a prominent role in the models of the 1990s and 2000s. The reputation of the industry (expected quality) is basically a public good into which firms invest a suboptimal amount, since the cost of these investments would almost exclusively be borne by them, while the benefits could be enjoyed by all companies. In these models self-regulation reduces this public good problem, leading to higher reputation and total profit at the industry level.

While these arguments demonstrate that self-regulation leads to better results than its absence, another important question is the relation of self-regulation to classical regulation. According to the literature, the advantage of self-regulation lies in the more efficient use of information, but its disadvantage is an increased probability of collusion, which can lead to deadweight loss (*OFT* [2009]). For a more profound understanding of the trade-off between classical and self-regulation one should focus more explicitly on the objective functions and constraints different actors face, which became a focus in the literature in the 2000s.

Self-regulation receives an increasingly greater role in the area of environmental protection. The reduction of pollution is in itself not in the interest of the industry, thus, self-regulation in this area is foremost motivated by preventing classical regulatory measures, such as the introduction of pollution quotas. To be able to model this, one also needs to model the political mechanisms deciding about the introduction of quotas, where the lobby efforts of parties are also influenced by the type of alternatives the opportunity of self-regulation represent for them.

In the following, we first present the classical models of self-regulation and then we move on to those models that yield a more in-depth analysis of the function of classical and self-regulations, and their relationship to the institutional system. We review the factors that, according to the literature, influence the efficiency of self-regulation.

Causes and models of self-regulation

The model-based literature of self-regulation started by an article of **Leland [1979]**. Its starting point was the problem of adverse selection presented by *Akerlof* [1970], and it examined whether there was any improvement if the regulator or the industry defined a quality threshold.

The model itself was also based on the model of Akerlof, which was structured in the following way. Consumers value quality, hence better quality pushes the demand curve outwards. The quality of produced goods by certain companies evolves exogenously, and higher-quality firms also face higher costs. The model is about an experience good, thus consumers are not able to assess the quality of the good or service before the purchase. Therefore, they are willing to pay a price corresponding to average quality, that is the expected quality of the good in the market. This leads to adverse selection: it is not worth for producers creating the best quality to enter the market, despite the willingness of consumers – under complete information – to pay the costs of a better quality product.

In this framework, *Leland* [1979] assesses whether it is possible to improve efficiency if a certain quality threshold is introduced. This means that only those producers can sell their products on the market that exceed a critical value. The social value of the threshold is that due to an increase in average quality, producers creating better quality goods return to the market; its social cost is that total quality exchanged on the market decreases. The welfare effect depends on how consumers evaluate better quality in comparison to greater quantity. The results show that it pays off to introduce such a threshold if consumers value quality highly in comparison to the cost of its production, and the elasticity of demand (with respect to quantity) is not too large.

The article analyses what quality standard will be defined by a self-regulatory organisation (SRO). Such an organisation – similarly to cartels – maximises industrial profits. The study of Leland does not address the inner workings of self-regulatory organisations: it simply assumes that they operate efficiently from the perspective of the collective interests of stakeholders. The study shows that if the unit cost function is strictly convex, and consumer demand for quality is linear or convex, then the self-regulatory organisation defines a higher than optimal quality threshold. The reason for this is that the self-regulatory organisation – similarly to other monopolies – tries to lower supply by all means at its disposal, and thereby generate monopoly profit.

Using simple tools, *Leland* [1979] also writes about the problem of moral hazard. He examines what happens when quality is endogenous. In this case, a public good problem arises. Since consumers cannot observe the quality of the product before purchase, the company investing into the quality of its product cannot access the total return on its investment, and thus, the investment will be lower than its socially optimal level. While this approach provides the idea for later models based on moral hazard, a precise modelling of moral hazard occurred only later, with the work of *Shapiro* [1986].

The study of **Shaked–Sutton [1981]** is another classic piece in the literature of self-regulation. It expands on *Leland's* [1979] model in many ways. On the one hand, it addresses consumer preferences pertaining to quality in more general terms: consumers are not only interested in general (expected) quality, but the distribution of the quality of those who work in the profession (for example, in health or legal counselling). On the other hand, it also models the labour market in detail, where the income of professional employees is determined endogenously. To this end, the authors use a certain general equilibrium model. The skills of potential workers are heterogeneous, and in equilibrium those chose, for example, the medical profession – as opposed to other professions –, who can provide better service than the quality threshold. Thus, the threshold defines the number of workers, that is, the size of profession, too.

Given their specific approach, the authors can also analyse some novel questions. They analyse all viable sizes of professions that are feasible in equilibrium. At the same time, they also examine the effect of the emergence of a new profession that provides lower standards than the original one. They find in the case of a single profession – similarly to *Leland* [1979] – that the threshold maximising the income of professionals is higher than the socially optimal one, meaning that if the definition of quality threshold was left to the profession, there would be fewer lawyers or doctors working.

In the case of two professions they examine cases where a new profession can appear – for example, paralegals – that allows the entrance of lower quality service providers. In the model, the technical condition for this is that paralegals should earn more than in their alternative professions.

Two regulatory settings may explain the emergence of a new profession. First, a professional organisation can be freely established. Second, the old profession – lawyers, doctors – may define the quality requirements for the new one. This is not at all unrealistic: it happens often that the representatives of the highest quality profession decide about the quality standards applicable to "lower" level professions, for example, doctors define professional requirements that must be met by nurses.

If representatives of the new profession can decide about the quality requirements applicable to them, then the new profession may also set the quality threshold too high, thus state regulatory authorities may have an interest to intervene and to set the quality threshold at the socially optimal level. In such interventions the entrance of a new profession clearly increases welfare, thanks to the wider range of choice and the lower rents enjoyed by the original profession. If however the requirements applicable to the new profession are chosen by the old profession, and financial transfers are possible between the two groups, then the quality threshold of the new profession will not be optimal, and the representatives of the original profession will take further rents from the generated revenue.

Thus, overall, this more general model of *Shaked–Sutton* [1981] confirms *Leland's* [1979] conclusions, according to which a profession functioning as a monopoly sets too high quality threshold. An important finding is that the appearance of a competing profession may be beneficial in the case of independent professions. If, however, the representatives of the old profession decide about the requirements of the new profession, then this leads to an increase of rents for the old profession.

Shapiro's [1986] model is the first important model that interprets quality regulation as a moral hazard problem. Originally, the group of producers is homogeneous, and it is up to their members to decide what qualifications they should obtain and what quality products they create (high or low). For more qualified producers it costs less to create high quality products: higher qualifications and higher-quality products therefore complement each other. Another important feature of the moral hazard-based model is that the state is not able to directly regulate the quality of the product, only one of its inputs: the qualification of the service provider.

It is important that there is opportunity in the model to develop reputation. The type of products created by service providers can only be observed after a while:

consumers are not able to ascertain the quality of services provided by young service providers, but they decide about the use of services offered by providers in their second career phase based on the provider's reputation in the first phase.

Without government intervention two types of efficiency losses arise on the market. *1.* Due to the initial phase characterised by moral hazard, there is less incentive to produce higher quality, than in the case with complete information, which reduces average quality. *2.* In the initial phase the allocation of higher and lower quality products is not optimal: the higher quality products do not necessarily reach those consumers who value quality, since the products are indistinguishable prior to consumption. This also implies that in comparison to the full information scenario, asymmetric information harms the situation of those consumers who value quality, and their surplus gets transferred to other consumers.

The state can intervene into market processes in various ways. One option is licencing, which means that the performance of an activity is only permitted above a certain qualification; this is basically input regulation. In this way, low-quality service providers obtain higher qualification than they would otherwise do in the absence of intervention. In effect, the marginal cost of higher quality is reduced, and supply is increased. Licencing thus increases average quality and decreases type 1 sources of efficiency loss.

Shapiro [1986] shows that licencing only leads to an increase in welfare if the reputation mechanism is not too strong. The introduction of licensing, however, does not lead to improvement in the Pareto-sense: due to a reduced marginal cost of quality, consumers with a high valuation of quality are the winners of intervention, while those with a low valuation of quality become the losers.

The second option of government intervention is issuing certificates by which the state – already at the beginning of the career – certifies the qualification of a service provider, and thus consumers get information about the properties of the service provider already in the first phase. This provides opportunity for signalling: service providers can signal their qualifications and through this, indirectly, the quality of their service. If there is sufficiently strong correlation between the qualification and the quality, then this mechanism can fully re-establish the social optimum. In other cases it can happen that high quality service producers need extremely high incentives to reveal their type. This excessive signalling can even lead to welfare loss.

The article of *Shapiro* [1986] is significant because it is the first one to present how the quality regulation of inputs can help reduce moral hazard related to quality. An important innovative element of the article is an emphasis on the role of reputation. If the reputation mechanism is strong and efficient in a profession, then this may in itself be enough to do away with moral hazard. Imperfect reputation implies, though, that producers can get only a part of social return from investment into a higher quality, and hence investment is suboptimal. In these models this latter effect represents the rationale of regulatory or self-regulatory intervention.

In Shapiro's model the issue is the individual reputation of the given service providers, and it is a problem that this can only be observed after a while. The study demonstrates that the quality threshold is advantageous from the perspective of society, but it does not deal with the question whether the collective organisation of the industry enables the creation of efficient self-regulation. This latter can also be motivated if the industry possesses a certain type of collective reputation, which is the sum of individual reputations, the average quality of the industry. This is implicitly included in the article of *Shapiro* [1986]: the average quality of young service providers can be interpreted as the reputation of the industry. But handling industrial reputation separately makes the drivers of self-regulation more explicit: if industrial reputation is a public good, then it is perfectly conceivable, that the contribution of individual service providers, from the perspective of the industry, is too low, and the establishment of a self-regulatory organisation could alleviate this public good problem.

The **Gehrig–Jost [1995]** model follows exactly the same line of thought. In their model companies operate as local monopolies, and with some probability, after a while, consumers move to a district of another service provider. Consumers who have moved do not know the service quality in the new district, therefore they can only form their expectations based on the quality provided by their previous provider. This is meant by reputation of the industry in the model: in every district, new consumers build on their experiences with other service providers. The moving of consumers of course also implies that certain companies can enjoy only a part of their investment in reputation, and thus, the investment falls behind the optimal degree from the perspective of the industry. In the model, the number of sedentary (non-moving) consumers are the source of the reputation mechanism. This is the reason why profit maximizing self-regulation can improve the quality of the product.

The main question asked by *Gehrig–Jost* [1995] is: In what cases is it expedient to choose self-regulation instead of classical govermental quality and price regulation? An important innovation of the model is that it highlights: the advantage of self-regulation is that market actors possess more information than regulators, but its disadvantage is increased market power, which can lead to a deadweight loss. The analysis demonstrates that if regulators and companies are equally informed, then from the perspective of society it is more expedient to use classical price or quantity regulation. If, however, the information available to the regulator is overly noisy, then self-regulation securing optimal quality leads to greater social welfare.

The research of **Tirole [1996]** describes a general model of collective reputation by modelling the collective reputation of an organisation (or a profession). Collective reputation is the sum of individual reputations. Collective reputation becomes an interesting question if the reputation of individuals is not only influenced by their own but also by their organisation's reputation. For example, if costumers conclude a contract with a representative (agent) of the industry, they cannot exactly check how many times this agent cheated in the past, but can know the general reputation of the industry, and with some probability can also find out if the given individual did not behave correctly in the past. Thus, individual and collective reputation simultaneously affect the expected payoffs of making business with a firm from the industry.

An important conclusion of the model is that it does not pay for individuals to behave correctly in companies with a bad reputation. The reason for this is that due to bad reputation, consumers are distrustful even toward those who were not caught as corrupt in the past. Therefore, members of these organisations can only get less profitable jobs even if they have never behaved corruptly. This may also lead to a situation that bad reputation prevails in such organisations where individuals from many generations work together. If a generation does not behave well, then in effect it is worth less to behave correctly for the next generation, thus the bad reputation of the organisation prevails.

While the article of *Tirole* [1996] does not directly address self-regulation, such an analysis of collective reputation demonstrates why self-regulation aimed at improving reputation might be important. His argument concerning the importance of reputation is particularly important for understanding the conditions under which self-regulatory organisations can function efficiently. Although, in subsequent works this dynamic question did not receive much attention, it still remains important.

The study of **DeMarzo–Fishman–Hagerty [2005]** examines more in-depth the issue that increased market power is the social cost of self-regulation. *Gehrig–Jost* [1995] also demonstrated this. The main innovation of the study is the consideration that for the efficient functioning of quality regulation, the regulator – be it the government or a self-regulating organisation – must perform costly investigations, and therefore, investigating every transaction cannot be efficient. For this reason the study is based on the Costly State Verification (CSV) framework proposed by *Townsend's* [1979] article.

DeMarzo et al.'s [2005] logic has been inspired by industries, such as the financial market, where consumers can only ascertain the expected return of their investment by means of costly assessments. In the model, the self-regulatory organisation clearly represents the interests of the industrial stakeholders and behaves as a monopoly in the control of service providers. This can be interpreted in a way that the self-regulatory organisation operates in the common ownership of industrial companies and its objective is not to maximize its own profit.

The model shows that service providers competing in prices can function as monopolies if the operation of industrial self-regulatory organisation endows the industry with monopoly power by regulation. Afterwards, the study examines the role of a government regulator as well. The authors show that in equilibrium the government authority does not even perform any investigations, but the threat of investigation can push the industry into the direction of perfect competition.

The main features of the classical models of self-regulation discussed here is summarised in *Table 3.*

TABLE 3 • **The main features of the classical models of self-regulation**

Study	Approach	Type of regulation	Main innovation	Efficiency of self-regulation
Leland [1979]	adverse selection	quality threshold	first model of self-regulation	too high quality threshold
Shaked–Sutton [1981]	adverse selection	quality threshold	general equilibrium, more professions	too high quality threshold
Shapiro [1986]	individual moral hazard	input regulation: licensing and certificates	emergence of moral hazard	licensing can be effective, certificates may lead to too high investment
Gehrig–Jost [1995]	industrial reputation	quality regulation	emergence of industrial reputation, the cost and benefit of self-regulation in comparison to classical regulation	if the self-regulatory organisation is more informed than the governmental regulator, self-regulation might be efficient
DeMarzo et al. [2005]	costly controls	costly controls	modelling of market forces created by self-regulatory organisations, complementary nature of self-regulatory organisations and authorities	a self-regulatory organisation controlling quality leads to cartel prices; the inclusion of authorities have a positive effect

Institutions and the functioning of the regulator

Articles written before the 2000s do not address the incentives that a self-regulatory organisation faces. It is generally assumed that enterprises establish such an organisation if they need one, and this will automatically and efficiently maximizes the industry's aims. This, however – similarly to cartels – does not happen automatically in the case of self-regulatory organisations, since it might be in the interest of such an organisation to diverge from the collective interests of the industry.

Kranton [2003] investigates this issue and points out that in the case of experience goods and repeated games there might be a need for a certain market power which makes it worth for companies to build a reputation that is associated with high quality production. To uphold high quality, there might be a need to limit entry or reduce price competition. This phenomenon can justify the notion that self-regulatory organisations should not only deal with quality control, but to some extent should also limit competition. The author demonstrates that the guilds of the Middle Ages in Europe and in the Middle East as well as modern American professional associations also functioned this way: they defined quality requirements toward professionals, and at the same time, limited competition.

In the models presented in the previous section as well as in the article of *Kranton* [2003], there was an automatic assumption that the self-regulatory organisation, as a body established by the companies of the industry, will maximise total industry profits. This, however, is not necessarily true. The self-regulatory organisation, as an entity recognised by the state and responsible for the realisation of certain social goals, can also have another objective function. In the analyses of Javier Núñez, for instance, the objective of the self-regulatory organisations is the development of its own reputation (*Núñez* [2001] and [2007]). In the models the development of reputation indicates that the self-regulatory organisation functions efficiently and can investigate firms at a low cost. While it is not clear why such an organisation would follow exactly this objective function, the analyses provide important insights on how results can differ if the self-regulatory organisation does not proceed as an agent of the companies.

An important advantage of these analyses is that the examination of self-regulatory and classical regulatory authority relationships became richer than in the approach where the self-regulatory organisation maximises profit, and the authority maximises some weighted sum of profit and consumer surplus. If there is substantial discrepancy between the objective functions of the two regulatory organisations, questions arise whether the functioning of two types of regulatory organisations substitute or complement each other.

Núñez [2001] also examines a mixed regulatory environment where there is a self-regulatory organisation parallel to a governmental regulatory authority, which also oversees quality. Both the self-regulatory organisation and the governmental regulator can perform investigations. Three scenarios are possible if a company is caught producing at too low quality: 1. the self-regulatory organisation voluntarily discloses misconducts, 2. it is the government supervision that discloses them, or 3. they will not be disclosed. The presence of the governmental regulator, on the one hand, directly reduces the optimal number of misconducts, and on the other hand, the threat of government investigation can encourage the self-regulatory organisation to conduct investigations more frequently, because this improves quality, and decreases the likelihood that a governmental investigation will reveal fraud, which would worsen the reputation of the self-regulatory organisation. This effect only applies to the number of investigations, but not to the disclosure of misconduct.

Núñez [2007] operates only one self-regulatory organisation, and examines what the effect is on efficiency if the companies can bribe the self-regulatory organisation. In the model, this means that companies producing lower quality and being caught during investigations pay money to the regulator, so that the latter does not reveal the result of the investigation, and thereby the company does not have to suffer the loss of consumer trust or the high cost of external legal sanctions. The self-regulatory organisation accepts the corruption offer if the offered amount is higher than the value of reputation gained from disclosure.

The conclusion is that the possibility of corruption – in certain cases – may increase the probability of fraud and decrease the probability of investigations. At the same time, even a corrupted self-regulatory organisation can be better than if there is no self-regulation at all, because the rent from corruption represents some level of motivation for investigating, which reduces misconduct. The effect on welfare is not clear, though, for investigations are costly.

While in the model of Núñez, the objective function of the self-regulatory organisation appears rather arbitrary, these types of objective functions can be better understood if there are more self-regulatory organisations competing with each other; in these cases it is indeed those self-regulatory organisations that can obtain higher shares which can more efficiently investigate the companies belonging to them. *Caglio–Pescatori* [2013] had built such a model that explicitly examined the functioning of competing self-regulatory organisations.

Caglio–Pescatori's [2013] model starts from and earlier model, that of *DeMarzo et al.* [2005] for securities which relies on the costly state verification framework. Their study focuses on the question that if there are multiple self-regulatory organisations present in an industry, then how competition between self-regulatory organisations affect 1. the number of investigations and compliance with contracts, and through this, 2. the broker-investor relationship as well as the participation of the investors.

The authors investigate the securities market of the United States, where there is a three-tier regulation in force. The US Securities and Exchange Commission (SEC) regulates the markets and the self-regulatory organisations (the stock exchanges) too. These stock exchanges oversee the broker-investor relationships, where brokers and broker companies are members of self-regulatory organisations. The supervision rights of self-regulatory organisations are regulated by laws. The self-regulatory organisations compete with each other for higher turnover, and therefore, it is questionable whether the competition reduces the likelihood of investigating.

Thus, the model analyses profit maximising self-regulatory organisations and stock exchanges in an explicit way. The main conclusion is that this type of competition hurts welfare, because stock exchanges gain market share with a reduced intensity of investigations (*race to the bottom*). According to the model, this negative situation would not unfold if one "monopolistic" self-regulatory organisation operated in the industry. The results did not change either, if the assumption on the heterogeneity of investors was changed, or expert investors with strategic behaviour were assumed.

Reiffen–Robe [2011] uses a similar model and examines what the difference is between the behaviour of profit-oriented self-regulatory organisations that pursue their own interests and others that seek the maximisation of total industry profits (that is, when the self-regulatory organisation is the joint property of the stakeholders of the industry). The self-regulatory organisation in joint property maximises the income of agents, while the profit-oriented self-regulatory organisation maximize the value of its shares. Since the profit-oriented self-regulatory organisation is less interested in the profit of agents, it imposes higher fines and conducts investigations more frequently to reveal misconducts. For profit-oriented self-regulatory organisations – due to more frequent controls – the introduction of innovations that reduce the unit-cost of controls also pays off better. The result is based on the logic that while a jointly-owned self-regulatory organisation will choose a minimal control level in line the participation constraint of consumers, a profit-oriented self-regulatory organisation will choose a maximal control level which is still in line with consumer participation.

The study also examines the effect of parallel functioning of the governmental regulation and the self-regulatory organisation. We have seen that the work of *DeMarzo et al.* [2005] pointed out that the frequency of investigations by a joint-property self-regulatory organisation is increased by the threat of government controls. Nonetheless, *Reiffen–Robe* [2011] draw attention to the fact that this threat does not matter for profit-oriented self-regulatory organisations, because the likelihood of them being controlled is already high enough in the absence of threats.

One can understand the functioning of governmental and industrial self-regulation better if one conceives regulation not as a one-shot decision implemented at a particular moment, but more realistically, as a process in itself. While the decision maker can lay down general rules, it is the authorities or self-regulatory organisations which have to work out their detailed implementation. Since this is about residual rule-making powers, according to *Grajzl–Murrell* [2007] a natural framework is represented by the theory of incomplete contracts elaborated in a study by *Grossman–Hart* [1986].

In the framework of **Grajzl–Murrell's [2007]** model one can endogenously examine the relationship of the governmental regulator and the self-regulatory organisation. The trade-off between the benefits and costs of self-regulation unfold similarly to the models based on incentives. The benefit of self-regulation – in contrast to central regulation – is that it is amendable with lower cost, and hence, is more flexible due to better informed stakeholders of the industry. The cost of self-regulation, on the other hand, is that industry stakeholders attach more weight to their own interests than what would be socially optimal.

The three main parameters of the model are uncertainty, the divergence between the interests of the consumers and the producers (polarisation), and the populism of the government which is represented by the weight of consumers in governmental decisions. The main results are the following. On the one hand, if uncertainty is higher, then self-regulation is more likely to be optimal from a social point of view, because in this case, flexibility has a higher value. The higher discrepancy between the interests of consumers and producers calls, however, for central regulation, since in this case, the biases associated with self-regulation are accompanied by too high social costs. Finally, government regulation is favoured if the government is more populist, because the selfishness of producers is evened out by the fact that governmental regulation would attach too great a weight to the interests of consumers.

The last point also makes it clear that the regulatory mechanism chosen by the government does not necessarily correspond to the socially optimal one. While increased uncertainty or higher polarisation affects likewise the choice made by the government, its populism increases the probability that central regulation will be created.

The authors demonstrate the empirical validity of the model by two case studies. The first one examines the difference between the Anglo-Saxon and the continental system. A number of studies confirm that self-regulation is more frequent in the Anglo-Saxon legal system.

The authors argue that in the centralised continental system the cost of subsequently changing regulations is higher both in the case of central regulation and in self-regulation. The benefit of self-regulation (higher flexibility) is, therefore, lower in the continental legal system, since given the high costs there would only be small changes anyway. On the other hand, the cost of self-regulation – the biases of self-regulators – are presumably similar in the two legal systems. Consequently, self-regulation is a more attractive opportunity in countries with Anglo-Saxon legal systems.

The authors also analyse the results with multinomial probit models, using the case of alcohol regulation. Their results confirm that there is a higher prevalence of self-regulation in countries with Anglo-Saxon legal systems than in continental countries of similar size and level of development.

The only exception in the analysis is the block of former socialist countries. Here the rate of self-regulation is particularly low, especially if their continental legal system is also taken into account. The authors explain this finding by claiming that there were strongly populist governments in power during the transition period, and inexperience in self-regulation also implied that the information advantage of self-regulation would not have been too strong either.

The other empirical example is the comparison of the progressive era of the United States and the *New Deal*. In the progressive era at the end of 19th century, the role of centralised regulation was significant, but self-regulatory institutions strengthened with the *New Deal.* The authors argue, this is explained by two factors. On the one hand, the progressive era was fundamentally characterised by stability, but after the Great Depression, uncertainty had strengthened. On the other hand, the perceived conflict between corporations and consumers was greater in the progressive era than in the *New Deal*, when exit from the Depression was a common goal. *Table 4* summarises the studies presented in this section.

TABLE 4 • **The self-regulatory and classical models in the context of institutional system**

Study	Approach	The objective function of the regulator	Main result
Kranton [2003]	dynamic game	total profit of the industry	In order to achieve high quality equilibrium, restricted competition might be needed
Núñez [2001]	dynamic game	reputation of the self-regulatory organisation	The regulator investigates too rarely and does not always disclose the results of investigations to the public; it helps if there is also a regulatory authority running in parallel
Núñez [2007]	dynamic game, opportunity for corruption	reputation of the self-regulatory organisation	If the company can corrupt the self-regulatory organisation that can reduce the probability of investigations, but a corrupt self-regulatory organisation is still better than no self-regulatory organisation at all.
Caglio–Pescatori [2013]	costly controls	profit of the self-regulatory organisation	The competition of more profit-oriented self-regulatory organisations leads to too few investigations in comparison to monopolistic self-regulatory organisations
Reiffen–Robe [2011]	costly controls	profit of the self-regulatory organisation or total profit of the industry	The frequency of investigations performed by a profit-oriented self-regulatory organisation are closer to optimal
Grajzl–Murrell [2007]	property rights theory	total profit of the industry	Higher uncertainty, lower polarisation of interests, and stronger populism of the government are in the favour of a self-regulatory organisation, in contrast to the governmental regulator

Factors influencing the efficiency of self-regulation

In this section, based on the presented literature, we summarise the factors that influence the efficiency of self-regulation.

Information asymmetry between the stakeholders of the industry and the regulatory authority • Policy materials on self-regulation primarily identify information advantage for the industry as the most important advantage of self-regulation (for example, *OFT* [2009]). Interestingly, information advantage is attributed an explicit role only in some of the articles dealing with the issue. This argument formally appears in the model of *Gehrig–Jost* [1995], where the self-regulatory organisation precisely monitors the companies' cost function, but the regulatory authority knows only the distribution thereof. The model demonstrates that when information is asymmetric, self-regulation can result in higher welfare than classical regulation.

In the model of *Grajzl-Murrel* [2007], the informational advantage of self-regulatory organisations arises in a property rights theory framework. Here, information advantage means that after the development of general regulations, a self-regulatory organisation can fine tune the regulation – in accordance with the changes in the environment – with lower expenses. The more fine-tuning is needed, that is, the more uncertainties exist concerning the exact parameters of the regulation during the creation of the original law, the higher the information advantage for the self-regulatory organisation.

Market power • The potential increase of corporate market power is often considered as the most important disadvantage of self-regulation. This problem can take two forms. One of the dangers can be that the organisation established for the cooperation of companies can facilitate collusion in terms of pricing, be it in the form of open cartels or tacit collusion. The micromodels examined in the chapter however grasp this phenomenon in a sense that if companies can jointly decide about a particular dimension of the product, then this decision in itself – without collusion on prices – significantly increases the market power of the companies.

This question received a lot of attention in the article of *Leland* [1979]. Leland demonstrates that a self-regulatory organisation maximizes its profits in a way that it stipulates a higher than socially optimal quality threshold. The same result is provided by a more general model of *Shaked–Sutton* [1981]. *Gehrig–Jost* [1995] also emphasises that the social cost of self-regulation is the increase of market power.

The model of *DeMarzo et al.* [2005] addresses mostly the problem whether it is sufficient for the development of a monopoly if industrial stakeholders, who exist within the framework of a single self-regulatory organisation, decide about quality – that is, they decide about the frequency of investigations for financial service providers included in the model.

Kranton's [2003] study approaches the problem from another perspective and points out exactly that a high quality equilibrium cannot even arise if the self-regulatory organisation founded on voluntary cooperation is not stable, and if the companies do not have adequate market power.

Andersson–Skogh [2003] reach a similar conclusion as well, and they draw some important policy conclusions. They argue that in the case of strongly experience goods such as, for example, insurance markets, the judicial enforcement of contracts can be extremely costly. Therefore, authorities do not necessarily have to step up against self-regulatory organisations even in cases when they significantly reduce competition; often it is enough to ease entry.

The relationship of the self-regulatory organisation and the governmental regulator • Most of the studies dealing with self-regulation in general regard the self-regulatory organisation and the governmental regulator as substitutes, that is, they examine under what condition it is optimal to replace one with the other. More recent research however often pose the question: To what extent is the parallel functioning of two regulators desirable? Whether, in case of parallel functioning, the advantages of both regulators can prevail, that is, the better information of industrial stakeholders can be harnessed without the increase of market power, or quite contrarily, it is the disadvantages of two solutions that prevail?

According to *DeMarzo et al.* [2005], the two types of regulators complement each other. Their model shows that in the financial markets, the threat of control by the central regulator increases the investigation activity of a jointly-owned self-regulatory organisation, because in this way, lower operation costs must be paid by the consumer. Similar result was reached in the model of *DeMarzo et al.* [2005], where the parallel functioning of regulatory authority motivates the self-regulatory organisation concerned about its reputation to perform investigations more frequently. *Reiffen–Robe* [2011] shows that this effect does not manifest itself in the case of profit-oriented self-regulatory organisations, as their frequency of investigations is already too high.

The number and type of self-regulatory organisations • It is evident from the literature that the number, internal functioning and objective function of self-regulatory organisations highly influences the efficiency of self-regulation.

Shaked–Sutton [1981] study first what effect the appearance of two "professions" has. According to the results of the study it is important to distinguish between the case when the second profession decides about its own quality threshold, and the case when the first profession determines the quality threshold for the new profession as well. In the first case, the new profession may define too high a quality threshold, while in the latter case, the quality threshold can be too low because of rent-maximisation by the original profession.

In the models of *Núñez* [2001] and [2007], the goal of a self-regulatory organisation is the improvement of its own reputation. To this end, the self-regulatory organisation may perform too few investigations, and might often not disclose the result of the investigation so as to protect the reputation of the organisation.

Reiffen–Robe [2011] compared the functioning of profit-oriented self-regulatory organisations and the ones in the joint property of service providers. The profit-oriented self-regulatory organisation investigates more frequently and introduces more innovations than the jointly-owned self-regulatory organisation, and therefore, is close to the social optimum. However, according to the results of *Caglio–Pescatori* [2012], the competition of profit-oriented self-regulatory organisations reduces the number of investigations.

SUMMARY

This paper presents a long line of regulatory alternatives, which go beyond straightforward government regulation. It demonstrates that the great variety of real-life market situations and the numerous available regulatory techniques have resulted in all kinds of regulatory solutions, most of which consist of some combination of various regulatory regimes. The discussion was restricted to various observed forms and variants of self-regulation and co-regulation especially the most prevalent ones, and those that incorporate some elements of other regulatory techniques such as the use of some market mechanisms, information provision agreements, etc.

Contrasting the use of each type of regulation against the legal systems, we saw that legislation based on the continental legal system was characterised more by centralised law-making and central regulation (case law in the Anglo-Saxon legal system entailed more uncertainty), and the Anglo-Saxon legal system provided more opportunities for the development of decentralised regulatory forms. Besides governmental regulation, the widespread use of self- and co-regulation formed an integral part in the decentralised (also called as "soft law") framework of Anglo-Saxon legal systems. These regulatory forms later served as examples for nations all over the world. At the same time, in the Anglo-Saxon countries, one can observe some temporal fluctuations in the demand for regulation as well as in the relationship between government regulation and self-regulation.

A more substantial review of the quality of regulations enabled a more in-depth analysis of regulations in terms of transparency, consultation mechanisms, institutional solutions, monitoring and progress checks. The process, which started in the '70s in the United States, have increasingly spread to other countries since the '80s, then use of impact assessment has been articulated by OECD recommendations, and they have become incorporated into the practice of the European Union as well.

By examining the various practices of self-regulation, we can establish that self-regulation often takes place in order to avoid governmental regulation and after significant shock events. In larger, more heterogeneous sectors, self-regulation is harder to apply, as it is easier for companies to evade it. Due to the costs of self-regulation, often external – economic, social, regulatory – incentives were needed to launch a regulation. The analysed cases of co-regulation suggest that regulation can be socially beneficial even if regulation defines the aims, but not the steps leading to it. For traditional regulations it is required that the aims should be clear, the effect of used means should be known, and sufficient resources should be available for monitoring and enforcement. If, however, the problem to be regulated is overly complex, and its details can hardly be known, or the objectives of the regulation are too diverse, co-regulation or self-regulation might be an appropriate choice. The feasibility of finally selected methods should not be considered in themselves, but they should be set against other viable alternatives.

The acceptance of self-regulation has especially weakened as a result of the 2008 crisis. Stiglitz cites Greenspan, who waivered in his faith in the opportunity of self-regulation and the rationalisation of market behaviour (*Stiglitz* [2009]). Nonetheless, others contend that the crisis provides an opportunity for the strengthening of self-regulation, especially in that sector which raises the most objections, and has triggered the most direct regulatory interventions, namely: the financial sector (*Omarova* [2011], *Schwartz* [2011]). According to the recommendations, there are two things that self-regulation can solve better than governmental regulation. One of them is the timely acquisition of market information, and another is the recognition and management of risks. According to Omarova, self-regulation, or specifically, co-regulation are the most appropriate methods to mitigate systemic risks. To this end, mutual self-insurance should be made compulsory for companies in the financial sector, so as to promote a sense of "common identity" between them. This system could fulfil a missing regulatory function in a complex, innovative and quickly changing industry, in accordance with – and as a complement to – existing governmental regulations. However, the main current of events following the crisis has demonstrably led to the prominence of governmental regulatory tasks, besides self-regulation, in certain areas (auditing, credit rating).

The diversity of motivations of self-regulation and the heterogeneity of institutional arrangements have led to the elaboration of various theoretical frameworks. The literature has pointed out that self-regulation can primarily function in those areas where the interests of corporations and society coincide: in this way self-regulation is not efficient in dealing with significant market power, but can help resolve asymmetric information problems. The literature investigating self-regulation demonstrates a fundamental conversion, namely, that self-regulation involves information advantage in comparison to classical regulation, but at the same time, it also gives an opportunity for companies to function in a way that may result in a deadweight loss.

The theoretical literature also makes it clear that the aim of self-regulatory organisations often diverges from those of the industry overall, and this is heavily influenced by organisational functioning, namely, by the roles companies play in the organisation, whether they are profit-oriented, and what role reputation-building takes in their aims. The literature has also examined the question when competition is beneficial between self-regulatory organisations, and when self-regulatory organisations and classical regulatory authorities complement or substitute each other.

REFERENCES

ACCC [2011]: Guidelines for developing effective voluntary industry codes of conduct. Australian Competition and Consumer Commission, http://transition.accc.gov.au/content/index.phtml/itemId/658186.

ACMA [2010]: Optimal conditions for effective self- and co-regulatory arrangements. Occasional Paper. Australian Communications and Media Authority.

Akerlof, G. [1970]: The market for "lemons": Quality Uncertainty and the Market Mechanism, Quaterly Journal of Economics, Vol. 84, No. 3, pp. 488–500.

Andersson, F.–Skogh, G. [2003]: Quality, self-regulation, and competition: the case of insurance. Insurance: Mathematics and Economics, Vol. 32, No. 2, pp. 267–280.

Bargeron, L.–Lehn, K.–Zutter, C. [2209]: Sarbanes-Oxley and Corporate Risk Taking. Journal of Accounting and Economics, Vol. 49, No. 1–2, pp. 34–52.

Bartle, I.–Vass, P. [2005]: Self-regulation and the regulatory state – a survey of policy and practice. CRI Research Report No. 17.

Becker, G. S. [1983]: A Theory of Competition Among Pressure Groups for Political Influence. The Quarterly Journal of Economics, Vol. 98, No. 3, pp. 371–400.

Brau, R.–Carraro, C. [2011]: The design of voluntary agreements in oligopolistic markets, Journal of Regulatory Economics, Vol. 39, No. 2, pp.111-142.

BRE [2011*a*]: One-in, One-out: Statement of New Regulation. HM Government.

BRE [2011*b*]: Sunsetting Regulations: Guidance. HM Government.

BRTF [1999]: Self-regulation. Interim Report. Better Regulation Task Force.

BRTF [2000]: Alternatives to State Regulation. Better Regulation Task Force.

BRTF [2003]: Imaginative hinking for Better Regulation. Better Regulation Task Force.

BRTF [2005]: Routes to Better Regulation. A guide to alternatives to classic regulation. Better Regulation Task Force.

Caglio, C.–Pescatori, A. [2013]: Competition among Exchanges and Enforcement Policy. IMF Working Paper, WP/13/37. http://www.imf.org/external/pubs/cat/longres.aspx?sk=40296.

Calcott, P. [2010]: Mandated self-regulation: the danger of cosmetic compliance, Journal of Regulatory Economics, Vol. 38, No. 2, pp. 167–179.

Coglianese, C.–Mendelson, E. [2010]: Meta-regulation and self-regulation. In: *Baldwin, R.–Cave, M.–Lodge, M.* (ed.): The Oxford Handbook of Regulation. Oxford University Press.

Cunningham, L. A. [2006]: Too Big to Fail: Moral Hazard in Auditing and the Need to Restructure the Industry Before it Unravels. Columbia Law Review, Vol. 106. Boston College Law School Research Paper, No. 108. http://ssrn.com/abstract=928482.

DeMarzo, P. M.–Fishman, M. J.–Hagerty, K. M. [2005]: Self-Regulation and Government Oversight. Review of Economic Studies, Vol. 72, No. 3, pp. 687–706.

Dodd–Frank... [2010]: Dodd–Frank Wall Street Reform and Consumer Protection Act. Public Law 111–203–July 21. 124 Stat. 1376. http://goo.gl/Lawh.

Dolan, P.–Hallsworth, M.–Halpern, D.–King, D.–Vlaev, I. [2010]: MINDSPACE: Influencing behaviour through public policy. Cabinet Office, Institute for Government.

Dyck, A.–Morse, A.–Zingales, L. [2010]: Who Blows the Whistle on Corporate Fraud? Journal of Finance, Vol. 65, No. 6, pp. 2213–2253.

EC [2010]: Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions. Smart Regulation in the European Union. COM(2010) 543 final, Brussels, 8.October 2010.

EC [2011*a*]: Proposal for the Directive of the European Parliament and of the Council amending Directive 2009/65/EC on the coordination of laws, regulations, and administrative provisions relating to undertakings of collective investment in transferable securities (UCITS) and Directive 2011/61/EU on Alternative Investment Funds Managers in respect of the excessive reliance on credit ratings. COM(2011) 746 final, Brussels, 15 November 2011.

EC [2011*b*]: Proposal for a regulation of the European Parliament and the Council amending Regulations (EC) No 1060/2009/EK on credit rating agencies. COM(2011) 747 final, Brussels, 15 November 2011.

EC [2011*c*]: Proposal for a directive of the European Parliament and the Council amending Directive 2006/43/EC on statutory audits of annual accounts and consolidated accounts. COM(2011) 778 final.

EC [2011*d*]: Proposal for a regulation of the European Parliament and the Council on specific requirements regarding statutory audit of public-interest entities. COM(2011) 779 final, Brussels, 30 November 2011.

Edwards, F. R. [2004]: U.S. Corporate Governance: What Went Wrong and Can It Be Fixed? In: *Borio, C.–Hunter, W. C.–Kaufman, G. G.–Tsatsaronis, K.* (eds.): Market Discipline Across Countries and Industries. MIT Press, pp. 237–255.

EESC [2005]: The Current State of Co-Regulation and Self-Regulation in the Single Market. European Economic and Social Committee, Brussels.

EPC [2006]: Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts, amending Council Directives 78/660/EEC and 83/349/EEC and repealing Council Directive 84/253/EEC (Text with EEA relevance) (OJ L 157., 9 June, p. 87).

EPC [2009]: Regulation (EC) No 1060/2009 of the European Parliament and of the Council of 16 September 2009 on credit rating agencies. OJ L 302. 17 November 2009.

EPC [2011]: Regulation (EC) No 513/2011 of the European Parliament and of the Council on the modification of regulation (EC) No 1060/2009. OJ L 145/30. 31. May 31. 2011.

Fleckinger, P.–Glachant, M. [2011]: Negotiating a voluntary agreement when firms self-regulate. Journal of Environmental Economics and Management, Vol. 62, No. 1, pp. 41–52.

GAO [2003]: Public Accounting Firms. Mandated Study on Consolidation and Competition. United States General Accounting Office Report to the Senate Committee on Banking, Housing, and Urban Affairs and the House Committee on Financial Services. GAO-03-864. http://www.gao.gov/new.items/d03864.pdf.

Gehrig, T.–Jost, P. [1995]: Quacks, Lemons, and Self Regulation: a Welfare Analysis. Journal of Regulatory Economics, Vol. 7, No. 3, pp. 309–325.

Glachant, M. [2005]: Voluntary agreements under endogenous legislative threats and imperfect enforcement. Ecole des Mines de Paris, manuscript.

Glachant, M. [2007]: Non-binding voluntary agreements. Journal of Environmental Economics and Management, Vol. 54, No. 1, pp. 32–48.

Grajzl, P.–Murrell, P. [2007]: Allocating Lawmaking Powers: Self-Regulation vs Government Regulation. Journal of Comparative Economics, Vol. 35, No. 3, pp. 520–545.

Green Paper [2010]: Green Paper. Auditing policy: the lessons of the crisis. Brussels, 13 October 2010. COM(2010) 561 final.

Grossman, S.–Hart, O. [1986]: The costs and benefits of ownership: A theory of vertical and lateral integration. Journal of Political Economy, Vol. 94, No. 4, pp. 691–719.

Hepburn, G. [2009]: Alternatives to traditional regulation. Report prepared for the OECD Regulatory Policy Division.

Hoogen, T. van den–Nowak, T. [2009]: The emergence and use of self-regulation in the European decision-making process: does it make a difference? In: *Hertogh, M.–Westerman, P.* (eds.): Self-regulation and the future of the regulatory state. University of Groningen, Groningen Centre for Law and Governance.

House of Lords [2011]: Auditors: Market concentration and their role. Vol. I. Report. Select Committee on Economic Affairs, 2nd Report of Session 2010–11, HL Paper 119–I, http://www.publications.parliament.uk/pa/ld201011/ldselect/ldeconaf/119/119.pdf.

Hu, S. [2011]: Convergence of Audit and Credit Rating Practices: Going Concern Ratings. International Journal of Disclosure and Governance, Vol. 8, No. 4, pp. 323–338.

Interinstitutional Agreement… [2003]: Interinstitutional Agreement on better lawmaking. European Parliament, Council, Commission (2003/C 321/01).

KISS, F. L. [2008]: Bevezetés a szabályozás gazdaságtanába [Introduction into regulatory economics]. In: *Valentiny, P.–Kiss F., L.* (eds.): Verseny és szabályozás, 2007 [Competition and regulation, 2007]. MTA KTI, Budapest, pp. 11–95.

KISS, F.–MAJOR, I.–VALENTINY, P. [2000]: Információgazdaság és piacszabályozás. [Information economy and regulation] Akadémiai Kiadó, Budapest.

KLEINER, M. M. [2006]: Licensing Occupations: Ensuring Quality or Restricting Competition? Upjohn University Press, Kalamazoo, MI.

KRANTON, R. E. [2003]: Competition and the Incentive to Produce High Quality, Economica, Vol. 70, No. 279, pp. 385–404.

LA PORTA, R.–LOPEZ-DE-SILANES, F.–SHLEIFER, A. [2008]: The Economic Consequences of Legal Origins. Journal of Economic Literature, Vol. 46, No. 2, pp. 285–332.

LEE, C. [2005]: Legal traditions and competition policy. The Quarterly Review of Economics and Finance, Vol. 45, No. 2–3, pp. 236–257.

LELAND, H. E. [1979]: Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards. Journal of Political Economy, Vol. 87, No. 6, pp. 1328–1346.

LYON, T. P.–MAXWELL J. W. [2003]: Self-regulation, taxation and public voluntary environmental agreements. Journal of Public Economics, Vol. 87, No. 7–8, pp. 1453–1486.

LYON, T. P.–MAXWELL J. W. [2011]: Self-Regulation and Negotiated Agreements: Complements or Substitutes? EAERE Conference paper, 18th Annual Conference, Rome.

MANDELKERN GROUP [2001]: Mandelkern Group on Better Regulation. Final Report.

MAXWELL, J. W.–LYON, T. P.–HACKETT, S. C. [2000]: Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism. Journal of Law and Economics, Vol. 43, No. 2, pp. 583–617.

MULLIGAN, C. M. [2009]: From AAA to F: How the Credit Rating Agencies Failed America and What Can Be Done to Protect Investors. Boston College Law Review, Vol. 50, No. 4, pp. 1275–1305.

NÚÑEZ, J. [2001]: A model of self regulation. Economics Letters, Vol. 74, No. 1, pp. 91–97.

NÚÑEZ, J. [2007]: Can Self Regulation Work? A Story of Corruption, Impunity and Cover-Up. Journal of Regulatory Economics, Vol. 31, No. 2, pp. 209–233.

OECD [1997*a*]: The OECD Report on Regulatory Reform: Synthesis Report. OECD, Paris.

OECD [1997*b*]: Regulatory impact analysis: best practices in oecd countries. OECD, Paris.

OECD [1999]: Voluntary approaches for environmental policy in OECD countries. OECD, Paris.

OECD [2001]: Regulatory Impact Analysis: Efficiency, Accountability, and Transparency. OECD, Paris.

OECD [2004]: Ex post evaluation of regulatory tools and institutions. OECD, Paris.

OECD [2009a]: Alternatives to traditional regulation. OECD, Paris.

OECD [2009*b*]: Indicators of regulatory management systems. OECD, Paris.

OECD [2010*a*]: Reducing the Risk of Policy Failure: Challenges for Regulatory Compliance. OECD, Paris.

OECD [2010*b*]: Competition and Credit Rating Agencies. DAF/COMP[2010)29. http://www.oecd.org/dataoecd/28/51/46825342.pdf

OECD [2012]: Recommendation of the Council of the OECD on Regulatory Policy and Governance. OECD, Paris.

OFT [2009]: The economics of self-regulation in solving consumer quality issues. Office of Fair Trading.

OFT [2010]: What does Behavioural Economics mean for Competition Policy? Office of Fair Trading.

OFT [2011]: Statutory audit. Market investigation reference to the Competition Commission of the supply of statutory audit services to large companies in the UK. OFT1357MIR. http://www.oft.gov.uk/shared_oft/markets-work/oft1357MIR

Ogus, A.–Carbonara, E. [2011]: Self-regulation. In: *Parisi, F.* (eds.): Production of Legal Rules. Edward Elgar, pp. 228–252.

Omarova, S. T. [2011]: Wall Street as Community of Fate: Toward Financial Industry Self-Regulation. University of Pennsylvania Law Review, Vol. 159, pp. 411–492.

PCAOB [2010]: Report on Observations of PCAOB Inspectors Related to Audit Risk Areas Affected by the Economic Crisis. PCAOB Release, No. 2010-006. 29 September 2010. http://pcaobus.org/Inspections/Documents/4010_Report_Economic_Crisis.pdf.

PCAOB [2011*a*]: Report on 2010 Inspection of Deloitte & Touche LLP (Headquartered in New York City, New York) Issued by the Public Company Accounting Oversight Board, 7 December 2011. http://pcaobus.org/Inspections/Reports/Documents/2011_Deloitte.pdf.

PCAOB [2011*b*]: Report on 2010 Inspection of Ernst & Young LLP (Headquartered in New York City, New York) Issued by the Public Company Accounting Oversight Board, 30 November 2011. http://pcaobus.org/Inspections/Reports/Documents/2011_Ernst_Young_LLP_US.pdf.

PCAOB [2011*c*]: Report on 2010 Inspection of KPMG LLP (Headquartered in New York City, New York) Issued by the Public Company Accounting Oversight Board, 8 November 2011. http://pcaobus.org/Inspections/Reports/Documents/2011_KPMG_LLP_US.pdf.

PCAOB [2011*d*]: Report on 2010 Inspection of PricewaterhouseCoopers LLP (Headquartered in New York City, New York) Issued by the Public Company Accounting Oversight Board, 8 November 2011. http://pcaobus.org/Inspections/Reports/Documents/2011_PricewaterhouseCoopers_LLP.pdf.

Petit, N.–Neyrinck, N. [2011]: Credit Rating Agencies and Competition Law. Competition Policy International, Antitrust Chronicle, Vol. 8. 30. August 2011.

Powell, M.–Czapracka, K. [2011]: Recent EU Antitrust Investigations into Financial Services – What Is the Scope for Antitrust Intervention? Competition Policy International, Antitrust Chronicle, Vol. 7. 28. July 2011.

Pritchard, A. C.–Puri, P. [2006]: The Regulation of Public Auditing in Canada and the United States: Self-Regulation or Government Regulation? Fraser Institute Digital Publication, http://www.fraserinstitute.org/research-news/display.aspx?id=13534.

Reiffen, D.–Robe, M. [2011]: Demutualization and Customer Protection at Self-Regulatory Financial Exchanges. Journal of Future Markets, Vol. 31, No. 2, pp. 126–164.

Roe, M. J. [2004]: The Inevitable Instability of American Corporate Governance. Harvard Law and Economics Discussion Paper, No. 493. http://ssrn.com/abstract=615561.

Romano, R. [2004]: The Sarbanes-Oxley Act and the Making of Quack Corporate Governance. Yale Law School, Center for Law, Economics and Public Policy, Research Paper, No. 297. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=596101.

Scarpa, C. [1999]: The Theory of Quality Regulation and Self-Regulation, In: *Bortolotti, B.–Fiorentini, G.* (eds.): Organized Interests and Self-Regulation: An Economic Approach. Oxford University Press, pp. 236–260.

Schwartz, S. L. [2011]: Financial Industry Self-Regulation: Aspiration and Reality. University of Pennsylvania Law Review, Vol. 159, pp. 293–302.

SEC [2011]: 2011 Summary Report of Commission Staff's Examinations of Each Nationally Recognized Statistical Rating Organization. http://goo.gl/j3Oh6.

Sen, M. [2011]: Non-State Regulatory Regimes. Understanding Institutional Transformation. Springer.

Shaked, A.–Sutton, J. [1981]: The Self Regulating Profession. Review of Economic Studies, Vol. 48, No. 2, pp. 217–234.

Shapiro, C. [1986]: Investment, Moral Hazard, and Occupational Licensing, Review of Economic Studies, Vol. 53, No. 5, pp. 843–862.

Spence, M. [1975]: Monopoly, Quality and Regulation, Bell Journal of Economics, Vol. 6, No. 2, pp. 417–429.

Staff Working Paper [2011]: Commission Staff Working Paper. Impact Assessment. Accompanying the document Proposal for a Directive of the European Parliament and of the Council amending Directive 2006/43/EC on statutory audits of annual accounts and consolidated accounts; Proposal for a Regulation of the European Parliament and of the Council on specific requirements regarding statutory audit of public-interest entities. SEC (2011) 1384 final, Brussels, 30 November 2011. http://ec.europa.eu/internal_market/auditing/docs/reform/impact_assesment_en.pdf.

Stiglitz, J. [2009]: Regulation and Failure. In: Moss, D.–Cisternino, J. (eds.): New Perspectives on Regulation. The Tobin Project, Cambridge, MA.

Stoltenberg, C.–Lacey, K. A.–Crutchfield, G. B.–Cuthbert, M. [2005]: A Comparative Analysis of Post-Sarbanes-Oxley Corporate Governance Developments in the US and European Union: The Impact of Tensions Created by Extraterritorial Application of Section 404. The American Journal of Comparative Law, Vol. 53, No. 2, pp. 457–491.

Tirole, J. [1996]: A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality). Review of Economic Studies, Vol. 63, No. 1, pp. 1–22.

TIS [2000]: Industry Self-Regulation in Consumer Markets. Report prepared by the Taskforce on Industry Self-regulation. Commonwealth of Australia.

Townsend, R. [1979]: Optimal Contracts and Competitive Markets with Costly State Verification. Journal of Economic Theory, Vol. 21, No. 2, pp. 265–293.

Valentiny, Pál [2008]: Hogyan mérik az ágazati szabályozó szervezetek hatékonyságát? [How is the efficiency of regulatory authorities measured?] Közgazdasági Szemle, 55. évf. 10. sz, pp. 891–906.

Valentiny, P. [2012]: Válságban a verseny – válságban a verseny? [Competition in crisis – is competititon in crisis?] In: *Valentiny, P.–Kiss, F. L.–Nagy, Cs. I.* (eds.): Verseny és szabályozás, 2011. [Competition and regulation 2011]. MTA KRTK KTI, Budapest, pp. 102–125.

White Paper [2001]: European governance. A white paper. COM(2001) 428 final.

White, L. J. [2012]: The Credit Rating Agencies: How Did We Get Here? Where Should We Go? Competition Policy International, Antitrust Chronicle, Vol. 4. 17 April 2012.

Zhang, I. X. [2005]: Economic Consequences of the Sarbanes-Oxley Act of 2002. AEI-Brookings Joint Center for Regulatory Studies, Related Publications 05-07. http://regulation2point0.org/wp-content/uploads/downloads/2010/04/phpEG.pdf.

• *Gergely Csorba* •

EMPIRICAL INDUSTRIAL ORGANIZATION IN HUNGARY: A SURVEY OF RESEARCH AND APPLICATIONS*

This paper surveys Hungarian applications of Empirical Industrial Organization. The articles and methods are grouped primarily based on the complexity of the data used, starting with the simplest. The paper also discusses how the results can be applied in the main areas of competition policy and economic regulation, in supporting analyses of market definitions, and in evaluating market power and the effects of market behaviour.

INTRODUCTION

This paper surveys the Hungarian literature in the area of Empirical Industrial Organisation (empirical IO for short), focusing especially on the possible practical applications of the results. In this case, "Hungarian" means that I present articles analysing Hungarian markets; in practice this also means that I only survey papers written by Hungarian authors.[1] These articles use empirical IO methods to varying degrees, and this is represented in my discussion of them.

A paper is considered to fall under the cateogry of empirical IO if it not only describes a market using basic statistics, but also aims to test hypotheses, based on economic models, analysing the relationships between various variables (usually, but not always via regression analysis). It is of course not possible to exactly define the boundaries of empirical IO; based on their broader topics and methods, there are many papers that would fit the bill in labour economics (for example, *Brown et al.* [2006]), agricultural economics (e.g. *Fertő* [2009]), economic geography (e.g. *Békés and Harasztosi* [2013]) or the economics of international trade (e.g. *Békés and*

* I am especially grateful to my previous co-authors *Dávid Farkas, Gábor Kézdi* and *Gábor Koltay,* who have, over the past years, greatly influenced the ideas presented here. I would also like to thank *András Kiss, László Kóczy, Balázs Muraközy, Péter Nagy, Zoltán Pápai,* and also editors *Ferenc Kiss* and *Pál Valentiny* for their invaluable comments on a previous version of this paper. The lists of references, compiled by *Éva Bálint,* appearing annually in the book series "Verseny és szabályozás" [Competition and Regulation] were very helpful. I have attempted to mention all relevant papers since 2005, and apologise for any omissions. I gratefully acknowledge financial support from the LP-004/2010 "Lendület" [Momentum] programme.

[1] For empirical IO papers by Hungarian authors regarding non-Hungarian markets, see for example Paizs [2009] and *Koltay* [2012*a*]. I currently have no knowledge of exclusively non-Hungarian authors focusing on a Hungarian market.

Muraközy [2012]). However, these papers are not traditionally classified as empirical IO, and are therefore not discussed in this survey.

Empirical IO papers aim to reveal the relationships governing market behaviour, using empirical methods within an economic framework. They can thereby confirm or refute either various hypotheses arising in everyday or regulatory thinking, or the results of theoretical models. It is important to note, however, that the method employed by the analyst will always remain arbitrary to some extent, and will rely on simplifying assumptions and the choice of theoretical model, which can influence the results. It is impossible, even in theory, to find a universally applicable method or functional form to test; these must always be determined by the available data and the trends apparent in it, as well as the research questions. And finally, we can only trust the robustness of the results (and we can never be 100 per cent sure), if as many methods as possible point to the same conclusion.

There can be several practical applications of the results of empirical IO, in the fields of competition policy and economic regulation among others. These applications typically concern one or more of the following three main questions.

1. *Relevant market definition:* the analysis of substitutability between potentially competing products and thereby the definition of the (product or geographical) market within which firms exert effective competitive pressure on each other.
2. *Evaluation of market power:* the analysis of whether a given firm is able to maintain a price that is above the assumed (so-called effective) competitive price level.
3. *Evaluation of the effects of market behaviour:* the analysis of how the behaviour (for example, an agreement or merger) of certain firms affected or is expected to affect market outcomes, competitors and consumers.

There are serious microeconomic considerations behind each of these – both theoretically and practically – relevant questions, which are however outside the scope of this survey.[2]

The following chapter provides a short, methodological and historical review of empirical IO in general. Then, I survey empirical papers based on the kind of data they use, starting with the simplest.[3] I do this firstly because the available data greatly influences the type and depth of the research questions that can be answered, and secondly because the structure of the data essentially provides a grouping of the applicable empirical methods as well.

[2] For further details and references, I recommend *Bishop and Walker's* [2010] book, which comprehensively discusses the theoretical background, the suitable empirical methods and several competition policy applications related to these topics.

[3] From a historical perspective, the models could be presented starting from demand estimation, however, in practice the quality of the data is key; and it is especially important to establish the limitations of the simpler methods as soon as possible.

1. *Analyses based on price data only:* price data is the easiest to access, as it is often publicly available or accessible from statistical databases (like the Central Statistical Office, KSH). The papers usually apply time series analyses. It is rare that these methods alone lead to conclusive results, but they can provide useful illustrative evidence for questions relating to market definition or market power.
2. *Analyses based on price data and simple statistical indicators:* price data can be complemented with data concerning the number of firms on the market or other aggregate concentration measures, sometimes even from publicly available sources.[4] These analyses typically use reduced form cross-sectional or panel estimation methods. It is both an advantage and a disadvantage of such methods that they examine the relationship between market performance and market structure directly, without deriving it from an underlying economic model in a strict sense. These analyses typically concern market power, and especially the evaluation of market behaviour.
3. *Analyses based on price data and detailed quantity data:* data on demand or costs is in most cases only available from firms or public institutions, therefore it is rarely used for research purposes only. If such data is available, a multitude of regression estimation methods can be used, including structural models. These methods typically make it possible to build theoretically grounded models based on empirical observations, or at least test hypotheses related to them. They can be used to analyse all three types of questions.

I will be very brief in introducing the theoretical models and econometric expressions and methods used in the surveyed papers. The interested reader will find the detailed descriptions in the referenced papers themselves.

A HISTORICAL REVIEW OF EMPIRICAL INDUSTRIAL ORGANISATION

Empirical IO aims to analyse the S*tructure-Conduct-Performance* (the so-called *SCP*) paradigm. Research in this field, which began in the 1950s, initially took a rather simplified view of these relationships, assuming that market structure, and the technological and entry barriers behind it, completely determined price, as well as other variables important for consumer welfare. Consequently the first empirical papers analysed the causal relationship between some measure of concentration (like the number of firms, or an index calculated from market shares), which described market structure, and prices; typically across several industries.[5] Such analyses led

[4] Sometimes data is available on margins or similar performance measures (such as profitability or innovations), instead of prices. The analyses can be conducted in a similar way and I refer to all these measures, for the sake of simplicity, as "prices".
[5] These are sometimes called *cross-industry analyses*.

to many erronous results, mainly due to the endogeneity between the dependent and the independent variables (that is, that causality runs both ways between structure and performance), and the fact that it is difficult for a competition agency, for example, to draw practical (or policy-related) conclusions about markets or market behaviour based on comparisons between different industries.[6] Analyses took a new direction in the second half of the 1980s, and *new empirical IO* was born.[7] These empirical investigations are firmly grounded in theoretical models of industrial organisation, which use modern economic (primarily, but not exclusively game theoretical) tools to describe firms' behaviour. *New empirical IO analyses* and *structural empirical IO models* are therefore often used as synonyms, but this is misleading: *new empirical IO* is part of a wider family, since many modern empirical IO papers estimate reduced form regressions. Over the past years, confidence in the "superiority" of structural models has been shaken in several areas,[8] and these is an ongoing debate among leading experts about whether, in some cases, it is sufficient or even better to use reduced form models.[9]

New empirical IO focuses primarily on analyses within given markets, and therefore leads to clearer and more easily applicable conclusions. More specific questions also enable the researcher to control for other independent variables, which eliminates several econometric problems; this, however, requires a lot of data. The specificity of the analysed questions often reverses the usual relationship between theory and applications: new methods used in empirical IO and published as research results are often developed because new problems were encountered when analysing a given market – for example, when consulting with agencies or firms in competition policy or regulatory cases.

ANALYSES BASED ON PRICE DATA

The academic community is often sceptical of empirical analyses based solely on price data, since the researcher is usually unable to use and control for the reactions to price changes.[10] It would be wrong to completely discount price analyses, however, because often price data is all that is available to the researcher, and they

[6] *Schmalensee* [1989] provides a comprehensive overview of this.

[7] This was first discussed by *Bresnahan* [1989], while *Berry and Reiss* [2007], and *Doraszelski and Pakes* [2007] provide more recent surveys. *Davis and Garces* [2010] give a detailed discussion of empirical methods and competition policy applications, mainly with European examples.

[8] See *Weinberg* [2011] and his references about the "errors" made in the popular area of merger simulations.

[9] See *Einav and Levin* [2010] and *Nevo and Whinston* [2010] for the debate in the Journal of Economic Perspectivesben.

[10] See *Werden and Froeb* [1993] for a thorough (though perhaps too strong) critique.

can be useful, for example, in establishing stylised facts; these can form the basis for hypotheses which may be investigated with more advanced methods later.

It is worth mentioning an empirical method which typically relies on price data, even though, in itself, it seldom leads to scientifically valuable results: the correlation between the prices of two products or firms. If two products belong to the same (relevant) market, then the correlation between their prices over time is expected to be high; otherwise, an opportunity for arbitrage would arise, the customers would take advantage of it, and the relative price would return to the equilibrium.[11] The stability of the relative price can also be investigated using econometric methods, so-called stationarity tests. The analysis of the "closeness of competition" between given firms can also be illustrated using correlation analysis.[12] The main drawback of correlation analyses is that there is no fixed threshold above which correlation can be said to be high enough; furthermore, it is important to control for factors (such as common costs), which can cause false correlation. This is usually achieved by differencing the time series.

The relationship between prices at different levels of a product chain is a well-researched topic. The typical approach is to conduct a so-called price-transmission analysis, which determines the pass-through by a downstream vertical level (retail, for example) of the price changes implemented by an upstream vertical level (wholesale, to continue the example). The main idea is that in the case of perfect competition, the pass-through for costs should stand at 100 per cent, and therefore any lower value indicates market power at the lower vertical level. The possible asymmetry of price transmission can also be analysed; if there is market power present, then the retail price may respond more to an increase in wholesale prices than to an identical decrease.

These hypotheses can be tested using regressions on the differenced time series of the price changes in the following simplified form:

$$\Delta p_t = \alpha + \beta_1 \times \Delta w_t \times D^+ + \beta_2 \times \Delta w_t \times D^- + \varepsilon_t,$$

where p_t and w_t are the retail and wholesale prices at time t, and D^+ and D^- are dummy variables, taking on a value of 1 if the wholesale price increased or decreased in the given period, and 0 otherwise. The β_1 and β_2 parameters shows the level of transmission, and the hypotheses to test are H_0: $\beta_i = 1$ (perfect transmission), and H_0: $\beta_1 = \beta_2$ (symmetric transmission).

Farkas et al. [2009] test these hypotheses for the wholesale and retail prices of gasonline. The level of price transmission is 0.98 for price increases, and 0.97 for

[11] A stable relative price / sufficiently high correlation is not, in itself, enough evidence of belonging to the same relevant market, however.

[12] These methods are demonstrated for gasoline markets in Chapter 5 of *Farkas et al.* [2009].

decreases; these coefficients do not differ significantly from each other, but they do from 1. Based on these results, the hypothesis of asymmetric price transmission can definitely be rejected. The hypothesis of complete (100 per cent) price transmission can be rejected in a statistical sense, but 98 per cent transmission can effectively be considered perfect.[13] These results therefore show no market power at the retail level.

The process of price adaptation can also be described using more complex, dynamic models, where the lagged dependent and independent variables, as well as the so-called error correction factors appear on the right-hand side of the estimated equation. Such error correction models provide a more detailed picture of price relationships, and enable the speed of transmission to be measured.[14]

An econometric method is also available to measure which vertical level affects which level's prices. The so-called Granger causality test may be capable of achieving this goal, using methods of time series analysis similar to those above. However, it is important to handle the results with care when interpreting them as evidence of market power, as there is no underlying microeconomic model behind the hypotheses. *Popovics and Tóth* [2006] use this method in a detailed analysis of the Hungarian milk product chain, looking at the prices at the production, processing and retail levels, and conclude that the price at the processing level Granger-causes both the production and the retail level prices (and the reverse does not hold), which they interpret as evidence of market power at the processing level.

ANALYSES BASED ON PRICE DATA AND SIMPLE STATISTICAL INDICATORS

There are several motivations for analysing the relationships between price data and structural measures. Firstly, they address one of the main issues of the structure-conduct-performance paradigm directly – namely, how market structure determines various performance indicators. Secondly, it is an important practical consideration that such analyses can be relatively easily conducted using publicly available databases. And finally, the results of such empirical investigations can typically still be understood by a professional audience of non-economists (like lawyers and decision-makers), and their results are therefore easier to implement than those presented later.

[13] Especially considering that the price of gasoline is available at most at a precision of one decimal, while the largest weekly wholesale price change is not greater than 15 forints.

[14] The paper by *Farkas et al.* [2009] conducts a simplified form of this analysis for the case shown in the previous paragraph, but the results changed only very slightly. Such so-called vector error correction models (also referred to as ECM or VECM models) are especially widespread in agricultural economics, see for example *Bakucs and Fertő* [2009].

Price-concentration analyses

A price-concentration analysis is a widely used method, which seeks to establish the relationship between prices and the level of concentration in a given industry. If there is a significant, positive relationship between price and concentration, then the concentration measure used may be a good indicator of market power, that is, if the level of concentration in the industry is high or is increasing (due to a merger, for example), then the probability of competitive concerns arising is higher.

The simple estimated equation in a price-concentration study is typically the following:

$$price = f(concentration, controls).$$

The name of the method is somewhat restrictive in that it is not only the price that can be explained by concentration, but also the margin or other performance indicators.[15] The use of the margin is typically recommended (although of course the data does not always allow for this), firstly because the structural behavioural equations derived from theoretical IO models usually refer to the margin (competitive interactions are better represented in the margin), and secondly because certain econometric problems, like endogeneity and in the case of time series, stationarity, can be better handled.[16]

Regression analyses are typically conducted on cross-sectional databases, making use of the cross-sectional variation in levels of concentration. Therefore, data on several separate markets is required; often, geographically separated markets are good candidates.[17] Of course, if there is variation over time in the concentration measures, panel methods can also be employed; this, however, partly overlaps with a method I will discuss later in the chapter.

Looking to the explanatory variables, there is no clear-cut answer concerning the correct concentration measure to use. C_1, C_4 and the Herfindahl–Hirschmann-index (HHI) are often used in the literature.[18] The results are easier to interpret if the number of competitors on the market is used as a concentration measure:[19] the C_4 measure is insensitive to the merger of the second and third largest firm, for example, while the change in the HHI is difficult to interpret. We can achieve even more useful re-

[15] *Halpern and Muraközy* analyse in this book the relationship between Hungarian firms' R&D activity and various concentration measures using a regression methodology, and find an upside-down U-curve.

[16] A time series of prices is usually non-stationary, while a time series of margins is more often so.

[17] It is important that we observe variation in prices. For example, if supermarkets employ uniform prices in their outlets, then it doesn't help that concentration is different in various regional markets – the price-concentration analysis cannot be conducted.

[18] The C_i measure is the simple sum of the market shares of the largest i firms, while the HHI is the sum of the squares of (some type of) market shares of all the firms on the market.

[19] In this case the hypothesis to test is a negative relationship: we expect that a decrease in the number of competing firms on the market leads to a price increase.

sults if the presence of the larger competitors is coded using dummy variables, as this makes it easier to handle the possible non-linear effects of changes in concentration.[20]

Control variables are variables that also influence the price or the margin, but whose effect we wish to partial out, in order to answer the main question: how the level of concentration in itself influences the price. There are two basic types of control variables: demand and supply controls. For example, the size of a given regional market (the number of inhabitants), or its purchasing power are demand controls, while the price of main inputs (like labour or real estate), or the density of competitors in the given region are supply controls.[21]

There are two factors that can bias the estimates: possibly omitted variables, and the endogeneity of the relationship between price and concentration. Unfortunately, typically neither problem can be eliminated completely (often due to a lack of data), but this does not mean that the results are meaningless; it is worth verifying them using multiple methods, testing their robustness. One way to alleviate the endogeneity concern is to use two-step estimation, with the first step investigating the effect of demand controls on the density of competitors, and the second step estimating the effect of this density on prices or margins, using an instrumental variable approach.[22]

Farkas et al. [2009] conduct a price-concentration study for regional retail gasoline markets, estimating the relationship between the margin and the number of firms. Using various specifications, a significant negative relationship is found, however, it is not significant in an economic sense: the presence of a further competitor decreases price by 0.3-0.6 forints, less than 1 per cent of the average retail price.

Price-concentration studies can be combined with the price transmission analyses presented in the previous chapter, to investigate whether the level of concentration or the composition of firms influenced the level of price transmission.[23] *Farkas et al.* [2009], in the analysis of the gasonline market mentioned above, find no such relationship between the level or asymmetry of transmission and the number of firms; *Koltay* [2012*b*] on the other hand analyses the pricing of each network of stations separately and finds a small degree of asymmetric transmission for certain networks.

The price-concentration studies discussed above may give the impression that the data limitations can be overcome, this is however often not the case for researchers: the data may be available, but it constitutes a business secret. Typical applications in this field are so-called *bidding studies*, where the markets are the

[20] It is very likely, for example, that if the number of competing firms decreases from three to two, there is a larger effect on price than if it decreases from seven to six.

[21] The strength of competition may be different, for example, if four competitors in a given area each have one, or if they each have five outlets.

[22] This is the method employed by *Békés et al.* [2011].

[23] The hypothesis is that in a market with many participants, competition is close to perfect, and therefore pass-through is (close to) 100 per cent, but in two-firm markets, for example, pass-through may be lower, which could indicate market power.

separate auctions or tenders, and the final price is compared to the number of firms submitting bids, or dummy variables showing their presence. This method is often used to measure the strength of competition between competitors and the pressure they exert on each other on so-called bidding markets.[24]

Impact assessments

The other main method measuring the result of changes in stuctural indicators identifies this effect using not the differences between markets, but the actual changes over time within a given market. Since these changes typically relate to entries and exits, the methods are sometimes called *event studies*, or *shock analyses*. However, in a broader sense they belong to the family of *impact assessments* used in many policy areas.[25]

The estimation strategy most often used for panel data in this area is based on the so-called *difference-in-differences* (or simply *"diff-in-diff"* or DID) method. This quasi-experimental approach applies when the researcher is able to observe various units (like markets and their prices) over time, some of which were exposed to some "treatment" (like an entry or a merger), and some of which were not. Therefore the effect of the given event (treatment) can be identified from the difference between the treated and control group (controlling, of course, for other factors). A panel database enables the use of cross-sectional and time fixed effects, which diminishes the omitted variable problem as well.

A paper by *Csorba et al.* [2011] applies a difference-in-differences approach to analyse the effects of two 2007 mergers, Agip-Esso and Lukoil-Jet, on retail prices in Hungarian local gasoline markets. The paper discusses the predictions of several IO models, for example that the prices of the merging companies increase more than those of their competitors; or that the price effect is larger on markets where the merging parties are each other's competitors. The fact that the two mergers took place almost simultaneously makes the identification of the effects more difficult, however, the variance in the companies' presence on the specific local markets enables the separation and estimation of the various effects. The analysis confirms several theoretically predicted asymmetric effects, but the *ex post* price effect of the mergers is minimal, although positive (according to the results, the price effect of each merger was smaller than 1 per cent).

Such models can be used to evaluate the *ex post* welfare effects of various policy interventions (in the previous example, the merger clearance decisions), and agen-

[24] *Csorba* [2008] discusses such a study in detail for the case of a Hungarian merger.

[25] These methods are especially widespread in labour economics, for example. The "In focus" chapter of the 2012 edition of "The Hungarian Labour Market" concerned this topic only (see *Kézdi* [2012]). *Imbens and Wooldridge* [2009] provide a thorough methodological survey.

cies can use them to assess planned mergers or interventions by analysing events from the past.[26] The method is also suitable for evaluating smaller scale changes, caused by specific market players: *Horváth et al.* [2013], for example, use a difference-in-differences method to assess how the prices of flats which participated in a large energy efficiency-increasing renovation changed compared to similar flats that did not. Their results show a treatment effect of close to 10 per cent in flats belonging to the renovated building.

ANALYSES BASED ON PRICE AND QUANTITY DATA

The typical area of empirical analysis between prices and quantities is demand estimation, and especially the estimation of own and cross-price elasticities, since these have many applications in competition policy and regulation. The most well-known application is the H*ypothetical Monopolist Test* and its variants (like Critical Loss Analysis), used in relevant market definition:[27] if the own-price elasticity estimated for a product or group of products is not low enough, then a hypothetical monopolist of this product group would not be able to profitably raise prices; therefore, the relevant market should be wider. Further products should be included in the hypothetically monopolised market until the repeated demand estimation yields a sufficiently low elasticity.

It is worth noting that due to a lack of data and the difficulties of estimation the need may arise to measure consumer behaviour directly, typically using survey methods. While these methods are not usually considered standard tools of empirical IO, their results can be widely used, especially in practical applications.[28]

[26] See, for example, *Ashenfelter et al.* [2006], which discusses the probably most well-known merger (Staples–Office Depot), where these econometric methods were used and seriously debated in American courts.

[27] The test is also called the HMT-test, or the SSNIP-test. See *Muraközy* [2010], which discusses hypothesis testing in telecommunication markets, for demand estimation methods used to implement the HMT and other tests. *Bölcskei* [2010] also surveys research questions arising in relation to telecommunications markets, and presents the empirical methods developed to answer them, as well as results for various countries.

[28] *Édes et al.* [2010] looked at the substitution between fixed and mobile telephone service providers, among other methods also using elasticites, pointing out the asymmetry in the direction of substitution. *Lőrincz and Nagy* [2011] used the results of a consumer survey to analyse the components of switching costs for various telecommunications services (fixed and mobile telephony, internet), and estimated their size. *Pápai et al.* [2011] conducted a critical loss analysis to test whether the package deals offered by telecommunications companies could be considered a separate relevant market. Finally, *Szolnoki and Tóth* [2008] provide an example for energy markets. The authors estimated a function for the switching behaviour of consumers of electricity, based on a household survey, and then used it, together with other market data, to calibrate a theoretical model.

Classic demand estimation

A regression for demand estimation takes the following simplified functional form:

$$q_i = f(p_{i,} p_{-i,} X_i) + \varepsilon_{i,}$$

where q_i is the quantity demanded of product *i*, p_i is product *i*'s price, p_{-i} is the price charged by potential competitors, and the *X* matrix contains the necessary control variables (income, or other demand- or supply-side controls, for example). The equation can be estimated for several functional forms; the choice between them must be determined by the data and the assumptions of the model used. A common choice (which still, however, needs to be justified to some extent by the analyst) is to use the variables' logarithms. In this case, the parameters estimated for p_i directly provide the own-price elasticity. Observations may be available for different consumers or consumer groups (or even settlements) in a give time period (cross-sectional form), for the same consumers over time (time series form), or for a combination of both (panel form).

As already discussed in the case of price-concentration analyses, the endogeneity between the dependent and independent variables can bias the estimates. While this problem was slightly less acute for price-concentration analyses where the structural indicators on the right-hand side changed quite slowly over time, it is very important in the case of demand estimation. One way to tackle this identification problem is to use the previously mentioned instrumental variable method. However, it is not easy to find good instruments (and good data for them), and there are consequently only very few Hungarian demand estimation analyses to be found.

Nagy et al. [2012] use a well-designed stepwise method to estimate the demand (elasticity) for fixed-line telephones. The demand estimation method takes advantage of the fact that subscribers faced different prices depending on whether they were located in Magyar Telekom's or Invitel's area of service, and this price difference was exogenous, since the consumers' current demand could not have influenced the assignment of concessions 20 years prior. Using this fact, the paper first calculates the difference in demand in settlements in Magyar Telekom and Invitel areas that are otherwise similar, second, estimates the price difference between the two, and finally calculates the arc elasticity of demand using the first two results. Using cross-sectional data from 2011, the estimated elasticity is low for both residential and business customers (between –0.1 and –0.2), far from the critical elasticity. Their panel estimations yield similar results, even though this estimation is better for controlling for unobserved heterogeneity between the regional markets. Based solely on the results of the demand estimation, one can draw the conclusion that fixed-line telephone services are a distinct relevant market, and the hypothesis of fixed-mobile substitution can be rejected.[29]

[29] The authors also estimated the elasticity of demand based on a consumer survey. They expected the calculated elasticity to be a little higher (–0.5 was their best estimate), but even that result is enough to reject the hypothesis of fixed-mobile substitutability.

Kézdi and Csorba [2012] also estimate the relationship between price and quantity data, investigating consumer lock-in in the Hungarian market for personal loans. The applied method is also similar to the difference-in-differences (DID) method introduced in the previous chapter: the authors compare the demand reactions of new consumers with those of old consumers, based on the assumption that the latter, who are locked in, can be considered a treated group, while the former can be their control group.[30] The various estimation results show that the old consumers' reaction to price changes is 70-80 per cent lower than that of the new consumers, which means that even the hypothesis of total lock-in (prohibitively high switching costs) cannot be rejected.

Demand estimation based on discrete choice models

One of the limitations of classic demand estimation techniques is that they hardly make any assumptions on the structure behind the factors influencing demand. Therefore, a large number of parameters must be estimated, which severely limits the applicability of the method. For example, if one wishes to estimate a complete demand system for 10 products, then, even without the control variablies, there would be $10^2 = 100$ parameters to estimate, causing serious identification problems.

One solution to this problem is to use a discrete choice model, where the consumers' main choice is not how much of a given product to purchase, but which supplier to choose.[31] Such models use a microeconomic model of consumer choice to derive linear demand equations. During estimation, their assumptions concerning substitution patterns translate into parameter restrictions, which significantly decrease the number of parameters that need to be estimated. The most common method is to assign products to groups (high and medium quality domestic and import products, for example), and estimate a "common" cross-price elasticity for substitution between and within the groups. It is worth noting that demand estimation based on discrete choice models is not the only possibility for estimating demand choices based on discrete choices, as demonstrated by the Hungarian papers discussed in the previous chapter. I will not discuss the further details of the approach based on discrete choice models, and refer the interested reader to a good survey provided in *Muraközy* [2010].

In general, the use of structural models, that is, the equations describing both demand and supply side behaviour are derived from theoretical models and then estimated, is most common in the case of estimation based on discrete choice mod-

30 In this case, for example, it was not possible to use adequate instruments, therefore the authors also used lagged price changes to estimate the demand reactions.

31 The primary, but not only form of a discrete choice is when demand is either zero or one; examples include automobiles, or most telecommunications services and public utilities.

els,. I have found only one application of this type of method, in *Molnár et al.* [2007], who analyse competition in a market with differentiated products, the market for residential financial products.[32] The paper estimates the own and cross-price elasticity for various specifications, and uses these to calculate optimal margins using the general model of competition for the supply side. The margins observed in the market are then compared to the equilibrium outcomes of two specific models of competition (Bertrand-competition and collusion). The paper thus effectively measures market power on the specific markets, and finds that the level of competition is quite low in the markets for most financial products; even the hypothesis of collusion cannot be rejected.

There are a few more examples for demand estimation beased on discrete choices: *Crawford and Molnár* [2008] analyse the effects of advertisements on the demand (and its elasticity) for Hungarian mobile telephone services, while *Tánczos and Török* [2007] present an application in the area of transportation economics by modelling the flow of traffic between Budapest and Győr.[33] *Koltay* [2012*a*] studies the German market, estimating the effect on consumer choices of the introduction of an eco-friendly brand in the market for hygiene products. He investigates how the results conform to various theoretical models describing the demand for common goods.

Analysing the supply side

Although they do not, in a strict sense, concern the relationship between prices (or some other performance indicator) and market structure or market behaviour, it is still worth discussing studies estimating production and cost functions, as well as production efficiences. The empirical methods employed typically seek to explain some output variable (production, cost, or productivity indicator) using the level of various inputs (or their price). The estimated equation is typically derived from the first order conditions of the firm's (or industry's) profit maximisation problem. There are only a handful of such papers in Hungary: *Reiff et al.* [2002] estimate production functions and various productivity indicators at an industry level, while *Bisztray et al.* [2010] estimate firms' energy efficiency in the case of water utilities.

[32] *Paizs* [2009] also estimates a structural model, however, he estimates the reaction functions of a specific theoretical model of competition, and not the equations for optimal behavioural in a discrete choice model. Furthermore, the paper estimates a model for the competition between European countries in determining excise taxes, which is a cross-market interaction.

[33] *Édes et al.* [2011] provide a general survey of the empirical methods for analysing substitutability between modes of transport.

CONCLUSION

This paper set out to survey Hungarian empirical IO analyses from the past decade, and also draw attention to the diversity of empirical methods that can be applied. The groups into which the methods have been sorted do not represent a ranking of quality: while it is true in general that more detailed databases enable the use of more complex empirical methods, this does not mean that the results will be more reliable (and especially not that they will be more easy to interpret in practice). Therefore it is important to be familiar with the various empirical methods, their advantages and their limitations, and to interpret the available facts according to several methods, if possible. This can be considered a type of robustness check.[34] I have shown several cases where relatively standard (reduced form) econometric methods were sufficient to conduct empirical analyses which could effectively assist in rejecting or verifying various hypotheses important in competition policy and regulation.

Surveying the Hungarian studies in empirical IO we can also draw the interesting conclusion that the majority of the authors is not or not only an academic. This confirms the common supposition that these studies are typically connected to practical application, and also that it is in institutions that are not foremost research facilities that authors encounter topics and databases which can be used to produce scientifically sound results. Four such institutions can be identified: the Regional Centre for Energy Policy Research (REKK), Infrapont Economic Consulting, the Hungarian Competition Authority (GVH) and the Central Bank of Hungary (MNB). We can only hope that these institutions can continue their scientific work, and that access to databases will improve so that in the future, research facilities can also focus more on modern empirical industrial organisation.

[34] As one of the reviewers of this paper aptly commented: "An empirical model is like the Hungarian language. It can be used to tell the truth, but also to lie."

REFERENCES

Ashenfelter, O.–Ashmore, D.–Baker, J.–Gleason, S.–Hosken, D. [2006]: Empirical methods in merger analysis: Econometric analysis of pricing in FTC v. Staples. International Journal of the Economics of Business, Vol. 13, pp. 265–79.

Bakucs L. Z.–Fertő I. [2009]: Marketing and Pricing Dynamics in the Presence of Structural Breaks. The Hungarian Pork Market. Journal of International Food and Agribusiness Marketing, Vol. 21, pp. 116–133.

Békés, G.–Harasztosi, P. [2013]: Agglomeration Premium and Trading Activity of Firms. Regional Science and Urban Economics, Vol. 43, pp. 51–64.

Békés, G.–Koren, M.–Zsohár, P. [2011]: Geographical pricing of retail petrol prices. MTA KTI Working Papers, MT-DP-2011/30.

Békés, G.–Muraközy, B. [2012]: Temporary Trade and Heterogenous Firms. Journal of International Economics, Vol. 87, pp. 232–246.

Berry, S.–Reiss, P. [2007]: Empirical models of entry and market structure. In: *Armstrong, M.–Porter R.* (ed.): Handbook of Industrial Organization, Vol. 3, North Holland, pp. 1845–1886.

Bishop, S.–Walker, M. [2010]: The Economics of EC Competition Law: Concepts, Application and Measurement. Sweet & Maxwell, London.

Bisztray, M.–Kis, A.–Muraközy, B.–Ungvári, G. [2010]: Energy Efficiency Analysis of Water and Wastewater Utilities Based on the IBNET Database, Regional Centre for Energy Policy Research Working Paper.

Bölcskei, V. [2010]: A review of the demand models of telephone services – with special regard to estimating substitution between land and mobile services. Economic Review, Vol. 57. No. 6, pp. 517–535.

Bresnahan, T. F. [1989]: Empirical studies of industries with market power. In: *Schmalensee, R.–Willig, R.* (ed.): Handbook of Industrial Organization, Vol. 2, North Holland, pp. 1011–1057.

Brown, D.–Earle, J.–Telegdy, Á. [2006]: The Productivity Effects of Privatization: Longitudinal Estimates from Hungary, Romania, Russia and Ukraine, Journal of Political Economy, Vol. 114, pp. 61–99.

Crawford, G.–Molnár, J. [2008]: The Empirical Consequences of Advertising Content in the Hungarian Mobile Phone Market. Manuscript, http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/crawford/research/adcontent.pdf.

Csorba, G. [2008]: A fúziókontroll módszertanáról: Dominancia- vagy versenyhatásteszt? [On the assessment criteria in merger analysis: Dominance vs Significant Lessening of Competition test] In: *Valentiny Pál–Kiss Ferenc László* (eds.): Verseny és szabályozás, 2007. [Competition and Regulation, 2007], Institute of Economics, Hungarian Academy of Sciences, Budapest, pp. 96–110.

Csorba, G.–Koltay G.–Farkas, D. [2011]: Separating the ex post effects of mergers: an analysis of structural changes on the Hungarian retail gasoline market. MTA KTI Working Papers, MT-DP-2011/18.

Davis, P.–Garces, E. [2010]: Quantitative Techniques for Competition and Antitrust Analysis. Princeton University Press, Princeton.

Doraszelski, U.–Pakes, A. [2007]. Framework for Applied Dynamic Analysis in IO. In: *Armstrong, M.–Porter R.* (ed.): Handbook of Industrial Organization, Vol. 3, North Holland, pp. 1887–1966.

Édes, B.–Bölcskei, V.–Lőrincz, L.–Nagy, P.–Pápai, Z. [2010]: A vezetékes és mobil távbeszélő szolgáltatások közötti helyettesítés vizsgálata [An examination of substitution between fixed and mobile telephone services]. In: *Valentiny Pál–Kiss Ferenc László–Nagy Csongor István* (eds.): Verseny és szabályozás, 2009. [Competition and Regulation, 2009] Institute of Economics, Hungarian Academy of Sciences, Budapest, pp. 204–242.

Édes, B.–Micski, J.–Pápai, Z. [2011]: Kereslet és intermodális helyettesítés a távolsági közlekedésben. [Demand and intermodal substitution in long distance transport]. Infrapont-paper. http://www.infrapont.hu/dokumentumok/Intermod%e1lis%20helyettes%edt%e9s_Infrapont_2011.pdf.

Einav, L.–Levin, J. [2010]: Empirical Industrial Organization: A Progress Report, Journal of Economic Perspectives, Vol. 24, pp. 145–162.

Farkas, D.–Csorba, G.–Koltay, G. [2009]: Prices and concentration on the Hungarian retail market for motor fuels. Economic Review, Vol. 56, No. 12, pp. 1088–1109.

Fertő, I. [2009]: Labour Market Adjustment and Intra-Industry Trade: The Effects of Association on the Hungarian Food Industry. Journal of Agricultural Economics, Vol. 60. No. 3, pp. 668–681.

Halpern, L.–Muraközy, B. [2012]: A verseny és a K + F összefüggései – elméleti megközelítések és számszerű eredmények_[The relationship between competition and R&D – theoretical approaches and quantitative results]. In: *Valentiny Pál–Kiss Ferenc László–Nagy Csongor István* (eds.): Verseny és szabályozás, 2011 [Competition and Regulation, 2011] Institute of Economics – Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budapest, pp. 13–37.

Horváth, Á.–Kiss, H. J.–McLean, A. [2013]: How does an Increase in Energy Efficiency Affect Housing Prices? A Case Study of a Renovation. ELTINGA Working Paper Series, No. 2.

Imbens, G.–Wooldridge, J. [2009], Recent Developments in the Econometrics of Program Evaluation. Journal of Economic Literature, Vol. 47, pp. 5–86.

Kézdi, G. (ed.) [2012]: In Focus. Evaluation of active labour market programs. In: *Fazekas Károly–Kézdi Gábor* (ed.): The Hungarian Labour Market, 2012. Institute of Economics – Centre for Economic and Regional Studies Hungarian Academy of Sciences & National Employment Non-profit Public Company Ltd. Budapest.

Kézdi, G.–Csorba, G. [2012]: Estimating consumer lock-in effects from firm-level data. Journal of Industry, Competition and Trade, November, http://link.springer.com/article/10.1007%2Fs10842-012-0149-2.

Koltay, G. [2012*a*]: How much does your environment matter? Estimating the effect of the Blauer Engel eco-label. European Association for Research in Industrial Economics, 39th Annual Conference, September 2–4, Rome.

Koltay, G. [2012*b*]: Not an average story: Asymmetric price transmission in the Hungarian gasoline retail market. 18th International Panel Data Conference, July 5–6, Paris.

Lőrincz, L.–Nagy, P. [2010]: Switching costs in telecommunications: Conclusions from a Hungarian survey. In: *Falch, M.–Markendahl, J.* (ed.): Promoting New Telecom Infrastructures: Markets, Policies and Pricing, Edward Elgar, Cheltenham, UK–Northampton, MA, pp. 175–193.Molnár, J.–Nagy, M.–Horváth, Cs. [2007]: A Structural Empirical

Analysis of Retail Banking Competition: the Case of Hungary. MNB Working Papers, 2007/1.

Muraközy, B. [2010]: Ökonometriai módszerek a vezetékes és mobiltávközlési szolgáltatások helyettesítésének vizsgálatára [Econometric methods for measuring the substation between fixed and mobile telecommunication services]. In: *Valentiny Pál–Kiss Ferenc László–Nagy Csongor István* (eds.): Verseny és szabályozás, 2009. [Competition and Regulation, 2009], Institute of Economics, Hungarian Academy of Sciences, Budapest, pp. 171–203.

Nagy, P.–Pápai, Z.–Muraközy, B.–Micski, J. [2012]: A vezetékes telefonszolgáltatás rugalmasságának vizsgálata a piacmeghatározás empirikus megalapozása céljából [Study on the elasticity of demand for fixed telephone services supporting empirical market definition]. Nemzeti Média- és Hírközlési Hatóság HF/25000–9/2012. számú határozatának II. Függeléke, http://nmhh.hu/dokumentum/153976/1es_piac_hatterv_ii_fuggelek_uzleti_titkot_nem_tartalmazo_valtozat.pdf.

Nevo, A.–Whinston, M. [2010]: Taking the dogma out of econometrics: Structural modeling and credible inference. The Journal of Economic Perspectives, Vol. 24, pp. 69–81.

Paizs, L. [2009]: Fiscal competition on the market for diesel fuel in the European Union. Economic Review, Vol. 56. No. 3, pp. 216–238.

Pápai, Z.–Lőrincz, L.–Édes, B. [2011]: Triple play as a separate market? Empirical findings and consequences to broadband market definition. International Telecommunications Society (ITS) 22nd European Regional ITS Conference, Budapest.

Popovics, P. A.–Tóth, J. [2006]: Price transmission and asymmetric price development in the vertical structure. Economic Review, Vol. 53. No. 4, pp. 349–364.

Reiff, Á.–Sugár, A.–Surányi, É. [2002]: Productive efficiency in the Hungarian industry. Hungarian Statistical Review, Vol. 7, pp. 45–74.

Schmalensee, R. [1989]: Inter-industry studies of structure and performance. In: *Schmalensee, R.–Willig, R.* (ed.): Handbook of Industrial Organization, Vol. 2. North Holland, pp. 951–1009.

Szolnoki, P.–Tóth, A. I. [2008]: Szolgáltatóváltás a magyar lakossági árampiacon 2008-ban [Switching in the Hungarian electricity market in 2008]. In: *Valentiny Pál–Kiss Ferenc László* (eds.): Verseny és szabályozás, 2007. [Competition and Regulation, 2007], Institute of Economics, Hungarian Academy of Sciences, Budapest, pp. 197–227.

Tánczos, L.–Török, Á. [2007]: Közúti közlekedési módválasztás modellezése Budapest és Győr között [Modelling individual choice in passenger travel between Budapest and Győr]. Közlekedéstudományi Szemle, Vol. 57, pp. 220–226.

Weinberg, M. [2011]: More Evidence on the Performance of Merger Simulations. American Economic Review, Vol. 101, pp. 51–55.

Werden, G.–Froeb, L. [1993]: Correlation, causality, and all that jazz: The inherent shortcomings of price tests for antitrust market delineation. Review of Industrial Organization, Vol. 8, pp. 329–353.

• *László Halpern–Balázs Muraközy* •

THE RELATIONSHIP BETWEEN COMPETITION AND R&D

Theoretical Approaches and Quantitative Results

This study is an attempt to examine and model the relationship between (the presence and intensity of) competition and corporate level R&D expenditures. Hungarian firm-level and industry-level data populate the empirical models. A brief summary of the history of research on the much-debated competition-innovation relationship is also offered. Results from the earliest studies seem to suggest that stronger competition generally results in lower levels of innovation, while the empirical results that were born in the 1990s showed some evidence for the opposite relationship. The seminal model of *Aghion et al.* [2005] indicated an inverted-U shaped relationship between the intensity of competition and the level of innovation. A discussion of the difficulties of measuring the pertinent variables and the relationship among them is followed by an elaborate investigation of the shape of the relationship. Based on our extensive empirical results, we conclude that the inverted-U shaped relationship can indeed be established in Hungary at the industry-level as well as at the firm-level. We also demonstrate that only certain types of indicators of the presence and intensity of competition seem to have had a detectable relationship with the innovative investments of firms.

INTRODUCTION

In economics, an issue of great interest concerns the factors affecting economic growth in the long run. Since the appearance of modern growth theory, the relative importance of capital accumulation and productivity growth has been continuously debated (the latter factor, at least partially, reflecting technological development). Since Robert Solow started his research programme in the 1950s[1], the key role of productivity growth has become evident. As shown by recent results, in the 20th century it contributed to annual average economic growth by over 1% (*Abramovitz–David* [2001]).

The Solow model, however, assumes that technological change depends on "exogeneous" factors, that is, factors outside the model. This is true inasmuch as the development of science and technology depends to a great extent on phenomena such as the knowable nature, complexity, or interrelatedness of natural laws.

[1] *Romer* [1996] in Chapter 1 gives a detailed and up-to-date description of the Solow model.

Endogeneous growth theory models, however, also take into account the fact that social institutions influence technological development (*Aghion–Howitt* [1998]). In market economies, technological development is attributed, above all, to the fact that firms decide to introduce new technologies in a decentralized way – in other words, they innovate. Therefore, in such economies the impact of social-economic institutions on technological development primarily means that institutions influence the firms' incentives to perform R&D and introduce innovations. Among the features of institutions and economic environment it is most probably competition whose impact on innovation and growth has been most thoroughly discussed by economists. This was due to the strong economic intuition that monopolies and highly competitive firms benefit from the development of production technology or the introduction of new products to different degrees.

This study seeks, above all, to describe the nature of this relationship. After discussing some dilemmas concerning the definition of innovation, it deals with the logic and predictions of the key theoretical models related to the issue. Then it offers details of empirical methods and key results on the relationship between innovation and competition, followed by the empirical analysis of the relationship between competition and the R&D expenditures of Hungarian firms. The study closes with a summary of conclusions.

THE CONCEPT OF INNOVATION

The economics literature places great emphasis on differentiating between innovation and research and development. Since Schumpeter's works were published, innovation has been interpreted as the actual implementation of an improvement, be it either the market introduction of a new product or the use of a new procedure during production. The former is *product innovation*, the latter is *process innovation*, and research and development may mean the development of either a product or a process. In other words, R&D is an input to the innovation process; research itself, however, is not innovation until its results appear on the market or in the production process (*Fagerberg* [2006]).

The definition of innovation given in the *Community Innovation Survey, CIS* of the European Union falls in line with the above:

- It follows that research and development (R&D) itself is not innovation but expenditure on innovation. And this is not the only expenditure of this kind. Innovation inputs also include when a firm purchases machines to implement its innovations or when managers make extra efforts to prepare the introduction of new processes or products. What is more, it is possible that the firm itself does not perform R&D activity yet it can still introduce new products or services relying, for example, on technology transfer.

This is especially true for small open economies and for countries which do not belong to the technology frontier. For such economies, implementing foreign technologies and products is a key to growth, and thus, it must be a focus of innovation policy. The significance of the issue is evidenced by the fact that while in Hungary approximately 10% of the firms that were included in the *Community Innovation Survey* (CIS) performed research and development activity on a continuous basis between 2004 and 2006, more than 30% of them introduced product or process innovation in the same period (*Halpern–Muraközy* [2010]). These data show that in Hungary, in most cases, innovation is performed without any formal research and development activity. It is evident, however, that such innovations require resources from the managers and employees of a firm. Yet such innovation efforts are not included in the R&D statistics, which means that in follower countries R&D statistics may significantly underestimate the actual innovation expenditures of firms.

THE THEORY OF THE RELATIONSHIP BETWEEN COMPETITION AND INNOVATION

The beginnings of research on the relationship between competition and innovation can be traced back, above all, to Josef Schumpeter's (1883–1950) research. In his early works, Schumpeter emphasized that new innovative entrepreneurs can break the "inertia" or "laziness" of large companies. The market entry and subsequent growth of such small enterprises explain the phenomenon of economic growth. Schumpeter termed this process "creative destruction."

Schumpeter's later works focused on the economies of scale that are achievable by big businesses in research and development and innovation. The difference between the two approaches can be interpreted in various ways. First, it can be regarded as a historical change: the growth of scientific knowledge generated economics of scale in research. Another interpretation is that the two Schumpeterian models describe different industries. In some industries, small firms carry innovative solutions (e.g. the Internet). In others, only large firms are capable of introducing innovations (e.g. the pharmaceutical industry), because of the high costs of introducing each innovation.

In and after the 1960s, based on Schumpeter's concepts and parallel with the appearance of game theory-based industrial organization models, a number of analyses have been published that examine the relationship between market structure and innovation with the aid of models of the strategic behaviour of firms. These models regard R&D as investment and, practically, do not differentiate between decisions on R&D and decisions on innovation. As a rule, the assumed decision making consists of two steps. First, a firm decides about the dimensions of its R&D

investment; second, it sells the new product (in case of product innovation) or, employing a new procedure, produces more efficiently (in case of process innovation). Competition starts playing a role in the second step. The type of the product market competition (e.g. Bertrand or Cournot competition) and its strength define the revenues of a given firm on the market. Firms make decisions about their R&D investment based on the profit they will realize in the second phase; that is, they compare the profits achievable with innovation and without innovation. For example, when strong competition decreases the amount of profit that is achievable by innovation, R&D investments and the innovation performance will be lower in competitive industries.

In the "standard" industrial organization model of innovation – regardless of the exact structure of the model – weaker competition (e.g. a monopoly) ensures higher profits for the innovating firm, and, therefore, the innovation level is expected to be higher (for a description, see *Aghion–Griffith* [2005], Chapter 1.1). When competition is stronger, the firm reaches fewer consumers, sets lower prices and, as a result, it becomes less profitable for it to invest in research. Therefore, the models introduced before the 1990s corroborated the existence of the Schumpeterian effect inasmuch as they showed that monopolized industries are innovative and thus are closer to social optimum.

It follows from these theoretical results that in monopolized industries technology develops faster, which means that competition policy has to choose between static and dynamic efficiency or, in other words, has to face a difficult trade-off. But in the 1990s empirical research came to a different conclusion: the research of *Geroski* [1990], [1995], *Nickel* [1996], and *Blundell et al.* [1999] evidenced that in a given industry the stronger the competition, the higher the productivity gains, that is, the stronger the incentive to innovative.

The seminal study of *Aghion et al.* [2005] was based on endogenous growth theory and worked with a different approach, relying on the heterogeneity of firms and on nonlinear relationships. Their model shows that the response of firms to competition can take many different forms. When competition intensifies, firms on the technology frontier increase their R&D, while firms lagging behind reduce their innovative effort. Taking into consideration the resulting industry dynamics, the authors highlight that there is an inverted-U shaped relationship between the strength of competition and innovation at the industry level: in case of weak competition they are positively related, while in case of strong competition the association will be negative.

The model presented by Aghion *et al.* involves multiple time periods. Technological development takes place step by step. The most developed technology is improving to the same degree in every period, independently of the firms modelled. Some firms have the most developed technologies, while others are some steps behind. When a firm introduces a successful innovation, it can take one step forward, otherwise it will start lagging behind even more. [2] This firm faces competitors that are always one step behind the most developed technology at the given time.

If the firm uses the most developed technology, then – given that the production costs of the competitors are necessarily higher – it is in the position to set monopoly prices or, rather, prices that correspond to the expenditure level of the other firms. In case the firm in question is two steps behind the most developed technology, its competitors will be ahead of it, so it will not be able to sell its product. In this model, competition is regarded as the profit level of a firm which employs technologies as developed as those of its competitors; the stronger the competition, the lower the profit of such a firm.

A key concept presented by the authors is that those firms that use the most developed technology respond differently to strengthening competition than those lagging behind. Innovation incentive is the amount their profit would (is expected to) rise if they increased their innovation expenditure. Profit levels achieved with and without innovation are to be compared. Non-linearity is caused by the fact that competition affects both kinds of profit.

If innovation proves to be successful, firms on the technology frontier will be able to produce with the new technology, otherwise other firms catch up to them. The stronger the competition, being caught up by competitors will be the more painful. Consequently, in response to stronger competition firms that use the most developed technology increase their R&D expenditure *to escape competition.*

The situation is reversed for firms that lag behind. They are assumed to be able to make only one step forward in the process technology development. By doing so, they can indeed catch up with other firms but have no chance to leave them behind. Stronger competition means that it is less attractive to catch up to others, and consequently R&D investment is less attractive. The reason for this is that in case of successful innovation the firm will realize lower profit when competing with other firms which are at the same level of development. This means that the firms lagging behind are influenced by the *Schumpeterian effect*:[3] Innovation is a decreasing function of the strength of competition.

Which effect is stronger? The answer depends on the ratio of industries in equilibrium where firms are close to each other (and, consequently, the competition

[2] A key sectoral factor is the intensity of competition in the product market, which is measured by the authors as the difference between the expenditure of firms that employ the most developed technology and that of other firms. It may be caused by various phenomena. Primarily, greater intensity is understood as greater substitutability among products produced in the given industry. Where competition (substitution) is stronger, a leader firm has a higher profit than a follower, as in such industries prices are more important for consumers. This definition falls in line with the logic employed above: the strength of the competition is linked to the profit from innovation (relative to profit achieved without innovation), and thus it may affect R&D expenditure.

[3] The term "Schumpeterian" here refers to the second phase of Schumpeter's scientific activity.

deterrence effect is strong) to industries where there is a greater difference between firms (and, therefore, the Schumpeterian effect is stronger). As competition strengthens, the level of innovative activity increases at first, and then starts to decrease. The relationship between the two phenomena resembles an inverted-U shape or a bell curve. The monopolies as well as the industries where competition is very strong tend to be less innovative than sectors with a low number of actors.

The model built by *Aghion at al.* [2009] in this kind of a framework indicates that the increased probability of the entry of new actors who use the most developed technology also has an inverted-U shaped relationship with innovative activity in the industry. In this model, the innovation efforts made by firms at the technological frontier intensify as the probability of entry by competitors increases. This *entry deterrence effect* is analogous with the competition escape effect in the previous model. Conversely, the innovation efforts of firms that use less developed technology decrease as the probability of entry increases. Their expected profit from successful innovation is lower if they are more likely to face a competitor which uses a more developed technology than theirs. It follows from the way the equilibrium ratio of sectors is defined that the relationship between competition (defined as the probability of entry) and the innovative effort is expected to take an inverted-U shape.[4]

The models discussed so far (with the exception perhaps of Schumpeter's model) are *neoclassical* models. The firms are assumed to be well informed and, based on their knowledge, they make perfectly rational decisions about innovation and everything else. In neoclassical models, even though the outcome of innovation decisions is uncertain, the firms are perfectly aware of the possible returns on investment and the probability of their occurrence.

Studies on *evolutionary models*[5] suggest that this kind of neoclassical model is not suitable for an adequate modelling of innovative behaviour as the payoff of innovation is basically uncertain and the actors are not likely to know the probability distribution of payoffs. This is Knightian uncertainty (*Knight* [1921]). Given the above, it is not justified to assume that firms make perfectly optimal decisions. Instead, they use a heuristic approach or some other, bounded rational decision making mechanisms when deciding about research fields and the amount to be invested. Among the firms that use various decision making procedures, those with higher profits grow faster, as they are in the position to invest more. In time, poor performers go bankrupt.

A key feature of these models is that firms differ from each other in various dimensions. As opposed to neoclassical models, they do not postulate that some

[4] Recently, several other studies have discussed the issue of entry and innovation: *Asker–Baccara* [2010], *Creane–Miyagiwa* [2009], *Grossman–Steger* [2007], *Kovac–Vinogradov–Zigic* [2010], *Miller* [2007].

[5] *Nelson–Winter* [2002] gives an overview of the main issues concerning evolutionary models. The first of the evolutionary models concerning innovation is *Nelson–Winter* [1982].

firms are one or two steps behind the others, rather, that firms in the market have different information and employ different decision making mechanisms. A related issue is path dependence: the situation of firms or industries that take different directions may differ radically.

Evolutionary logic sheds light on the fact that industry productivity can grow not only when a firm introduces new products or processes, but also when the market share of firms with good "genes" or expert knowledge grows, while that of worse performers decreases. (Some of them leave the market). *Motta* [2004] (pp. 55–64) presents a simple model of this kind.

EMPIRICAL METHODS AND EMPIRICAL RESULTS

Measuring the relationship between competition and innovation

Research on the relationship between competition and innovation raises several problems (*Aghion–Griffith* [2005] Chapter 1.2.2). *First,* besides the strength of competition there are several *other factors* that define how much a firm or an industry invests in innovation. These variables may easily be correlated with competition and, for analytical purposes, they must be taken into account.

Second, the relationship between competition and innovation in an industry is not a one-way causal relationship; rather, it is *simultaneous,* which means that innovation also influences market structure. In general, panel data are needed to handle simultaneity. When such data are available – and it is assumed that market structure is pre-determined (that is, innovation in a given period of time affects only future market structure) – the issue of simultaneity can be handled with the use of lagged explanatory variables. Weaker assumptions are needed when, for the purposes of analysis, exogenous changes of economic policy and regulation (e.g. free trade agreements) are used as instrumental variables (e.g. *Aghion et al.* [2005]).

Third, a major issue is that of *measurement errors in explanatory variables*. The analyst is interested in the impact on innovation of competition. The indicators that describe the market structure (number of firms, concentration, etc.) do not measure competition directly. This problem is aggravated by the fact that in open economies external competition needs also be taken into account in some way. Consequently, in theory it is more practical to use an indicator that has a more direct link with competitive pressure than market structure does. Such indicators include the Lerner-index or some other indicator of the market power of firms.

Fourth, the selection of the dependent variable *(a measure of innovation)* is not an easy task either. As referred to above, R&D activity is an input to innovation, not a measure of innovation itself. While for large companies it may have a strong correlation with innovation, smaller firms may introduce important innovations without spending on R&D. As for the outputs of innovation, the number of patents is the measure most widely used. The main problem with this is that not every patent is of equal significance or entails the same level of innovation. Therefore, researchers often decide to weight the number of patents by the number of times it has been referred to in another patent (*Jaffe* [1986]). Theoretically variables showing innovation output (from innovation surveys) are better measures than the numbers of patents. This holds especially true for countries which are not at the technology frontier, therefore, the majority of innovations do not entail patent registration. In practice, however, in most countries these indicators are available only on a relatively small sample of firms and, therefore, fail to reflect the total innovative performance of the economy. When there are no available indicators that directly show spending on and results of innovation, then innovation may be approximated with variables indirectly related to innovation. Such variables include, for example, the productivity of the firm in question (labour productivity or total factor productivity, TFP). Nevertheless, productivity gains depend on several other variables beside technology. For example, it is often difficult to filter out the effect of the economies of scale.

The nature of the measurement of innovation also affects the appropriate estimation methods. For R&D expenditures, for example, the value is zero for a large number of firms; consequently, a tobit model is to be used. When we ask which firm introduced innovation, then probit or logit models may be used.

Fifth, as *Aghion at al.* [2005] state, the relationship between competition and innovation *is not necessarily linear* (Chapter 3.1). According to *Aghion–Griffith* [2005], the results of some early studies contradict each other from time to time, as the authors did not consider this possibility and examined only the linear effects of the competition variable. Nonlinearity is to be dealt with by using quadratic terms or nonparametric models.

Empirical results

Ahn [2002] and *Aghion–Griffith* [2005] offer a summary of the specialized literature on innovation published in the 1990s. These empirical studies failed to corroborate the Schumpeterian hypothesis that the presence of large firms or a greater concentration may lead to higher levels of innovation. A number of studies state that there is a strong positive relationship between competition in the product market and productivity. Further research has shown that the effect of different changes in the economic environment – regulatory changes, greater exposure to global competition, the introduction of competition for non-profit enterprises – justify that competition contributes to productivity, wealth and long-term growth. It is also pointed out that it often takes a long time for enterprises and consumers to adjust to a new context and for the competition to fully exercise its positive impact on efficiency.

Major articles of the 1990s include those by *Geroski* [1990], [1991], [1994], *Blundell et al.* [1995], [1999]. These studies examined the firm-level and industry-level panel data of the 1970s and 1980s, and revealed that competition has a positive impact on innovation. *Pohlmeier* [1992], taking into consideration the fact that this is a simultaneous relationship, found – instead of the theoretically assumed positive relationship – a negative relationship between market concentration and product and process innovation in 2,200 German firms by 1984. *Crépon at al.* [1996] analyzed the 1991 data of approximately 10,000 firms. Results on the relationship between market concentration and innovation differed depending on which innovation indicator was used. When the number of patents and other performance indicators of innovation were used, a negative relationship was established with market concentration, while in the case of the sale of new products a positive relationship was found. As for R&D investment, no relationship was established.

As mentioned in the theoretical summary: *Aghion et al.* [2005] showed that, theoretically, an inverted-U shaped relationship is possible between competition and innovation. Aghion and his colleagues performed empirical studies which established the inverted-U shaped relationship between product market competition (measured with the Lerner index) and innovation (measured with the number of patents). As it was referred to above, in a later study they described a similar theoretical relationship between the probability of entry and the level of innovation (*Aghion et al.* [2009]). Positive relationship was also indicated by panel data on UK firms for the period between 1987 and 1993. The effect of market entry analyzed at the four-digit industry level (especially foreign market entry) is positive in industries where the UK is on the technology frontier, and weak or negative in industries which lag behind. In line with the theoretical model, the results indicate that the relationship between competition and innovation may also be affected by the distance to the technological frontier.

Later, other studies also corroborated the hypothesis of the inverted-U shaped relationship. For example, *Tingvall–Polsdahl* [2006] quantified such a relationship between competition (measured by the Herfindahl index) and innovation on data gathered in Sweden between 1990 and 2000; however, no significant results were found for the price-cost margin. *Brouwer–Van der Wiel* [2010] succeeded in establishing a clear positive relationship between competition and total factor productivity for Dutch industries. In addition, for the Netherlands – at least for the manufacturing industry – these authors provided evidence for the inverted-U shaped relationship between competition and innovation, in other words, for the fact that competition (if significantly stronger than observed) has a negative impact on productivity because of the lower innovation expenditure. The reverse relationship did not show up in the data, which means that the intensity of competition does not decrease because of innovation.

From the models establishing an inverted-U shape, it can be concluded that the shape of the relationship is influenced by the distance of the firms of a country to the technological frontier. *Acemoglu et al.* [2006] studied, among others, this issue and observed a positive correlation between the cross-sectional productivity and R&D expenditure of a country, as well as between the distance to the technological frontier and R&D expenditure. The growth rate of countries where – due to high barriers to entry – competition is weak falls more sharply when the country in question gets closer to the technological frontier than the growth rate of countries with strong competition. The weakness of competition exercises its adverse effects in countries which are close to the technological frontier. *Lee* [2009] came to similar conclusions. Relying on the data of more than 1,000 Canadian, Japanese, South-Korean, Taiwanese, Indian and Chinese businesses, he concluded that the way firms respond to competitive pressure depends on the level of their technological expertise: firms at a higher level step up their R&D efforts, while those at a lower level reduce them.

To sum up, the empirical results of the last two decades have corroborated that competition has a basically positive impact on innovation. Nevertheless, numerous problems with the measurement and empirical methodology have not been properly solved. The creation of targeted corporate databases on innovation is a huge step forward, yet problems (such as measuring competition in an industry, the management of international relations or the adequate consideration of lagged effects) still persist.

THE RELATIONSHIP BETWEEN COMPETITION AND INNOVATION IN HUNGARY

Data

Our major data source was the database of the Hungarian Tax Authority, more specifically, the data from the balance sheets and profit and loss accounts of firms with double-entry accounting from the period between 1992 and 2006. *Table A1* in the Appendix shows the distribution of the firms contained in the database (by industry and size). From 2000 onwards, sampling has been designed to ensure that all large companies and exporting firms are included without exception; however, numerous smaller firms were omitted. The firms in the sample represent more than 90% of employees, turnover and export. As only a very low number of micro firms perform R&D activities, firms with less than 5 employees were excluded from the sample. We perform our analysis in the manufacturing, as the relationship between competition and innovation is easier to measure and interpret in this industry than in services. As in certain cases we used lagged variables as well, we restricted the sample to those firms that were included in the database both in 2003 and 2005.

Finally, for the purpose of data cleansing, we excluded from the analysis firms with a negative value added.

The database contains four-digit NACE number industry classification of the firms, the number of employees and the balance sheet data. Unfortunately, NACE industries do not necessarily correspond to markets as interpreted in industrial organization or competition policy. An industry may consists of several separate markets or a firm may perform productive activity in more than one industry, which then may result in a certain bias during the measurement of the effects of competition. We approximate the innovation efforts of businesses with R&D expenditures between 2003 and 2005.

Table *A2* of the Annex shows the summary statistics of the explanatory variables. *R&D intensity* measures the firm's R&D expenditure relative to its turnover. *Value added* is calculated from the balance sheet. *Labour productivity* is the ratio of corporate value added and the number of employees. *Capital intensity* is the value of tangible assets per employee. Data include information on foreign ownership share. We created a binary dummy variable. When its value is 1, it indicates that at least 10% of the company's equity is owned by foreigners. Data also give information on the firm's export activity. Again, we created a binary dummy variable. When its value is 1, it indicates that the firm performs export activities.

Variables that measure competition can also be defined from the database. We calculated the indicator C_3 to show the share of the three largest companies from the industry's turnover. The *Hirschman–Herfindahl index,* calculated on the basis of turnover, is an alternative measure for concentration. As mentioned above, the concentration variables often fail to measure the market power accurately. Therefore, we also used the indicator *ROA (return on assets)* to show the ratio of a firm's pre-tax profit to its assets.

Other indicators in the competition database of the Hungarian Competition Authority (Hungarian acronym: GVH) were also used as alternative indicators of the strength of competition.[6]

Models

The question is: What is the impact of competition on the innovation of firms? In our basic model, the (firm- or industry-level) R&D activity is the dependent variable, while the explanatory variables include measures of competition as well as control variables.

Three models have been estimated. In the first one, industry-level R&D intensity was modelled with industry competition variables and other explanatory variables. *Aghion et al.* [2005] employed a similar industry-level analysis.

[6] http://www.gvh.hu/gvh/alpha?do=2&st=1&pg=54&m5_doc=5635&m251_act=4.

$$R\&D\ intensity_{j,\,2005} = \alpha + \beta competition_{j,\,2005} + \gamma X_{j,\,2005} + \varepsilon_{j,\,2005} \qquad (1)$$

where j indicates industries; the time index indicates the fact that we used cross-sectional data of the year 2005 for the analysis; $R\&D\ intensity_{j,\,2005}$ is the industry's average R&D intensity; $competition_{j,\,2005}$ is an indicator of the competition; $X_{j,\,2005}$ contains other sectoral-level explanatory variables (labour productivity, capital intensity); and $\varepsilon_{j,\,2005}$ is random error. In the model, β shows the impact of competition on the R&D intensity of the industry.

We run the other two models at the firm level. In the first case, the dependent variable indicates whether the firm in question performed R&D activity in 2005. As the dependent variable is binary, we used a probit model.

$$P(RD_{i,\,2005}) = F(\alpha + \beta competition_{j,\,2005} + \gamma X_{j,\,2005} + \delta Z_{i,\,2005} + \varepsilon_{j,\,2005}) \qquad (2)$$

where i stands for the firm and j for the firm's industry, as the competition variable can be interpreted at the industry level. $X_{j,\,2005}$ contains industry control variables (binary industry dummies based on two-digit codes). $Z_{i,\,2005}$ contains some features of the firm (labour productivity, size and capital intensity). As the dependent variable is binary, the model can be interpreted as the probability of R&D activity. The function $F(x)$ is the normal distribution function.

In the last model, the dependent variable is the firm-level R&D intensity:

$$RD_{i,\,2005} = \alpha + \beta competition_{j,\,2005} + \gamma X_{j,\,2005} + \delta Z_{i,\,2005} + \varepsilon_{j,\,2005} \qquad (3)$$

As a large number of firms do not perform R&D activity – and, thus, the dependent variable is zero for them – this equation is estimated with a tobit model.

The *first* question is: How can we take into account the various variables that may possibly influence the dependent variable? A major problem may arise when at the industry level the nature of technology is such that it is related to the competition variable. Industrial technology is approximated by productivity and capital intensity. Here, the impact of the competition variable is identified from the comparison of the industries which use similar technologies. In the firm-level models, the heterogeneity of the industry is depicted with the aid of two-digit industrial codes. We address the issue of firm heterogeneity with the introduction of size dummies and variables measuring firm productivity, export status, foreign ownership and capital intensity. In the firm-level regressions some explanatory variables are industry-level variables, which may cause heteroskedasticity. We handle this with clustered standard errors.

The *second* question pertains to the issue of endogeneity; in other words, the fact that innovation in a given year is determined simultaneously with market structure. They mutually influence each other. To handle this, we ran all regressions with lagged explanatory variables (from 2003). Since innovation in year 2005 cannot affect the market structure variables in year 2003, we hope that the coefficients thus derived exhibit a causal relationship.

The *third* question is that of measuring competition. First, we performed all measurements with three competition variables. Two of them approximate the market structure, while ROA approximates the profit margin. After that, the firm-level regression is run on all variables of the competition statistics database of the Hungarian Competition Authority.

The *fourth* question pertains to the variable that reflects innovation. In this respect, the best solution would be to use the definition of innovation given in the Community Innovation Survey (CIS) of the European Union. However, as it is available only for a relatively small sample of firms, we decided to use the R&D value of year 2005, which was available for all firms.

Finally, as *Aghion et al.* [2005] (Chapter 3.1) emphasize this, the relationship between competition and innovation is not necessarily linear. To examine this relationship, we also estimated the model using a quadratic specification. The inverted-U shape is corroborated if the coefficient of the linear term is positive and that of the quadratic term is negative.

Results

Table *A2 of the Annex* contains the key summary statistics. It demonstrates that out of the 7,575 firms of the sample, only 256 (3.4%) performed R&D activities in 2005. By international comparison, this rate is very low, but – as mentioned in the first section – the true rate of innovative firms was higher. Approximately one-fifth of the firms in the sample were in foreign ownership and more than half of them performed export activity.

The first glimpse on the relationship between competition and innovation is given in *Table 1*. Based on the strength of the competition, we categorized the four-digit industries into four quartiles. In each column, we used a different competition indicator for the purpose of categorization. The numbers indicated in the

TABLE 1 • **Concentration and average R&D intensity, 2005**

NACE4 industries	Concentration (C_3)	Herfindahl index	ROA
1st quartile	0.096	0.096	0.092
2nd quartile	0.136	0.280	0.138
3rd quartile	0.350	0.196	0.140
4th quartile	0.043	0.053	0.257
F-test	4.24	2.36	1.1
p-value	0.006	0.072	0.348

Note: The Table shows the average R&D intensity in the industry quartiles defined on the basis of competition indicators. The *F-test* examines the hypothesis that these are equal.

Table show the average R&D intensity of the firms that fall into the given quartile. The bottom rows show the results of the F-test whose null hypothesis, that the industries belonging to each quartile exhibit the same average R&D intensity, could not be rejected.

The results shown in the table indicate significant differences between the quartiles for the two concentration indicators. This pattern falls in line with the model and empirical results of *Aghion et al.* [2005]: the relationship is depicted by an inverted-U shaped curve. R&D intensity is the highest in those industries where competition is of medium strength. However, there is no significant difference between the R&D intensity of the sectors in terms of the quartiles defined on the basis of ROA.

The relationship between average R&D intensity of the industries and competition was also examined with the industry-level regression shown in equation (1). The results are presented in *Table 2*. This Table contains three equations for all the three competition indicators. The first equation contains only the competition indicator. The second equation contains industry productivity and capital intensity as well, and thus takes into consideration the technological features of the industry. In the third equation, by including the square of the competition indicator, competition is allowed to have a non-linear impact on the dependent variable.

TABLE 2 • **Impact of competition on the R&D intensity in the industry**

Variable	OLS	Extended	Quadratic	OLS	Extended	Quadratic	OLS	Extended	Quadratic
Concentration (C_3)	0.042 (0.075)	0.006 (0.068)	1.685*** (0.575)						
Concentration² $(C_3)^2$			–1.313*** (0.439)						
Herfindahl index				–0.126** (0.053)	–0.154** (0.067)	0.506* (0.293)			
Herfindahl index²						–0.668** (0.296)			
Average ROA in industry							0.214 (0.385)	0.126 (0.344)	1.624 (1.330)
ROA²									–3.657 (2.531)
Labour productivity		0.001 (0.002)	0.002 (0.002)		0.002 (0.002)	0.002 (0.003)		0.001 (0.002)	0.001 (0.002)
Log capital intensity		0.051 (0.047)	0.052 (0.046)		0.053 (0.048)	0.041 (0.047)		0.051 (0.047)	0.049 (0.046)
Constant	0.127*** (0.047)	0.070 (0.073)	–0.377** (0.188)	0.201*** (0.047)	0.12** (0.052)	0.044 (0.061)	0.141*** (0.031)	0.066 (0.068)	–0.003 (0.103)
Number of observations	231	231	231	231	231	231	231	231	231
R^2	0.000	0.007	0.031	0.005	0.015	0.026	0.001	0.008	0.018

Note: The dependent variable is the average R&D intensity of the industry (in percentage). The observation units are four-digit industries.

* Significant at the 10 percent level, ** significant at the 5 percent level, *** significant at the 1 percent level.

In line with the descriptive statistics, for C_3 and for the Herfindahl index the results show an inverted-U shaped relationship between competition and R&D intensity. The results for the ROA indicator are not significant. To examine robustness, we ran the same regressions on industry-level data aggregated to three digits, and the results were the same. To handle the issue of simultaneity between competition and innovation, we performed the calculations with lagged explanatory variables as well, and came to the same conclusions.

Overall, it was found that the industry-level data support the hypothesis of the inverted-U shaped curve. The low explanatory power of the models, however, indicates that (albeit competition does have an impact on R&D expenditure) technological and other differences between industries play a much more decisive role.

We now turn to the firm-level regressions. *Table 3* shows our estimation results for equations (2) and (3). The dependent variable is the probability of a positive

TABLE 3 • **Impact of competition on the R&D of firms**

Variable	Probit	Tobit	Probit	Tobit	Probit	Tobit
Concentration (C_3)	0.007**	3.191				
	(0.003)	(1.267)				
Herfindahl index			0.005	1.923		
			(0.004)	(1.374)		
Average ROA in industry					0.010	2.725
					(0.015)	(5.316)
Labour productivity	0.000	0.038	0.000	0.040	0.000	0.042
	(0.000)	(0.048)	(0.000)	(0.049)	(0.000)	(0.051)
Log capital intensity	0.001**	0.506**	0.001***	0.526**	0.002***	0.55**
	(0.001)	(0.235)	(0.001)	(0.239)	(0.001)	(0.242)
	0.03***	5.882***	0.03***	5.846***	0.03***	5.813***
Size: 25–50	(0.006)	(1.373)	(0.007)	(1.369)	(0.007)	(1.367)
Size: 50–250	0.073***	7.825**	0.073***	7.81***	0.073***	7.774***
	(0.010)	(1.336)	(0.011)	(1.335)	(0.011)	(1.332)
Size: 250	0.269***	11.079***	0.274***	11.116***	0.276***	11.136***
	(0.035)	(1.804)	(0.036)	(1.809)	(0.036)	(1.817)
Exporter	0.006***	2.617**	0.007***	2.695**	0.007**	2.751**
	(0.002)	(1.058)	(0.002)	(1.077)	(0.002)	(1.080)
Foreign ownership > 10 %	–0.003***	–1.325**	–0.003***	–1.27**	–0.003**	–1.285**
	(0.001)	(0.646)	(0.001)	(0.642)	(0.001)	(0.646)
Constant		–22.37***		–21.315***		–21.24***
		(4.150)		(3.921)		(3.880)
Observations	7,125	7,575	7,125	7,575	7,125	7,575
Pseudo R^2	0.342	0.218	0.339	0.216	0.338	0.215
Log Likelihood	–726.1	–1267	–728.8	–1271	–729.6	–1272

Note: The dependent variable of the probit models indicates whether the firm in question performed R&D activity in 2005. The dependent variable of the tobit models show the firms' R&D intensity (in %). For the probit models, the table shows the average marginal effects at the sample mean. We calculated competition variables for four-digit NACE industries. Regressions also include two-digit industry dummies. We clustered standard errors at the industry level.
* Significant at the 10 percent level, ** significant at the 5 percent level, *** significant at the 1 percent level.

R&D expenditure in the probit columns, and the firms' R&D intensity in the tobit columns. For the probit model, the table contains the average marginal effect of the variables at the sample mean. The equations contain two-digit industry dummies as well; however, the table does not show the point estimates for them.

Larger firms with higher capital intensity that perform export activity have a higher-level innovative activity. An interesting result is that productivity does not affect R&D decisions when size and capital intensity are taken into consideration. Another surprising finding is that foreign-owned firms *ceteris paribus* perform less R&D than those owned by Hungarians. Our study *Halpern–Muraközy* [2010] established neither a negative nor a positive impact on the data taken from the Community Innovation Survey of the European Union. This can be attributed to the fact that the innovation expenditure of foreign business in Hungary exhibit only a weak correlation to the innovations implemented in Hungary.

As for the competition indicators, only the concentration indicator (C_3) is significant. It is positive, which means that firms in the relatively more concentrated industries are more likely to perform R&D, and their R&D intensity is higher. The Herfindahl index and ROA are not significant.

As indicated by the industry-level regressions, a possible reason is that the relationship between competition and innovation is not linear. *Table 4* shows the results of the estimation which involves a quadratic term in the equation. With regard to the competition indicators, an inverted-U shaped relationship was established for the concentration indicator (C_3) and the Herfindahl index. In this model, no significant effect was revealed for ROA. *Table A3* of the Appendix shows the results achieved with the use of lagged explanatory variables. The results of these specifications are similar to those of the previous estimations, but the coefficients of the competition variables are not as significant. For the other variables, the results are the same.

It has been mentioned above that concentration indicators are not necessarily the best tools to measure the strength of the competition actually affecting the market. Therefore, it is of great importance to examine which competition indicators are linked to the innovation efforts of firms and to what extent. For this purpose, we estimated equation (2) for 70 further competition indicators of the competition statistics database of the Hungarian Competition Authority. For the purposes of estimation – to handle the problem of simultaneity – we used the lagged values (of year 2003) of the indicators. *Table A4* of the Appendix shows the marginal effect of the competition indicators and their squares in the sample mean for the R&D binary value.

The results corroborate the conclusion that the concentration variables showing the share of the biggest companies are in an inverted U-shaped relationship with the innovation efforts of firms. The results were affected only to a small degree by whether concentration was calculated on the basis of assets or turnover. The strength or the direction of the relationship is not affected by whether the indicator shows the share of the three, five or ten largest firms. Interestingly, for

TABLE 4 • **Non-linear impact of competition on the R&D of firms**

Variable	Probit	Tobit	Probit	Tobit	Probit	Tobit
Concentration (C_3)	0.027** (0.011)	13.659** (5.306)				
Concentration² $(C_3)^2$	–0.018* (0.009)	–9.178** (4.382)				
Herfindahl index			0.019* (0.011)	9.721** (4.624)		
Herfindahl index²			–0.018 (0.012)	–10.009** (5.098)		
Average ROA in industry					0.009 (0.038)	2.964 (13.054)
ROA²					0.006 (0.090)	–0.687 (30.215)
Labour productivity	0.000 (0.000)	0.042 (0.049)	0.000 (0.000)	0.044 (0.051)	0.000 (0.000)	0.042 (0.051)
Log capital intensity	0.001** (0.001)	0.516** (0.237)	0.001*** (0.001)	0.524** (0.238)	0.001*** (0.001)	0.551** (0.242)
Size: 25–50	0.03*** (0.006)	5.91*** (1.372)	0.03*** (0.006)	5.859*** (1.374)	0.03*** (0.007)	5.813*** (1.366)
Size: 50–250	0.071*** (0.010)	7.803*** (1.325)	0.073*** (0.011)	7.796*** (1.338)	0.073*** (0.011)	7.774*** (1.331)
Size: 250 <	0.268*** (0.035)	11.106*** (1.804)	0.272*** (0.036)	11.085*** (1.812)	0.277*** (0.036)	11.135*** (1.814)
Export	0.006*** (0.002)	2.543 (1.038)	0.006*** (0.002)	2.636** (1.062)	0.007*** (0.002)	2.752** (1.077)
Foreign ownership 10 % <	–0.003*** (0.001)	–1.386** (0.659)	–0.003*** (0.001)	–1.331** (0.647)	–0.003** (0.001)	–1.285** (0.646)
Constant		–24.656*** (4.765)		–21.825*** (4.039)		–21.249*** (3.912)
Observations	7,125	7,575	7,125	7,575	7,125	7,575
Pseudo R^2	0.343	0.220	0.340	0.217	0.338	0.215
Log Likelihood	–724.4	–1265	–727.6	–1268	–729.6	–1272

Note: The dependent variable of the probit models indicates whether the firm in question performed R&D activity in 2005. The dependent variable of the tobit models show the R&D intensity of firms (in %). For the probit models, the table shows the average marginal effects at the sample mean. We calculated competition variables for four-digit NACE industries. Regressions also include two-digit industry dummies. We clustered standard errors at the industry level.
* Significant at the 10 percent level, ** significant at the 5 percent level, *** significant at the 1 percent level.

the concentration indicators that measure domestic consumption, no significant impact has been established. Similarly, the Herfindahl index (from the database of the Hungarian Competition Authority) does not exhibit a significant relationship with innovation expenditure.

Among other indicators, it is the industrial dynamics variable that has a significant impact on R&D expenditure: the intensity of entry and exit is in a convex

(U-shaped) relationship with the firms' R&D probability. These variables can be regarded as the measures of the threat of entry. It is in connection with this variable, that the model of *Aghion et al.* [2009] sets forth a prediction that is contrary to our results.[7]

As for the financial variables, the return on equity (ROE) is in a concave (albeit not inverted-U shaped) relationship with innovation efforts. Falling in line with the above calculations, the ROA obtained from the competition statistics database of the Hungarian Competition Authority is not significant, either. Finally, the presence of foreign-owned firms, the indicator is also shown to have an inverted-U shaped relationship with competition.

In sum, the empirical results show that in Hungary there is a detectable inverted-U shaped relationship between competition and innovation at both the industry and firm levels. Competition increases innovation, yet R&D intensity is somewhat lower in industries where competition is very strong than in industries with medium-strength competition. The result can be interpreted as a causal impact inasmuch as lagged explanatory variables yield the same result. The analysis of a wide range of competition indicators evidences the importance of the method of measuring competition: the concentration indicators, the industrial dynamics, ROE and the ratio of foreigners are in significant relationship with the probability of R&D activity in firms.

CONCLUSIONS

This study presents an overview of the key theories and empirical results related to the relationship between competition and innovation. Our study contributes to the ongoing debates in Hungary by sharing information on empirical results with respect to this relationship.

In recent decades, research on innovation has been calling attention to the requirement that the inputs to and the results of innovation need to be distinguished and dealt with separately. The difference between the two is important, especially in countries which are not among the technologically most advanced ones in the world. For instance, the number of Hungarian firms that introduced innovations in 2006 was three times as high as the number of those that performed R&D activity continuously in the preceding years (*Halpern–Muraközy* [2010]).

Theoretical models explaining the relationship between competition and innovation have a long history. Schumpeter's theory holds that large firms often perform R&D more efficiently and, as a result, some market power is required for a firm to implement a large number of innovations. Important new developments were pre-

[7] However, this relationship can be explained by other circumstances. For example, it is possible that the entry and exit rates are higher in countries with several, geographically segmented markets.

sented by *Aghion et al.* [2005]. Their empirical models yielded inverted-U shaped relationships, which suggest that innovative activity is lower in firms that operate either in highly concentrated or highly competitive industries than in firms in moderately competitive sectors of the economy.

The measurement of the relationship between competition and innovation raises a number of problems. In addition to the difficulties related to measuring the explanatory and dependent variables, another grave issue is presented by the simultaneity of the relationship between competition and innovation.

As evidenced by research in the 1990s, growing competition, in general, strengthens corporate innovation. In the 2000s, some authors came to the conclusion that the relationship is non-linear but an inverted-U shaped relationship can be frequently established.

We analyzed an extensive set of data on Hungarian firms, based on methods we derived from leading international literary sources. The main conclusion from our efforts is that the inverted-U shaped relationship can indeed be established at the industry-level as well as at the firm-level. By applying several competition indicators to our models, we also discovered that only certain types of indicators of the presence and intensity of competition seem to have had an impact on the innovative investments of firms.

REFERENCES

Abramovitz, M.–David, P. [2001]: Two Centuries of American Macroeconomic Growth from Exploitation of Resource Abundance to Knowledge-Driven Development. Stanford Institute for Economic Policy Research, Policy Paper, No. 01–005.

Acemoglu, D.–Aghion, P.–Zilibotti, F. [2006]: Distance to Frontier, Selection, and Economic Growth. Journal of the European Economic Association, Vol. 4, No. 1, pp. 37–74.

Aghion, P.–Bloom, N.–Blundell, R.–Griffith, R.–Howitt, P. [2005]: Competition and Innovation: An Inverted U Relationship. The Quarterly Journal of Economics Vol. 120, pp. 701–728.

Aghion, P.–Blundell, R.–Griffith, R.–Howitt, P.–Prantl, S. [2009]: The Effects of Entry on Incumbent Innovation and Productivity, The Review of Economics and Statistics, Vol. 91, pp. 20–32.

Aghion, P.–Griffith, R. [2005]: Competition and Growth, Reconciling Theory and Evidence. MIT Press. Cambridge, MA.

Aghion, P.–Howitt, P. [1998]: Endogenous Growth Theory, MIT Press, Cambridge MA.

Ahn, S. [2002]: Competition, Innovation and Productivity Growth: A Review of Theory and Evidence, OECD Economics Working Paper, No. 317.

Asker, J.–Baccara, M. [2010]: Subsidies, Entry and the Distribution of R&D Investment, International Journal of Industrial Organisation, Vol. 28, pp. 254–270.

Blundell, R.–Griffith, R.–Reenen, J. Van [1995]: Dynamic Count Data Models of Technological Innovation. Economic Journal, Vol. 105, pp. 333–344.

Blundell, R.–Griffith, R.–Reenen, J. Van [1999]: Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms. Review of Economic Studies Vol. 66, pp. 529–554.

Brouwer, E.–Van der Wiel, H. [2010]: Competition and Innovation: Pushing Productivity Up or Down. CentER DP No. 2010-52.

Creane, A.–Miyagiwa, K. [2009]: Forgoing Invention to Deter Entry. International Journal of Industrial Organisation, Vol. 27, pp. 632–638.

Crépon, B.–Duguet, E.–Kabla, I. [1996]: Schumpeterian Conjectures: A Moderate Support from Various Innovation Measures, In: *Kleinknecht, A.* (ed.): Determinants of Innovation. Macmillan, London.

Fagerberg, J. [2006]: Innovation. A Guide to the Literature. In: *Fagerberg, J.– Mowery, D.– Nelson, R.* (ed.): Oxford Handbook of Innovations, Oxford University Press, Oxford, pp. 1–28.

Geroski, P. [1990]: Innovation, Technological Opportunity, and Market Structure. Oxford Economic Papers, Vol. 42, pp. 586–602.

Geroski, P. [1991]: Market Dynamics and Entry. Basil Blackwell, Cambridge.

Geroski, P. [1994]: Market Structure, Corporate Performance and Innovative Activity. Oxford University Press, Oxford.

Geroski, P. [1995]: Market Structure, Corporate Performance and Innovative Activity. Oxford University Press, Oxford.

Grossman, V.–Steger, T. M. [2007]: Anti-competitive Conduct, In-house R&D, and Growth, European Economic Review, Vol. 52, pp. 987–1008.

Halpern, László–Muraközy, Balázs [2010]: Innovation and Corporate Performance in Hungary. Economic Review, Vol. 57, No. 4, pp. 293–317.

Jaffe, A. B. [1986]: Technological Opportunity and Spillovers of R & D: Evidence from Firms' Patents, Profits, and Market Value. The American Economic Review, Vol. 76, No. 5, pp. 984–1001.

Knight, F. H. [1921]: Risk, Uncertainty and Profit. Houghton Mifflin Company, Boston, New York.

Kovac, E.–Vinogradov, V.–Zigic, K. [2010]: Technological Leadership and Persistence of Monopoly Under Endogenous Entry: Static versus Dynamic Analysis. Journal of Economic Dynamics and Control, Vol. 34, No. 8, pp. 1421–1441.

KSH [2006]: Innovation, 2004 [Innováció, 2004] By: *Zsuzsanna Szunyogh* Központi Statisztikai Hivatal, Budapest, http://portal.ksh.hu/pls/ksh/docs/hun/xftp/idoszaki/innovacio/innovacio04.pdf.

Lee, C.Y. [2009]: Competition Favours the Prepared Firm: Firms' R&D Responses to Competitive Market Pressure. Research Policy, Vol. 38, pp. 861–870.

Miller, D. A. [2007]: Invention under Uncertainty and the Threat of ex Post Entry. European Economic Review, Vol. 52, No. 3, pp. 387–412.

Motta, M. [2004]: Competition Policy. Theory and Practice, Cambridge University Press.

Nelson, R.–Winter, S. [1982]: An Evolutionary Theory of Economic Change, The Belknap Press of Harvard University Press, Cambridge, MA.

Nelson, R.–Winter, S [2002]: Evolutionary Theorizing in Economics. Journal of Economic Perspectives, Vol. 16, No. 2, pp. 23–46.

Nelson, R. R. [1995]: Recent Evolutionary Theorizing about Economic Change, Journal of Economic Literature, Vol. 33, No. 1, pp. 48–90.

Nickell, S. [1996]: Competition and Corporate Performance. Journal of Political Economy, Vol. 104, pp. 724–746.

Pohlmeier, W. [1992]: On the Simultaneity of Innovations and Market Structure. Empirical Economics, Vol. 17, pp. 253–272.

Romer, D. [1996]: Advanced Macroeconomics. McGraw-Hill, New York, 1st edition.

Solow, R. [1956]: A contribution to the Theory of Economic Growth. Quarterly Journal of Economics, Vol. 70, pp. 65–94.

Tingvall, P. G.–Polsdahl, A. [2006]: Is There Really an Inverted U-shaped Relation between Competition and R&D? Economics of Innovation and New Technology, Vol. 15, pp. 101–118.

Vives, X. [2008]: Innovation and Competitive Pressure, The Journal of Industrial Economics Vol. 56, No. 3, pp. 419–469.

APPENDIX

TABLE A1 • **Sample size**

Industries	Number of employees in sample				Total
	1–25	25–50	50–250	250 <	
	employees in company				
Manufacture of food products and beverages	549	222	279	61	1111
Manufacture of tobacco products	1	0	2	2	5
Manufacture of textiles	137	45	48	10	240
Manufacture of wearing apparel	186	85	115	24	410
Tanning and dressing of leather	50	26	41	9	126
Manufacture of wood and wood products	316	72	47	6	441
Manufacture of pulp, paper and paper products	68	27	33	8	136
Publishing and printing	372	76	63	9	520
Manufacture of coke, refined petroleum products and nuclear fuel	0	0	0	3	3
Manufacture of chemicals, chemical products	107	34	44	19	204
Manufacture of rubber and plastic products	319	107	124	18	568
Manufacture of non-metallic mineral products	172	55	60	19	306
Manufacture of basic metals	40	17	38	10	105
Manufacture of fabricated metal products	738	246	177	15	1176
Manufacture of machinery and equipment	442	137	167	29	775
Manufacture of office machinery and computers	19	7	7	3	36
Manufacture of electrical machinery and apparatus n.e.c.	132	45	66	47	290
Manufacture of radio, television and communication equipment and apparatus	79	23	30	25	157
Manufacture of instruments	192	38	47	7	284
Manufacture of motor vehicles, trailers and semi-trailers	51	28	37	37	153
Manufacture of other transport equipment	35	11	11	8	65
Manufacture of furniture	273	81	64	6	424
Recycling	25	7	8	0	40
Total	4,303	1,389	1,508	375	7,575

TABLE A2 • **Summary statistics of key variables**

Dummy	Value of variable	
	0	1
Performs innovation	7,319	256
Foreign ownership > 10 %	6,025	1,550
Exports	3,503	4,072

Continuous variables	Number of observations	Average	Median	Standard Deviation
R&D intensity for all firms	7,575	0.001	0.000	0.009
R&D intensity for firms that perform R&D	7,575	0.022	0.007	0.044
Labour productivity	7,575	3.419	2.292	4.834
Added value	7,575	498.25	49.56	7265.98
Capital intensity	7,575	4.318	2.182	7.805

TABLE A3 • **Nonlinear impact of competition on the R&D of firms (lagged explanatory variables)**

Variable	Probit	Tobit	Probit	Tobit	Probit	Tobit
Concentration (C_3)	0.010 (0.008)	6.262* (3.464)				
Concentration² $(C_3)^2$	–0.005 (0.008)	–4.248 (3.522)				
Herfindahl index			0.023** (0.012)	11.838** (4.970)		
Herfindahl index²			–0.030* (0.016)	–16.616** (6.764)		
Average ROA in industry					0.014 (0.042)	8.188 (19.164)
ROA²					0.109 (0.286)	41.334 (130.920)
Labour productivity	0.000* (0.000)	0.093 (0.062)	0.000 (0.000)	0.090 (0.061)	0.000 (0.000)	0.089 (0.061)
Log capital intensity	0.002*** (0.001)	0.685** (0.287)	0.002*** (0.001)	0.697** (0.289)	0.002*** (0.001)	0.715** (0.295)
Size: 25–50	0.027*** (0.006)	5.19*** (1.086)	0.027*** (0.006)	5.18*** (1.091)	0.026*** (0.006)	5.125*** (1.089)
Size: 50–250	0.068*** (0.009)	7.483*** (1.122)	0.067*** (0.009)	7.448*** (1.124)	0.068*** (0.009)	7.457*** (1.135)
Size: 250 <	0.290*** (0.032)	11.118*** (1.698)	0.291*** (0.032)	11.15*** (1.715)	0.3*** (0.032)	11.313*** (1.758)
Export	0.006*** (0.002)	2.518** (1.038)	0.006*** (0.002)	2.582** (1.048)	0.006*** (0.002)	2.559** (1.044)
Foreign ownership > 10%	–0.004*** (0.001)	–1.622** (0.766)	–0.004*** (0.001)	–1.625** (0.762)	–0.004*** (0.001)	–1.585** (0.757)
Constant		–21.67*** (4.035)		–21.105*** (3.880)		–21.195*** (3.990)
Observations	7,125	7,575	7,125	7,575	7,125	7,575
Pseudo R^2	0.358	0.223	0.358	0.223	0.358	0.223
Log Likelihood	–708.3	–1,259	–708.2	–1,260	–708.5	–1,260

Note: The dependent variable of the probit models indicates whether the firm performed R&D activity in 2005. The dependent variable of the tobit models shows the R&D intensity of firms (in %). For the probit models, the Table shows the marginal effects for the sample mean. We calculated competition variables for four-digit NACE industries. Regressions also include two-digit industry dummies. The explanatory variables are from 2003. We clustered standard errors at the industry level.
* Significant at the 10 percent level, ** significant at the 5 percent level, *** significant at the 1 percent level.

TABLE A4 • **Non-linear impact of competition indicators on the R&D activity of firms (with NACE4 quadratic term)**

Competition indicator	Beta	Standard error	Beta squared	Standard error
Number of firms	–0.00098	0.00037**	0.00000	0.00000**
Concentration (C_3) on the basis of net sales	0.02706***	0.00780**	–0.00022	0.00007***
Concentration (C_3) on the basis of total assets	0.02094	0.00739**	–0.00015	0.00006**
Concentration (C_5) on the basis of net sales turnover	0.02647***	0.00868***	–0.00019	0.00007**
Concentration (C_5) on the basis of total assets	0.02457	0.00862**	–0.00017	0.00007**
Concentration (C_{10}) on the basis of net sales	0.02830	0.01130**	–0.00018	0.00008**
Concentration (C_{10}) on the basis of total assets	0.02936	0.01237**	–0.00018	0.00009*
Relative standard deviation of shares on the basis of net sales	0.02166	0.00848**	–0.00020	0.00008**
Relative standard deviation of shares on the basis of total assets	0.01609	0.00798*	–0.00012	0.00007
HHI on the basis of net sales	0.00009	0.00007	–0.00000	0.00000
HHI on the basis of total assets	0.00010	0.00007	–0.00000	0.00000
C_3 on the basis of domestic consumption (hypothesis 1)#	0.01384	0.0967	–0.00007	0.00008
C_3 on the basis of domestic consumption (hypothesis 2)#	–0.00436	0.00823	0.00000	0.00011
C_5 on the basis of domestic consumption (hypothesis 1)#	0.01687	0.01043	–0.00009	0.00009
C_5 on the basis of domestic consumption (hypothesis 2)#	–0.00683	0.00774	0.00005	0.00010
HHI on the basis of domestic consumption (hypothesis 1)#	0.00004	0.00007	0.00000	0.00000
HHI on the basis of domestic consumption (hypothesis 2)#	–0.00026	0.00015	0.00000	0.00000**
Domestic consumption	0.00000	0.00000**	–0.00000	0.00000**
Domestic consumption (% of net sales)	0.00067	0.00040	–0.00000	0.00000
Import categorized on the basis of products (% of domestic consumption)	0.00523	0.00782	0.00002	0.00007
Share of large firms in industry sales	0.00804	0.00504	–0.00006	0.00005
Share of medium firms in industry sales	–0.00584	0.00628	0.00006	0.00007
Share of small and micro firms in industry sales	0.00155	0.00823	–0.00009	0.0012
Share of large firms in industry total assets	0.00648	0.00497	–0.00004	0.00005
Share of medium firms in industry total assets	–0.00785	0.00627	0.00009	0.00007
Share of small and micro firms in industry total assets	–0.00710	0.00766	0.00001	0.00011
Ratio of the turnover of small firms to that of large firms	–0.04793	0.03223	0.00083	0.00107
Share of import in the industry total supply	0.00075	0.00038	–0.00000	0.00000
Number of firms entering the market in the given year	–0.01014	0.00315***	0.00004	0.00002**
Number of firms exiting the market in the given year	–0.01375	0.00485**	0.00009	0.00003**
Ratio of entering firms in year *t*	–0.00792	0.02294	–0.00012	0.00071
Ratio of exiting firms in year *t*	0.02923	0.02922	–0.00041	0.00119
Drop-out rate in year *t*	0.01442	0.01448	–0.00019	0.00024
Net turnover of the sale of dissolved firms in year *t* (% of total industry turnover in year *t*)	–0.06895	0.04960	0.00286	0.00264
Assets of dissolved firms in year *t* (% of total industry assets in year *t*)	–0.05556	0.01707***	0.00143	0.00045***
Net turnover of the sale of new entrants in year *t* (% of total industry turnover in year *t*)	0.00269	0.05794	–0.00268	0.00576
Assets of new entrants in year *t* (% of total industry assets in year *t*)	0.00442	0.03847	–0.00238	0.00272

Competition indicator	Beta	Standard error	Beta squared	Standard error
Profitability of exiting firms compared to the profitability of those that stay in the market	0.00001	0.00001	0.00000	0.00000
Productivity of exiting firms compared to the productivity of those that stay in the market	–0.00194	0.00118	0.00000	0.00001
Number of firms not included in the sample	–0.00393	0.00141**	0.00001	0.00000*
Industrial output price index	–0.26605	0.31791	0.00133	0.00158
Domestic sales price index	0.55400	0.38505	–0.00275	0.00189
Export sales price index	0.00030	0.10754	0.00005	0.00055
EBIT ratio	–0.00124	0.01770	0.00043	0.00077
EBITDA ratio	0.00918	0.02381	–0.00007	0.00080
Return on equity before tax (ROE1)	–0.00439	0.00153**	–0.00000	0.00000**
Return on equity after tax (ROE2)	–0.00430	0.00148**	–0.00000	0.00000**
Balance sheet earnings on equity (ROE3)	–0.00358	0.00165*	–0.00000	0.00000**
Return on capital employed (ROCE)	0.01231	0.01209	–0.00059	0.00037
Return on Sales (ROS)	–0.00700	0.01690	0.00085	0.00069
Return on Investment (ROI)	0.00353	0.00310	–0.00002	0.00004
Return on Asset (ROA)	0.01141	0.01235	–0.00029	0.00094
Return on Invested Capital (ROIC)	0.01106	0.01449	–0.00068	0.00063
Industry loss (% of net turnover)	–0.09287	0.03208**	0.00687	0.00207***
Gross added value per capita	0.01702	0.02602	0.00015	0.00071
Gross added value per unit labour cost	0.00142	0.00223	–0.00000	0.00000
Relative standard deviation of gross added value per capita	0.00741	0.01058	0.00003	0.00014
Relative standard deviation of gross added value per unit labour cost	–0.00891	0.00967	0.00014	0.00012
Simple arithmetic mean of gross added value per capita	0.24613	0.08342***	–0.01921	0.00641***
Simple arithmetic mean of gross added value per unit labour cost	–0.00040	0.00030	0.00000	0.00000***
Total factor productivity (TFP) in industry	0.00926	0.05413	–0.00175	0.00278
Relative standard deviation of total factor productivity	–0.00030	0.00997	–0.000003	0.00016
Simple arithmetic mean of total factor productivity of firms in industry	0.01017	0.01809	–0.00006	0.00020
Productivity of smaller firms compared to that of large firms	–0.00191	0.00474	–0.00002	0.00007
Numerator of the indicator of the relationship between profitability and productivity	0.00260	0.00598	0.000000	0.00005
Denominator of the indicator of the relationship between profitability and productivity	–0.00135	0.02207	0.00028	0.00094
Export share in industry total demand	0.00042	0.00067	0.00000	0.00000
Renewal of assets on the basis of implemented investments	–0.00570	0.01742	0.00036	0.00036
Rate of foreign ownership in subscribed capital	0.01691	0.00693**	–0.00015	0.00007**
Net turnover of sales in industry	0.00000	0.00000	–0.00000	0.00000
Size of industry	0.00081	0.00202	–0.00000	0.00001
Cost disadvantage ratio	0.00470	0.00456	–0.00003	0.00003

Note: For each variable, the probit model referred to in *Table A3* was estimated. The table shows the average marginal effect of the competition indicator at the sample average. Competition variables were calculated for four-digit NACE industries. Regressions also include two-digit industry dummies. The explanatory variables are from 2003. We clustered standard errors at the industry level.
For the description of the two hypotheses, see: http://www.gvh.hu/gvh/alpha?do=2&st=1&pg=54&m5_doc=5635&m251_act=4.
* Significant at the 10 percent level, ** significant at the 5 percent level, *** significant at the 1 percent level.

• *Ferenc László Kiss* •

INTER-FIRM COMPARISON AND DECOMPOSITION OF PRODUCTIVITY GAINS FOR REGULATORY PURPOSES*

In this study, temporal changes in firm-level productivity, referred to as "corporate productivity gains" are decomposed into causal factors, such as cost-saving technological changes, output growth in the presence of economies of scale, changes in input prices, and the effect of input price changes on the firm's demand for inputs. The decomposition is then applied to inter-firm comparisons of productivity. Inter-firm differences in productivity gains are decomposed into the same causal factors as the productivity gains themselves. Following a brief description of the economic concepts and variables that are associated with the concept of productivity, an empirical study – a comparative analysis and decomposition of productivity gains in two real-life regulated companies – offers an opportunity to assess the practical problems of measurement and comparison, and to introduce some useful indexing and econometric tools. The empirical study demonstrates that inter-firm productivity comparison and decomposition can indeed be successfully achieved, and that they can play a useful role not only in corporate management, but also in the regulation of imperfect markets.

THE COMPARATIVE ANALYSIS OF PRODUCTIVITY

To a considerable extent, the operating conditions of companies that sell their products on imperfect, regulated markets are created, influenced and managed by the regulators themselves. This involves tremendous social responsibilities, which include not only the task of preventing regulated firms from rent-seeking anti-competitive practices and various other socially detrimental activities that may arise from their market power, privileges, etc., but also the duty of doing everything in their power to facilitate the efficient operation of the firms they regulate in order to ensure the supply of

* I am indebted to three excellent colleagues and dear friends of mine, *Bernard J. Lefebvre, Robert E. Olley* and *Shafi A. Shaikh.* The present work draws from a major Canadian project, which was carried out under Olley's chairmanship and my technical leadership. At the request of the Canadian Ministry of Communications, I investigated productivity comparisons within the framework of this project. I worked extensively with Shaikh on turning "dirty" accounting and managerial data into meaningful economic variables, and with Lefebvre on building various models of productivity decomposition. This article re-visits some of the results of my work. It emphasizes the usefulness of productivity comparisons and decomposition, hoping that it would help re-focus attention on the all-important but lately somewhat ignored topic of productivity, and introduce more quantitative analytical tools into economic analysis in my native Hungary.

regulated markets with the widest possible selection of products and services of the best possible quality at the lowest possible cost. These requirements apply to monopolies as well as any kind of imperfectly competitive markets. In order to fulfil their task, regulators are expected by society at large to have a thorough knowledge of not only the demand but also the supply side of the markets they regulate. The economic efficiency and financial well-being of regulated companies must be monitored, analysed, evaluated and corrected if necessary. The regulator cannot possibly accomplish this if it is not known – among other things – how productive the regulated firms are, how quickly their productivity improves over time, what the causes of their productivity gains are and how much improvement can reasonably be attributed to each cause. Productivity measurement and analysis are important regulatory tasks.

Managers of both regulated and unregulated companies also show a keen interest in productivity as one of the two main *endogenous* factors that determine corporate profits. It was recognised as early as before World War II that the size of, and changes in, corporate profit depend rather crucially on management decisions concerning changes of productivity and output prices. Productivity studies became one of the basic tools of short-term operational planning and budgeting as early as in the late 1960s. Productivity indices also became *ex ante* targets in addition to being *ex post* attributes of corporate performance. Budget- or plan-implicit productivity gains were derived to show how the fulfilment of the annual corporate budget or plan would improve productivity and profits. Assessments of the reality and reasonableness of the budget- or plan-implicit productivity gain resulted in annual productivity targets, and budgets or plans were modified if necessary to meet the productivity target. Simple measurements of actual annual productivity gains and *ad hoc* data analyses were no longer sufficient. Target setting required more knowledge. Corporate analysts were increasingly turning to sophisticated econometric models to understand the causes and consequences of productivity improvements.

Traditional managerial and regulatory knowledge of corporate productivity was based on simple index numbers showing annual changes in the relationship between the volumes of inputs and the outputs they were producing. Productivity measures were compared to each other in various ways. *Ad hoc* comparisons were made most often within the boundaries of a given firm, involving a comparison between its own past and present, or past and future, or present and future performance. These are the *intra-firm comparisons.* Comparisons with past performance are indispensably useful but may also be misleading. From a regulatory point of view, the greatest risk in comparing productivity performances over time is that the comparison may lead to an erroneous assessment of the economic performance or productive capability of a regulated firm. It has been a common problem throughout the history of regulation, that the regulator (and/or the management of the regulated firm) simply and mechanically assumed that the expected growth rate in the firm's productivity should be equal to some (usually the average) past growth rate. This is tantamount to disregarding changes in operating conditions and their impact on productivity. When future conditions differ from past conditions, the productivity growth in the future will also differ from that in the past.

Productivity comparisons are often made with other firms. *Ad hoc inter-firm* comparisons can accomplish more than revealing where the productivity growth rate was higher, lower or the same in any given period. They may point to various causes and consequences of observed inter-firm differences but also carry a considerable risk of mistaken conclusions. Superficial comparisons are often built upon an implicit underlying assumption; *viz.,* if a comparable other firm has achieved a certain rate of output growth then a similar rate of improvement ought to be expected at one's own firm. This assumption may turn out to be correct or incorrect, but it is definitely harmful if the analysts carrying out the comparison do not explore in sufficient detail the factors affecting productivity, the differences between them and the effects of those differences.

While *ad hoc* comparisons may indeed direct the attention of the analyst toward some factors that affect productivity, they do not allow the quantification of their effects. Quantitative analyses of the thus identified effects cannot be made. For this reason, comparisons should not be made without proper decomposition. Decompositions not only *identify* the causes of improvements in productivity but also *quantify* their effects. Changes in productivity result from the combined effect of a number of economic variables such as growth in the output of the firm or cost saving technological changes. What makes the analysis particularly revealing and useful is that the total effect on productivity of each causal variable can be broken down into two components. One component is the magnitude or *rate of change* in the size of the variable. If output growth improves productivity then it will make a difference whether the output growth rate is 3 percent or 10 percent. Secondly, the variables that affect productivity produce their impact with certain *intensity.* The higher the intensity the greater the effect on productivity. Decompositions distinguish changes in the *size* of an explanatory variable from the *intensity* with which the explanatory (causal) variable influences productivity.

The econometric models that began to emerge in the 1970's facilitated a major new development in productivity analysis. This paper extends the method of analysing corporate productivity performance by using econometric models. It advances in two directions. First, a *joint multi-firm* econometric cost model is constructed for two (and possibly more) firms, whose productivity performances are to be compared. Second, using the estimated parameters of this joint cost model, a *causal decomposition* of inter-firm productivity differences is conducted, quantifying the inter-firm difference that is due to each causal factor. As we shall see in the next sections, the most important causes are cost-saving technological changes and the exploitation of economies of scale, when the volume of production increases.[1]

[1] It is also possible to decompose temporal changes in productivity by *consequence*.

Decomposition also plays a key role in *forecasting* productivity. It is equally important for the regulator and the regulated firm to have an idea of how much improvement in productivity they can reasonably expect as a consequence of expected future operating conditions, or some specific forthcoming change in the operation of the firm. Mergers and acquisitions, among other things, typically generate various productivity-altering organisational and other changes in the operations of affected firms.

TEMPORAL CHANGES IN PRODUCTIVITY

We look at changes in corporate *total factor productivity.* As the name implies, this concept recognizes the firm's output as the result of the combined productive services (inputs) of *all* of the firm's factors of production. Changes in total factor productivity may occur in time or space. Temporal changes refer to progress in total factor productivity within a given firm between two points or periods in time. Temporal changes are usually referred to as *productivity gains.* Spatial changes, on the other hand, show the difference between the total factor productivities of two firms. Both types of changes are measured by *proportional volume changes*, defined as the natural logarithms of input and output *volume indices.*

Let us first define the variables used in the measurement of productivity![2] The temporal *proportional change* of total factor productivity (the productivity gain) of a firm is defined as

$$\dot{\phi} = \dot{q} - \dot{x}, \tag{1}$$

where $(\dot{q})$ denotes the temporal continuous proportional change in total output, and $(\dot{x})$ is the temporal continuous proportional change in total input. Temporally *continuous* proportional changes are expressed as Divisia volume indices. For the output, the time-continuous Divisia volume index is

$$\dot{q} = \sum_{i=1}^{n} \frac{P_i Q_i}{R} \dot{q} = \sum_{i=1}^{n} r_i \dot{q}_i, \tag{2}$$

where n outputs exist,

$\dot{q}_i = \dfrac{dQ_i}{dt}\dfrac{1}{Q_i} = d\ln Q_i$ is the temporal proportional change in the i-th output,

$R = \sum_{i=1}^{n} P_i Q_i$ is total revenue and r_i stands for the i-th output's revenue share.

[2] In addition to total factor productivity, the concept of productivity may be expanded to the so-called *partial productivity* measures. These are termed "partial" because they show the relationship between the firm's total output and *only one* category of its factor inputs. The most frequently investigated partial productivity measures are labour and capital productivity, but measures for material and sometimes even for other input categories also exist. Inter-firm comparisons and decompositions of partial productivity measures are not investigated in this article.

In practice, the analyst is forced to work with data that show *discrete* temporal changes between two points or periods in time. For this reason, a discrete approximation must be found for the time-continuous Divisia indices. The Törnqvist index is such an approximation. With discrete changes in outputs from a given period $(t - 1)$ to period t, the Törnqvist output volume index is

$$\dot{q}_i = \sum_{i=1}^{n} \bar{r}_{it} \ln\left(\frac{Q_{it}}{Q_{i,t-1}}\right), \tag{3}$$

where $\bar{r}$ shows each i-th individual output's average revenue share, which is the simple arithmetic mean of its revenue shares in the two compared periods, that is

$$\bar{r}_{it} = \frac{r_{it} + r_{i,t-1}}{2}. \tag{4}$$

Inputs are treated in the same manner as outputs. The temporally continuous proportional change in total input is expressed by the Divisia volume index

$$\dot{x} = \sum_{j=1}^{m} \frac{W_j X_j}{C} \dot{x}_j = \sum_{j=1}^{m} s_j \dot{x}_j, \tag{5}$$

where m inputs exist,

$\dot{x}_j = \dfrac{dX_j}{dt}\dfrac{1}{X_j} = d \ln X_j$ is the temporal proportional change in the j-th input,

$C = \sum_{j=1}^{m} W_j X_j$ is total cost, and s_j denotes the j-th input's cost share.

Its discrete approximation, the Törnqvist input volume index describes the change in total input from period $(t - 1)$ to period t as

$$\dot{x}_t = \sum_{j=1}^{m} \bar{s}_{jt} \ln\left(\frac{X_{jt}}{X_{j,t-1}}\right), \tag{6}$$

where $\bar{s}$ denotes the average cost share of the j-th input, that is[3]

$$\bar{s}_{jt} = \frac{s_{jt} + s_{j,t-1}}{2}. \tag{7}$$

As stated above, proportional volume changes are defined as the natural logarithms of input and output *volume indices*. This offers a choice of representation to the analyst. Indeed, some authors use index numbers, and not proportional changes, for the measurement of changes in input and output volumes. Expressed with the

[3] For partial productivity measures, the Divisia and Törnqvist output volume indices remain the same as in equations (2) and (3), respectively, but the input volume indices (equations (5) and (6), respectively), must be re-defined. The W_j input prices and X_j input volumes (and consequently the $\bar{s}_j$ cost shares and $\dot{x}_j$ proportional input changes) refer only to the individual labour, capital, or material inputs that are included in the partial measure.

aid of volume indices, temporal change in the total factor productivity of the firm, the *productivity index,* becomes

$$\hat{\Phi} = \frac{\hat{Q}}{\hat{X}}, \tag{8}$$

where $\hat{Q}$ and $\hat{X}$ are the temporal volume indices of total output and total input, respectively. For discrete changes between the consecutive periods $(t-1)$ and t, the Törnqvist volume indices of output and input are

$$\hat{Q}_t = \prod_{i=1}^{n} \left[\frac{Q_{it}}{Q_{i,t-1}}\right]^{\bar{r}_{it}}, \tag{9}$$

and

$$\hat{X}_t = \prod_{j=1}^{m} \left[\frac{X_{jt}}{X_{j,t-1}}\right]^{\bar{s}_{jt}}. \tag{10}$$

The Törnqvist volume indices of output and input are the weighted geometric means of the ratios of volume changes occurring in two consecutive periods in all outputs and in all inputs, respectively. The respective weights are the average revenue shares of individual outputs, and the average cost shares of individual inputs.

The temporal proportional change in total factor productivity, the *productivity gain* is defined as the natural logarithm of the *productivity index; i.e.,*

$$\dot{\phi} = \ln \dot{\Phi} = \ln \hat{Q} - \ln \hat{X}. \tag{11}$$

INTRA-FIRM DECOMPOSITION

Causal decompositions may be carried out within one firm. The phenomena that cause total factor productivity to change are identified and classified, and a certain part of the productivity gain is assigned to each cause. Successful identification of the causes requires extensive knowledge of certain basic economic characteristics of the firm's production process. Such knowledge may be derived from econometric production models.

Production functions, various forms of cost functions and profit functions may be selected as models. This study works with *total cost* functions. A cost function is a suitable analytical vehicle for estimating and describing the basic economic characteristics of the production process that are needed for productivity analysis. Production functions are not used here because they do not allow us to analyse the economic characteristics of multi-output production processes. Profit functions are omitted due to data problems.

For simplicity, and strictly for introductory purposes, intra-firm decomposition is attempted first with the aid of *single-output* cost functions. These are the simplest possible models we can use to demonstrate the essence of the decomposition in a transparent manner. Once we gain a basic insight, we switch to multi-output cost function specifications to gain a more detailed understanding of the various impacts on productivity of different categories of output.

The single-output case

It is common knowledge that a total cost function is capable of describing the production technology of a firm. The simplest general form of a total cost function assumes that a single output is produced, and that technological changes are exogenous; *i.e.,*

$$C = g(W_1, \ldots, W_m, Q, T), \tag{12}$$

where C is the total economic cost of production, m inputs are employed by the firm, W denotes input prices, and T represents exogenous technological changes, whose measurement will be discussed later, when we describe our empirical models.

Technological changes cause increases in productivity by saving costs.[4] The temporal productivity gain generated by technological changes, $(\dot{B})$, which we shall henceforth term the *technological effect,* is equal in size (but opposite in sign) to the temporal *shift* of the cost function generated by the technological changes,

$$\dot{B} = -\frac{\partial \ln C}{\partial t}. \tag{13}$$

This effect can be estimated using the elasticity of the total cost with respect to technological changes. The estimated cost function yields estimates of this elasticity, while the proportional change in technology can be expressed using data on technological changes.[5]

$$\dot{B} = -\left[1 - \frac{\partial \ln C}{\partial \ln T}\right] \frac{d \ln T}{dt} = \varepsilon_{CT} \dot{T}. \tag{14}$$

Output causes increases in productivity, when it grows in the presence of economies of scale.[6] The temporal productivity gain generated by output growth, $(\dot{E})$, which we shall henceforth term the *output effect,* therefore depends on the degree of economies of scale and the growth rate of output. Economies of scale constitute a basic technological property of the firm. Their degree can be derived from the estimated total cost function, where it is the inverse of the output elasticity of cost $\left(\varepsilon_{QX} = \varepsilon_{CQ}^{-1}\right)$. Increases in the volume of output contribute to increases in productivity, when $0 < \varepsilon_{CQ} < 1$. The proportional change in output is calculated from the firm's output data. The output effect is therefore

$$\dot{E} = \left[1 - \frac{\partial \ln C}{\partial \ln Q}\right] \frac{d \ln Q}{dt} = \xi_{CQ} \dot{q}, \tag{15}$$

where $\xi_{CQ} = 1 - \varepsilon_{CQ,}$ and $\varepsilon_{CQ} = \partial \ln C / \partial \ln Q$ is the output elasticity of cost.[7]

[4] The technological effect reflects the immediate and short-term cost-saving effects of technological changes. However, technological changes also may have long-term cost-saving effects, which gradually emerge over time with or without increases in the scale of production. We shall return to these effects when discussing estimation results.

[5] Notice that $\varepsilon_{CT} = -\partial \ln C / \partial \ln T$. Since the elasticity is always negative, ε_{CT} is always positive if technological changes reduce cost.

[6] Or it decreases in the presence of diseconomies of scale. This case, however, will not be discussed here.

[7] Notice that $\xi_{CQ} = 1 - \partial \ln C / \partial \ln Q$ is positive when there are economies of scale; *i.e.,* when $0 < \partial \ln C / \partial \ln Q < 1$.

Equation (15) shows that the proportional temporal productivity change, the productivity gain, of a firm, represented by its total cost function, can be expressed as the sum of the products of *1.* the temporal proportional change in each independent variable affecting productivity (and the cost) and *2.* the cost elasticity with respect to the same independent variable. If the cost function is in the form of a regression equation containing an error variable, decomposition proceeds as follows:

$$\dot{\phi} = \varepsilon_{CT}\dot{T} + \xi_{CQ}\dot{q} + \dot{R}, \tag{16}$$

where $\dot{R}$ represents the *residual* productivity change; *i.e.,* that portion of the observed (actual) productivity change, which is not explained by the first two expressions on the right-hand side of equation (16). This equation decomposes the change in productivity into causal components. This is the simplest form of decomposition. The productivity gain is caused by three effects: *1.* technological effect, *2.* output effect, and *3.* residual effect:

$$\dot{\phi} = \dot{B} + \dot{E} + \dot{R}. \tag{17}$$

As stated above, in this simplest possible case, the firm's technology is described by a cost function with a single output and a single exogenous technological change. When technology is represented in a more elaborate fashion, in more detail and with greater precision, the decomposition of the productivity gain also becomes more elaborate. When it is recognized in the specification of the cost function, that the firm produces more than one output, instead of the single output effect shown in equation (17), as many output effects are generated as the number of outputs; *i.e.,*

$$\dot{\phi} = \dot{B} + \sum_{i=1}^{n} \dot{E}_i + \dot{R}, \tag{18}$$

where the individual output effects $\dot{E}_i$ appear as the product of *1.* the proportional change in output *i* and *2.* the cost elasticity with respect to the same output *i,*

$$\dot{E}_i = \xi_{CQi}\dot{q}_i. \tag{19}$$

The explicit inclusion of complex technological changes in the model complicates the decomposition of productivity gains in a variety of ways, especially if input-neutral (Hicksian) technological changes appear contemporaneously with changes that affect the input structure and/or result in input or output augmentation. Such complex models are outside the scope of this study.

The multi-output case

Most firms produce more than one product or service. When there is more than one output, changes in the output structure presumably have an effect on productivity. In order to be able to estimate the effects of output structure in addition to the effects discussed above, a multi-output cost model must be introduced into the analysis. The single output in the cost function in equation (12) is now replaced by n number of individual output variables:

$$C = g(W_1, \dots, W_m, Q_1, \dots, Q_n; T). \tag{20}$$

By taking the total time derivative of this cost function, introducing Shepard's lemma whereby $\partial g/\partial W_j = X_j$, and rearranging the resulting equation, we obtain

$$\dot{B} = \sum_{i=1}^{n} \varepsilon_{CQi}\dot{q}_i - \sum_{j=1}^{m} s_j \dot{x}_j, \tag{21}$$

where $\dot{B}$ expresses the extent of the shift in the cost function that is caused by the technological change, and $\varepsilon_{CQi} = \frac{\partial g}{\partial Q_i}\frac{Q_i}{C} = \frac{\partial \ln C}{\partial \ln Q_i}$ denotes the elasticity of cost with respect to the i-th output.

Comparing equation (21) with the Divisia indices in equations (1), (2) and (5), it becomes obvious that the temporal cost shift expresses the productivity gain without distortion if $\varepsilon_{CQi} = r_i \; \forall i$, that is, when the revenue share of each output equals the cost elasticity with respect to the same output, resulting in

$$\frac{P_i Q_i}{R} = \frac{\partial g}{\partial Q_i}\frac{Q_i}{C}. \tag{22}$$

This equality can materialise only if *1.* the price of each output equals its marginal cost ($P_i = MC_i = \partial g/\partial Q_i, \forall i$), and *2.* total revenue is equal to total cost ($R = C$). Both conditions are met if the production process is characterised by constant returns to scale and if the input and output markets of the firm are perfectly competitive. The output markets of regulated firms cannot, however, be said to be perfectly competitive. Furthermore, regulated firms tend to be engaged in network-based production, therefore their technology may well be characterized by economies of scale. It is precisely these properties that warrant their regulation. For this reason, the productivity gains of regulated firms may differ in size from the temporal proportional shifts in their cost. Using equation (22), the difference can be expressed as

$$\dot{\phi} - \dot{B} = \sum_{i=1}^{n} \left[\frac{P_i Q_i}{R} - \frac{MC_i Q_i}{C}\right] \dot{q}_i. \tag{23}$$

The productivity gain can be expressed in a very instructive way from equation (23) if we add and subtract $(P_i Q_i)/C$, and rearrange the right-hand side of the equation.[8] We obtain

$$\dot{\phi} = \sum_{i=1}^{n} \left[(P_i - MC_i)\frac{Q_i}{C}\right] \cdot \dot{q}_i + \sum_{i=1}^{n} [(P_i Q_i)(R^{-1} - C^{-1})] \cdot \dot{q}_i + \dot{B}. \tag{24}$$

[8] The consequences of non-marginal cost pricing were first analysed by *Denny, Fuss & Everson* [1979].

For a multi-output model, this is the starting point of the decomposition of productivity gains. The equation demonstrates that the productivity gain equals the change in cost $(\dot{\phi} - \dot{B})$ if the firm has constant returns to scale and marginal cost pricing. $MC_i = 0$, but $R - C < 0$ if the firm has marginal cost pricing but there are economies of scale. $P_i - MC_i \neq 0$, but $R = C$ if the firm has some sort of cost-covering zero-profit constraint, as in the case of average cost pricing or Ramsey pricing. The first and second items on the right-hand side of the equation are activated if there are economies of scale and – for this reason or independently from this – prices do not equal marginal costs.

The concept of average cost may be meaningful for some firms if the great majority of their production costs are output specific. If average cost is meaningful then useful information can be generated by decomposing the *price – marginal cost* difference into a *price – average cost* and an *average cost – marginal cost* difference; *i.e.,*

$$P_i - MC_i = (P_i - AC_i) + (AC_i - MC_i). \tag{25}$$

Substituting equation (25) into equation (24) the following formula is obtained:

$$\begin{aligned}\dot{\phi} = \sum_{i=1}^{n}\left[(P_i - AC_i)\frac{Q_i}{C}\right]\cdot \dot{q}_i + \sum_{i=1}^{n}\left[(AC_i - MC_i)\frac{Q_i}{C}\right]\cdot \dot{q}_i + \\ + \sum_{i=1}^{n}[(P_iQ_i)(R^{-1} - C^{-1})]\cdot \dot{q}_i + \dot{B}.\end{aligned} \tag{26}$$

The rather lengthy and complicated structure of this equation may be simplified by replacing the three multipliers of $\dot{q}_i$ on the right-hand side by ZA_i, ZM_i and ZR_i, as in

$$\dot{\phi} = \sum_i ZA_i\dot{q}_i + \sum_i ZM_i\dot{q}_i + \sum_i ZR_i\dot{q}_i + \dot{B}. \tag{27}$$

SPATIAL CHANGES IN PRODUCTIVITY

As suggested in the introductory section, there are two kinds of inter-firm productivity comparisons. On the one hand, the productivity "levels" of two firms may be compared by composing spatial volume indices, showing which one is more productive and quantifying the difference between their productivity "levels". On the other hand, each of the compared firms has a time series of temporal productivity gains, and these can be compared as well by composing temporal volume indices for both, showing which one improves productivity faster and quantifying the relationship between their "speeds". The first case is that of *level comparison,* and the second case is that of *speed comparison.*

Let us take a quick look at the spatial indices! The spatial index of productivity is defined analogously with the definition of the temporal index. The only difference is that in a spatial index, the price and volume data for periods $(t-1)$ and t are replaced by the data obtained for Firms A and B. Thus, the spatial proportional change (difference) in productivity is

$$\dot{\phi}_{AB} = \sum_{i=1}^{n} \frac{1}{2}\left[\frac{P_{iA}Q_{iA}}{\sum_{i=1}^{n} P_{iA}Q_{iA}} + \frac{P_{iB}Q_{iB}}{\sum_{i=1}^{n} P_{iB}Q_{iB}}\right] \ln\left(\frac{Q_{iA}}{Q_{iB}}\right) - \\ - \sum_{j=1}^{m} \frac{1}{2}\left[\frac{W_{jA}X_{jA}}{\sum_{i=1}^{n} P_{iA}Q_{iA}} + \frac{W_{jB}X_{jB}}{\sum_{j=1}^{m} W_{jB}X_{jB}}\right] \ln\left(\frac{X_{jA}}{X_{jB}}\right). \tag{28}$$

The measurement of spatial differences in productivity and their inter-firm decomposition are outside the scope of this study.[9] Our task is limited to "speed comparisons" and the decomposition of inter-firm differences in temporal productivity gains.

The difference between the productivity gains of two firms (A and B) can be expressed as

$$\Delta\dot{\phi}_t = \dot{\phi}_{At} - \dot{\phi}_{Bt} = (\dot{q}_{At} - \dot{x}_{At}) - (\dot{q}_{Bt} - \dot{x}_{Bt}) = \Delta\dot{q}_t - \Delta\dot{x}_t, \tag{29}$$

where t refers to productivity gain in period t, showing change relative to t–1.

INTER-FIRM DECOMPOSITION

Equation (16) demonstrates the simplest decomposition of productivity gains. It shows that temporal increases in productivity are generated by three effects: *1.* a technology effect, *2.* an output effect and *3.* a residual effect. When two firms, A and B, are compared, this kind of decomposition can be performed for the temporal productivity gains of both firms. Furthermore, it is also possible to *decompose the differences* in the productivity gains of the two firms as

$$\Delta\dot{\phi}_t = \dot{\phi}_A - \dot{\phi}_B = \left(\dot{B}_A - \dot{B}_B\right) + \left(\dot{E}_A - \dot{E}_B\right) + \left(\dot{R}_A - \dot{R}_B\right) = \Delta\dot{B} + \Delta\dot{E} + \Delta\dot{R}. \tag{30}$$

Temporal productivity gains always refer to specified time periods, such as years. In this study we use annual data. The productivity *gain* of year t is understood as the productivity *level* in year t, expressed as a change over the productivity *level* in the previous year $t-1$. The comparison and decomposition are also done annually. However, for simplicity, the references to time have been omitted from the following lengthy decomposition equations of this section.

Just as the Divisia indices that assume temporally continuous changes were approximated for discrete changes by Törnqvist indices, we once again need

[9] The spatial comparison of productivity "levels" of firms raises many severe practical problems. It is very difficult, sometimes impossible, to ensure consistency and comparability for the technology, input and output data of different firms. *Kiss* [1984] discussed some of the problems of measurement and comparison.

a discrete approximation of the continuous changes assumed in equation (30). *Kiss* [1981], [1983] discussed some issues of discrete approximations. The simplest possible solution is recommended in the current study. For differences in technological effects, the equation becomes

$$\dot{B}_A - \dot{B}_B = \overline{\dot{T}}\Delta\varepsilon_{CT} + \overline{\varepsilon}_{CT}\Delta\dot{T} = \frac{(\dot{T}_A + \dot{T}_B)}{2}(\varepsilon_{CT_A} - \varepsilon_{CT_B}) + \frac{(\varepsilon_{CT_A} + \varepsilon_{CT_B})}{2}(\dot{T}_A - \dot{T}_B). \tag{31}$$

We approximate the difference in output effects in a similar way; *i.e.,*

$$\dot{E}_A - \dot{E}_B = \overline{\dot{Q}}\Delta\xi_{CQ} + \overline{\xi}_{CT}\Delta\dot{Q} = \frac{(\dot{Q}_A + \dot{Q}_B)}{2}(\xi_{CQ_A} - \xi_{CQ_B}) + \frac{(\xi_{CQ_A} + \xi_{CQ_B})}{2}(\dot{Q}_A - \dot{Q}_B). \tag{32}$$

Changes in input prices also play a role in the productivity performance of firms by influencing the firm's demand for factor inputs. Let us investigate this role! *Temporal* proportional changes in individual inputs (such as labour, capital and materials) can be decomposed into causal components in a similar fashion. The decomposition of change in the j-th input is

$$\dot{X}_{jt} = \varepsilon_{jT}\dot{T} + \varepsilon_{jQ}\dot{Q} + \sum_{i=1}^{m}\varepsilon_{ji}\dot{W}_i + \dot{R}_j, \tag{33}$$

where ε_{jT}, ε_{jQ} *and* ε_{ji} show the elasticity of the j-th input with respect to technology, output and the i-th input price, respectively, and $\dot{W}_i = d\ln W_i/dt$ denotes the proportional change in the i-th input's price. Simplifying equation (33), we obtain equation (34) as

$$\dot{X}_j = \dot{B}_j + \dot{E}_j + \sum\dot{W}_j + \dot{R}_j. \tag{34}$$

In this case, the *spatial* decomposition of inter-firm difference becomes

$$\begin{aligned}\Delta\dot{X}_j = \dot{X}_{jA} - \dot{X}_{jB} = (\dot{B}_{jA} - \dot{B}_{jB}) + (\dot{E}_{jA} - \dot{E}_{jB}) + (\sum\dot{W}_{jA} - \sum\dot{W}_{jB}) + \\ +(\dot{R}_{jA} - \dot{R}_{jB}) = \Delta\dot{B}_j + \Delta\dot{E}_j + \sum\dot{W}_j + \Delta\dot{R}_j.\end{aligned} \tag{35}$$

This equation shows that the difference between the two firms with respect to the proportional change in the volume of the j-th input can be decomposed into *1.* a technological effect, *2.* an output effect, *3.* an input-price effect and *4.* a residual effect. The input-price effect $\Delta\sum\dot{W}_j = (\sum\dot{W}_{jA} - \sum\dot{W}_{jB})$ can be further divided into an own-price effect $\varepsilon_{jj}\dot{W}_j$ and $(m - 1)$ number of cross-price effects (when $j \neq i$).

Now we return to productivity. It has been shown that if the decomposition of temporal proportional changes in productivity as defined by equation (27) can be performed for both firms, the two temporal decompositions can be used to identify the inter-firm differences in causal components. Where there is more than one output, the first three additive factors on the right-hand side of equation (27) must be repeated for each output. Assuming for simplicity that both companies produce two outputs (α and β) and using equation (27) as a point of departure, the inter-firm decomposition of the sources of temporal changes takes the following, rather lengthy, form:

$$\begin{aligned}\Delta\dot{\phi} = \dot{\phi}_A - \dot{\phi}_B = (ZA_{\alpha A}\dot{q}_{\alpha A} - ZA_{\alpha B}\dot{q}_{\alpha B}) + (ZM_{\alpha A}\dot{q}_{\alpha A} - ZM_{\alpha B}\dot{q}_{\alpha B}) + \\ +\left(ZA_{\beta A}\dot{q}_{\beta A} - ZA_{\beta B}\dot{q}_{\beta B}\right) + \left(ZM_{\beta A}\dot{q}_{\beta A} - ZM_{\beta B}\dot{q}_{\beta B}\right) \\ +(ZR_{\alpha A}\dot{q}_{\alpha A} - ZR_{\alpha B}\dot{q}_{\alpha B}) + \left(ZR_{\beta A}\dot{q}_{\beta A} - ZR_{\beta B}\dot{q}_{\beta B}\right) + \left(\dot{B}_A - \dot{B}_B\right)\end{aligned} \tag{36}$$

Equation (31) is particularly instructive because it shows how the $\dot{B}_A - \dot{B}_B$ difference can be decomposed into two causal components: *1.* the difference between the proportionate changes in the variables and *2.* the difference between their cost elasticities. The same decomposition can be performed for the remaining differences. The seven differences contained in equation (36) thus decompose the inter-firm difference in productivity gain into a total of nine well-defined and clearly characterised components. The tenth component is the inter-firm difference due to the unexplainable residual effect. The categories emerging from the decomposition are summarised in *Table 1.*

TABLE 1 • **Components of inter-firm differences in productivity gains**

Explanatory factor	Formula
1. An increase in the production of output α	$\Delta\dot{q}_\alpha\left(\overline{Z}A_\alpha + \overline{Z}M_\alpha + \overline{Z}R_\alpha\right)$
2. Economies of scale specific to output α	$\Delta ZM_\alpha\overline{\dot{q}}_\alpha$
3. The non-average-cost price of output α	$\Delta ZA_\alpha\overline{\dot{q}}_\alpha$
4. Increase in the production of output β	$\Delta\dot{q}_\beta\left(\overline{Z}A_\beta + \overline{Z}M_\beta + \overline{Z}R_\beta\right)$
5. Economies of scale specific to output β	$\Delta ZM_\beta\overline{\dot{q}}_\beta$
6. The non-average-cost price of output β	$\Delta ZA_\beta\overline{\dot{q}}_\beta$
7. Profit/loss due to non-cost-covering pricing	$\Delta ZR_\alpha\overline{\dot{q}}_\alpha + \Delta ZR_\beta\overline{\dot{q}}_\beta$
8. Technological changes	$\Delta\dot{T}\overline{\varepsilon}_{CT}$
9. Technology elasticity of cost	$\Delta\varepsilon_{CT}\overline{\dot{T}}$
10. Residual increase in productivity	—

Changes in the output structure of Firm A may differ from the changes in the output structure of Firm B. The differences result from the diverging growth rates of outputs α and β in the two companies. Their effects are shown in Items 1 and 4 of the table. We may, however, also give a broader interpretation to the effects of changes in output structure. The first three items show the total effect of output α on inter-firm differences of productivity gains, while the next three items give the same information for output β. Item 7 equals zero if the two firms use cost-covering average cost pricing. If the prices of one of the outputs exceed its average cost and generate surplus revenue beyond the cost while the prices of the other output are lower than the average cost and therefore generate a loss, *i.e.,* internal cross-subsidisation takes place, the total effect of non-average-cost pricing will be the sum of Items 3, 6 and 7. With two outputs, this may be very simple; *e.g.,* one output – say β – may generate a profit while the other output – say α – may generate a loss for both firms. If we look at Firm A or Firm B separately, the loss effect is due to the non-average-cost prices of output *α,* and the profit effect is due to the non-average-cost prices of output *β.* The definitions used for the decomposition can be modified so that they reflect this situation. However, we cannot "allocate" the profit-loss effect if more than two outputs exist, unless we obtain more information than what can be reasonably assumed to be available, and construct a more complex multi-output model.

A REPRESENTATIVE INTER-FIRM COMPARISON[10]

In the remainder of this article it is demonstrated – with the aid of *actual firm-level data* on output and input prices and volumes as well as technological changes – how inter-firm comparisons and decompositions of annual productivity gains can be successfully conducted. Two firms have been chosen for the empirical study, mainly because their technologies were sufficiently similar, and their data were publicly available. Since the purpose of this demonstration is the illustration of the process, their names, locations, and the chosen period of observation are not revealed.

Let us introduce the data! The two companies make the same products. The output volume of Firm *B* surpassed that of Firm *A* by a great deal but *A*'s growth rate (14 percent per annum on average) was substantially higher than *B*'s (8 percent per annum on average). A relatively fast process of catching up is witnessed. The output growth rates are displayed in Table *A1* in the *Appendix*. At the start of the observation period, Output *α* had a 33 percent revenue share in Firm *A*, and 57 percent in Firm *B*. During the observation period, this revenue share declined to 30 percent in Firm *A* and to 48 percent in Firm *B,* while the share of Output *β* increased from 62 percent to 66 percent in Firm *A,* and from 37 percent to 48 percent in Firm *B,* demonstrating that Firm *B* underwent a faster structural change than Firm *A.* The third Output *g* had low revenue shares, which did not change significantly over time; it decreased from 5 percent to 4 percent in both companies. Most of the output effects were therefore generated by Outputs *β* and *α*. *Figure A1* in the *Appendix* shows that very different forces acted upon the markets of the two companies. During the 12-year period of observation, there were only seven years when the growth of the output of the two companies accelerated or decelerated in parallel.

A phenomenon of some importance with respect to productivity performance is that the faster growth of Firm *A* was accompanied by greater annual fluctuations. The standard deviation of the former (10.7) is almost twice as large as that of the

[10] In order to maintain focus on the problems and solutions associated with concepts, measurements and analytical tools that regulatory and corporate productivity analysts encounter in their normal practice, we strive to divert attention from the representative firms themselves. They are just an illustration. This paper is not about them; it is about the principles, methods, and the practice of productivity analysis.

latter (5.8). Large variability in the growth rates of the outputs was accompanied by an almost equally large variability in the growth rates of the inputs. The input growth rates are shown in *Table A2* in the *Appendix*. Their standard deviation was 10.1 for Firm *A* and 4.7 for Firm *B*. The capital inputs of both companies displayed a linear increase over time but there was substantial fluctuation in the growth of material input and, for Firm *A*, in the growth of labour input as well. As revealed by *Figure A2* in the *Appendix*, not only the output markets but also the input markets of the two companies were affected by quite different forces. During the 12-year period, there were only four years when the input growth of the two companies accelerated or decelerated in parallel.

The high annual fluctuations in outputs and inputs do not lead to high fluctuations in the annual productivity gains if there is a strong correlation between them. *Figures A3* and *A4*, however, show a weak correlation for both companies. During the 12-year period, the increase in output and input accelerated or decelerated simultaneously in Firm *A* and Firm *B* in only five of the years. Consequently, as shown in *Figure A5*, the annual productivity growth rates are characterised by a great deal of fluctuation. The annual proportional changes in productivity are displayed in *Table 2* below and in *Figure A4* in the *Appendix*. As an illustration of the degree of fluctuations, the table also shows the extent of deviation from the mean.

Technological changes presumably did not influence the annual fluctuations in productivity gains to a significant extent. There are two reasons for this: the first one applies to Firm *B* and the second one mainly to Firm *A*. First, technological changes took place at a fast pace but were relatively evenly distributed in time for Firm *B* (see

TABLE 2 • **Changes in productivity in Firm *A* and Firm *B* $(100\ \dot{\phi}_t)$**

	Annual proportional change			Deviation from the mean		
Year (t)	*A*	*B*	*A – B*	*A*	*B*	*A – B*
1	9.39	3.54	5.85	–3.04	0.28	–3.33
2	3.72	4.20	–0.48	2.63	–0.38	3.00
3	3.61	0.47	3.14	2.74	3.35	–0.62
4	8.78	6.07	2.71	–2.43	–2.25	–0.19
5	6.70	5.17	1.53	–0.35	–1.35	0.99
6	12.67	5.65	7.02	–6.32	–1.83	–4.50
7	1.60	8.45	–6.85	4.75	–4.63	9.37
8	–1.42	2.33	–3.75	7.77	1.49	6.27
9	3.77	1.14	2.63	2.58	2.68	–0.11
10	8.61	2.36	6.25	–2.26	1.46	–3.73
11	9.70	3.49	6.21	–3.35	0.33	–3.69
12	9.47	3.93	5.54	–3.12	–0.11	–3.02
13	5.89	2.90	2.99	0.46	0.92	–0.47
Mean	6.35	3.82	2.52	0.00	0.00	0.00

Table A3 and *Figure A10).* Their standard deviation has a low value for this firm. The same observation however would not apply to Firm *A*, where the technology variable shows high annual fluctuation and a large standard deviation. Second, as *Figure A8* indicates, there is a weak correlation between annual changes in technology and productivity for Firm *A*. This suggests that technological changes did not have significant immediate or short-term cost-saving effects. *Figure A9* shows a stronger correlation pointing to more intensive effects of technological changes in Firm *B*.

There are two important further observations. First, the introduction new technologies was completed much faster in Firm *A* than in Firm *B*. Faster output growth rate and better financial position appear to have played a role in this. Second, although technological progress and output growth show the same cyclic behaviour, the technology cycles follow the output cycles with some delay.

SPECIFICATION OF THE EMPIRICAL MODELS

Initially the technologies of the two firms are analysed separately in *firm-specific* production models. Results from the models that are specified and estimated independently for Firm *A* and Firm *B* are expected to provide guidance for an assessment of whether the statistical quality and economic meaningfulness of the parameter estimates can be improved by building *shared models*. If – as in our case – results from the estimated firm-specific models suggest that there is a need and room for significant improvement, shared models are attempted.

A thorough investigation of technological properties and management behaviour suggested that the production processes of both firms could be expressed with the aid of either the single-output or the multi-output total cost functions of equations (12) and (20), respectively.

As mentioned before, the two companies were producing the same three outputs. However, as a first approximation, it was assumed for the sake of simplicity that only one output – an aggregate of the three – was produced. The simplicity of the single output model makes it especially suitable for illustrating the basic characteristics of comparisons and decompositions. This assumption was discarded in our second approximation, and a multi-output model was built.

Equations (12) and (20) are capable of describing either a cost minimising or a profit maximising firm. The initial set of cost models was based on the assumption of pure cost minimising corporate behaviour. Cost minimisers endogenously determine input volumes in order to produce exogenously determined volumes of output volumes at minimum cost, subject to exogenous input prices. The exogeneity of input prices was a reasonable assumption because both firms purchased their inputs on markets which could safely be characterised as perfectly competitive. Their output on the other hand was neither purely exogenous nor purely endogenous.

On the one hand, regulation imposed on both firms some obligations to satisfy demand generated by prices that had been strongly influenced by regulation, thereby *de facto* defining or at least strongly influencing market size. On the other hand, due to their market power and the light-handed nature of regulation, the management of both firms could influence output prices and volumes to a considerable degree. Recognising the resulting endogeneity of outputs, and as an alternative to pure cost minimisation, profit maximising corporate behaviour, subject to endogenous output volumes, was also assumed in the second set of cost models.[11]

Technological changes were assumed exogenous in both sets. The assumption of exogeneity seems essentially reasonable. The main driving force of the observed technological changes was the digital revolution itself, which left very limited technological choices for the companies. It was clear to both firms that their business success was to a large extent a function of how rapidly and efficiently they managed to exploit the constantly emerging new technological possibilities.[12]

The total cost function is specified as a so-called *transcendental* function. It is generated by the flexible, second-order Taylor series local expansion of the general-form neoclassical cost function. The transcendental specification has been selected in order to avoid *a priori* constraints on technological properties. The mathematical shape of a transcendental function is determined by the data rather than by such constraints.

First we assume a single homogeneous output, include the prices of three homogeneous inputs (labour, capital and materials), and apply a temporal index series of cost-saving exogenous technological changes. The firm-specific transcendental cost function relying on these assumptions is

$$\begin{aligned} C= {} & \alpha_0 + \alpha_1 w + \alpha_2 r + \alpha_3 m + \alpha_4 Q + \beta T + 1/2(\gamma_{11} w^2 + \gamma_{22} r^2 \\ & + \gamma_{33} m^2 + \gamma_{44} Q^2 + \beta_T T^2) + \gamma_{12} wr + \gamma_{13} wm + \gamma_{23} rm + \gamma_{14} wQ + \\ & + \gamma_{24} rQ + \gamma_{34} mQ + \beta_1 wT + \beta_2 rT + \beta_3 mT + \beta_Q QT, \end{aligned} \tag{37}$$

where C denotes the total economic cost of production; w, r and m are the prices of labour, capital and material inputs, respectively; Q represents the volume of the single output; and T denotes the technological index.

The usual parametric restrictions to impose first-order homogeneity in input prices are applied.[13] Following the usual practice of econometric cost analyses, under the assumption of cost minimisation the cost function is estimated as part of

[11] The profit maximising cost models have been excluded from this paper.

[12] Since certain elements of technological changes are highly dependent on management decisions, such elements should be regarded as endogenous. However, due to limitations in size, this study is not extended to more complex, detailed treatments of technological changes.

[13] The imposition of $\Sigma_j \alpha_j = 1$; $\Sigma_j \gamma_{ji} = \Sigma_j \beta_j = 0$ ($i = 1, \ldots, 4$; $j = 1, \ldots, 3$) results in first-degree homogeneity of the cost function in input prices. This common sense requirement ensures that if all input prices are raised by the same percentage then production cost undergoes an identical percentage increase.

a simultaneous equation system in which the application of Shephard's well-known lemma yields two cost share equations.[14] The parameters are estimated in a modified version of the procedure originally developed by *Zellner* [1962] for the estimation of "seemingly unrelated" regression equations.

The productivity performances of any two companies may be compared to each other in order to establish which one has higher or faster improving productivity. Firm-specific cost models may also reveal what causes generated how much cost saving and improvement in productivity. Inter-firm comparisons can be made. It is not required that technological or other similarities exist between the two firms. The quality of the estimated parameters can, however, often be improved in a situation where technological and other similarities between the compared firms allow the building of a *common technology model*. When relatively few observations are available, for instance, the use of this model increases the degrees of freedom and thus contributes to the "sharpening" of the parameter estimates. Building a common technology model often leads to a significant improvement, when the two companies operate in the same industry.

Common technology is represented by a *common or "shared" total cost function* which allows both similarities and differences to exist between the two firms. Technological similarities are revealed by forming and testing a set of constraining null hypotheses that express various equivalences between their technologies. The test results allow us to describe a technology some parts of which are shared by the two companies, while other parts are not. The least constrained shared cost function allows differences between Firm A and Firm B in each of the parameters of the cost function, but assumes that the same variance-covariance matrix applies to both companies. Binary dummy variables allow each parameter of the shared cost function to be firm-specific. The use of all possible dummies ($D_A = 1, D_B = 0$) results in the following lengthy specification:

$$\begin{aligned} C = {} & \alpha_0 + \alpha_{0A}D_A + (\alpha_1 + \alpha_{1A}D_A)w + (\alpha_2 + \alpha_{2A}D_A)r + (\alpha_3 + \alpha_{3A}D_A)m + \\ & + (\alpha_4 + \alpha_{4A}D_A)Q + (\beta + \beta_A D_A)T + 1/2((\gamma_{11}+\gamma_{11A}D_A)w^2 + (\gamma_{22} + \gamma_{22A}D_A)r^2 + \\ & + (\gamma_{33} + \gamma_{33A}D_A)m^2 + (\gamma_{44} + \gamma_{44A}D_A)Q^2 + (\beta_T + \beta_{TA}D_A)T^2 + (\gamma_{12} + \gamma_{12A}D_A)wr + \\ & + (\gamma_{13} + \gamma_{13A}D_A)wm + (\gamma_{23} + \gamma_{23A}D_A)rm + (\gamma_{14} + \gamma_{14A}D_A)wQ + \\ & + (\gamma_{24} + \gamma_{24A}D_A)rQ + (\gamma_{34} + \gamma_{34A}D_A)mQ + (\beta_1 + \beta_{1A}D_A)wT + \\ & + (\beta_2 + \beta_{2A}D_A)rT + (\beta_3 + \beta_{3A}D_A)mT + (\beta_Q + \beta_{QA}D_A)QT. \end{aligned} \tag{38}$$

The firm-specific models of equation (37) are estimated first. The next two sections discuss the results for Firm A and Firm B, respectively.

[14] In order to preserve the non-singularity of the variance-covariance matrix, one equation – here the material cost share – was omitted. Under the assumption of profit maximisation, revenue share equations would also appear in the system of simultaneous equations.

ESTIMATION RESULTS FOR FIRM *A*

Annual observations are available for both companies, but for Firm *A* we had to use relatively short, 15-year time series because the company's productivity reports covered only a decade and a half. In line with our expectations and with the results of studies performed by other analysts before us, the estimation of equation (37) failed due to insufficient degrees of freedom. For the most general form, the unconstrained equation (37), we obtained non-significant estimates for several necessarily non-zero parameters, while the estimates were unreasonably high for some of the basic economic characteristics. A lengthy exploration of the various constraining null-hypotheses, however, provided some useful results. A likelihood ratio tests showed that three hypotheses, namely $\gamma_{44} = \beta_Q = \beta_T = 0$, could not be rejected. Some estimation problems (incorrect curvatures and signs) remained, however, even after introducing the constraints. Our further analyses revealed that the input structure of Firm *A* was quite stable throughout the period. This phenomenon suggested that the observed small changes in the input structure were probably a consequence of the changes in input prices, *i.e.,* the production technology was homothetic and the technological changes were input-neutral. These conclusions led to the testing of hypotheses $\beta_j = 0$ and $\gamma_{j4} = 0$, which gave noteworthy results. Although the hypotheses were rejected, their inclusion resulted in a statistically acceptable and economically meaningful model, which offered useful information for the specification of the shared *A-B* model. The following parameter estimates were obtained (t-values are given in brackets underneath the coefficients):

$$\begin{aligned} C - m = \underset{(10.9)}{0.87} - \underset{(-3.8)}{0.081}D_s + \underset{(94)}{0.335}(w-m) + \underset{(178)}{0.530}(r-m) + \underset{(22)}{0.615}Q - \underset{(-1.15)}{0.141}T + \\ + \underset{(1.99)}{0.144}(1/2w^2+1/2m^2-wm) - \underset{(-2.64)}{0.076}(wr-wm-rm+m^2) + \underset{(3.9)}{0.103}(1/2r^2+1/2m^2-rm), \end{aligned} \tag{39}$$

where D_s is a binary dummy variable showing the effect of major structural changes in one of the years of the observation period.

Every variable is logarithmically transformed in our model because hypotheses on the linear and Box-Cox transformations were rejected for every variable. We furthermore used the usual parameter constraints ensuring input price homogeneity of degree 1 of the cost function because our test results did not allow us to reject the hypothesis.

The first-order technological parameter was non-significant. This result surfaced in most models for Firm *A*, suggesting that technological changes did not have a substantial direct (immediate) cost-saving effect, that is, they did not play a recognisable role in explaining changes in the company's inputs, costs and productivity. The degree of economies of scale was high and constant. The output elasticity of cost was estimated at $\varepsilon_{CQ} = 0.61$, and thus the derived input elasticity of output, the scale elasticity was $\varepsilon_{QX} = 1.63$. The input price elasticities of cost showed little change over time. Production costs were most sensitive to capital prices and least sensitive to material prices. Demand for all three inputs was price-inelastic. For the relationships between inputs, the estimates suggested complementary between labour and material, while labour-capital as well as capital-material substitution was indicated. As this brief summary of results shows, the cost function satisfied the theoretical requirements of economic rationality, and contained no unreasonable economic properties. Various further null-hypotheses were applied to the model, but all of them were rejected. The model as shown in equation (39) can therefore be regarded as the final outcome of our explorations.

ESTIMATION RESULTS FOR FIRM *B*

The time series provided by the productivity report of Firm *B* go back all the way to the end of World War II. The oldest data had to be disregarded, however, because the technology used during the early years was fundamentally different from the one in use during the 15-year period we considered for Firm *A*. First we determined where to cut the time series. An investigation into past technological changes revealed that in one rather short time period, technological changes of so fundamental a nature occurred that it seemed reasonable to separate the long time series into periods of "old" and "new" technologies. Several cost models were estimated for various periods of observation. Based on test results, years of the "old" technologies were cut out. After discarding the observations of the years that preceded the great technology change, we still managed to lengthen the 15-year period shared with Firm *A* by 12 more years. The cost functions were then estimated with both the 27-year and the 15-year-long data sample. The parameter estimates and the economic characteristics gained from them differed only to a negligible extent but the estimates of the longer period were more efficient in a statistical sense and we decided to work with them. Tests conducted to determine variable transformations in all specifications, including the unconstrained translog cost function shown in equation (37) yielded the following final conclusions for Firm *B:*

- Linear transformation was rejected for every variable.
- Logarithmic transformation could not be rejected for the output variable and the technology index.
- Box-Cox transformation was applied to all other variables. Variable transformation parameters were obtained in the $0 < \lambda < 1$ interval.
- The hypothesis of a homothetic production process could not be rejected.

Estimation results:

$$C - m = \underset{(2.26)}{0.009} + \underset{(145)}{0.297(w-m)} + \underset{(253)}{0.542(r-m)} + \underset{(26)}{0.576Q} - \underset{-4.1)}{0.588T} + \tag{40}$$

$$+ \underset{(11)}{0.130(1/2w^2 + 1/2m^2 - wm)} - \underset{(-14)}{0.16\,(wr - wm - rm + m^2)} +$$

$$+ \underset{(14)}{0.241(1/2r^2 + 1/2m^2 - rm)} - \underset{(-20)}{0.302(w-m)T} + \underset{(22)}{0.34(r-m)T} +$$

$$+ \underset{(2.11)}{0.141(1/2Q^2)} + \underset{(2.45)}{1.43(1/2T^2)} - \underset{(-2.94)}{0.734QT}$$

The statistical properties of the cost function satisfy the requirements set by economic theory and reasonable expectations. The economic characteristics are rational and reasonable. The annual estimates of scale elasticity are generally high. They increased slightly at the beginning of the period then stayed at the elevated level for several years, after which a slight decline was observed, and finally their value remained constant for the last few years of the period. During the last ten years, the scale elasticity of *A* was slightly higher than the scale elasticity of *B* but the difference is not statistically significant.

The technology elasticity of cost was negative. The estimated annual values appeared reasonable and corresponded roughly to results from engineering type investigations. Another interesting result that also matches engineering type estimates is that in the last two thirds of the period, technological changes (involving primarily the digitisation and computerisation of increasingly network-based production processes) increased the technology elasticity of cost. In other words, the direct cost-saving effect of technological changes seems to have increased.

The estimates of input price elasticities of cost were essentially the same as those obtained for Firm *A* for all three inputs. Once again, production costs were most sensitive to capital prices and least sensitive to material prices. As in Firm *A*, demand for all three inputs was insensitive to input prices; the price elasticity of capital input was somewhat lower here than in Firm *A*, the price elasticity of material somewhat higher, while the price elasticity of labour was the same. Inputs were shown as substitutes to each other, except in the second part of the period, when labour and capital were complementary. Null-hypotheses of numerous further possible relationships were tested but we always rejected the constraints they implied. The model described in equation (40) can therefore be considered to be the final outcome of our exploration.

The cost models estimated separately for the two firms show some profound technological similarities. This warrants the testing of *common technological* hypotheses in estimated *shared cost function*.

ESTIMATION RESULTS FROM THE SHARED ($A+B$) MODEL

For each parameter, we tested the null hypothesis that there is no difference between the two firms, *i.e.,* that the parameter estimate of the binary variable D_A does not significantly differ from zero. It was obvious, however, that not even the increased number of observation points of the shared cost function can provide sufficient degrees of freedom for the simultaneous estimation of the large number of parameters that appear in equation (38) even in the simplified case where the cost function is restricted to being homogeneous of degree 1 in input prices. Testing the equality of parameters for A and B required a large number of hypotheses which had to be tested in several steps.

Four constraints offered themselves as a point of departure. First, since the cost share of capital input was the same for the two firms in the year around which the Taylor-series expansion of the function was done, it seemed sensible to use the constraint $\alpha_{2A} = 0$, *i.e.,* to test the hypothesis that the two companies had the same first-order capital parameters. Second, as the parameters of the cost functions estimated separately suggested that the first-order output parameters were also equal, we also tested the constraint $\alpha_{4A} = 0$. The results of the cost functions estimated separately for A and B suggested the third and fourth constraints: $\gamma_{14A} = \gamma_{24A} = 0$. The constraints were applied individually and also in combination. The results were discouraging. Some parameters proved to be of poor statistical quality and unacceptable from an economic point of view (incorrect sign and unreasonable magnitude) indicating that the specification was far from being able to provide a reasonable representation of the firms' technologies.

An examination of the results revealed that most of the estimation problems were rooted in three parameters (γ_{44}, β_Q *and* β_T). The realisation of this led to a reassessment and some correction of the output and technology data and prompted an investigation of the behaviour of the three parameters in the presence of various constraints. We returned to the results of the firm-specific models once again and established that none of the three parameters was significantly different from zero for Firm A. After introducing constraints in the forms of $\gamma_{44A} = -\gamma_{44}$, $\beta_{QA} = -\beta_Q$ and $\beta_{TA} = -\beta_T$, there was a dramatic improvement in the estimates.[15]

Further constraints could also be implemented because the second-order parameters were not statistically significant allowing us to introduce zero-constraints. Indeed, the hypotheses of equality between the corresponding second-order parameters of the two firms $\gamma_{11A} = \gamma_{12A} = \gamma_{22A} = 0$ could not be rejected based on the

[15] The results of the tests were somewhat contradictory because the likelihood ratio test indicated that the hypotheses could be rejected while the t-statistic suggested that they could not. However, the results improved to such an extent that we elected to retain the hypotheses and use the constrained cost function as the starting point for further investigation.

likelihood ratio test. When the constraints were applied, all of the parameters of the cost function became highly significant.

The results also indicated that even more constraints could be introduced. Since the second-order parameter estimates of the input-technology interaction were close to zero for Firm A, the hypotheses $\beta_{1A} = \beta_1$ and $\beta_{2A} = \beta_2$ were tested. Neither could be rejected. At this point, we reached the limit of constraining the cost function. All further constraints were rejected. The model shown in equation (41) can therefore be considered to be the final outcome of our exploration of the single-output shared cost function

$$\begin{aligned}
C - m = {} & \underset{(0.74)}{0.004} - \underset{(-249)}{1.899}D_A - \underset{(-5.4)}{0.082}D_s + (\underset{(149)}{0.296} + \underset{(10.8)}{0.029}D_A)(w-m) + \\
& + \underset{(360)}{0.542}(r-m) + \underset{(32.6)}{0.600}Q - (\underset{(-4.47)}{0.650} + \underset{(5.33)}{0.556}D_A)T + \\
& + \underset{(8.9)}{0.122}(1/2w^2 + 1/2m^2 - wm) - \underset{(-10.1)}{0.129}(wr - wm - rm + m^2) + \\
& + \underset{(12.5)}{0.214}(1/2r^2 + 1/2m^2 - rm) - \underset{(-16.7)}{(0.325 - 0.325D_A)}(w-m)T + \\
& + \underset{(20.9)}{(0.373 - 0.373D_A)}(r-m)T + \underset{(2.44)}{0.010}(w-m)Q - \underset{(-5.83)}{0.022}(r-m)Q + \\
& + \underset{(2.67)}{(0.300 - 0.300D_A)}\, 1/2Q^2 - \underset{(-3.04)}{(1.153 - 1.153D_A)}QT + \underset{(2.36)}{(2.566 - 2.566D_A)}\, 1/2T^2.
\end{aligned} \tag{41}$$

where all variables, with the exception of the dummies, appear in a logarithmically transformed form.

Equation (41) describes the greatest possible extent of similarities between the production technologies of the two companies. We built these similarities into the shared cost function in a statistically and economically justifiable way and learned a great deal more from the shared models than what we knew having estimated only single-firms models. The structural information thus gained can be used to complete the decomposition and forecasting of changes in the two companies' input volumes and productivity. Before doing that, let us sum up what we have learned!

For both companies and for each year, the cost function satisfies the behavioural requirements that neoclassical production theory poses for production costs. Reasonable estimates were obtained for input demand and the relationships between the three input categories. The own-price elasticities of input demand have *a priori* correct negative signs. Demand for each of the three inputs is inelastic with respect to its own price. The inputs substitute for each other with the exception of complementary relationship between capital input and material input in Firm A.

There is a high degree of economies of scale for both companies. For Firm A, the annual estimates vary within the narrow range of $\varepsilon_{QX} = 1.65 - 1.67$. The reasonableness of this estimate can be tested by looking at the relationship between the annual output growth rates and the annual productivity growth rates. As can be seen in *Figure A6* in the *Appendix*, this relationship is relatively stable for Firm *A; i.e.,* the scale elasticity is approximately constant. For Firm *B*, the annual estimates of scale elasticity follow the same characteristic path as the estimates derived from the firm-specific model and available previous estimates by other analysts. The relatively stable degree of economies of scale is explained for both companies by the opposite forces of the scale-economies-exhausting effects of output growth on the one hand, and the scale-economies-increasing effects of technological changes on the other hand. It seems that the two forces roughly counterbalanced each other. An increase in production normally reduces the degree of economies of scale if technology remains unchanged.[16] However, those technological changes whose cost saving effects expand as the volume of output increases lead to increases in the degree of economies of scale. The rapid growth and the rapid technological progress appeared to be more or less in balance during the observation period for Firm *A*. For Firm *B*, however, we may argue that the effects of rapid technological progress exceeded the effects of the gradually decelerating growth of the firm during the observation period.

The annual estimates of the technology elasticity of cost are not completely satisfactory. For Firm *A*, the annual estimates remain constant over time at the value of $\varepsilon_{CT} = -1$. This is considered reasonable. For Firm *B*, however, the absolute values of the *a priori* correctly negative estimates are slightly higher than what we could accept as reasonable. Finally, when the annual estimates are based on the longer 27-year sample period, they show a trend. A weaker impact on costs during the first 12 years is followed by a temporally increasing trend. If, however, the model overestimates the technology elasticity of costs, it will also underestimate its counterpoint, the degree of economies of scale. A comparison with the results of the firm-specific model indeed appears to support the suspicion that the shared model somewhat underestimated the scale elasticity of Firm *B* (while somewhat overestimating it for the years preceding the shared period.)

The shared cost function described in equation (38) and the estimation results shown in equation (41) rely on the assumption that the products of both companies can be aggregated into a single output. The single-output model gave statistically valid and economically reasonable results, which are often perfectly suitable for cost analysis and the decomposition of temporal changes in productivity and of inter-firm differences in those temporal changes. But there are two hidden dangers in using single-output models. First, the estimates may be biased if the output aggregate does not exist. Second, information with respect to individual output categories may be needed for both regulatory and management purposes. Impacts

[16] This phenomenon tends to exist for established firms that have been operating for a long time. With new or young firms, however, the opposite phenomenon may also occur, *i.e.,* an increase in the size of production may be accompanied by an increase in the degree of economies of scale. Our firms *A* and *B* are old, established, large enterprises.

by the output aggregate need to be decomposed into individual output effects. This can be achieved by estimating *multi-output* shared cost functions. A multi-output model offers valuable information: its estimates show the roles of individual outputs in productivity growth, and reveal similarities and differences in the interactions between technological changes and individual outputs. When faced with output-augmenting technological progress, it is an especially useful feature that we can assign output-specific effects.

Both companies have three main output categories (α, β and γ, see *Table A1*, where their temporal proportional changes are shown). With these inserted into equation (38), the number of first- and second-order parameters waiting to be estimated increases to such an extent that it exceeds the number of observations, therefore the parameters of the three-output model cannot be estimated. In our case the number of observations is insufficient to gain efficient estimates even for two-output models. Decomposition of productivity growth rates and inter-firm differences in productivity growth rates are accomplished using the single-output model. To overcome the difficulty caused by the insufficient number of observations, it is advisable to base the productivity measurement and analysis on at least quarterly, and preferably on monthly, data.

DECOMPOSITION OF PRODUCTIVITY GAINS

During the 13-year observation period,[17] the average annual productivity gain (proportional change in productivity) in Firm *A* was 2.52 percentage points higher than the corresponding rate in Firm *B*. We now attempt to find out why. The inter-firm difference is decomposed into several causal components displayed in *Table 3*. The component which is generated by economies of scale can be further decomposed according to the following formula:

$$\dot{E}_A - \dot{E}_B = \frac{(\dot{Q}_A + \dot{Q}_B)}{2}(\xi_{CQ_A} - \xi_{CQ_B}) + \frac{(\xi_{CQ_A} + \xi_{CQ_B})}{2}(\dot{Q}_A - \dot{Q}_B). \qquad (42)$$

The firm-specific and the shared models show strong similarities with respect to our most important empirical result: the fairly large – 2.52 percentage point – difference between the two firms' productivity growth rates was almost entirely (in 95-96 percent) the consequence of Firm *A* growing more rapidly than Firm *B*. The remaining effects were individually negligibly small, even if combined. As they showed very small values, we did not deem it necessary to further decompose the technology effects.

[17] For a period of 13 years a total of 12 growth rates can be computed, since growth in the first year is not known.

TABLE 3 • **The decomposition of *average proportional* productivity change**

Effects		Firm *A* Rate	Firm *A* Percent distribution	Firm *B* Rate	Firm *B* Percent distribution	*A* + *B* Rate	*A* + *B* Percent distribution
Firm-specific models							
Productivity growth	$\dot{\phi}$	6.34	100	3.82	100	*2.52*	100
Technological effect	$\dot{B}$	0.69	11	0.75	20	*–0.07*	–3
Growth effect	$\dot{E}$	5.45	86	3.49	91	*1.98*	79
Due to economies of scale	$\overline{\dot{Q}}\Delta\xi$	–	–	–	–	*–0.45*	–17
Due to output increase	$\overline{\xi}\Delta\dot{Q}$	–	–	–	–	*2.43*	96
Residual effect	$\dot{R}$	0.20	3	–0.42	–11	*0.61*	24
Shared model							
Productivity growth	$\dot{\phi}$	6.34	100	3.82	100	*2.52*	100
Technological effect	$\dot{B}$	0.45	7	0.84	22	*–0.39*	–15
Growth effect	$\dot{E}$	5.66	89	3.30	86	*2.36*	94
Due to economies of scale	$\overline{\dot{Q}}\Delta\xi$	–	–	–	–	*–0.04*	–1
Due to output increase	$\overline{\xi}\Delta\dot{Q}$	–	–	–	–	*2.40*	95
Residual effect	$\dot{R}$	0.23	4	–0.32	–8	*0.55*	22

In Firm *A*, 86–89 percent of productivity growth is due to the rapid increase in the volume of the firm's output in the presence of substantial degrees of economies of scale. The direct cost saving due to technological changes is responsible for only 7–11 percent of the actual productivity gain. We mentioned the possibility of such a result when we showed that there was a strong correlation between the annual growth rates of Firm *A*'s output and productivity, while the technological changes correlated rather weakly with changes in productivity during the observation period. We then surmised that the main reason for the introduction of the new technologies was probably not the immediate and short-term cost saving. We later added that technological progress was more likely driven by the expected positive effects of new technologies on economies of scale. It was expected that cost savings due to the introduction of new technologies would gradually emerge and increase over time as the volume of output increased over a longer period of time.

The estimates seem to support this reasoning. It is an interesting result that the technology of Firm *A* had a weaker effect on productivity growth *(i.e.,* caused a smaller immediate cost reduction) than the technology of Firm *B* even though technological progress was faster in *A* than in *B*. The explanation may be that Firm *A*'s markets, and hence its output and revenues grew at a very high rate. Fast output growth forced – and rapidly increasing revenues allowed – the introduction of new technologies at an ambitious pace, which – precisely because of its ambitious nature – resulted in extra costs and thus curbed the extent of immediate and short-term cost savings.

Turning to Firm *B*, 86–91 percent of productivity growth is due to the increase in output in the presence of economies of scale. The immediate and short-term cost-saving effects of technological changes are responsible for 21–22 percent of the average productivity gain. This finding is consistent with previous estimates available from the company.

The residual effect left unexplained by the model is fairly high, especially for Firm *B*, where it represents 8–11 percent of productivity growth. This phenomenon is due to the high variability of annual productivity gains. In a small sample – a 13-year period in our case – highly variable individual residuals can have a strong effect on the mean residual. High temporal variability of productivity gains is a widely observed phenomenon. Corporate reports show that productivity growth – with, it is safe to claim, few exceptions – tends to proceed at an uneven pace over time. There are several reasons for this. An especially important reason is that temporarily unused capacities are necessarily created during investments, since capital input exhibits high degrees of indivisibility. Unused capacities temporarily decrease the annual productivity growth rate, and when the unused capacities are finally utilised, their presence accelerates the annual growth in productivity.

DECOMPOSITION OF INPUT GROWTH RATES

Input growth rates are one of the two components of productivity gains. *Table* 4 offers their decomposition in firm-specific and combined *A* + *B* models. Average annual growth rates are broken down to a technology effect, an output growth effect, and as many input price effects as the number of inputs. We work with three input categories: labour, capital and material. Output growth exerts the most important effect on the volumes of inputs. In fact, its effect is so important that it exceeds that of the growth rate in productivity. This is possible only if the combined other effects are negative in the sense that they make input volumes to decline. As can be seen in *Table 4,* this is what happens for all three inputs.[18] This is not surprising, given that the majority of input prices increased and the majority of technological changes had an input-reducing effect. The input-saving effect of technological changes shows substantial annual fluctuation. For Firm *B*, there was a year when the annual capital-saving effect was as low as zero, but in another year a saving as high as 2.19 percentage points was achieved. The latter is a quite exceptional figure, but our examination of the events of that year convinced us that it was a valid estimate.

[18] For all inputs in Firm *A* and for labour input in Firm *B*.

TABLE 4 • **Decomposition of the average growth rates of inputs**

		Firm *A*		Firm *B*		*A* + *B*	
Effects		Rate	Percent distribution	Rate	Percent distribution	Rate	Percent distribution
Labour input	$\dot{L}$	8.73	100	2.69	100	*6.04*	100
Technological effect	$\dot{B}_L$	–0.45	–5	–2.19	–81	*1.74*	29
Growth effect	$\dot{E}_L$	8.93	102	5.16	192	*3.77*	62
Labour price effect	$\dot{w}_L$	–2.28	–26	–3.16	–117	*0.88*	15
Capital price effect	$\dot{r}_L$	1.36	16	0.73	27	*0.63*	10
Material price effect	$\dot{m}_L$	1.22	14	1.28	48	*–0.06*	–1
Residual effect	$\dot{R}_L$	–0.05	–1	0.87	32	*–0.92*	–15
Capital Input	$\dot{K}$	7.30	100	4.69	100	*2.61*	100
Technological effect	$\dot{B}_K$	–0.45	–6	0.00	0	*–0.45*	–17
Growth effect	$\dot{E}_K$	7.93	109	4.57	97	*3.36*	129
Labour price effect	$\dot{w}_K$	0.69	9	0.64	14	*0.05*	2
Capital price effect	$\dot{r}_K$	–0.62	–8	–0.44	–9	*–0.18*	–7
Material price effect	$\dot{m}_K$	–0.19	–3	0.03	1	*–0.22*	–8
Residual effect	$\dot{R}_K$	–0.06	–1	–0.11	–2	*0.05*	2
Material input	$\dot{M}$	8.01	100	6.31	100	*1.70*	100
Technological effect	$\dot{B}_M$	–0.45	–6	–1.20	–19	*0.75*	44
Growth effect	$\dot{E}_M$	9.86	123	5.55	88	*4.31*	254
Labour price effect	$\dot{w}_M$	2.92	36	3.68	58	*–0.76*	–45
Capital price effect	$\dot{r}_M$	–0.92	–11	0.09	1	*–1.01*	–59
Material price effect	$\dot{m}_M$	–2.25	–28	–2.44	–39	*0.19*	11
Residual effect	$\dot{R}_M$	–1.15	–14	0.63	10	*–1.78*	–105

Increases in input prices substantially reduced the growth rates of inputs, especially those of labour, because labour was the most price sensitive input. Increases in labour prices generated an average annual decline of 2.28 percentage points in the use of labour in Firm *A*, and 3.16 percentage points in Firm *B*. Material input also showed high degrees of price sensitivity: 2.25 and 2.44 percentage points annually on average for Firms *A* and *B*, respectively. Cross-price elasticities tended to be smaller than own-price elasticities. One exception was the price of labour input, which had a substantial effect on the use of material inputs in both companies.

Averaging the annual rates over the 13-year observation period, Firm *A* increased the volume of labour input 6 percent faster, capital input 2.6 percent faster and material input 1.7 percent faster than Firm *B*. The differences between input growth rates are due mainly to the faster increase of output volumes in Firm *A* than in Firm *B*. Relative to Firm *B*, the input-saving effect of technological progress in Firm *A* accelerated the increase in labour and material inputs but decelerated the growth of capital.

Inter-firm differences in input prices had only a mild effect on the differences in input growth rates in the case of labour, an especially mild effect with respect to capital, but an outstandingly strong effect on material. Owing to faster increases in input prices and higher price elasticities of demand for inputs, overall the input price effect caused the average annual growth rate of productivity of Firm *A* to be 1.5 percentage points lower than the corresponding indicator of Firm *B*.

The unexplained residual effect is very high for material input. This is caused mostly by the high annual fluctuation in the material input of Firm *A* that is also apparent in *Table A2.* In contrast, since the capital input increased relatively steadily over time for both companies, the residual effect is negligibly low for capital. Five causal factors are shown in *Table 4.* They could be further decomposed into *size* effects and *intensity* effects. Such decomposition, however, would fall beyond the scope of the present study.

CONCLUSION

As noted in the introduction, doing everything in their power to facilitate the efficient operation of the firms they regulate is one of the most important duties of socially responsible regulators. They cannot carry out their duty, unless they understand productivity. They must measure, compare and analyse corporate productivity in various ways, using an arsenal of economic and econometric analytical tools.

Historically, productivity studies were developed first by regulated monopolies for their own use as well as for regulatory purposes. Beginning in the 1960's, regulators of monopolies made extensive use of them for several decades, especially following the world-wide spread of price cap regulation. However, with the advent of the competitive era; *i.e.,* the introduction of competition into formerly monopoly markets, productivity studies were forced into the background by not more important but more urgent problems of imperfectly competitive markets. However, it is easy to see the reasons why productivity is even more important for competitive companies and regulators of imperfectly competitive markets than it used to be for monopolies and their regulators.

Productivity analysis is also an important and useful tool in the hands of corporate management. The competitiveness and market position of regulated as well as unregulated companies, the price and quality of their products and services, and ultimately their profitability all depend on how rapidly they may be able to improve their productivity. They also must study and understand productivity.

From a corporate as well as a regulatory point of view, a renaissance of productivity studies in the not so distant future would be very much in order. This paper is a small contribution showing how certain useful productivity analyses, particularly those that involve inter-firm comparisons and causal decompositions, could be conducted for management and regulatory purposes.

REFERENCES

Denny, M.–Fuss, M.–Everson, C. [1979]: Productivity, employment and technical change in Canadian telecommunications. Manuscript, Final report to the Department of Communications.

Denny, M.–Fuss, M.–Waverman, L. [1981]: The measurement and interpretation of total factor productivity in regulated industries. In: *Cowing, T. G. & Stevenson, R.* (eds.): Productivity measurement in regulated industries. Academic Press, New York. pp. 179–218.

Kiss, F. [1982]: Forecasting productivity growth. Proceedings of the Fourth International Conference on Forecasting and Public Utilities, Paris, June 27–30.

Kiss, F. [1983]: Productivity growth in Bell Canada. In: *Courville, L., de Fontenay, A.–Dobell, R.* (eds.): Economic Analysis of Telecommunications. North-Holland, Amsterdam, New York, Oxford, pp. 83–113.

Kiss, F. [1984]: On inter-firm productivity comparisons. In: *Olley, R. E.–Le, C. D.* (eds.): Total Factor Productivity of Canadian Communications Carriers. Final Report to the Department of Communications, Ottawa.

Kiss, F.–Karabadjian, S.–Lefebvre, B. [1983]: Economies of scale and scope in Bell Canada. In: *Courville, L.–de Fontenay, A.–Dobell, R.* (eds.): Economic Analysis of Telecommunications. North-Holland, Amsterdam, New York, Oxford, pp. 55–82.

Kiss, F.–Olley, R.–Lefebvre, B. J. [1983]: Inter-firm productivity growth comparisons in the Canadian telecommunications industry. Proceedings of the Third International Symposium on Forecasting, Philadelphia, June 5–8.

Kiss, F.–Lefebvre, B. J. [1987]: Econometric models of telecommunications firms: A survey, Revue Economique, Vol. 38. No. 2. Paris, pp. 307–374.

Shephard, R. W. [1970] Cost and production functions. Princeton University Press, Princeton, N. J.

Zellner A. [1962] An efficient method for estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, Vol. 57. pp. 348–368.

APPENDIX

TABLE A1 • **Annual proportional changes* in the output volumes of Firms *A* and *B*** $(100\,\dot{q}_{it})$

	Output α			Output β			Output γ			Total output		
Year	*A*	*B*	*A – B*	*A*	*B*	*A – B*	*A*	*B*	*A –+ B*	*A*	*B*	*A – B*
1	10.8	7.5	3.4	15.2	13.9	1.3	–3.2	7.7	–10.9	12.9	9.9	3.0
2	9.2	6.7	2.5	12.7	8.5	4.2	14.6	8.2	6.3	11.7	7.5	4.1
3	9.3	6.4	2.9	10.8	4.4	6.3	7.1	7.0	0.1	10.1	5.7	4.5
4	11.2	7.5	3.7	16.5	12.1	4.4	3.6	9.8	–6.2	14.3	9.5	4.8
5	9.0	7.6	1.5	15.5	15.0	0.5	9.3	–18.4	27.7	13.3	10.0	3.3
6	14.3	8.3	6.0	21.0	12.8	8.2	11.9	12.5	–0.6	18.7	10.3	8.4
7	15.2	7.8	7.4	19.0	13.0	6.0	–0.9	18.9	–19.7	17.2	10.4	6.8
8	13.9	6.0	7.9	7.9	8.5	–0.6	8.5	18.6	–10.1	9.6	7.5	2.1
9	11.3	5.1	6.2	11.6	8.3	3.4	10.3	12.5	–2.3	11.5	6.7	4.7
10	12.7	4.2	8.4	17.4	11.3	6.1	14.7	29.2	–14.5	15.8	8.3	7.6
11	15.1	3.2	11.9	18.6	8.3	10.3	10.0	10.5	–0.4	17.2	5.8	11.4
12	17.2	4.9	12.3	19.0	10.4	8.7	19.2	13.6	5.5	18.5	7.9	10.7
13	16.1	3.4	12.7	12.5	10.7	1.8	18.1	11.4	6.8	13.8	7.2	6.6
Mean	12.7	6.0	6.7	15.2	10.5	4.7	9.5	10.9	–1.4	14.2	8.2	6.0

* The proportional changes in output are defined by equation (3).

TABLE A2 • **Annual proportional changes* in the input volumes of Firms *A* and *B*** $(100\,\dot{x}_{jt})$

	Labour			Capital			Materials			Total input		
Year	*A*	*B*	*A – B*	*A*	*B*	*A – B*	*A*	*B*	*A – B*	*A*	*B*	*A – B*
1	2.4	1.7	0.7	6.7	5.7	1.0	–5.0	17.5	–22.5	3.5	6.4	–2.9
2	13.1	2.1	11.0	6.1	4.9	1.2	2.4	0.0	2.4	8.0	3.3	4.6
3	8.0	–1.0	8.9	6.6	4.8	1.8	2.6	17.2	–14.6	6.5	5.2	1.3
4	2.4	0.1	2.3	6.8	5.5	1.3	9.0	2.3	6.7	5.5	3.5	2.1
5	10.7	4.6	6.1	4.5	4.7	–0.2	4.2	5.7	–1.5	6.6	4.9	1.7
6	8.3	4.1	4.2	3.5	4.8	–1.4	10.8	5.2	5.6	6.0	4.7	1.4
7	21.8	–2.7	24.4	10.7	5.8	5.0	18.0	–2.5	20.5	15.6	1.9	13.7
8	3.2	4.0	–0.8	14.1	5.2	8.8	19.7	7.1	12.6	11.0	5.1	5.9
9	11.6	4.1	7.5	8.6	5.5	3.1	–5.9	8.8	–14.7	7.7	5.6	2.1
10	5.0	7.4	–2.4	5.3	3.8	1.5	20.7	9.2	11.4	7.2	5.9	1.3
11	3.7	2.6	1.1	6.9	2.7	4.2	18.2	1.0	17.2	7.5	2.4	5.2
12	17.6	1.8	15.8	7.4	3.7	3.7	–3.2	8.6	–11.8	9.0	3.9	5.1
13	5.7	6.1	–0.4	7.8	3.8	4.0	12.9	2.0	10.9	7.9	4.3	3.6
Mean	8.7	2.7	6.0	7.3	4.7	2.6	8.0	6.3	1.7	7.8	4.4	3.5

* The proportional changes in input are defined by equation (6).

TABLE A3 • **Indices of technological change for Firms *A* and *B***

	Year 0 = 1,0		Year 5 = 1,0		Previous year = 1,0		$100\,\dot{T}_t$	
Year	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
0	1.00	1.00	0.37	0.58	–	–	–	–
1	1.05	1.12	0.39	0.64	1.05	1.12	4.88	11.10
2	1.27	1.27	0.47	0.73	1.21	1.13	19.20	12.55
3	1.31	1.37	0.48	0.79	1.03	1.08	2.59	7.83
4	2.23	1.54	0.83	0.89	1.71	1.13	53.44	11.98
5	2.69	1.74	1.00	1.00	1.21	1.12	19.02	11.73
6	2.97	2.07	1.10	1.19	1.10	1.19	9.62	17.44
7	3.94	2.34	1.46	1.35	1.33	1.13	28.49	12.29
8	4.94	2.67	1.83	1.54	1.25	1.14	22.48	13.24
9	6.90	2.94	2.56	1.69	1.40	1.10	33.44	9.77
10	8.12	3.16	3.01	1.82	1.18	1.07	16.31	7.23
11	9.47	3.43	3.51	1.98	1.17	1.09	15.32	8.20
12	10.04	3.55	3.73	2.04	1.06	1.03	5.92	3.26
13	13.24	4.17	4.92	2.40	1.32	1.17	27.66	16.09

FIGURE A1 • **Annual proportional changes in output in Firms *A* and *B***

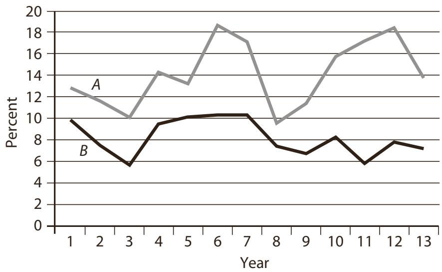


FIGURE A3 • **Annual proportional changes in output and input in Firm *A***

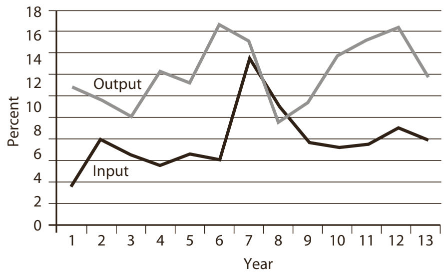


FIGURE A2 • **Annual proportional changes in input in Firms *A* and *B***

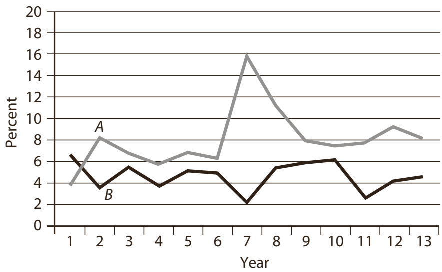


FIGURE A4 • **Annual proportional changes in output and input in Firm *B***

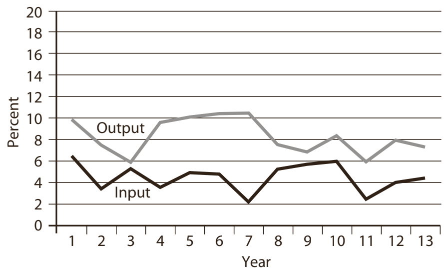


FIGURE A5 • **Annual productivity gains in Firms *A* and *B***

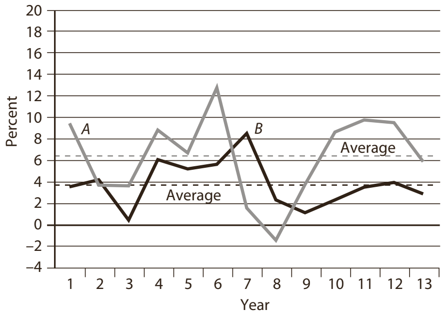


FIGURE F8 • **Annual changes in the technology index and productivity of Firm *A***

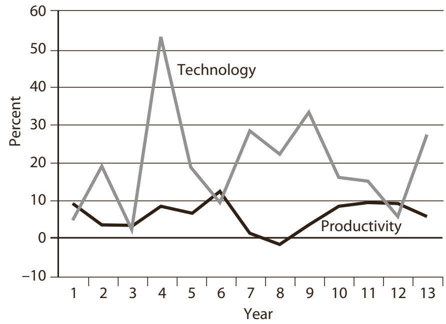


FIGURE A6 • **Annual changes in the output and productivity of Firm *A***

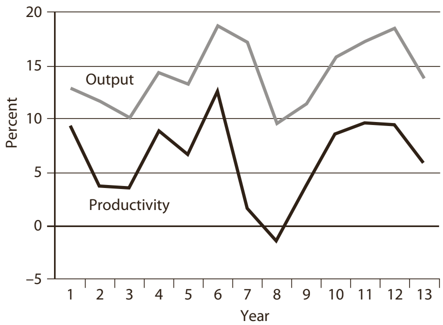


FIGURE 9 • **Annual changes in the technology index and productivity of Firm *B***

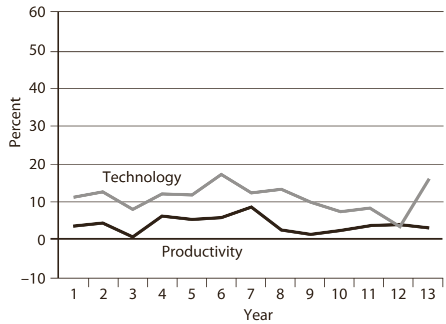


FIGURE F7 • **Annual changes in the output and productivity of Firm *B***

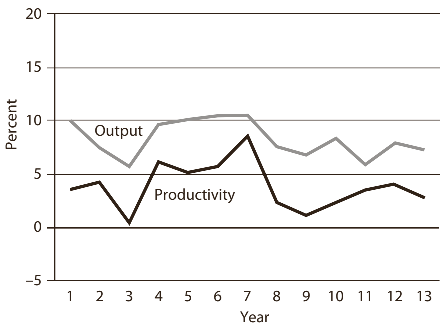
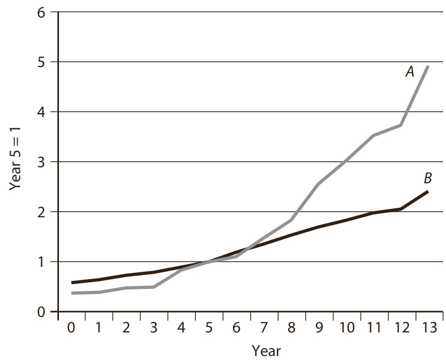


FIGURE F10 • **The technology indices of Firms *A* and *B***



• *Csongor István Nagy* •

THE FRAMEWORK OF NATIONAL PRICE REGULATION IN THE EUROPEAN UNION, WITH SPECIAL EMPHASIS ON LIBERALIZED MARKETS*

The paper analyses the requirements erected by EU competition rules, free movement (internal market) law and market liberalization rules against national price regulation. The paper summarizes the main EU law requirements against national price regulation in liberalized markets and assesses, from a critical perspective, the ECJ's jurisprudence.

EU COMPETITION RULES AND NATIONAL PRICE REGULATION – REGULATED PRICES AS RESTRICTIONS OF FREE COMPETITION

EU competition rules surprisingly do not contain provision on national price regulation. EU competition law does not interdict national rules distorting competition in the market, these are, in general and in themselves, not prohibited; EU law only contains specific bans in this regard. EU law prohibits only particular distorting measures but contains no general prohibition.

EU law contains specific regimes on state action: state aid law, sectoral regulation (market liberalization), as well as a certain immunity for services of general economic interest (*Nagy* [2010] pp. 32–34); from these, sectoral regulation will be addressed in detail. The concept of services of general economic interest (as embedded in Article 106 TFEU) provides an exception (immunity) to an existing legal obligation. The applicability of state aid rules (Articles 107–109 TFEU) may be considered as a theoretical possibility with respect to national price regulation. However, the ECJ established, very early, in *van Tiggele*[1] that state aid rules do not apply to national price regulation, taking into account that the benefit is not conferred by the Member State, and cannot be traced back to public resources.

Article 107(1) TFEU prohibits "any aid granted by a Member State or through State resources in any form whatsoever which distorts or threatens to distort competition by favouring certain undertakings or the production of certain goods shall, in so far as it affects trade between Member States, be incompatible with the internal

* This paper is based on research carried out in the frame of OTKA PD-101612 research program (Hungarian National Scientific Research Fund).

[1] Case 82/77 *van Tiggele* (1978) ECR 25.

market." Accordingly, a subsidy may qualify as illegal if it is "granted by a Member State or through State resources".

In *van Tiggele*, the ECJ held that national price regulation cannot qualify as a state aid as the benefit accruing from the regulated prices is not related (directly or indirectly) to the state budget (paras 24–25). The price floor established for distilled spirits in the Netherlands, in essence, protected traders to the cost of consumers. It could be argued that if, in the absence of regulated prices, the distilled spirits had been sold at a lower price, the regulated prices conferred a monetary benefit on traders in value of the difference between the minimum price and the market price (provided, of course, that the former would have been higher than the latter).

The ECJ established that it does not qualify as a benefit "granted by a Member State or through State resources", if a Member State sets out minimum retail prices to the detriment of the consumers. Although in *van Tiggele* the ECJ dealt with minimum prices protecting traders, the ruling may be, mutatis mutandis, applicable also to price caps protecting purchasers as the benefit is (similarly) not "granted by a Member State or through State resources".

The backbone of EU competition rules comprises of provisions applicable to enterprises. These, in themselves, should not be applicable to Member States, since their addressees are undertakings (save the state or a public entity engages in market conduct). This proposition has to be confined. Notably, if reading competition rules governing market conduct in conjunction with the loyalty clause encapsulated in Article 4(3) TEU[2] and the attached jurisprudence of the ECJ, it can be established that these antitrust rules, through the intermediation of the loyalty clause, do erect requirements against national governments.

On the basis of the loyalty clause, Member States must refrain from adopting or maintaining measures that may make the competition rules (applicable to undertakings) ineffective. It goes counter to this principle, if a Member State prescribes or supports acts that fall foul of EU competition law or reinforces the effects of these, or deprives its own law of its official character through conferring rule-making or legislative power on market operators. A Member State may not force or encourage (assist) an enterprise to violate EU competition rules (*Nagy* [2008] pp. 25–26). Thus, Articles 101–102 TFEU, indirectly though, do establish limits as to state action.

The ECJ condemned numerous Member States for being disloyal, because they encouraged or backed undertakings to conclude agreements restricting competition or reinforced the effects of such agreements or they compelled dominant undertak-

[2] "Pursuant to the principle of sincere cooperation, the Union and the Member States shall, in full mutual respect, assist each other in carrying out tasks which flow from the Treaties. The Member States shall take any appropriate measure, general or particular, to ensure fulfilment of the obligations arising out of the Treaties or resulting from the acts of the institutions of the Union. The Member States shall facilitate the achievement of the Union's tasks and refrain from any measure which could jeopardise the attainment of the Union's objectives."

ings to abuse their dominant position.[3] It has to be stressed, however, that the ECJ, in *Cullet v Centre Leclerc*, established very early that national price regulation, in itself, does not breach Article 101 TFEU applied in conjunction with the loyalty clause (para 18). Furthermore, Article 106 TFEU establishes an exception to this indirect duty in respect of services of general economic interest. Undertakings providing such services "shall be subject to the rules contained in the Treaties, in particular to the rules on competition, in so far as the application of such rules does not obstruct the performance, in law or in fact, of the particular tasks assigned to them."

In *Höfner v Macrotron*,[4] Macrotron hired two employment agents to find a suitable sales manager. However, Macrotron did not engage the person selected by the agents and, as a consequence, refused to pay agents' fee (paras 2 & 11). At the relevant time, according to German law, employment intermediation came under the exclusive competence of the federal employment office (paras 3–6). Notwithstanding this regulation, numerous enterprises operated in this market, providing consultancy and agency services concerning the recruitment of corporate executives and managers. Though the federal employment office tolerated this 'grey market', these agency contracts were invalid as they violated the federal employment office's legal monopoly and, thus, German law. This should also have been the fate of the above agreements (paras 8–10).

The ECJ, after establishing that the federal employment office was an undertaking as it pursued market activities, examined whether Germany breached the loyalty clause through conferring a legal monopoly on the federal employment office. It established that national law is incompatible with EU law, if it creates a plight where the enterprise cannot avoid violating Article 102 TFEU. The mere fact that a Member State creates a dominant position i.e. it confers a legal monopoly on an enterprise, does not fall foul of Article 102 TFEU, unless the enterprise is under the necessity of abusing its dominance. According to Article 102(b) TFEU, the limitation of production, markets or technical development qualifies as an abuse if it occurs to the prejudice of consumers The ECJ established that Germany pushed the federal employment office, which qualified as an undertaking from the perspective of competition rules, into violating of Article 102(b) TFEU. Germany conferred a legal monopoly on the office but the latter was not capable of satisfying the demand, while market operators were, due to the legal prohibition, prevented from providing the service concerned (paras 27 and 29–31). The same reasoning was used, in essence, by the Court in *Job Centre II.*[5]

[3] The ECJ pronounced this principle very early, in Case *G.B.-INNO-B.M.* See Case 13/77 *NV GB-INNO-BM* (1977) ECR 2115. As to judicial practice see *Nagy* [2008] 110–131.
[4] Case C-41/90 *Klaus Höfner and Fritz Elser v Macrotron GmbH* (1991) ECR I-1979.
[5] Case C-55/96 *Job Centre coop. arl* (1997) ECR I-7119, paras 29–35.

In *Ambulanz Glöckner*,[6] the ECJ condemned Germany for forcing and encouraging an undertaking to reserve an ancillary market through preventing market entry. According to German law, patient transport services (both emergency and non-emergency) came under the responsibility of the *Länder*, the administrative districts of each region (*Land*) and municipalities, which, however, could provide these public services through licensed non-profit medical aid organizations *(Sanitätsorganisationen).* These organizations were supervised by the *Länder* and the administrative districts, which gave directions and bore the costs (para 4). According to the rules, other organizations could also be authorized in addition to the medical aid organizations. As a matter of practice, this could take place only if the medical aid organizations' capacities were fully utilized and accordingly, this statutory requirement created a *de facto* monopoly position (paras 6–8).

The dispute emerged after the license of Ambulanz Glöckner was not prolonged because the local medical aid organizations had surplus capacities and were making losses (para 13).[7]

The ECJ reiterated the holding of *Höfner v Macrotron*. Although a dominant position conferred by national special or exclusive rights is, in itself, not incompatible with EU competition rules, a Member State violates these rules, if the exercise of these special or exclusive rights incites the enterprise to abuse its dominant position or creates a situation like this (para 39). It qualifies as an abuse, if a dominant undertaking, without any objective reason, reserves for itself an ancillary activity. It may infringe competition rules, if the extension of the dominance of an undertaking disposing of special or exclusive rights is the result of a state measure (para 40). This is the case, if there is a sufficiently high probability that taking into account the economic characteristics of the market in question, this prevents enterprises from other Member States from providing ambulance transportation services or from establishing themselves there (operative part of the judgment). Such conduct is regarded as a restriction under Article 102(b) TFEU (para 43).

These cases may provide guidance as to price regulation although they only dealt with the creation of a legal monopoly. Price controls are acts of public authority and, as such, immune from the competition rules enshrined in Articles 101 and 102 TFEU (which apply to undertakings). However, the Member State infringes the principle of loyalty if national price regulation compels firms to breach EU competition law, by way of example through abusing their dominant position, Though regulated prices

[6] Case C-475/99 *Firma Ambulanz Glöckner v Landkreis Südwestpfalz* (2001) ECR I-8089.

[7] The earlier regulation permitted the involvement of for-profit organizations as regards non-emergency patient transportation services, and Ambulanz Glöckner was authorized to provide such services. Subsequently, the relevant rules were amended and both emergency and non-emergency services were brought under the scope of the same regulatory regime, which initially applied to emergency patient transportation services.

do not bring about special or exclusive rights, by way of example, excessively low prices, in extreme cases, may result in the restriction of output or the deterioration of quality or may lead to sales at loss.

In summary, EU competition rules (aside from the rules on market liberalization, which will be addressed below) do not contain substantial limits as to national price regulation. Regulated prices are not regarded as state aid as they cannot be traced back to budgetary resources. Theoretically regulated prices may compel an enterprise to violate EU competition rules (e.g. restriction of output, sales below costs) but this scenario may occur only in extreme cases.

NATIONAL PRICE REGULATION AND THE LAW OF THE INTERNAL MARKET: REGULATED PRICES AS BARRIERS TO FREE MOVEMENT

EU law interdicts national customs duties and measures having equivalent effect and creates a customs union (Articles 28–31 TFEU). Likewise, EU law prohibits quantitative restrictions and measures having equivalent effects to quantitative restrictions (Articles 34–35 TFEU), though Member States may adopt such measures with reference to local legitimate ends (Article 36 TFEU).

The jurisprudence of the ECJ defines measures having equivalent effects to quantitative restrictions (MEQR) extremely widely. The judicial practice distinguishes between product-bound measures and state acts regulating selling arrangements. The former relate to the physical composition and appearance of the product (such as content, packaging), while the latter refer to the way products are marketed (e.g. Sunday trading rules, licensing requirements).

The treatment of product-bound measures is very strict and was established in *Dassonville*, even non-discriminatory measures are prohibited, if they restrict market access: "[a]ll trading rules enacted by member states which are capable of hindering, directly or indirectly, actually or potentially, intra-community trade are to be considered as measures having an effect equivalent to quantitative restrictions."[8]

This extremely wide, all-embracing notion was confined in *Keck et Mithouard*[9] (which emerged from the French rules prohibiting sales at loss), where the ECJ introduced the concept of 'selling arrangements', holding that, as opposed to product-bound measures, rules on selling arrangements are considered to be measures having equivalent effects to quantitative restrictions only if they are, in law or in effect, discriminatory.

[8] Case 8/74 *Procureur du Roi v Benoît and Gustave Dassonville* (1975) ECR 837, para 5.
[9] Joined Case C-267 and 268/91 *Keck and Mithouard* (1993) ECR I-6097.

- "the application to products from other Member States of national provisions restricting or prohibiting certain selling arrangements is not such as to hinder directly or indirectly, actually or potentially, trade between Member States within the meaning of the Dassonville judgment (…), so long as those provisions apply to all relevant traders operating within the national territory and so long as they affect in the same manner, in law and in fact, the marketing of domestic products and of those from other Member States." (Para 16.)

Member States may maintain restrictions with reference to local legitimate ends under two doctrines. In *Cassis de Dijon*,[10] the ECJ created an exception within the concept of MEQR, holding that non-discriminatory measures justified by the local public interest are not MEQR.

- "Obstacles to movement within the community resulting from disparities between the national laws relating to the marketing of the products in question must be accepted in so far as those provisions may be recognized as being necessary in order to satisfy mandatory requirements relating in particular to the effectiveness of fiscal supervision, the protection of public health, the fairness of commercial transactions and the defence of the consumer." (Para 8.)

Second, Article 36 TFEU contains a statutory exception to the prohibition of MEQR.

- "The provisions of Articles 34 and 35 shall not preclude prohibitions or restrictions on imports, exports or goods in transit justified on grounds of public morality, public policy or public security; the protection of health and life of humans, animals or plants; the protection of national treasures possessing artistic, historic or archaeological value; or the protection of industrial and commercial property. Such prohibitions or restrictions shall not, however, constitute a means of arbitrary discrimination or a disguised restriction on trade between Member States."

One of the major differences between the doctrine of mandatory requirements and Article 36 TFEU is that the former only applies to indistinctly applicable (i.e non-discriminatory) measures, while under Article 36 TFEU even discriminatory measures can be justified. Furthermore, under *Cassis de Dijon*, any local legitimate end may justify a restriction while Article 36 TFEU enumerates the public interest goals that may be used for this purpose.

Accordingly, national price regulation qualifies, in principle, as a measure on selling arrangements and is, hence, prohibited if (directly or indirectly; actually

[10] Case 120/78 *Rewe-Zentral AG and Bundesmonopolverwaltung Für Branntwein (Cassis de Dijon)* (1979) ECR 649.

or potentially) restricts inter-state trade and is (in law or in effect) discriminative. In short, national price regulation entails concerns under the law of the internal market, if it thwarts the market access of foreign products or enterprises and fails to comply with the requirement of equal treatment. As a matter of terminology, it has to be noted that the ECJ's judgment in *Keck et Mithouard*, where the distinction between product-bound measures and rules on selling arrangements was established, was rendered on 24 November 1993 and beforehand, the ECJ used the *Dassonville* formula.

In *Tasca*,[11] Italy set a cap on the consumer prices of certain sugar varieties. The ECJ held that this, in itself, did not qualify as a MEQR, However, it might have had such an effect if the price-cap made the sale of foreign products more difficult or impossible. This is the case if the maximum price is so low that it makes the sale of import products unprofitable (para 13). This proposition was endorsed in *SADAM*[12] and *GB-Inno-BM*.[13]

In *van Tiggele,*[14] the ECJ dealt with price floors. The Netherlands established minimum prices for distilled spirits. The ECJ considered that while price regulation normally does not go counter to free movement, in certain cases it may.

- "[I]mports may be impeded in particular when a national authority fixes prices or profit margins at such a level that imported products are placed at a disadvantage in relation to identical domestic products either because they cannot profitably be marketed in the conditions laid down or because the competitive advantage conferred by lower cost prices is cancelled out." (Para 14.)

The prohibition of sale at loss places equal burdens on imported and domestic products and thus may not qualify as a MEQR (para 16.).

If a Member State sets a maximum margin as a particular amount (and not as a price-cost ratio), it may entail no detrimental effects as to the potentially cheaper import products, as in the case at stake where the retail margin made up only a relatively insignificant part of the final retail price (para 17.).

Contrary to the above, the setting of a price floor as a particular amount, which applies indistinctly both to imported and domestic products, may have a negative impact on the former, "in so far as it prevents their lower cost price from being reflected in the retail selling price" (para 18).

If different price-setting methods are used as to imported and domestic products, this may qualify as a MEQR. In *Roussel Laboratoria*[15] the ECJ established that it is

[11] Case 65/76 *Tasca* (1976) ECR 291.
[12] Joined Cases 88–90/75 *SADAM* (1976) ECR 323, para 15.
[13] Case 13/77 *NV GB-INNO-BM* (1977) ECR 2115, para 52.
[14] Case 82/77 *van Tiggele* (1978) ECR 25.
[15] Case 181/82 *Roussel Laboratoria* (1983) ECR 3849.

contrary to Article 34 TFEU if a Member State fixes prices of import products on the basis of the price charged by the producer of the pharmaceutical products in the country of origin, while the price is frozen for domestic products on the level at a given reference date (para 25).

In *Cullet v Centre Leclerc,*[16] French law established minimum retail prices for fuels exclusively with reference to the prices of French refineries. The latter were covered by a price limit although this was, in principle, set in accord with European prices, if the latter were more than 8% higher or lower than these prices, the price limit was determined by the cost prices of the French refineries. As a matter of practice, French refineries sold at the highest prices permitted by the law (para 5). The ECJ found that this price-setting methodology is adverse to foreign products as it deprives them of the competitive advantage they may enjoy due to their potentially lower prices (para 29).

In *Roelstraete,*[17] Belgian law regulated the retail margin of meat. It was determined, as a fixed sum, the margin a butcher could include in the price per kilogram. Retail prices consisted of the average weighted purchase price calculated on the basis of the receipts of the preceding four weeks, the maximum gross profit margin (determined as a fixed sum) and the value-added tax (para 2).

The ECJ considered that the Belgian legislation, through fixing the sum of the margin applicable to both domestic and imported products, discouraged traders from importing meat from other Member States. Traders had to cover the import expenses from this margin, that is, the extra costs attached to importation decreased the profitability of import products in comparison to domestic products (paras 21–22). The same line of interpretation was followed by the Court in *Lefèvre,*[18] as to a similar French regime (para 13).

In *Vocarex,*[19] Belgian law prohibited sales at a low margin (sales at loss were prohibited and excessively low margins were considered to be sales at loss) (para 3). The ECJ found that this price regulation was not a MEQR.

In summary, national price regulation violates the rules of the internal market (free movement), if it distinctively restricts the entry of foreign products into the market of another Member State. Accordingly, a price cap is incompatible with the internal market, if it prevents or thwarts the importation of more expensive foreign products (which may be more expensive due to the importation costs). A price floor is counter to the law of the internal market if it restricts the importation of foreign products through depriving them of their competitive advantage consisting in lower prices. The prohibition of sale at loss (or at a negligible margin), in principle, does not infringe the law of the internal market.

[16] Case 231/83 *Cullet v Centre Leclerc* (1985) ECR 305.
[17] Case 116/84 *Henri Roelstraete* (1985) ECR 1705.
[18] Case 188/86 *Lefèvre* (1987) ECR 2963.
[19] Case C-63/94 *Vocarex* (1995) ECR I-2467.

The ECJ's jurisprudence on retail minimum prices can be criticized on the ground that price floors do not make traders disinterested in purchasing cheaper foreign products. From an economic perspective, retail price floors do not make the trader interested in purchasing and reselling the more expensive local products. Instead, the trader's interest is to keep on purchasing the cheaper foreign merchandise and to resell it at a higher margin. Though the minimum price is secured by the law, the trader will try to cut input costs to a minimum to maximize his profit. The trader is obviously not interested in sharing the guaranteed margin with domestic producers through purchasing the more expensive local brand. However, the ECJ's concerns are valid from a consumer perspective. The Court very probably had in mind the situation where both the foreign and the domestic product is on the shelves of the local convenience store and the final consumer has to choose one. If they are of the same or similar price, the consumer would very likely opt for the local (well-known) brand as opposed to the plight where the price of the imported product is significantly lower than that of the domestic merchandise. In this case, the lower price may be a reason to overcome the loyalty to the local brand.

MARKET LIBERALIZATION RULES AND PRICE REGULATION: REGULATED PRICE AS THE HINDRANCES TO MARKET OPENING

The ECJ's jurisprudence (in particular *Federutility* and *Enel*)[20] suggests that in industries covered by market liberalization regimes (electronic communications, electricity, natural gas, railway, postal services), Member States may not introduce or maintain price regulation, unless they are specifically authorized or obliged. Although these regimes permit states to adopt and maintain such regulatory measures with reference to public services, this price-control is confined both in terms of scope and time.

Natural gas

Directive 2009/73/EC governing the internal natural gas market treats network services (access to the network) as natural monopoly and, accordingly, subjects them to regulated prices which are to be set by the national regulatory authority. On the contrary, the energy product itself is not subject to price control, the production and trade of natural gas are regarded as activities prone to competition. The only provision that, in respect of the latter, makes mention, in the context of public services, of regulated prices is Article 3(2) of the Directive. The Directive on internal natural gas market (in line with its predecessor, Directive 2003/55/EC) does not pronounce

[20] Case C-242/10 *Enel*.

natural gas to be universal service which is very probably due to the circumstance that the majority of European consumers do not use it (see *Cremer et al.* [1998] p. 7.). Nonetheless, the Directive establishes public service obligations in Article 3(2) and consumer protection requirements in Article 3(3). These provisions enable Member States to introduce universal service also in respect of natural gas, albeit the national playing field was confined by the ECJ in *Federutility*.[21]

In this case, Italy set regulated prices on the basis of reference prices, justified by the absence of workable competition and the interests of final consumers. The ECJ established that this can be maintained only under certain conditions although Member States may regulate retail prices after the opening of the market (that is, 1 July 2007) (paras 17–24).

The ECJ examined the legality of price regulation by investigating whether the liberalization of the market implies the exclusion of regulated prices. Namely, the Directive on the internal natural gas market contains no specific prohibition on price regulation. The ECJ also considered whether the purpose of assuring public services may justify the exceptional regulation of prices (in case the first question is answered in the affirmative). It is to be noted (again) that the Directive on the internal natural gas market does not contain the concept of universal service (contrary to the Directive on the internal electricity market), hence, natural gas universal service has no independent legal basis and can be deduced only from the Directive's provision on public services and the doctrine of services of general economic interest (as included in the TFEU).

Directive 2009/73/EC provides for the regulation for certain tariffs (e.g. network fees) and apart from these, it contains no express provision as regards national price controls. It neither allows nor prohibits regulated prices. The ECJ, however, deduced this prohibition from the Directive's purpose (market liberalization and the competitive natural gas market – para 19) and the consumers' right to freely choose the service provider.[22]

- "Although it is not explicitly stated in that provision [Article 23(1)(c) of Directive 2003/55], or indeed in any other provisions of that directive, that the price for the supply of natural gas must, as from 1 July 2007, be determined solely by the operation of supply and demand, that requirement follows from the very purpose and the general scheme of that directive, which, as its 3rd, 4th and 18th recitals state, is designed progressively to achieve a total liberalisation of the market for national gas in the context of which, in particular, all suppliers may freely deliver their products to all consumers." (Para 18.)

[21] Although in this case the ECJ interpreted the previous natural gas directive, there is no difference between the two directives as to the provisions concerned. For the sake of simplicity, if possible, the references in this paper are made to Directive 2009/73/EC, although the ECJ's judgment refers to Directive 2003/55/EC.

[22] Articles 2 and 23 of Directive 2003/55/EC and para 17 of the judgment.

Afterwards, the ECJ addressed the Directive's provision on public service[23] to ascertain the conditions of derogating from the above general prohibition. It held that "it follows from the wording of Article 3(2) that measures adopted on the basis thereof must be adopted in the general economic interest, be clearly defined, transparent, non discriminatory and verifiable, and guarantee equality of access for EU gas companies to national consumers" (para 22). Member States have no automatic right to regulate prices with reference to public services and this can be done only in case of necessity and the regulatory intervention has to be restricted both in terms of scope and time. Article 3(2) of the Directive permits state intervention only by reason of general economic interest and makes an express reference to TFEU 106.

Member States have a very wide margin of appreciation as to the definition of the general economic interest, and local peculiarities and policy considerations may legitimately gain ground in the frame of this. "In that context, Member States are entitled, while complying with the law of the Union, to define the scope and the organisation of their services in the general economic interest"; "[i]n particular, they may take account of objectives pertaining to their national policy". (Para 29.) EU law follows a minimalist approach as to the definition of services of general economic interest, "requirements concerning the public service must be capable of being interpreted, subject to the observance of the law of the Union, 'on a national basis', and 'taking into account national circumstances'." (Para 30.)

- "It follows from the above that Directive 2003/55 allows Member States to assess whether, in the general economic interest, after 1 July 2007, it is necessary to impose on undertakings operating in the gas sector public service obligations in order, in particular, to ensure that the price of the supply of natural gas to final consumers is maintained at a reasonable level having regard to the reconciliation which Member States must make, taking account of the situation in the natural gas sector, between the objective of liberalisation and that of the necessary protection of final consumers pursued, as mentioned in paragraphs 18 and 20 of this judgment, by the Union legislature." (Para 32.)

It is to be noted that though Member States have, indeed, a wide give as to the definition of services of general economic interest, the ECJ's judgment implies that the power to regulate prices is not automatic, can be exercised only in case regulatory intervention is justified and, if they do, they have to set out their reasons, in particular because under Article 3(11) of the Directive on the internal natural gas market Member States are obliged to inform the Commission about all measures adopted for the provision of public services and about their potential impact on domestic and international competition.

[23] Article 3(2) of Directive 2003/55/EC.

In the case at stake, the ECJ considered the protection of final users against the market power of service providers to be a legitimate general economic interest and sanctioned the endeavor to pursue "a general economic interest consisting in maintaining the price of the supply of natural gas to final consumers at a reasonable level having regard to the reconciliation which Member States must make, taking account of the situation in the natural gas sector, between the objective of liberalisation and that of the necessary protection of final consumers pursued by [the] Directive."[24]

Regulatory intervention, taking the form of price regulation, shall be proportionate. It follows from the requirement of proportionality that its ambit has to be limited to the domain where it is necessary (material scope); may be maintained only for a limited period of time (temporal scope) and may cover only users who truly need protection (personal scope). Public service obligations set out by the Member State "may compromise the freedom to determine the price for the supply of natural gas only in so far as is necessary to achieve the objective in the general economic interest which they pursue and, consequently, for a period that is necessarily limited in time." (Para 33.)

- "First, such an intervention must be limited in duration to what is strictly necessary in order to achieve its objective, in order, in particular, not to render permanent a measure which, by its very nature, constitutes an obstacle to the realisation of an operational internal market in gas. (…)" (Para 35.)

  "Secondly, the method of intervention used must not go beyond what is necessary to achieve the objective which is being pursued in the general economic interest." (Para 36.)

  "If, following those verifications, it were to be shown that such an intervention is capable of being justified in that way, the requirement of proportionality would imply in particular that it be limited in principle to the price component directly influenced upwards by those specific circumstances" (Para 38.)

  "Thirdly, the requirement of proportionality must also be assessed with regard to the scope ratione personae of the measure, and, more particularly, its beneficiaries." (Para 39.)

In the context of limits in terms of time, the ECJ highlighted that a mere reference to the national legislation's provisory nature is not sufficient.

- "[T]he referring court should examine whether and to what extent the relevant national law requires the administration to make a periodic re-examination, at close intervals, of the need for it to intervene in the gas sector and the manner of its doing so, having regard to the development of that sector." (Para 35.)

[24] Operative part of the judgment.

The ECJ made a very important remark regarding the possible beneficiaries of price regulation. While the Court recognized enterprises as well as consumers may benefit from the regulatory intervention, it also stated that the public service obligations, in principle, do not fulfil the requirement of proportionality. On the other hand, not only vulnerable but all consumers may be sheltered through price regulation.

- "In that regard, it should be emphasised that that requirement does not prevent 'reference prices' for the supply of natural gas, such as those at issue in the main proceedings, from being applied to all customers whose consumption of natural gas is above a certain threshold rather than being limited to the circle of those, expressly referred to in Article 3(3) of Directive 2003/55, who must necessarily be protected on account of their vulnerability.' (Para 40.)

  "If, as some of the applicants in the main proceedings maintain before the Court, the definition of 'reference prices' for the supply of natural gas, such as those at issue in the main proceedings, applies also to undertakings irrespective of their size, which it is for the referring court to verify, it should be noted that Directive 2003/55 does not in principle exclude the possibility that the latter may also benefit, as final consumers of gas, from the public service obligations which Member States may adopt in the context of Article 3(2) of that directive. The 26th recital of that directive states in particular that measures taken by the Member States to protect final consumers may differ according to whether they are addressed to households or to small- and medium-sized undertakings." (Para 41.)

  " In that case, however, it would be necessary to take account, in assessing the proportionality of the national measure in question, of the fact that the situation of undertakings is different from that of domestic consumers, the objectives pursued and the interests present being not necessarily the same and also of objective differences between the undertakings themselves, according to their size." (Para 42.)

- "In those circumstances, apart from the specific case, referred to at the hearing, of management companies of apartment blocks, the requirement of proportionality referred to above would not, in principle, be complied with if the definition of 'reference prices' for the supply of natural gas, such as those at issue in the main proceedings, were to benefit individuals and undertakings in an identical manner, in their capacity as final consumers of gas." (Para 43.)

The public service obligations have to be clearly defined, transparent, non-discriminatory and verifiable, and shall guarantee equal access for European gas companies to consumers (para 44). In respect of equal treatment, it has to be inquired

- "…whether, having regard to the whole of the measures which may have been taken in that area by the Member State concerned, the definition of 'reference prices' for the supply of natural gas, such as those at issue in the main proceedings, which applies in an identical manner to all undertakings supplying natural gas, must nevertheless be regarded as discriminatory." (Para 45.)

  It would be so if such intervention were to lead in reality to imposing the financial burden arising "from the intervention mainly on some of those undertakings, in this case those not also carrying on the business of producing/importing natural gas." (Para 46.)

It has to be noted that in *Federutility*, the ECJ faced a general, industry-wide price regulation, which covered one subgroup of market operators. However, this does not qualify as *strictu sensu* universal service price regulation where the Member State appoints a universal service provider (supplier of last resort, default supplier, public service operator etc.) whose prices are fixed.

Para 33 of the ECJ's judgment refers to "the freedom to determine the price for the supply of natural gas" and makes no mention of cases where a Member State ensures the availability of natural gas at a given price through appointing a universal service provider. It is dubious whether the holding of *Federutility* encompasses only general industry-wide price regulation or it extends also to prices secured through a universal service provider. This is a pivotal question: although natural gas is not considered to be an EU universal service, quite a few Member States characterize it as such. In Hungary, Act XL of 2008 on the supply of natural gas, in Section 34(1), repeats the corresponding provisions of electricity universal service, with the difference that in the natural gas sector universal services extends only to those users who are already connected to the infrastructure. In Spain, natural gas equally qualifies as universal service and reasonable prices are secured. Spain terminated the regulation of retail prices and provides cost-based prices to small consumers in the form of supplier of last resort prices (*IEA* [2009] p. 64 and p. 70, *CNE* [2010] pp. 12–13, *CNE* [2011] pp. 124–126).

Finally, it is to be noted that this Italian regime survived, with strings attached, the ECJ's preliminary ruling procedure and remained in force (*Cavasola–Ciminelli* [2012] 114.).

Electricity market

The regulatory and statutory pattern of Directive 2009/72/EC on the internal electricity market is, in essence, in line with the Directive on the internal natural gas market, with the significant difference that electricity is considered to be an EU universal service. Accordingly, the Directive on the internal electricity market treats network services (access to the network) as a natural monopoly and subjects them to regulated prices, which are to be set by the national regulatory authority, while the energy product itself is not subject to price control. The production and trade of electricity are regarded as activities prone to competition. The possibility of price regulation is mentioned both in the provision on public services, in Article 3(2), and in the provision on universal service, in Article 3(3); the latter embraces "the right to be supplied with electricity of a specified quality within their territory at reasonable, easily and clearly comparable, transparent and nondiscriminatory prices." The institution of universal service enshrined in Article 3(3) implies the automatic right and duty of Member States to intervene (contrary to public service obligations embedded in Article 3(2) which tie regulatory intervention to conditions). The legal basis of the regulatory intervention is given, though the method and extent – notwithstanding the wide margin of appreciation of Member States – is subject to EU law restraints.

Due to this conceptual difference, the ECJ's jurisprudence on natural gas can be extrapolated to the electricity market only outside the ambit of universal service and raises the question "does the ECJ's ruling in *Federutility* provide guidance as to the electricity market?" The system and structure of the natural gas and electricity directives are very similar and the goal of market liberalization may be deduced in both schemes and the provisions on public services are, in essence, identical. This general parallelism was approved also by the ECJ in *Enel*,[25] where it referred to *Federutility* as guidance; AG Cruz Villalón, in his opinion in *Enel*, mentions *Federutility* as the "the first case-law concerning those rules on public service obligations in the energy sector – provides a direct precedent for the interpretation of Article 3(2) of Directive 2003/54" (para 2). Accordingly, the analysis of *Enel*, in essence, follows the structure and pattern established in *Federutility* (paras 32, 39 & 50).

It must be stressed that the ruling in *Federutility* provides no guidance as to the construction of the electricity universal service, though it is good law outside this domain. It is important to emphasize this again, since the above regulatory differences between the two directives are quite often disregarded and not infrequently the position emerges in the scholarship that *Federutility* is directly applicable to the electricity sector (*Energy Community* [2012] pp. 5–6.).

As regards the electricity market, it was the *Enel* judgment that pronounced that, in the absence of a specific authorization to regulate prices, regulatory price controls are, in principle, prohibited but may be justified with reference to general economic interests. In this case, Italy obliged, as regards dispatching and balancing services, electricity companies disposing of installations essential to the operation and security of the electricity system to make supplies to the distribution system operators at fixed prices.

[25] Case C-242/10 *Enel Produzione SpA v Autorità per l'energia elettrica e il gas* (2011) ECR I-13665.

When analyzing the case, the ECJ simply skipped the detailed examination of whether, in the absence of express authorization, national price regulation is contrary to the Directive on the internal electricity market. The Court took this prohibition as given, and moved to analyze the exception to this prohibition, that is the pre-conditions of regulatory intervention. Accordingly, the Court examined the existence of three requirements:

- "Legislation providing for such intervention must pursue an objective in the general economic interest and be consistent with the principle of proportionality", as well as "[s]uch obligations must also be clearly defined, transparent, non-discriminatory and verifiable, and must guarantee those undertakings equality of access to national consumers." (Para 48.)

According to the ECJ, "the electricity dispatching service is a public service designed to ensure that, within the national transmission system, the supply of electricity matches demand, thereby guaranteeing security and continuity in the energy supply)" (para 51). Accordingly, the ECJ considered that "rules governing essential installations pursue a general economic interest objective" (para 54). The requirement of proportionality is met, if regulation is "appropriate for securing the objective which it pursues" and does not "go beyond what is necessary in order to attain it" (para 55). As to suitability, the Court noted that the national regulation "applies solely in cases where there is only one generating unit which, owing to its technical features and the speed with which it can vary its power output, is capable of supplying the resources needed to meet the dispatching requirements" (para 57), "these are generating installations which are strictly necessary and vital in order to meet dispatching service requirements" (para 62). As to whether the intervention went beyond what was necessary (para 63), it established that the regime appeared to secure fair remuneration for operators owning such installations (para 68). Finally, the ECJ established that the intervention was limited in terms of time, because "the list of essential installations is annually reviewed and updated, it would appear that installations are not kept on it for more than a limited period" (para 75).

In electricity reasonable prices are part of the universal service package. The Directive on the internal electricity market gives no guidance as to when can prices be regarded as reasonable, conferring a wide margin of appreciation on Member States.[26] In 2010, 16 out of 27 Member States had in force some form of price regulation in the house-hold segment.[27] The usual mechanism was to appoint a universal service provider (supplier of last resort, public service provider) and to cap prices.

[26] *Cameron* [2005] p. 25.

[27] *ERCEG* [2010]. As to the determination of end-user regulated price see *ERCEG* [2010] p. 11.

SUMMARY

This paper analysed Member States' give as to price regulation from the perspective of the EU internal market and competition rules.

National price control is incompatible with the internal market, if it thwarts free movement (that is, the market access of foreign products, services or the freedom of establishment). National price regulation, as a set of rules relating to selling arrangements, goes counter to the law of the internal market, if it entails differential treatment as to foreign and domestic products. Accordingly, a price-cap falls foul of the law of free movement, if it prevents the importation of more expensive foreign products (which may be more expensive because of the importation costs). Likewise, a price floor infringes the law of the internal market, if it deprives foreign products of their competitive advantage consisting in lower prices. On the other hand, the prohibition of sale at loss, at cost-price or at minuscule margin, in principle, does not violate the law of the internal market.

Only the rules of market liberalization, in essence, can be regarded as substantial restrictions as to national price regulation from EU competition rules. This implies that the industries not covered by sectoral (market opening) regimes are free from these restraints. Price regulation may not be considered to be state aid as it does not concern the public budget. Although there is a theoretical possibility that through determining regulated prices a Member State incites or compels an enterprise to infringe competition rules (e.g. restricting output, selling at loss), as a matter of practice, this can be established only in extreme cases.

The ECJ established, in the context of energy market liberalization, that, as a general principle, if the applicable sectoral regime does not specifically authorize a Member State to regulate prices, national price control is, in principle, illegal, and can be maintained only with reference to the general economic interest. As a corollary, national price regulation may be introduced and maintained only if it is justified, and has to be proportionate in terms of scope and time. This doctrine may be extrapolated to other liberalized markets.

The ECJ established in *Federutility* that market opening excludes price regulation in itself, that is, in liberalized markets it is irreconcilable with EU law's command of market opening to regulate prices. This implies the proposition that in case of public services (services of general economic interest) price controls may be, tough exceptionally, maintained, however, certain requirements apply. Member States have no subjective right to regulate prices as this would go counter to the concept of liberalized market. Furthermore, no matter how long the road to workable competition, price regulation – at least a general application – must be provisory.

The prohibition established in *Federutility* is very interesting if put in the light of the industries that have remained intact from the liberalization waves of the preceding decades (that is, most part of the EU's economy). In liberalized markets,

price regulation, in the absence of specific authorization, is considered to be prohibited, if it restricts competition. Outside this domain, national price regulation is prohibited, if it restricts importation in addition to the restriction of competition. Here, national price control is regarded as unlawful, if it restricts imports (or export) thus restricting free movement. This is the case, if it makes the market entry of foreign products impossible or less attractive. This bifurcation may lead to self-contradiction. In liberalized markets price controls may not be maintained once competition becomes workable. In sectors where no liberalization rules were introduced, presumably because these were not needed, as they were liberalized and competitive markets, the regulation of prices cannot be challenged on the basis of EU law, unless it is an obstacle to free movement.

It is also important to note that the limits set in *Federutility* do not apply to services qualifying as universal service. This concept was not analysed by the ECJ in *Federutility*, as this case was based on the Directive on the internal natural gas market which does not contain this concept (contrary to the Directive on the internal electricity market). In case of universal service, if it is codified in the market liberalization regime at stake, there is an express EU law requirement to secure the fairness (affordability, reasonableness) of prices. Regulatory intervention motivated by the purpose of universal service is 'self-justifying'.

In *Federutility*, the ECJ established that, since there is no universal service as to natural gas, price regulation has to be justified with reference to public services. The status of electricity is, however, different, due to the presence of the regulatory concept of universal service. The electricity universal service encompasses the right to be supplied with electricity of a particular quality at "reasonable, easily and clearly comparable, transparent and nondiscriminatory prices"[28] and the right to reasonable prices creates an automatic possibility and duty to regulate prices, without a demand for justification. Accordingly, the regulatory intervention carried out for the purpose of universal service demands no such justification, it is automatically legitimate, provided it does not transgress the EU regulatory framework. EU law demands the provision of universal service, including the requirement of reasonable prices.

Another very important point is that while in *Federutility* Italy introduced industry-wide price regulation, the customary mechanism of providing universal service is the appointment of a universal service provider, which is subject to universal service requirements, including the requirement of reasonable prices. In this scheme, formally, there may be no regulated price. It is another question that the economic effects of the general industry-wide price-cap and those of the universal service price (which is applied solely to the appointed universal service provider(s) but not enforced on all market operators) are similar (in fact, the same). In the latter case, though alternative service providers may, from a legal perspective, charge higher

[28] Article 3(3).

prices, due to the presence of the universal service price (capped by the regulator), they very probably would not be able to sell these products.

It is to be noted though that the foregoing is only half-true and a little bit over-simplified. Service providers compete along various parameters, not only with price, but with quality. This includes the quality of the physical product but also the quality of the customer service, billing and customer-relations. Regulated-prices have straight-jacketing effects on alternative market operators. Still, it has to be noted that notwithstanding the above economic equivalence between industry-wide price regulation and the universal service provider's prices, the ECJ may easily come to the conclusion that the prices enforced on the universal service provider (which are some sort of a social transfer) may not share the fate of the regulated prices condemned in *Federutility*. The pricing freedom of alternative (non-universal) service providers is, at least from a legal perspective, not subject to restriction.

REFERENCES

Cameron, P. [2005]: Completing the internal market in energy: an introduction to the new legislation. In: *Cameron, P.* (ed.): Legal aspects of EU energy regulation. Oxford University Press, Oxford.

Cavasola, P.–Ciminelli, M. [2012]: Italy. In: *Ninane, F.–Ancel, A.–Eskenazi, L.* (ed.): Gas Regulation in 32 Jurisdictions Worldwide. Getting The Deal Through, pp. 111–116.

CNE [2010]: Spanish Energy Regulator's Annual Report to the European Commission, 2010. Comisión Nacional de Energía.

CNE [2011]: Spanish Energy Regulator's Annual Report to the European Commission, 2011. Comisión Nacional de Energía.

Cremer, H.–Gasmi, F.–Grimaud, A.–Laffont, J.-J. [1998]: The Economics of Universal Service: Practice. The Economic Development Institute of the World Bank, http://info.worldbank.org/etools/docs/library/64601/practice.pdf.

Energy Community [2012]: Regulated Energy Prices in the Energy Community: State of Play and Recommendations for Reform. A Report by the Energy Community Secretariat, http://www.energy-community.org/pls/portal/docs/1568177.PDF.

ERCEG [2010]: Status Review of End-User Price Regulation as of 1 January. 2010. European Regulators Group For Electricity & Gas: http://www.energy-regulators.eu/portal/page/portal/EER_HOME/EER_PUBLICATIONS/CEER_PAPERS/Customers/Tab1/E10-CEM-34-03_price%20regulation_8-Sept-2010.pdf.

IEA [2009]: Energy Policies of IEA Countries: Spain. Review, International Energy Agency.

Nagy, Csongor István [2008]: A közösségi és a magyar kartelljog kézikönyve [Handbook of Community and Hungarian Law on Restrictive Agreements]. HVG–Orac, Budapest.

Nagy, Csongor István [2010]: Versenypolitika. In: *Kengyel, Ákos* (ed.): Az Európai Unió közös politikái [Competition Policy in The European Union's Common Policies]. Akadémiai Kiadó, Budapest.

• *Iván Major–Károly M. Kiss* •

INTERCONNECTION AND INCENTIVE REGULATION IN NETWORK INDUSTRIES

The price regulation of network industries has changed tremendously all over the world recently. Theoretical contributions specifically advocate and telecommunications, energy and other market regulators in various parts of the world practice cost-based pricing for inter-firm network access services. Cost-based pricing is performed under the assumption that the regulator has perfect information regarding the costs of producing the services. We show that – under fairly general conditions – cost-based pricing creates incentives for regulated firms *not* to improve their efficiency. Allowing for information asymmetry between the regulator and the regulated firms, we find that incentive regulation will eliminate the adverse effect of cost-based pricing on the firms' efficiency and on social welfare.

INTRODUCTION

The regulation of network industries has received ever greater attention during the current financial and economic crisis than before. This paper addresses the regulation of interconnection prices for firms with interconnected networks with perfect, and with imperfect and asymmetric information. It unites two separate lines of previous analyses. On the one hand, important works by *Armstrong–Doyle–Vickers* [1996], *Laffont–Rey–Tirole* [1998a,b); *Carter–Wright* [1999, 2003] and *Armstrong* [2002], as well as studies by *De Bilj–Peitz* [2002], *Peitz* [2005] and numerous others address the issue of interconnection and termination charges under the assumption that the regulator has perfect information about the true costs of providing inter-firm network access services. On the other hand, the literature is equally extensive on the nature and consequences of asymmetric cost information between the regulator and the regulated firm. The seminal work on regulating a firm with unknown costs was written by *Baron–Myerson* [1982]. Important contributions were made, among others, by Laffont – Tirole [2000] and *Laffont–Martimort* [2002].[1] However, we are aware of only a few studies that combined these two lines of investigation.[2] Some authors did not see the need for doing so. For example, *Armstrong* [2002] noted that "While it is clear that imperfect regulatory knowledge of costs and

[1] *Armstrong–Sappington* [2007] offer an overview of the issues of imperfect information in regulated industries.

[2] See, for instance, *Sappington* [1980], *Stefos* [1990] and *Blackmon* [1994].

the potential for cost reduction has an important impact on regulatory policy, the interaction of these features with the access pricing problem does not often seem to generate many new insights" (p. 380). Believing that perfect regulatory knowledge was an adequate assumption, Armstrong then went on to propose that the regulator should base inter-firm network access prices on "estimated efficient costs," or costs computed from engineering models, or benchmarking.

The principle of cost-based pricing has long dominated the regulatory approaches to pricing end user services. In addition to the major carriers' own cost models, North American regulators required the construction of elaborate service cost simulation models for various levels of service aggregation as early as the 1970s. When, after opening up the market to competition in telecommunications and other utilities markets, the regulation of inter-firm network access prices became a regulatory task of critical importance, cost-based pricing quickly found these industries as a new field of application. Regulators began to demand that network operators provide access to their network for other service companies for charges that were based on long run incremental costs.

Many difficulties are inherent in this approach.[3] We show in this paper that cost-based pricing may signal incentives to firms *not* to improve the efficiency level of interconnection. The adverse effects of cost-based price regulation work through two channels. First, even if the regulator had perfect information about the service providers' call termination costs and based termination charges on those costs, service providers would not be induced to attain high efficiency because a higher efficiency in network interconnection would not result in higher profits for them. This is a direct consequence of the complex cross-price effects in inter-firm services.[4] Second, the adverse effect of cost-based pricing on service providers' efficiency is exacerbated if the regulator's information about the firms' cost is imperfect.

In reality, regulators can never perfectly know the true costs of network access services. More is involved than the informed party's unwillingness to disclose private information, or biases due to the unavoidable arbitrariness of some elements of cost allocation. The firm and the regulator may have some misperceptions about what the other party knows or infers from the information they both possess. For instance, a firm may assume, albeit mistakenly, that the regulator is also aware of some specific information about efficient operation that the firm previously acquired. Consequently, the firm would expect the regulator to incorporate this piece of information into his regulatory decision, although this will not, in fact, occur. Thus, the firm would adjust its output decision to a false assumption. *Madarász* [2007]

[3] *Laffont–Tirole* [2000] briefly mention the shortcomings of cost-based pricing.

[4] *Hansen* [2005] addresses a similar issue that he labels "tariff mediated network externalities." He shows that a low-cost firm will attain a smaller, while a high-cost firm will have a larger market share in equilibrium than would be rational from an efficiency point of view.

labelled this kind of assumption "information projection." The opposite may also happen. The firm may ignore important portions of cost accounting information and assume that the regulator is equally ignorant. According to Madarász, this is "ignorance projection." Cost-based pricing may give rise to simultaneous cases of "information projection" and "ignorance projection". As a result, cost-based pricing may do more harm than good.

Regulatory agencies have recently also recognised some of the weaknesses of a cost-based regulatory design and they started applying "bottom-up" benchmark models in their effort to find efficient prices. Bottom-up models establish the lowest feasible level of costs for each element of the network and then aggregate these cost components up to the level of end user services. We show in this paper that bottom-up benchmarking is not a solution for regulatory games if the firms operate with different marginal costs. In addition, bottom-up cost models may also fare poorly if one of the parties has private information. The more informed party – usually the service provider – can rightfully claim that its actual conditions of business operation differ widely from what the regulator assumes when it constructs an efficiency model of a hypothetical company. These disputes between the firm and the regulator usually end up in the courts where the regulator can rarely win his case. We use the example of the telecommunications industry, but our findings can be easily generalised for other network industries such as energy distribution, transportation, water supply, and postal services, where there is two-way network interconnection between firms.

Our point of departure is a model of customers' choice between service providers, similar to the one presented by *Laffont et al.* [1998*a*], [1998*b*]. We diverge from their model on one important point: we assume that a customer's valuation of network size and each customer's demand for calls are additively separated. Our assumption is supported by empirical observations that indicate that customers may assign a greater value to a service provider with a larger network than to a service provider with a smaller network, but each customer's actual demand for calls will depend on the calling price rather than on the firms' size. In our model, the firms' market share will also be a function of customers' demand for intranet and for off-net calls, but our approach renders it feasible to derive analytical results and conclusions. In addition, we relax the assumption of Laffont et al. about perfect information in the later part of the paper, and develop a different model in which asymmetric information between service providers and regulators is assumed. We shall show that incentive regulation with imperfect information is not merely a more realistic assumption than assuming perfect information of the regulator, but that it also eliminates the adverse effects of cost-based pricing on the firms' efficiency.

The regulatory model of interconnection with imperfect information conveys important policy implications. We demonstrate that incentive regulation extends the proper incentives to firms to improve efficiency and that it results in a smaller social welfare loss than cost-based pricing or bottom-up cost accounting. Principal-agent models of price regulation are more "knowledge intensive" but less time consuming with regard to monitoring the costs of different services than cost accounting. Most importantly, a regulatory mechanism that takes into account the existence of asymmetric information between the regulator and the regulated firm induces cooperation between the contracting parties, while cost-based pricing inevitably brings about conflict between the regulator and the regulated firm.

The structure of the paper is as follows. The assumptions are outlined in section 2. The benchmark case of regulation with perfect information and cost-based pricing is presented in section 3. Section 4 is devoted to the description of a model of incentive regulation with two different efficiency types of the regulated firms. We solve the model in section 5. The results are compared to those of other pricing policies, and the main conclusions are drawn in section 6.

ASSUMPTIONS

Two firms (denoted by subscripts 1 and 2) are assumed to operate in a market of telecommunications services. They offer differentiated services to subscribing end users and in doing so, they compete in prices. For simplicity's sake, we shall work with a one period model where end users do not migrate between service providers, that is, we do not deal with the issue of switching costs. Subscribers initiate and receive intra-firm and inter-firm calls. Intra-firm calls are initiated and terminated in the same network, while inter-firm calls are terminated in the other network. There are three kinds of prices: *subscription* fees f_1 and f_2 that customers must pay in order to gain access to the network of firms 1 and 2, respectively; usage sensitive *intra-firm* calling prices p_1 and p_2; and usage sensitive *inter-firm* calling prices $\hat{p}_1$ and $\hat{p}_2$.[5] Inter-firm calling prices $\hat{p}_1$ and $\hat{p}_2$ include *termination* charges a_1 and a_2, respectively. These are paid by each firm to the other firm for using the other firm's network in order to terminate inter-firm calls. There are no separate transit charges since there are only two networks. All subscription and calling prices are unregulated. Termination charges are subject to regulation.

Subscribers have an identical valuation $0 < V(s_i) < s_i$ for belonging to network i of size s_i, where s_i is the number (mass) of subscribers who subscribe to network i. $\sum_i s_i$ is normalised to one and it also denotes the market share of firm i. Hence, for two firms, $s_1 + s_2 = 1$. For simplicity's sake, subscriber valuation is given by $V(s_i) = s_i V$.

A customer chooses between the two networks based on his valuation of network size and on the monetary utility, $u(p, \hat{p}, f)$, he can gain from using the services of each network. We assume that the customer's valuation of network size and his

[5] The total price a customer pays for subscribing to a network then for using its services is similar to the two-part tariff introduced in *Carter–Wright* [2003] and in *Valletti–Cambini* [2005].

monetary utility from using the network are additive in his total utility. The intuition behind this assumption is that a customer's decision of how many calls he will make depends only on the price of placing calls. Network size matters when a customer chooses a service provider because the size of the network will affect his utility through the intranet and off-net calling prices he expects to pay. Market shares will be functions of customers' total utility and may be derived from a simple, slightly modified price competition model of consumer choice *à la* Hotelling.

The representative customer's demand for intranet calls is given by $d(p)$, while the mass of a customer's inter-firm calls is $\hat{d}(\hat{p})$. A subscriber's consumer surplus from a mass of $d(p)$ intranet calls is denoted by $v(p)$. It is assumed that $v'(p) \equiv -d(p)$ Similarly, a subscriber's consumer surplus from a mass of $\hat{d}(\hat{p})$ inter-firm calls is denoted $\hat{v}(\hat{p})$, and $\hat{v}' \equiv -\hat{d}(\hat{p})$ by assumption. We assume that subscribers' choice between networks is also influenced by their non-price preferences for service providers. A customer's "distance" from his most preferred network will be denoted θ. We assume that θ is uniformly distributed on the unit interval between firm 1 and firm 2: $\theta \in [0,1]$ may be understood as the factor of substitution between network 1 and network 2. Thus, a subscriber's total utility – or total consumer surplus, denoted *CS* – from choosing network 1 or network 2 becomes

$$\begin{aligned} U_1 &= CS_1 = s_1 V + v_1(p_1) + \hat{v}_1(\hat{p}_1) - f_1 - \theta, \text{ or} \\ U_2 &= CS_2 = (1-s_1)V + v_2(p_2) + \hat{v}_2(\hat{p}_2) - f_2 - (1-\theta) \end{aligned} \tag{1}$$

The marginal subscriber between networks 1 and 2 will be the person for whom

$$s_1 V + v_1(p_1) + \hat{v}_1(\hat{p}_1) - f_1 - \theta^* = (1-s_1)V + v_2(p_2) + v_2(\hat{p}_2) - f_2 - (1-\theta^*)\,, \tag{2}$$

or

$$v_1(p_1) + \hat{v}_1(\hat{p}_1) - f_1 - s_1(1-V) = v_2(p_2) + \hat{v}_2(\hat{p}_2) - f_2 - (1-s_1)(1-V)\,, \tag{3}$$

where θ^* denotes the marginal customer's distance from firm 1. (With uniform distribution, $\theta^* = s_1$ will also hold.)

The indifference condition in (2) gives

$$\begin{aligned} s_1 &= \frac{v_1(p_1) - v_2(p_2) + \hat{v}_1(\hat{p}_1) - \hat{v}_2(\hat{p}_2) + f_2 - f_1}{2(1-V)} + \frac{1}{2}, \text{ and} \\ s_2 &= 1 - s_1 = \frac{v_2(p_2) - v_1(p_1) + \hat{v}_2(\hat{p}_2) - \hat{v}_1(\hat{p}_1) + f_1 - f_2}{2(1-V)} + \frac{1}{2} \end{aligned} \tag{4}$$

Service providers operate with constant, but different marginal costs in each segment of the service. This assumption implies that it is in a society's interest to purchase and use the services of the less efficient company, too, for it still increases social welfare. Consequently, uniform regulated prices would not be feasible in the short run because the less efficient company would incur losses and it would cease to provide a socially useful service. Fixed costs are disregarded because they do not affect the optimal level of service.

c_i^F denotes the marginal (unit) cost of connecting a new subscriber to network i.[6] Firm i (i = 1, 2) incurs a total marginal cost of $c_i = c_i^O + c_i^T$ by providing on-net (intranet) calls to its own subscribers – where c_i^O denotes the marginal cost of call origination, while c_i^T labels the marginal cost of call termination – but the firm incurs only the unit cost c_i^T by terminating the off-net calls for subscribers of the other firm, respectively.

Firm i's total profit from serving a mass of s_i customers with on-net calls and a mass of s_j customers with inter-firm calls can be written as

$$\pi_i = \underbrace{s_i\Big((p_i - c_i^O - c_i^T)d_i(p_i) + (\hat{p}_i - c_i^O - a_j)\hat{d}_i(\hat{p}_i) + f_i - c_i^F\Big)}_{\text{profit from internal subscribers}} + \underbrace{s_j(a_i - c_i^T)\hat{d}_j(\hat{p}_j)}_{\substack{\text{profit from call termination}\\ \text{for external customers}}}. \quad (5)$$

The total profit for the whole industry thus becomes

$$\Pi = \pi_1 + \pi_2 = s_1\Big((p_1 - c_1^O - c_1^T)d_1(p_1) + (\hat{p}_1 - c_1^O - c_2^T)\hat{d}_1(\hat{p}_1) + f_1 - c_1^F\Big) + \quad (6)$$
$$+ s_2\Big((p_2 - c_2^O - c_2^T)d_2(p_2) + (\hat{p}_2 - c_2^O - c_1^T)\hat{d}_2(\hat{p}_2) + f_2 - c_2^F\Big)$$

REGULATING INTERCONNECTION WITH PERFECT INFORMATION: COST-BASED PRICING FOR CALL TERMINATION

It is assumed to be in a society's interest to control the firms' monopoly power over interconnection in order to foster competition in end user services. In fact, such regulation exists in numerous countries, where the regulator sets an upper limit on the call termination charge a. We assume that the regulator wants to maximise social welfare (W) – measured as total consumer surplus plus total industry profit – in the regulated segment of the market, subject to some constraints. The regulator's valuation over gross economic surplus is concave with the usual properties: $W' > 0$, $W'' \leq 0$. Thus, the regulator's objective function can be written as

$$\max_{a_i, a_j} W = \max_{a_i, a_j}\{s_1 CS_1 + s_2 CS_2 + \Pi\}, \quad (7)$$

where Π is total industry profit as described in (6) and CS_i is the net consumer surplus enjoyed by a subscriber to network i.

When firms find their optimal calling prices (p_i, $\hat{p}_i$) and subscription fee (f_i) by maximising profits, they take into account the termination fee a_i that will be set by the regulator. The first order conditions of the companies' profit maximum in equation (5) are as follows:

[6] By this, we implicitly assume that service providers cannot extract all consumer surplus from new subscribers accessing their network.

$$\frac{\partial \pi_i}{\partial p_i} = \frac{\partial s_i}{\partial p_i}\tilde{\pi}_i + s_i d_i + s_i\left(p_i - c_i^O - c_i^T\right)\frac{\partial d_i}{\partial p_i} = 0, \tag{8}$$

$$\frac{\partial \pi_i}{\partial \hat{p}_i} = \frac{\partial s_i}{\partial \hat{p}_i}\tilde{\pi}_i + s_i \hat{d}_i + s_i\left(\hat{p}_i - c_i^O - a_j\right)\frac{\partial \hat{d}_i}{\partial \hat{p}_i} = 0, \tag{9}$$

$$\frac{\partial \pi_i}{\partial f_i} = \frac{\partial s_i}{\partial f_i}\tilde{\pi}_i + s_i = 0, \tag{10}$$

where $\tilde{\pi}_i = \left(p_i - c_i^O - c_i^T\right)d_i\left(\hat{p}_i - c_i^O - a_j\right)\hat{d}_i + f_i - c_i^F$ is firm *i*'s profit from one of its own customers. Using these conditions and the market share equation in (4), we have

$$p_i^* = c_i^O - c_i^T, \tag{11}$$

$$\hat{p}_i^* = c_i^O - a_j, \tag{12}$$

$$f_i^* = \frac{v_i(p_i) - v_j(p_j) + \hat{v}_i\left(\hat{p}_i(a_j)\right) - \hat{v}_j\left(\hat{p}_j(a_i)\right) + 2c_i^F + c_j^F}{3} + 2(1-V). \tag{13}$$

Substituting equation (13) into equation (4) yields the following market shares:

$$s_i^* = \frac{v_i(p_i) - v_j(p_j) + \hat{v}_i\left(\hat{p}_i(a_j)\right) - \hat{v}_j\left(\hat{p}_j(a_i)\right) + c_j^F + c_i^F}{6(1-V)} + \frac{1}{2}. \tag{14}$$

Since the regulator knows how firms solve their optimisation problem, he will use the firms' profit maximising prices to obtain optimal termination fees that will maximise total social welfare.[7] Substituting equations (11) and (12) into the regulator's objective function in (7) gives

$$\begin{aligned} W = {} & s_1\left[(a_2 - c_2^T)\hat{d}_1 - c_1^F + v_1(p_1) + \hat{v}_1\left(\hat{p}_1(a_2)\right)\right] + \\ & + s_2\left[(a_1 - c_1^T)\hat{d}_2 - c_2^F + v_2(p_2) + \hat{v}_2\left(\hat{p}_2(a_1)\right)\right] \end{aligned}, \tag{15}$$

where

$$(a_2 - c_2^T)\hat{d}_1 - c_1^F + v_1(p_1) + \hat{v}_1\left(\hat{p}_1(a_2)\right) = w_1,$$

and

$$(a_1 - c_1^T)\hat{d}_2 - c_2^F + v_2(p_2) + \hat{v}_2\left(\hat{p}_2(a_1)\right) = w_2 \tag{15a}$$

are economic surpluses at the firms' profit maximising prices per each subscriber in networks 1and 2, respectively.

Note that $w_1 = w_2$ must hold, otherwise the regulator would alter the termination charges in a way that would direct customers away from the network that yields lower economic surplus and toward the other network that offers a higher economic surplus per customer. For instance, if $w_1 > w_2$, then the regulator should reduce a_2,

[7] Our results would not change if the regulator established the cost-based termination fee at $a_i = c_i^T$ and firms maximised profits by knowing the regulated termination charges.

the termination fee he had set to firm 2 (and/or he should increase a_1) in order to direct customers away from network 2 and toward network 1. However, a reduction of a_2 will reduce economic surplus at network 1. The adjustment of termination fees continues until $w_1 = w_2$. From this result and from $s_1 + s_2 = 1$ it follows that a_i will maximise total social welfare in equation (15) if total net surplus per consumer, $w_i = (a_j - c_j^T)\hat{d}_i - c_i^F + v_i(p_i) + v_i\left(\hat{p}_i(a_j)\right)$ attains its maximum at a_i^*. The first order condition of social welfare maximum is

$$\frac{\partial w_i}{\partial a_i} = \hat{d}_j + (a_i - c_i^T)\frac{\partial \hat{d}_j}{\partial a_i} - \hat{d}_j = 0, \tag{16}$$

which yields

$$a_i^* = c_i^T. \tag{17}$$

Based on the above results we can formulate our first proposition.

PROPOSITION 1 • *Regulated cost-based pricing of call termination cannot be reconciled with competitive (unregulated) calling prices and subscription fees if companies operate with different marginal costs. Cost-based call termination prices will punish the efficient firm for its market share and its subscription fee will be smaller, consequently, its profit will be lower than if this firm remained less efficient. Thus, cost-based pricing of call termination will extend a "perverse" incentive to service providers that they should not offer call termination at efficient costs.*

Proof • *It is easy to see from equation* (14) *that describes the market shares of the firms, that firm i's market share increases in its own termination charge a_i, but its market share is a decreasing function of the other firm's termination charge a_j:*

$$\frac{\partial s_i^*}{\partial a_i} > 0 \ \ and \ \ \frac{\partial s_i^*}{\partial a_j} < 0. \tag{18}$$

Equation (17) *above shows the profit maximising call termination charges. Since $a_i < a_j$ because $c_i^T < c_j^T$ by assumption, it follows from equations* (15) and (18) *that $s_i^* < s_j^*$.*

In addition, it can be seen from equations (11), (12) *and* (13) *that give the profit maximising calling charges and subscription fees, that firms will earn positive profits only on subscription. It is obvious from equation* (13) *that firm i's profit maximising subscription fee f_i^* increases in its own termination charge a_i, but it decreases in the other firm's termination charge a_j.*

$$\frac{\partial f_i^*}{\partial a_i} > 0 \ \ and \ \ \frac{\partial f_i^*}{\partial a_j} < 0 \tag{19}$$

Consequently, if $a_i < a_j$ because $c_i^T < c_j^T$, then $f_i^* < f_j^*$. Since $\frac{\partial \pi_i}{\partial f_i} > 0$ in the profit equation (5), *it follows from $f_i^* < f_j^*$ that $\pi_i(a_i^*) < \pi_j(a_j^*)$.* ■

Proposition 1 shows that the efficient firm will lose, while the inefficient firm will gain in terms of market share and profits with cost-based pricing of call termination. Consequently, it is not in a firm's interest to invest in efficiency improvements. In case the regulator does not know the companies' true costs of providing interconnection, or cost information is "noisy" – that is, the regulator can know the firms' true costs only with some probability – the lack of the regulator's perfect information will exacerbate the effects of "perverse" incentives under cost-based pricing as we shall show in the next sections.

REGULATION IN THE PRESENCE OF MORAL HAZARD AND ADVERSE SELECTION

Firms do not have the incentive to improve their cost efficiency if regulators of telecommunications companies exercise cost-based pricing of network interconnection (call termination) as we have shown above. In addition, cost-based pricing requires that regulators possess perfect (noiseless) information of each firm's termination costs. However, information about the companies' effort level to increase their efficiency and about the true cost of service provision is the firms' valuable private information that the firms are not willing to disclose voluntarily. If the regulator incurs large expenses with collecting detailed firm level information and he cannot be certain that the information he acquired is reliable, the adverse effects of cost-based pricing are exacerbated. We shall show that the regulator can induce effort from the companies to improve their cost efficiency and attain true information revelation of the firms if he accepts the fact that his information on the firms' costs is limited. In other words, society will be better off if the regulator implements a regulatory regime that we call "incentive regulation" than what can be attained under cost-based pricing.

In the second part of the paper, we discuss the regulatory design for network interconnection (call termination), when the companies' effort to improve efficiency as well as the efficiency level of their termination service, c_i^T, constitute the firms' private information. We would obtain similar results if we assumed that the regulator obtains information about the firms' true termination costs with some probability, while it has positive probability that the firms' cost information is "noisy". That is, a company's true cost of call termination may be $\underline{c^T}$, but the regulator may also obtain cost information of $\overline{c^T} = \underline{c^T} + \omega$, where ω denotes the term of random error with a given probability distribution. Note that cost-based pricing assumes that firms' termination cost is either $\underline{c^T}$ or $\overline{c^T}$, each with probability 1. The regulator and the regulated firms play a static game in our model. Companies can improve their cost efficiency by additional effort costs.[8] The regulator has the right to offer a contract

[8] A thorough analysis of moral hazard followed by adverse selection can be found in *Laffont–Martimort* [2002] pp. 269–294.

menu to the firms.[9] This "regulation game" has a unique Nash equilibrium in each case presented below. The firms themselves play another pricing game within the regulation game that also has a unique equilibrium, as will be shown below.

We assume that two companies offer telecommunications services in the market. Each firm's efficiency level of call termination may have two different values: it may be "high," $\underline{c_i^T}$, or "low," $\overline{c_i^T}$, where the lower bar and the upper bar indicate low marginal cost (high efficiency) and high marginal cost (low efficiency), respectively. It follows from the definition of efficiency that $\underline{c_i^T} < \overline{c_i^T} \quad i = 1, 2$. The distance between firm *i*'s two efficiency levels is $\Delta c_i^T = \overline{c_i^T} - \underline{c_i^T}$.

The regulator does not know the true value of c_i^T, but he knows from past experience that the firms' efficiency may be high or low. The firms can improve their efficiency level with effort. Assuming that the firms' effort (*e*) can be "high" or "low," $e \in \{e^h, e^\ell\}$,[10] we denote the firm's effort costs $\psi(e)$ as $\psi(e^h) = \psi$, and $\psi(e^\ell) = 0$, respectively.

The regulator can observe the volume (and quality) of service by each firm, but he has only probabilistic knowledge about the companies' effort and efficiency level. These conditional probabilities are based on past experience. The conditional probability of high efficiency if the firm exerted a high effort is given as

$$\nu^h = P(i = h | e^h) = \frac{P(h \cap e^h)}{P(e^h)} .$$

The probability of low efficiency with high effort then becomes $1 - \nu^h$. Similarly, the conditional probability of high efficiency with low effort is ν^ℓ, hence the conditional probability of low efficiency with low effort becomes $1 - \nu^\ell$). We assume that the company is always capable of improving its efficiency level by exerting effort. However, the actual realisation of the efficiency level is a stochastic variable. When the company decides on effort – it may, for instance, invest in an efficiency enhancing technology – it cannot be certain that the effort will reap the expected efficiency level. We assume that the conditional probability of high efficiency is strictly increasing with effort: $\nu^h > \nu^\ell$. The difference between the conditional probabilities of high efficiency with respect to high and low effort is $\Delta \nu^h - \nu^\ell$. We also assume that high effort is always socially optimal, i.e.

$$\Delta \nu (\Delta W^h - W^\ell) \geq \psi, \tag{20}$$

where W^h and W^ℓ are the total economic surpluses from interconnection (inter-firm call termination) with the firms' high and low effort, respectively. Before elaborating the model of incentive regulation, we briefly present the regulatory contract with perfect regulatory information as a benchmark case.

[9] It could be the other way around: the firms may design and offer the contract menu and the regulator may accept or reject their offer.

[10] We could have assumed a continuous level of effort as we could have had a continuum of types, but it would have rendered the analysis technically more complex without adding to the important results. (See, for instance, *Laffont–Martimort* [2002] pp. 185–186.

We assume that the firms are *risk neutral*, but they are protected by *limited liability*. Under such assumptions, it is not in the firms' interest to reveal their true type and exert high effort. Nevertheless, firms may be induced to reveal their type and exert high effort by the creation of an "information rent," which is allocated by the regulator between the regulated firms. Such an information rent can be financed from a "service provision fund." Firms may pay to or receive payment from this fund. If firm i receives a transfer payment τ per customer in addition to the termination fee it obtains from the other service provider for terminating inter-firm calls, then the firm's net utility per customer becomes

$$u_i(\tau_i, a_i) = \tau_i - \left(c_i^T - a_i\right)\hat{d}_j. \tag{21}$$

The schedule of contracting between the firm and the regulator is as follows:

1. "Nature" sets the probability distributions of the efficiency types conditional on effort. The regulator and the firms learn these probability distributions.
2. The regulator offers a contract menu $\left\{\left(\underline{\tau_i}, \underline{a_i}, \underline{\hat{d}_j}\right), \left(\overline{\tau_i}, \overline{a_i}, \overline{\hat{d}_j}\right)\right\}$ for each combination of effort level and efficiency type for each firm i (i = 1, 2). The lower and upper bar variables stand for efficient outcomes and inefficient outcomes, respectively.
3. The firm decides on its effort level without revealing the decision, which thus remains private information.
4. Having selected an effort level, its efficiency type is set as a stochastic function of the firm's effort. (Note that even the firm is unable to know its efficiency type for certain.)
5. The firm delivers the interconnection (call termination) service, customers pay the termination charge as a fraction of the inter-firm calling price, and firms settle the net balance of mutual interconnection charges among themselves according to the rule that has been specified by the regulator.

Additional contracting conditions are set for a firm by its participation constraint, limited liability constraints, and the adverse selection and moral hazard incentive constraints. We assume that the reservation utility of the firms, $\underline{u_i^0}\left(\underline{\tau_i}, \underline{a_i}\right), \overline{u_i^0}(\overline{\tau_i}, \overline{a_i})$, equals zero for all efficiency types. The constraints are introduced below.

Participation constraint • Since the regulator intends to induce high effort by the firm by assumption, the participation constraint is associated only with high effort. It is

$$\nu^h \underline{u_i} + \left(1 - \nu^h\right)\overline{u_i} \geq 0\,.[11] \tag{22}$$

[11] Note that u does not have a superscript index. We assume that the regulator prefers high to low effort; consequently, participation must be ensured only for firms exerting high effort. When the superscript index is omitted, the variable or probability always refers to high effort.

Limited liability constraints • We assume that the firm does not possess disposable assets to finance any loss. This is not as strong an assumption as it appears to be. We could allow a loss, say *L*, which would affect our equations with a constant term, but it would not have any substantial effect on the model. The limited liability constraint of the firm with high efficiency is

$$\underline{u_i} \geq 0, \tag{23a}$$

and the limited liability constraint of the firm with low efficiency is

$$\overline{u_i} \geq 0. \tag{23b}$$

Adverse selection incentive compatibility constraints • These constraints ensure that the firm does not mimic another type of efficiency, which is different from its true type, because its utility cannot be higher with lying than with revealing the truth (its true efficiency level). (One may call these "Do not lie!" constraints.) The incentive constraints of the highly efficient firm are

$$\underline{u_i} \geq \overline{u_i} + \Delta c_i^T \overline{\hat{d}_j}, \tag{24a}$$

while the incentive constraints for the firm with low efficiency become

$$\overline{u_i} \geq \underline{u_i} - \Delta c_i^T \underline{\hat{d}_j}, \tag{24b}$$

where $\Delta c_i^T = \overline{c_i^T} - \underline{c_i^T},\ i = 1,2$ denotes the difference between high and low marginal costs of call termination.

Moral hazard incentive compatibility constraint • The moral hazard incentive constraint induces the firm to exert high effort provided that high effort is desirable for society. (One may call these "Do not cheat!" constraints.) In other words, the moral hazard incentive constraint ensures that the expected utility of the firm cannot be lower with high than with low effort. The incentive constraint is

$$\nu^h \underline{u_i} + (1 - \nu^h)\overline{u_i} - \psi \geq \nu^\ell \underline{u_i} + (1 - \nu^\ell)\overline{u_i} \Rightarrow \Delta\nu \left(\underline{u_i} - \overline{u_i}\right) \geq \psi. \tag{25}$$

The regulator's objective function • Since the regulator does not possess perfect information about the firms, he must give up some of his benefits in order to induce effort and true revelation. The regulator's lost benefit becomes the firm's information rent. The information rent has two parts. The first part is the firm's limited liability rent, for the firms must be able to charge a higher interconnection fee than what the regulator would otherwise accept because of the firms' limited liability constraint. The second part is the "adverse selection" rent, which acts to induce true revelation of the firms' efficiency type. The regulator's objective function becomes

$$E(W) = \nu^h \left(\underline{W} - \underline{u_1} - \underline{u_2}\right) + (1 - \nu^h)(\overline{W} - \overline{u_1} - \overline{u_2}) \tag{26}$$

with constraints (22)–(25), where *W* is the social welfare function as given by equation (15).

The relevant constraints •The analysis of constraints reveals that we need to deal only with the limited liability constraints of the less efficient firm (23*b*), the adverse selection constraints of the efficient firm (24*a*), the moral hazard constraint (25), and the following *monotonicity constraint* (derived from the adverse selection constraints):

$$\underline{\hat{d}_j} \geq \overline{\hat{d}_j}\,. \tag{27}$$

The wider the gap between the regulated interconnection fee a_i, $i = 1, 2$, and its first best optimum, the larger the lost economic surplus will be. Consequently, the information rent of the inefficient type must be kept at minimum by the regulator. It then follows from the limited liability constraint of the inefficient firm that

$$\overline{u_i} = 0 \text{ must hold.} \tag{28}$$

The information rent of the efficient type will be affected by the relative strength of the effect of adverse selection and moral hazard. Different constraints may be binding depending on the probability distribution of efficiency types and effort level, and on the magnitude of the effort cost. The regulator faces a trade-off between the information rent, resulting from the adverse selection and limited liability constraints, and the allocative efficiency of the firm with different efficiency types. In certain cases, it makes sense for the regulator to distort the output level of the firm downwards (i.e., away from the first best level of output) in order to save a portion compensation for the information rent of the more efficient type. We show that the downward distortion of output becomes smaller and smaller as the problem of moral hazard is exacerbated.

OPTIMAL CONTRACT MENUS WITH DIFFERENT BINDING CONSTRAINTS

We need to discuss three different cases that are distinguished by the relative magnitude of the information rent and the effort cost. Notably, it will depend on the relative magnitude of the information rent and effort cost which constraints of the different efficiency types will be binding. We only present the first case in detail, when the information rent exceeds the effort cost. Then we outline only the final results of the other two cases, for the technical analysis goes along the same lines in all cases.

Case (a) • It is assumed that the information rent that a firm can extract with high efficiency is not less than the cost of inducing effort, that is, comparing (24*a*) and (25) we obtain

$$\Delta c_i^T \overline{\hat{d}_j^{SB}} \geq \frac{\psi}{\Delta \nu}, \tag{29}$$

where the second best outcome of interconnection services is denoted by $\hat{d}_i^{SB}$.

The following result is obtained from (29):
If the cost of inducing effort of the efficient firm is smaller than the firm's information rent, then the adverse selection incentive constraint of the efficient firm (24*a*) *is binding:*

$$\underline{u_i} = \Delta c_i^T \overline{\hat{d}_j}. \tag{30}$$

The first order conditions of the regulator's welfare maximisation problem yield optimal charges of call termination with different efficiency types. Substituting (28) and (30) into the regulator's objective function in (26) we get

$$E(W) = \nu^h \left(\underline{W} - \Delta c_1^T \overline{\hat{d}_2} - \Delta c_2^T \overline{\hat{d}_1} \right) + \left(1 - \nu^h\right)\left(\overline{W}\right). \tag{31}$$

The first order conditions of call termination charges yield

$$\begin{aligned} \frac{\partial E(W)}{\partial \underline{a_i}} &= \nu^h \left(\underline{a_i} - \underline{c_i^T} \right) \frac{\partial \underline{\hat{d}_j}}{\partial \underline{a_i}} = 0 \\ \frac{\partial E(W)}{\partial \overline{a_i}} &= \nu^h \Delta c_i^T \frac{\partial \overline{\hat{d}_j}}{\partial \overline{a_i}} + \left(1 - \nu^h\right) \left(\overline{a_i} - \overline{c_i^T} \right) \frac{\partial \overline{\hat{d}_j}}{\partial \overline{a_i}} = 0, \end{aligned} \tag{32}$$

which sets the following termination fees:

$$\underline{a_i} = \underline{c_i^T} \text{ and } \overline{a_i} = \overline{c_i^T} + \frac{\nu^h}{1 - \nu^h} \Delta c_i^T. \tag{33}$$

Our conclusion is that the different information rents that must be paid to high and to low types, respectively, differ to the extent that is sufficiently large to induce the high effort of all firms. In such cases, the optimal contract menu looks the same as the contract that the regulator would offer in case of pure adverse selection.

Case (b) • It is assumed that the cost of inducing effort is higher than the information rent of the efficient type, but is lower than this information rent would be if the output of the less efficient firm were not reduced below its first best level, i.e.,

$$\Delta c_i^T \overline{\hat{d}_j^{SB}} < \frac{\psi}{\Delta \nu} \leq \Delta c_i^T \overline{\hat{d}_j^{FB}}, \tag{34}$$

where $\overline{\hat{d}_j^{FB}}$ is the first best level of optimum output.

The adverse selection incentive constraint (24*a*) *and the moral hazard incentive constraint* (25) *will equally bind in case of the high efficient firm:*
$\underline{u_i} = \Delta c_i^T \overline{\hat{d}_j}$ *as in* (30) *and*

$$\underline{u_i} = \frac{\psi}{\Delta \nu}, \tag{35}$$

so that equation (35) can be re-written as

$$\Delta \nu \Delta c_i^T \overline{\hat{d}_j} - \psi = 0, \tag{36}$$

and the regulator's objective function becomes

$$E(W) = \nu^h \left(\underline{W} - \Delta c_1^T \overline{\hat{d}_2} - \Delta c_2^T \overline{\hat{d}_1} \right) + \left(1 - \nu^h\right)\left(\overline{W}\right) + \lambda \left(\Delta \nu \Delta c_1^T \overline{\hat{d}_2} - \Delta \nu \Delta c_2^T \overline{\hat{d}_1} - \psi \right). \tag{37}$$

The first order conditions yield

$$\underline{a_i} = \underline{c_i^T} \text{ and } \overline{a_i} = \overline{c_i^T} + \left(\frac{\nu^h}{1-\nu^h}\lambda\Delta\nu\right)\Delta c_j^T \tag{38}$$

where $\lambda > 0$ is the Lagrange multiplier of equation (36).

The results indicate that exacerbated moral hazard results in a larger information rent of the efficient firm in Case (b) than in Case (a). The regulator cannot substantially reduce the information rent by deteriorating allocative efficiency, i.e., by reducing the level of service of the low efficiency type. Consequently, it is sensible to cut back the output of the less efficient firm to a lesser extent. As the first order conditions show, the efficient firm will produce at its first best optimum level. The regulator will distort the output level of the inefficient company downward as in Case (a), but it follows from (38) that this distortion will now be smaller. Consequently, $\overline{a_i}$ is now smaller than in Case (a), and the information rent of the efficient firm under Case (b) will exceed the information rent of the same firm under Case (a). The regulator must pay higher information rent for the gain in allocative efficiency.

Case (c) • It is assumed that the cost of inducing effort is larger than the information rent accrued by the efficient type:

$$\Delta c_i^T \overline{\hat{d}_j^{FB}} < \frac{\psi}{\Delta\nu}. \tag{39}$$

The moral hazard constraint (25) *and the limited liability constraint* (23b) *are binding;*

$$\overline{u_i} = 0 \text{ and } \underline{u_i} = \frac{\psi}{\Delta\nu}.$$

The problem of moral hazard is so pervasive – the cost of inducing effort is so high – that it renders the reduction of the information rent of the more efficient type unfeasible by distorting the output level of the less efficient type downwards. Consequently, each type will produce at its first best level. The regulator's objective function in (26) becomes

$$E(W) = \nu^h\left(\underline{W} - \frac{\psi}{\Delta\nu}\right) + \left(1-\nu^h\right)\underline{W}\,. \tag{40}$$

Solving the first order conditions obtains

$$\underline{a_i} = \underline{c_i^T} \text{ and } \overline{a_i} = \overline{c_i^T}\,. \tag{41}$$

Substituting the results of the three cases into the firm's profit functions in (6), our second proposition is formulated.

PROPOSITION 2 • *Cost-based pricing rewards low efficiency in call termination services in terms of profits, while incentive regulation provides the proper incentives to firms: the companies' higher effort to increase efficiency reaps larger profits.*

Proof • *Efficient types can charge lower, while inefficient types can charge higher termination fees with incentive regulation. But the adverse effect of termination charges will be compensated for the efficient type through the information rent it obtains. The source of this information rent is a direct transfer of revenues from the inefficient to the efficient firm.* ■

As the analysis demonstrates, incentive regulation does not come without a cost. The cost of inducing effort is inversely related to the allocative inefficiency of the firms with different efficiency types in mixed models if moral hazard precedes adverse selection.

DISCUSSION AND CONCLUSIONS

Our most important conclusion is that incentive regulation does not have a perverse effect on the regulated firms' profit and efficiency, while cost-based regulation does have such an effect. Cost-based pricing of call termination ultimately rewards the less efficient types of regulated firms. In contrast, when the regulator offers the regulated firm an incentive-based contract menu, the efficient firm will earn higher profits, while the less efficient firm's profit will be zero. These results suggest that incentive regulation puts an additional burden on the regulator, for he must reallocate a fraction of the termination charge between the less efficient and the most efficient firm. However, this difficulty may not materialise, since firms normally pay each other only the net balance of interconnection charges.

The regulator needs to compare and contrast three possible cases if adverse selection and moral hazard are both present. Regulated firms of both efficiency types provide their service at the first best, Pareto-efficient level in Case (c). The efficient type produces the first best level of output in all other cases as well, but the output level of the less efficient type is downward biased in Cases (a) and (b). In these cases, the regulator is forced to distort allocative efficiency in order to induce information revelation and high effort from any type of regulated firm.

The cost of inducing effort is larger relative to the information rent in Case (b) than in Case (a), and the regulator distorts the output level of the less efficient type downward to a lesser extent in Case (b) than in Case (a). As the cost of inducing effort keeps increasing, as in Cases (b) and (c), the downward distortion of the output level of the less efficient type becomes smaller and smaller. The service levels of firms of different efficiency types come closer and closer to their Pareto-efficient level as the benefit (what the firm can acquire in return for revealing private information) becomes smaller and smaller relative to the effort cost. Consequently, it is less and less necessary and sensible for the regulator to offer an information rent to the firm for information revelation. As the distortion of allocative efficiency becomes smaller, the interconnection charge is also reduced.

Firms are induced to remain inefficient if termination charges are cost-based. In incentive regulation, the regulator transfers a certain amount of information rent from total economic surplus in order to induce effort for efficiency improvement. As a result of true cost revelation, allocative efficiency among firms improves and consumer surplus increases.

REFERENCES

Armstrong, M.–Doyle, C.–Vickers, J. [1996]: The Access Pricing Problem: A Synthesis. Journal of Industrial Economics, 44, pp. 131–150.

Armstrong, M. [1998]: Network Interconnection in Telecommunications. Economic Journal, 108, pp. 545–564.

Armstrong, M. [2002]: The Theory of Access Pricing and Interconnection. In: Cave, M. – Majudmar, S.–Vogelsang, I. (eds.): Handbook of Telecommunications Economics, Vol. 1, North Holland, Amsterdam, pp. 295–384.

Armstrong, M.–Sappington, D. E. M. [2007]: Recent Developments in the Theory of Regulation. In: *Armstrong, M.–Porter, R.* (eds): Handbook of Industrial Organization, Vol. III. North Holland, Amsterdam, pp. 3–137.

Baron, D.–Myerson, R. [1982]: Regulating a Monopolist with Unknown Costs. Econometrica, 50, pp. 911–930.

Blackmon, G. [1994]: Incentive Regulation and the Regulation of Incentives. Kluwer, Boston.

Carter, M.–Wright, J. [1999]: Interconnection in Network Industries. Review of Industrial Organization, 14, pp. 1–25.

Carter, M.–Wright, J. [2003]: Asymmetric Network Interconnection. Review of Industrial Organization, 22, pp. 27–46.

De Bilj, P.–Peitz, M. [2002]: Regulation and Entry into Telecommunications Markets. Cambridge University Press, Cambridge.

Fudenberg, D.–Tirole, J. [2000]: Customer Poaching and Brand Switching. The RAND Journal of Economics, Vol. 31. No. 4, pp. 34–657.

Hansen, B. E. [2005]: Network Competition when Costs are Heterogeneous. Manuscript. BI Norwegien School of Management, Oslo.

Laffont, J. J. [1994]: The New Economics of Regulation Ten Years After. Econometrica, 62, pp. 507–537.

Laffont, J. J.–Martimort, D. [2002]: The Theory of Incentives – The Principal-agent Model. Princeton University Press, Princeton, NJ.

Laffont, J. J.–Rey, P.–Tirole, J. [1998*a*]: Network Competition: I. Overview and Nondiscriminatory Pricing. The RAND Journal of Economics, 29, No. 1, pp. 1–37.

Laffont, J. J.–Rey, P.–Tirole, J. [1998*b*]: Network Competition: II. Price Discrimination. The RAND Journal of Economics, 29, No. 1, pp. 38–56.

Laffont, J. J.–Tirole, J. [1993]: A Theory of Incentives in Procurement and Regulation. The MIT Press, Cambridge, MA.

Laffont, J. J.–Tirole, J. [2000]: Competition in Telecommunications. The MIT Press, Cambridge, MA.

Madarász, K. [2007]: A Theory of Information Projection with Applications to Employment Relations and Social Attitudes. Manuscript, University of California, Berkeley.

Peitz, M. [2005]: Asymmetric Access Price Regulation in Telecommunications Markets. European Economic Review, 49, pp. 341–358.

Sappington, D. E. M. [1980]: Strategic Firm Behavior under a Dynamic Regulatory Adjustment Process. Bell Journal of Economics, 11, No. 1, pp. 360–372.

Stefos, Th. [1990]: Regulating without Cost Information. International Economic Review, 31, No. 4, pp. 1021–1025.

Valletti, T. M.–Cambini, C. [2005]: Investments and Network Competition. The RAND Journal of Economics, 36, No. 2, pp. 446–467.

Villas-Boas, J. M. [1999]: Dynamic Competition with Customer Recognition. The RAND Journal of Economics, 30, No. 4, pp. 604–631.

• *László Paizs* •

INCENTIVE PROBLEMS IN THE HUNGARIAN ENERGY-BALANCING MECHANISM

This paper examines the functioning of the balancing market of the electricity sector in Hungary. Balancing energy is an ancillary service, which is used by the transmission system operator (TSO) to guarantee the continuous supply of electricity. The TSO resolves unforeseen imbalances by calling on power plants in real time to increase or decrease their production (called upward and downward regulation). In order to comply with the balancing mechanism and settlement process, market participants organize into so-called balancing groups led by the balancing responsible party (usually a trader or supplier). Based on the forecast of the balancing group's day-ahead production and consumption, the balancing responsible party (BRP) prepares the schedule of the balancing group, forwards it to the TSO, and then settles the imbalances with the TSO resulting from any deviation from the announced schedule. In our study we examine the question of how current balancing energy and imbalance prices affect the incentives on suppliers to keep their portfolio balanced. Taking only the price difference between negative and positive imbalance prices into consideration, we can say that the incentive on suppliers to avoid imbalances is very strong in the Hungarian market. However, we also show that because of the asymmetrical penalties for being long versus short, suppliers are inclined to under-contract energy on the wholesale market. Finally, our analyses also reveal that the current structure of the purchase and settlement price of balancing energy motivates the public utility wholesaler (the BRP for the public utility balancing group) to nominate more than its expected load.

INTRODUCTION

Since the liberalization of the Hungarian electricity market in 2003, it has been functioning as a dual market: there is an *open market* for authorized consumers, and a *public utility* market with prices set by the authorities. The market was liberalized gradually, in several phases. Beginning in January 2003 all consumers with electricity consumption higher than 6.5 GWh, and as of July 2004 all industrial consumers – together representing 70 percent of all consumers – have had the opportunity to choose their supplier freely.

The first phase of liberalization was characterized by great consumer activity. By the end of 2004 the consumption of consumers opting for an open market reached

20 percent of the total domestic consumption. The import competition resulting from the liberalization of external trade played the main role in the fast-paced expansion of this competitive market segment. The import share of open market consumers in the electricity supply was over 60 percent of the total in 2004. After the opportunities brought by import-based growth ran out, the competitive market slowed down significantly. While the share of open market consumption reached 25 percent by the end of 2005 and it was as high as 30 percent by the end of 2006, there was a sharp drop to 20 percent at the beginning of 2007, when open market consumers returned *en masse* to public utilities.

In spite of the strong activity on the demand side and the savings realized by large consumers – which were very significant at the beginning – the competition in Hungary is characterized by severely distorted market conditions. The long-term power purchase agreements between MVM (Hungarian Electricity Ltd.) and power generators present the biggest problem. Based on these agreements 65 percent of domestic power generation and 80 percent of domestic electricity sales are controlled by MVM. This means that although the ownership structure of the Hungarian power generation market is fragmented, the market itself is overwhelmingly dominated by MVM. We have to add that due to the restrictions on international competition, domestic conditions play a more important role in the development of power market competition than in other markets.

This paper examines the functioning of the balancing market of the electricity sector in Hungary. It is necessary to have a balancing energy market in order to control and financially settle unforeseen imbalances in the electricity market. The company responsible for the reliability of the electric power system called Hungarian Transmission System Operator, or MAVIR for short, is in charge of the purchase and settlement of balancing energy. MAVIR prepares the system schedule from the schedules that have been submitted by BRPs. This schedule contains the planned generation-consumption balance of the country for every 15 minute. If there is any deficiency in the system balance – for instance, because the actual generation by plants is less than the agreed amount – the TSO can restore the balance by drawing on reserves, and then charge the costs of balancing to those participants who failed to meet their schedule.

THE PROCUREMENT OF BALANCING ENERGY

In 2006, the Hungarian TSO purchased around 1300 MW reserve capacity from domestic power plants, which is approximately 20 percent of the annual peak consumption (approximately 6300 MW). The reserves have varying levels of response time, i.e. how fast they can be made available. Deficits are typically balanced out by calling upon the reserve that can be made available at the shortest notice, and if the deficit is very big and/or lasts for a longer period, cheaper and larger reserves with slower response times gradually replace the fast ones. The total cost of reserves utilized by MAVIR is approximately HUF 25bn, which increases the cost of every KWh supplied by about 0.7 Hungarian forints.

Just like the energy market, the procurement of balancing power can also be organised as a competitive market. One of the most frequently used methods is when the TSO holds auctions at regular intervals (once a year, daily, or every hour) to procure reserves of the required quantity and composition for upward and downward regulation.

In the current model MAVIR procures control reserves at daily auctions. Within each reserve type participants submit two-part bids – containing capacity fee and energy price – for both upward and downward regulation. Settlement is based on the offer price: winning bidders get the capacity fee they bid for, and if activated they receive the energy fee contained in the bid submitted.

The main obstacle to competition on the balancing energy market (among others) is the long-term purchase agreement system. The dominance of MVM is made even more obvious on this market due to import competition being excluded (very restricted). MAVIR covers its regulation reserve needs almost exclusively from MVM: in 2005 MVM's share was as high as 95 percent. The regulatory authority currently addresses this problem by requiring MVM to submit bids to the balancing energy market that do not exceed the settlement prices of its contracts with generators.

THE SETTLEMENT OF BALANCING POWER

Market participants organize into so-called balancing groups, and the TSO settles the real-time imbalances financially with the balancing responsible parties (BRPs). The BRPs calculate what the members (power plants, traders, consumers) of their respective balancing group inject into the system as well as what they take out, and forward the balanced schedule to the system operator one day in advance. The settlement of balancing energy is based on the difference between the scheduled amount and the actual amount generated (loaded into the system) and consumed (withdrawn). The cost or revenue of balancing for the balancing group splits between the members of the balancing group.

The settlement price of balancing energy

The settlement of balancing energy, with those participants who require settlement, can be done in two ways. In the so called single-price system, the same price – though with the opposite sign – is applied for both the negative and positive imbalances. In the so called double price accounting system negative and positive imbalance prices are different; the former is higher than the latter. To settle balances in the United States they tend to use the single price system, while in European countries they prefer the double price system (*Glachant–Saguan* [2007]).

Double imbalance prices provide a strong incentive for market participants to keep their own position in balance. Since the energy price charged for negative imbalances is typically higher than the day-ahead price, and the energy price paid for positive imbalances is typically lower, the system penalizes both taking more or taking less. The measure of "penalty" for both negative and positive imbalances (B_{negative} and B_{positive}) can be expressed as the following:

$$B_{\text{negative}} = NIP - P$$
$$B_{\text{positive}} = P - PIP,$$

where P is the market price (for example the relevant power exchange price), NIP (negative imbalance price) is the settlement price of negative imbalances, and PIP (positive imbalance price) is the settlement price of positive imbalances. The cost of being short can therefore be measured as the difference in costs between buying energy on the balancing market and the wholesale market. The cost of any deviation from the submitted schedule can be reduced by keeping to the schedule more accurately, i.e. reducing the standard deviation of imbalances.

An additional feature of the double-pricing system is that often the price charged for imbalances does not only depend on the market player's own balance (be it positive or negative), but also on the direction of its balance relative to the overall status of the system (same direction/opposite direction). In countries that have their own electricity exchanges, the settlement of imbalances that are in the same direction as that of the system are based on the cost of balancing services, while the settlement of imbalances that are in the opposite direction are based on day-ahead power exchange prices. By using day-ahead prices the exposure of market parties to balancing risks is lower, as normally the price of the day-ahead market is lower (higher) than price of negative imbalance (positive imbalance) of the balancing market. Using less "penalizing" settlement prices in the case of imbalances in the opposite directions is justifiable, as participants with opposite direction balances actually decrease the real-time costs of the balancing of the whole system.

In the absence of an organized market in Hungary we have no reference price, which we could refer to when settling imbalances that are in the opposite direction

TABLE 1 • **The double imbalance price scheme In Hungary**

		MAVIR (system imbalance)	
		deficit (net upward regulation)	surplus (net downward regulation)
Balancing group	negative (short)	negative imbalance price	negative imbalance price, if there was upward regulation, otherwise 0
	positive (long)	positive imbalance price, if there was downward regulation, otherwise 0	positive imbalance price

as the system balance. Currently the way the settlement price is set does not depend on whether the direction of individual deviation is the same as that of the system deviation. A player with a negative balance will be charged the price of negative imbalance (NIP) even if the system has a net surplus, provided that in the settlement period the system operator performed both downward and upward regulation. If there was no regulation in the opposite direction to the net system balance, then the settlement price will be equal to zero.

The rules of setting imbalance prices recently changed in Hungary. Earlier the unit price of balancing energy was fixed, the settlement prices of positive and negative deviations were determined by multiplying the public utility wholesale prices with fixed factors.[1] In order for balancing energy to correspond more to the costs of regulation performed by the TSO, as of July 1, 2006 the fixed price system was replaced by cost-based pricing. Under the new scheme the settlement price of imbalances is based on the average procurement cost of balancing services.

TABLE 2 • **The settlement price of imbalances in Hungary in 2006** (HUF/KWh)

Period	The settlement price of negative imbalances	The settlement price of positive imbalances
January 2006–June 2006		
Peak period*	22.65	0.88
Off-peak period*	11.23	0.00
Average**	13.47	0.26
July 2006		
Average***	15.28	0.00

* Unit price imposed by the authorities.
** Fifteen-minute settlement price average, assuming that in 90 percent of the settlement periods there was some downward or upward regulation.
*** Fifteen-minute settlement price average.

An assessment of the Hungarian imbalance price system

The Hungarian balancing mechanism will be assessed against the requirements considered necessary to achieve healthy operation of the balancing energy market. We examine how much the current system of imbalance prices encourage suppliers to avoid imbalances, how big is the risk it poses to market participants, whether it helps minimise overall balancing costs, and how much room it leaves for arbitrage or other undesirable gaming.

[1] The peak and off-peak period unit prices of negative imbalances were respectively 1.3 times the peak and off-peak period public utility wholesale prices set out in the regulation. The price of positive imbalances in peak and off-peak periods were equal to the pro-rata average of peak and off-peak period public utility wholesale prices, while in peak periods they were equal to 0 (*Network Code* [2006]).

As the costs of system-level balancing are usually considerable (see below in the study), market participants need to be encouraged to cover their generation and consumption as accurately as possible. For example, the more accurate forecast of generation and load and better incentive mechanisms within the balancing group can help to meet individual schedules more accurately.

However, it is important to see that the measures taken by BRPs to avoid imbalances are costly, even if these costs are borne not at the level of centralised system control, but by the market participants themselves. Therefore, the minimisation of system balancing costs cannot be considered the ultimate goal. If this happened, it would lead to a very high degree of individual balancing, which in turn would increase the cost of individual balancing too much. Theoretically, the pricing system of the balancing energy market can be considered optimal if it provides incentives for individual balancing of market participants to such an extent that its marginal cost is exactly the same as the marginal cost of system level balancing.

IMBALANCE PRICE SPREAD

The economic cost of being short or long depends very much on the difference between the negative and positive imbalance prices. To demonstrate the potential impact of the spread between negative and positive imbalance prices, let us consider a BRP owning only load (i.e. a supplier). Assuming that the supplier's real-time deviation from its schedule is a random variable following normal distribution, and that the opportunity cost of buying or selling balancing energy are identical ($NIP - P = P - PIP = B$). The average value of the imbalance costs of the supplier can be expressed as the following:

$$2B\frac{\sigma}{\sqrt{2\pi}}, \tag{1}$$

where σ is the standard deviation of imbalances, and *2B* is the difference between the negative and positive imbalance prices (see *Appendix 1* for a detailed description of the calculation). So if a supplier can predict the consumption of its clients with a five percent margin of error, and the difference between imbalance prices is EUR 50/MWh, then the average cost of balancing for the supplier will be approximately EUR 1/MWh. If this price difference increases, then the projected balancing cost will increase, and so will the incentive to keep their portfolio balanced.

The difference between imbalance prices in Hungary has been EUR 50–60/MWh in the past 18 months, which is very high, compared to price differences abroad (*see Table 3* and *Figure 1*). The price system of Hungarian balancing energy can therefore be considered very penalizing. We have to add that expensive balancing energy may act to stunt the development of competition especially in the early phases of liberalization.

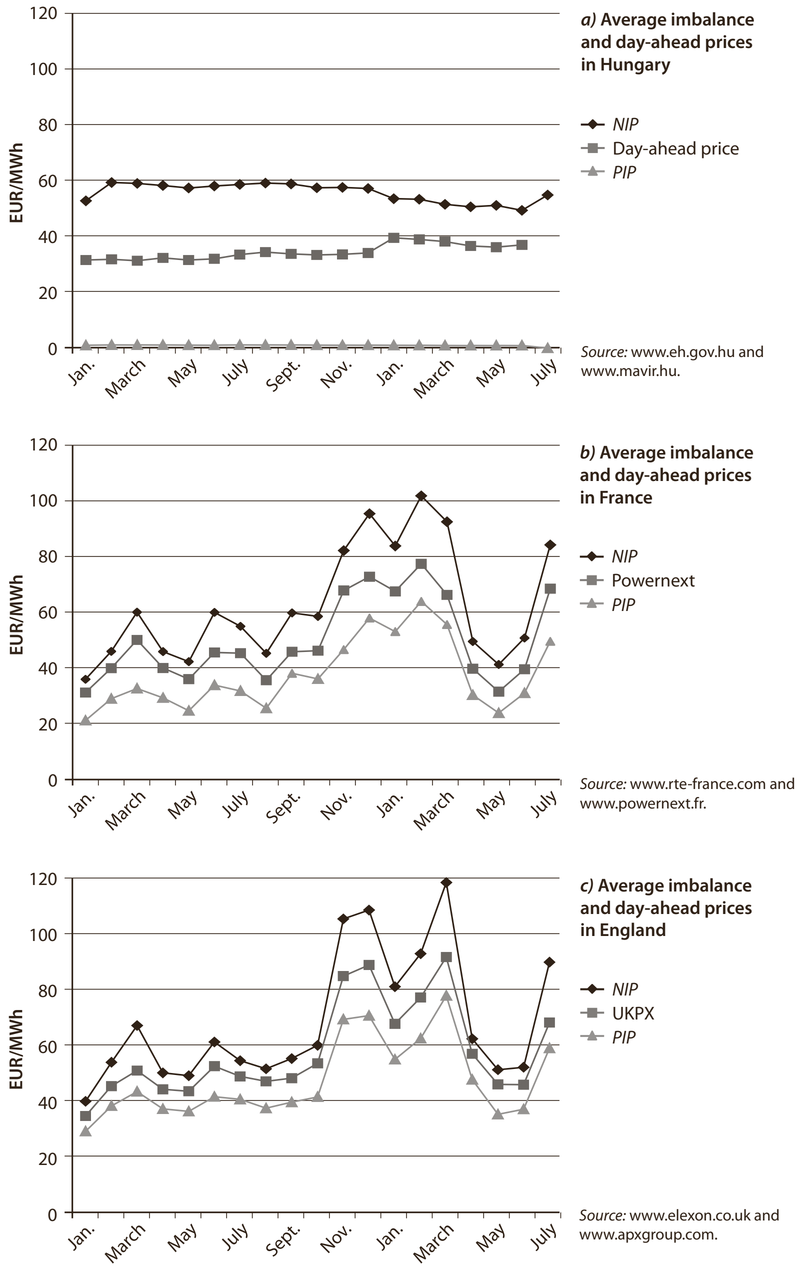


FIGURE 1 • **Average imbalance and day-ahead prices on the Hungarian, French and English electricity markets** (January 2005 – July 2006)

TABLE 3 • **Imbalance prices in European countries, 2005 (EUR/MWh)**

Country	The average price of negative imbalances	The average price of positive imbalances	Price difference
Austria	51	24	27
Belgium	56	12	44
Czech Republic	21	0	21
Denmark	36	27	9
United Kingdom	55	39	16
Finland	32	27	5
France	50	45	5
Greece	44	44	0
Netherlands	69	28	41
Ireland	69	60	9
Poland	37	24	13
Hungary	40	0	40
Germany	70	2	68
Norway	29	29	0
Italy	102	23	79
Portugal	58	23	35
Sweden	32	28	4

Source: EC [2005]: Report on progress in creating the internal gas and electricity market, COM(2005) 568 final, Commission of the European Communities p. 67.

Suppliers that have recently entered the market naturally have a smaller clientele than their more established competitors. Due to the fact that it is much harder to plan the schedule, service to smaller consumer portfolios can only be provided if we make greater use of balancing energy. Balancing is consequently a greater burden to smaller suppliers than to big ones. Hence, the high exposure to balancing risk may act to prevent new players from entering the market.

Asymmetrical penalties

Figure 1 illustrates yet another peculiar aspect of balancing energy pricing in Hungary. While in England and France the position of negative and positive imbalance prices with respect to the wholesale price can be said to be symmetrical, in Hungary the position of imbalance prices compared to wholesale price are asymmetric. The tendency that the difference between the wholesale price and the positive imbalance price significantly exceed the difference between negative imbalance price and wholesale price can be observed since January 2006.

In the price structure of Hungary, the cost of settling a long position is significantly higher than that of a short positions ($P - PIP > NIP - P$). It is easy to see how this encourages suppliers to under contract energy on the wholesale market and thus avoiding long positions.

Let us look at the behaviour of a supplier when $P - PIP > NIP - P$. When making decision on the wholesale purchase of energy, the supplier is faced with the following problem. If he cuts the wholesale purchase of energy by one unit, he gains the saving of the power exchange price, lessens his exposure to the *PIP* (i.e. the likelihood that he ultimately goes long and needs to sell the surplus at PIP), while raising its exposure to the NIP (i.e. the risk that it subsequently goes short and needs to make up the shortfall at the PIP). Defining p^s = the probability that the position of the supplier is ultimately short, this can be expressed as

$$\text{Gain} = P - p^s NIP - (1 - p^s)PIP. \quad (2$$

To minimize the cost of his wholesale energy purchase *plus* the expected costs of balancing, the supplier will reduce his purchasing of energy down to the point where there is no further gain from cutting down on trading, i.e.

$$P - p^s NIP - (1 - p^s)PIP = 0, \quad (3)$$

that is

$$p^s = \frac{1 - (PIP/P)}{(NIP/P) - (PIP/P)}. \quad (4)$$

If we put the averages of wholesale and balancing energy prices for the first six months of 2006 into the equation (P = 9.85 HUF/KWh, NIP = 13.47 HUF/KWh, PIP = 0.28 HUF/KWh), we get the following: the optimal probability of a supplier going short in the given period was approximately 0.73. Therefore, a supplier can minimize his expected costs in the period studied by purchasing only so much energy as to result in a short position for 73 percent of the settlement periods.

In *Appendix 1* we provide a detailed assessment of the optimal *degree* of suppliers under-contracting. Our calculations show that in the first six months of 2006 the optimum degree of under-contracting for a supplier was on average equal to 0.6 times the standard deviation of the consumption forecast error σ.[2] So assuming for example that the supplier has a demand forecast error standard deviation of 5 percent, then according to our estimation he must have contracted only 97 percent of its projected consumption. If all participants are inclined to under contract, then of course the whole system will also tend to be "under-contracted". On the basis of this example, and assuming that there are altogether four suppliers on the market, all having equal shares, the system imbalance has a mean of –3 percent, and a standard deviation of 2.5 percent ($= 5\% / \sqrt{4}$) (see *Figure 2*).

[2] In addition to trying to minimize the costs there are of course other factors (for example arbitrage) that can drive the behaviour of market participants. Naturally, when calculating the optimal degree of under-contracting we did not take these factors into consideration.

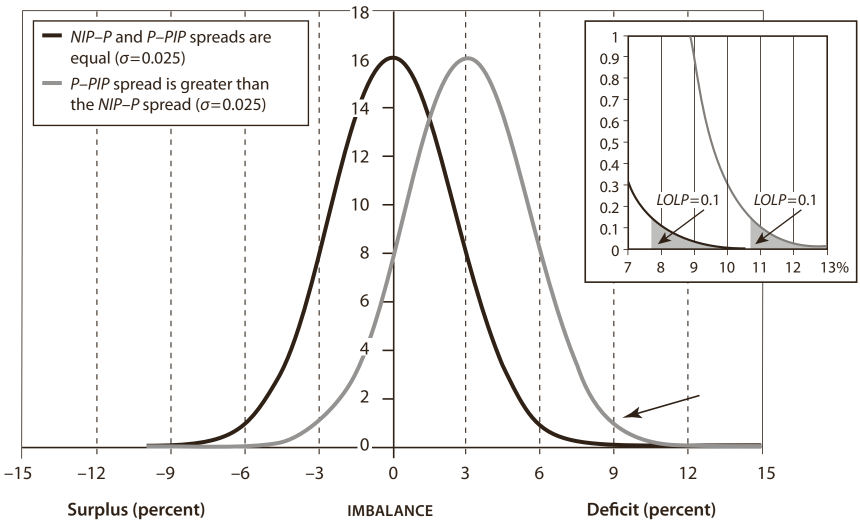


FIGURE 2 • **Density function of system imbalance when the *NIP–P* and *P–PIP* spreads are equal and when they are not**

The current balancing mechanism with unequal *NIP* – *P* and *P* – *PIP* spreads imposes high system balancing costs compared to the costs that would be incurred by a symmetrical spread around the day-ahead price (assuming that the difference between *NIP* and *PIP* is the same under the two pricing regimes). Since the frequency and the maximum value of negative imbalances increases, the system operator has to keep larger generation capacities in reserve for upward regulation.

In order to quantify this latter effect, let us compare the amount of upward regulation capacity needed to achieve a *LOLP*[3] of 0.1 percent under the existing pricing scheme as well as under a pricing scheme with equal *NIP* – *P* and *P* – *PIP* spreads (NIP_s = 16.445 HUF/KWh and PIP_s = 3.255 HUF/KWh, that is $NIP_s - P = P - PIP_s$ = 6.595 HUF/KWh and $NIP_s - PIP_s = 2 \times 6.595$ HUF/KWh). As shown in the close-up view in *Figure 2*, under the current pricing scheme the need of the system for upward regulation reserve reaches 10.5 percent of load, while with symmetrical spread around P it would only reach 7.5 percent.

Market dominance and "gaming"

Table 4 shows the developments in regulating and balancing energy prices in the past 18 months.

[3] *LOLP* = Loss of Load Probability

TABLE 4 • **Regulating and balancing energy prices in Hungary** (HUF/KWh)

	Regulating market		Balancing market	
Period	incremental energy price	decremental energy price	negative imbalance price	positive imbalance price
January 2005–January 2005	8.2	0	13.90	0.23
February 2005–December 2005	8.8	0	15.50	0.25
January 2006–June 2006	12.9	0	14.40	0.24
July 2006–August 2006	14.3	0	14.30	0.00 ??
September 2006			20.48	0.90 ??

Note: the figures indicate the weighted average of the peak and off-peak prices values. Following the terminology of the Network Code we refer to the procurement side of the balancing energy market as "regulating market", and the settlement side of it as "balancing market".

One of the most striking features of the Hungarian balancing energy market is the extremely low price of positive imbalances. Even considering that the key feature of the double pricing system is to incentivise more accurate planning, the size of penalty imposed on long positions seems to be unwarranted. The main reason for the low positive imbalance price is the fact that the Hungarian TSO receives a very low price for the surplus energy from the providers of downward regulation. In other words, the price of decremantal energy is very low (0 HUF/KWh) in the regulating market. This is due to the monopolistic structure of the regulating market, i.e. the fact that only MVM offers bids for decremental energy to the TSO.[4] Besides having the ability to influence the prices of regulating services, MVM are also capable of manipulating the overall system balance. This is due to the large size of the public utility balancing group managed by MVM. The latter's consumption accounts for 65-70 percent of the total domestic consumption. This enables MVM to shift the system balance in a direction that is favourable to the company's own interest.

Next we show that in the present pricing system MVM is encouraged to nominate more than its expected load (i.e. declare a larger than-anticipated load). Let us assume that MVM nominates more than its expected load and this brings the market into surplus. What happens in such situations? First, MVM as the single provider of regulating services will be called upon by the TSO to decrease output from its generators. The zero price of decremantal energy means that MVM pays nothing to the TSO for reducing its output. Second, MVM as the balance responsible party of the public utility balancing group will have a positive imbalance volume, which will be settled at the positive imbalance price of 0.24 HUF/kWh. Hence the net profit for MVM from pushing the system into a surplus is 0.24 HUF/kWh times the positive imbalance volume of the public utility balancing group. This arbitrage between the two markets could only be prevented if the positive imbalance prices were lower than the prices for decremental energy.

[4] Under sufficiently competitive conditions, the price for decremantal energy would come close to the variable cost of the least efficient operating plant (i.e. the plant with highest marginal cost).

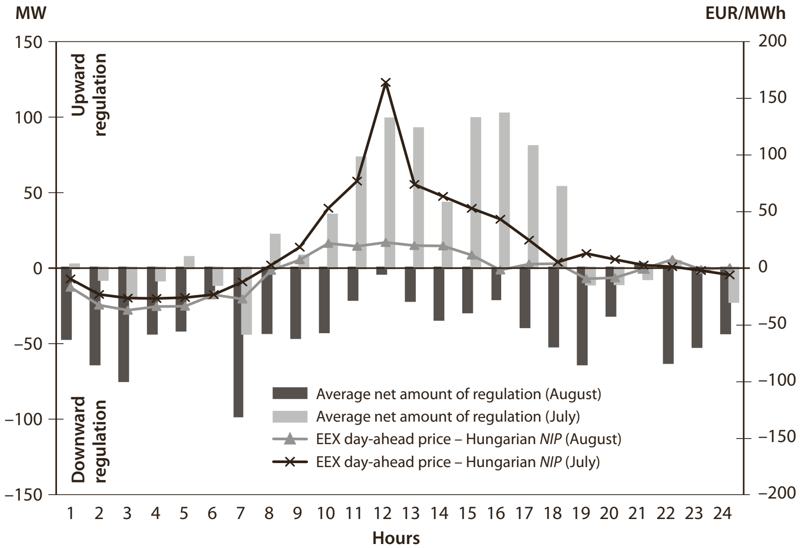


FIGURE 3 • **Direction of regulation versus the *NIP – P* spread in the Hungarian electricity market in July and August 2006**

Although we have no data about the development of the position of the public utility balancing group, the available aggregate data on the system imbalance are in line with the prediction that MVM continuously nominates more than its expected load. In 2005, the total volume of upward regulation was 384 GWh, while that of downward regulation was 545 GWh, meaning that in the course of the past year the system balance was more often in surplus than in deficit. This, in combination with the predicted under-contracting behaviour of suppliers other than MVM (i.e. ones serving the free segment of the market), suggests that MVM indeed keeps nominating more than the expected load of the public utility balancing group.

The 15 minute increment data on system balance pertaining to July and August of 2006 that were on the Hungarian TSO's webpage also suggest over-nomination by the MVM. As seen in *Figure 3,* in each hour of the day in August the TSO purchased on average more downward regulation than upward regulation. In contrast, in July the direction of regulation was predominantly upward. In *Figure 3* we also show for each hour of the day the estimated average spread between negative imbalance price and day-ahead price[5]. We can observe that in July the market price in the peak hours significantly exceeded the settlement price for negative imbalances, creating a huge incentive for arbitrage between the two markets. We suspect that in this period suppliers serving the free segment of the market generated large deficits in their scheduled portfolio, thereby making it impossible for MVM to push the system balance into surplus.

[5] This latter was calculated on the basis of the EEX day-ahead prices.

CONCLUSIONS

The balancing mechanism plays a central role in the wholesale electricity market. On the one hand, it is crucial in maintaining network stability and, on the other, it allows for a market-based settlement of imbalances between network users and the TSO. However, designing a balancing mechanism is a very complex task. The price system needs to incentivise the BRP to stay in balance, needs to minimize the social costs of balancing and at the same time needs to be robust against activities that threaten the functioning of the market.

In our study we examined the incentive properties of the current Hungarian energy balancing mechanism. We showed that the large spread between the NIP and the PIP creates a strong incentive for BRPs to reduce their imbalances resulting from inaccurate forecasts. On the other hand, we also show that, due to the existence of asymmetric penalties in the price system, BRPs have an incentive to under contract in the wholesale market as a hedge against real-time long positions and the associated higher imbalance costs. Finally, we demonstrated that MVM, who is responsible for the balancing of the public utility balancing group, has a strong incentive to nominate more that its expected load. This is due to the inconsistency in the pricing of decremantal energy versus positive imbalances, which is allowing for a positive spread between the negative imbalance price and the decremantal energy price. As a monopoly provider of downward regulation and the leader of the largest balancing group, MVM can push the system balance into surplus and then earn a positive profit from selling energy in real time to the TSO.

REFERENCES

EC [2005]: Report on progress in creating the internal gas and electricity market, COM(2005) 568 final. Commission of the European Communities.

Glachant, J. M.–Saguan, M. [2007]: An Institutional Frame to Compare Alternative Market Design in EU Electricity Balancing. Cambridge Working Papers in Economics, 07-24. Electricity Policy Research Group, 07-11, January 21. http://www.electricitypolicy.org.uk/pubs/wp/eprg0711.pdf.

Network Code [2006]: The network code of the Hungarian electricity system http://www.mavir.hu/domino/html/www/mavirwww.nsf/vAllPages/DC4ACF4CE5388E77C1256F1F0044C1E7/$FILE/KSZ20071120_20071128.pdf.

APPENDIX

1. The imbalance cost of a free market supplier

The aim of this appendix is to provide an estimate of the imbalance cost of a free market supplier under the assumption that $NIP - P = P - PIP$; i.e. its cost exposure to negative and positive imbalances are the same. (Note that under such prices a supplier has no incentive to nominate differently from its expected load.) Let $B = NIP - P$ $(= P - PIP)$ and x a random variable denoting the supplier's imbalance position. Then the expected cost of imbalances can be expressed as follows:

$$K = -B \int_{-\infty}^{0} x f(x) dx + B \int_{0}^{\infty} x f(x) dx,$$

where $f(x)$ is the probability density function of x. If x follows a normal distribution with mean 0 and standard deviation σ, then

$$K = -B \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} x e^{\frac{x^2}{2\sigma^2}} dx + B \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} x e^{\frac{x^2}{2\sigma^2}} dx,$$

that is

$$K = -B \frac{1}{\sigma\sqrt{2\pi}} \left[-\sigma^2 e^{\frac{x^2}{2\sigma^2}} \right]_{-\infty}^{0} + B \frac{1}{\sigma\sqrt{2\pi}} \left[-\sigma^2 e^{\frac{x^2}{2\sigma^2}} \right]_{0}^{\infty}$$

of which

$$K = 2B \frac{\sigma}{\sqrt{2\pi}}.$$

So if for example a supplier has a demand forecast error standard deviation of 5 percent and the difference between the negative and positive imbalance prices *2B* is EUR 50/MWh, then the imbalance cost of the supplier will be approximately EUR 1/MWh.

2. The optimal degree of under-contracting for a free market supplier

The aim of this appendix is to provide an estimate of the degree of under-contracting by a supplier in the Hungarian electricity market. Q denotes the supplier's customers' consumption. Real-time consumption then equals the expected amount of consumption plus the forecast error, θ, which is distributed according to $F(\theta)$. To meet its customers' demands, the supplier purchases $(1 + u)Q$ amount of energy on the wholesale market, where u stands for the degree of under-contracting.

If the real-time consumption is less than the contracted energy, i.e. $\theta < u$, then the surplus energy is sold at the *PIP* (positive imbalance price). The supplier's loss can be expressed as the following:

$$K = (u - \theta)(P - PIP)$$

If the real-time consumption is greater than the contracted energy, i.e. $\theta > u$, then the missing energy has to be covered in the balancing market at the NIP (negative imbalance price). The loss made by the supplier will be

$$K = (u - \theta)(NIP - P).$$

The expected cost of imbalances is then K, where

$$K = (P - PIP)\int_{-\infty}^{u}(u - \theta)f(\theta)d\theta + (NIP - P)\int_{u}^{\infty}(\theta - u)f(\theta)d\theta,$$

where f is the probability density function of the forecast error. The optimal degree of under-contracting is then u that minimizes K:

$$\frac{\partial K}{\partial u} = F(u)(NIP - PIP) - (PIP - P),$$

So the value of u solves

$$F(u) = \frac{NIP - P}{NIP - PIP} = \frac{1 - (PIP/P)}{(PIP/P) - (NIP/P)},$$

where F is the probability distribution of the forecast error θ. Given F, the optimal degree of under-contracting u can be solved. Assume that F is normal with mean 0 and standard variation σ. In the first six months of 2006, the average day-ahead price P was 9.85 HUF/KWh, the average NIP 13.47 HUF/KWh and the average PIP 0.28 HUF/KWh. Using these values we get

$$F(u) = 0.2737.$$

If F is normal, this can be expressed as

$$F(u) = 1 - \Phi\left(-\frac{u}{\sigma}\right) = 0.2737,$$

from which

$$\Phi\left(-\frac{u}{\sigma}\right) = 0.7263,$$

i.e.

$$-\frac{u}{\sigma} = 0.6,$$

i.e.

$$u = -0.6\sigma.$$

Our calculation shows that a supplier should on average be 0.6σ under-contracted in the first half of 2006. Assuming that the supplier had a consumption forecast error standard deviation of 5 percent, he should optimally cover only 97 percent of its expected consumption on the wholesale market.

• *András Kiss* •

THE EFFECT OF THE REGIONAL INTEGRATION OF ELECTRICITY MARKETS ON THE MARKET POWER OF POWER PLANTS*

The purpose of this paper is to construct a short-term economic model for the wholesale electricity market in the Central and Eastern European region – assuming conditions after a complete opening-up of the market. Among the inputs we provide an estimate for the generating capacities available and the cost of generation, demand as well as the transmission network data. The advantage of our modelling approach is that we simultaneously take into consideration the spatial structure of the electricity market and the capability of dominant companies to control prices. Our main conclusions: 1. at the current stage of market integration, major electricity generators are very powerful market players; 2. tighter market integration reduces the chances of abuse of market dominance and prices; however 3. even complete market integration cannot sufficiently limit the power of electricity generators. However, the practical importance of our modelling results cannot be assessed appropriately without determining how realistic they are.

INTRODUCTION

One of the key issues of the liberalisation of the electricity market in our region is *Can real competitive markets develop with the current ownership structure?* and *What threat does the market dominance of certain players pose?* The supply side of the electricity sectors in Central and Eastern European countries is quite concentrated: the overwhelming majority of generating capacities is concentrated in the hands of one or just a few players. This is one of the reasons why it is often argued that effective competitive markets are less likely to develop within a country. So in order to take advantage of the competition between electricity generators we need some kind of regional integration.

* The first, more detailed version of this paper was produced in the Central and Eastern European Energy Market (C3EM) Research Project conducted by the Regional Centre for Energy Policy Research (REKK) in 2005–2006 (*Kiss et al.* [2006]). The numerical model applied in the original version was constructed by the co-authors Julián Barquín and Miguel Vázquez (Universidad Pontificia Comillas, Madrid), and the author would like to express his deep gratitude to them. The author would also like to thank the following individuals: Zoltán Sulyok (MAVIR Hungarian Independent Transmission Operator Company Ltd) and REKK staff for their contribution and valuable suggestions in the course of the study.

In our paper we look at this issue using a numerical model. We have studied seven neighbouring countries in the region: Austria (AT), the Czech Republic (CZ), Croatia (HR), Hungary (HU), Romania (RO), Slovakia (SK) and Slovenia (SI).[1] After the description of: the structure and the workings of the model, generating capacities and costs, demand, and cross-border capacities, we try to find the equilibrium of the model in a competitive market environment and an environment characterized by strategic behaviour. After this we look at what changes can be expected from tighter integration regarding key variables, primarily: prices. At the end of the study we draw some conclusions from the numerical modelling exercise.

REGIONAL MARKET MODEL

We can look at the applied market model from four aspects: market demand, generating technology, spatial structure and corporate behaviour. We look at all four of these issues in detail below. We present not only the theoretical background but the data and estimates used for the numerical simulation.

Market demand

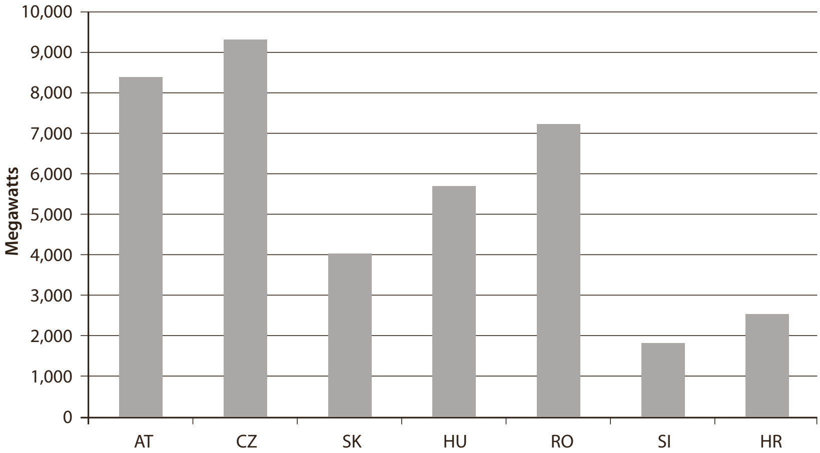
The demand for electricity is represented by an aggregate demand curve for each of the seven countries. It is a well-known fact that the electricity consumption of a country changes by the minute. However, we are not concerned with such temporal fluctuations, as our model is static by nature. Instead we have to record how demand changes at a specific point in time – which is typically the winter peak period – in relation to the market price of electricity. *Figure 1* shows the winter peak load of the various countries in the region.

As we do not have appropriate data to estimate the demand curve, we have to make various assumptions regarding the shape and position of the curve. To make things simple we chose a linear function, which can be described perfectly using three (easy-to-understand) data items.

The first is the demanded quantity, which has been described above, the second is the corresponding market price, which for the sake of simplicity we assume to be 30 EUR/MWh on every market.

By this we have defined a point on the demand curve. The steepness of the curve (the third data item) can be described by the elasticity of demand. Generally, the elasticity of demand for electricity is quite low: it is hard for consumers to find a substitute for the product.

[1] In brackets we use the abbreviations of the UCTE (Union for the Co-ordination of Transmission of Electricity). These abbreviations are used in the figures for the different countries.



Source: UCTE, own calculations.

FIGURE 1 • **Estimated winter peak demand (maximum system load) in the countries studied**

As no hard data are available we are forced to fall back on assumptions: we assume the elasticity of the demand to be –0.1 in every country (at specified demand points). Based on this, for example, a ten percent increase in price (short term) decreases consumption by one percent.[2]

Generating technology

There are numerous primary energy sources available for generating electricity, the most important ones being coal, natural gas, hydropower and nuclear power. As we are modelling short-term competition, we will concentrate on only production marginal costs.

As a good approximation, it can be assumed that, with regard to any given technology, the marginal cost of electricity generation at different production levels fluctuates within a very small range; therefore we assume the marginal cost to be constant.[3]

[2] Due to the nature of the linear function form, demand elasticity continuously changes along the demand function (higher prices mean higher demand elasticity). In the 20–50 EUR/MWh price range, which is interesting for modelling purposes, actual price elasticity is somewhere between –0.06 and –0.18. (However, we must consider that we have no reason to prefer constant price elasticity to the linear function form.)

[3] The average cost of generation is of course not constant because of fixed costs. However, since we are concerned with short-term supply decisions, we treat fix costs (e.g. labour costs and capital costs) as sunk costs, which do not influence the optimal supply decisions of power plants.

In order to estimate the marginal costs, first we need to determine the cost of fuel required to generate 1 MWh of electricity. Here we can set out in two different directions. We can take the observed total fuel-consumption (and related costs) of power plants and can project it on the quantity of electricity generated, or we can estimate the technology-based marginal cost of electricity generation from the energy conversion efficiency of generators and the fuel prices observed in the specific regions.

Although the first approach (using actual cost data) may seem more tempting theoretically, this method cannot be applied in practice – to the level of consistency required by the modelling example – due to the fact that such data is considered sensitive from a business perspective. On the other hand the advantage of the technology-based estimation is not only that significantly less data is required but also that there is a higher level of consistency inherent in the procedure: even if we are mistaken about the actual level of costs, the marginal costs of power plants in relation to one another remain consistent.[4]

We have aggregated the marginal cost curves resulting from technology estimation by country, as shown in *Figure 2.*

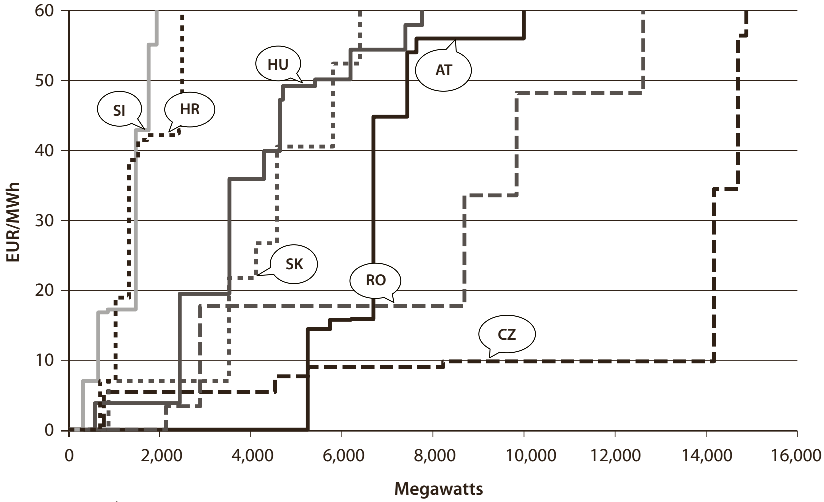
Figure 2 shows the available generating capacities and their costs as well as the load in peak periods. The bubbles with country codes indicate the point on the supply curve where the domestic demand can be met within the specific country (price elasticity has been ignored here). This way we can see the "international competitiveness" of the electricity sector of each country. The lower the bubble of a country is positioned, and the flatter the supply curve continues towards the right, the more the country's power plants are able to export cheaply to the regional market. In this regard the Czech and Romanian power plants are at an advantage.

Spatial structure

Since we are modelling quite a large regional market, the question arises whether the spatial structure has any significant effect on market equilibrium, and if so, how to take it into consideration.

Electricity is transmitted over long distances through high voltage transmission grids. The actual cost of transmission (the heat loss resulting from the resistance of power lines) are insignificant for the purposes of the model. However, the capacity

[4] In the case of hydropower we need to take a somewhat different approach, as the potential energy of water has no price as such. Of course it is true that we cannot generate electricity tomorrow with the water we use today, so we may not realise tomorrow's revenues this way. However, to estimate alternative costs we would need to have a fully dynamic market model, which is far beyond the scope of our study. As the second best solution we assume the marginal cost of hydro power to be zero; however, we reduce the amount of electricity that can be generated to the level of the annual average capacity utilization.



Source: Kiss et al. [2006].

FIGURE 2 • **Aggregate marginal cost curves**

constraints of power lines cannot be ignored: if the load exceeds the capacity limit, the lines simply burn (which the transmission system operators do not allow to happen).

The structure of transmission networks had to be simplified for the purposes of the model. In our model all parts of the transmission network, within each country, were simplified to a single node, and between two neighbouring nodes (countries) there are no more than one cross-border link drawn. Every consumption and generation takes place at the nodes and the transmission of electricity (trade) occurs through the limited capacity lines connecting them. By marking countries with a single node, we assume that congestion can only occur on the interconnectors.[5] *Figure 3* shows a stylized drawing of the modelled region. We will concentrate on the interconnectors denoted by solid lines (and the countries located at the end of these lines) explicitly.

Our capacity constrained electricity transmission model is further „complicated" by the laws of physics pertaining to current: Kirchhoff's junction and loop rules. The former is interpreted in a relatively intuitive way from an economic aspect in our model: the sum of all electricity flowing to a node (generation + import) is equivalent to the electricity flowing from the node (consumption + export). However, the loop rule is not in compliance with the general view on transportation of goods: free route choice does not apply to electricity!

[5] In the case of Austria for instance this assumption is not always true; therefore it cannot be considered a perfect approximation of the real situation.

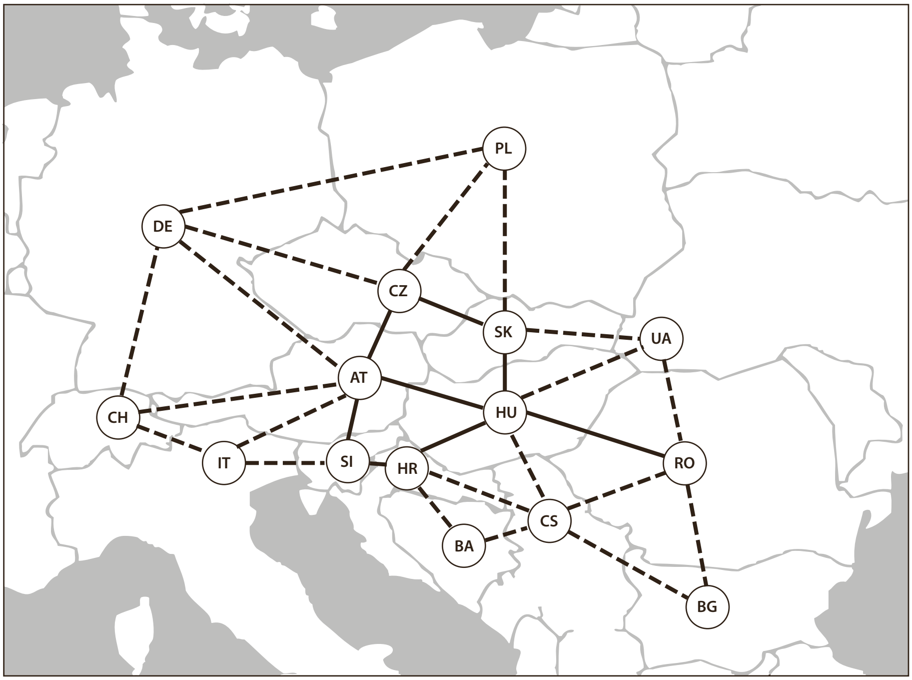


FIGURE 3 • **The spatial structure of the regional market**

On a network with parallel paths between two nodes, electricity flows along all parallel paths between the two nodes. Furthermore, the amount of electricity flowing through the specific network lines is distributed (roughly) in inverse proportion to the resistance of each path.

Let us take the transaction of 100 MW from Hungary to Austria as an example. If we want to consider its actual physical effect on the specific cross-border lines, we find that only one third of the 100 MW actually flows through that line from Hungary to Austria, the rest takes parallel routes through Slovakia, the Czech Republic, Croatia and Slovenia before reaching Austria – but if we look at the map we can see that the transaction also effects the Polish-German, or the Swiss-Italian borders as well. Of course the further the route is the less electricity flows through there.

The effect that electricity transmission between two nodes has on a line can be described using so called PTDF matrices, which are used on a regular basis by transmission system operators.[6] The current European cross-border capacity dis-

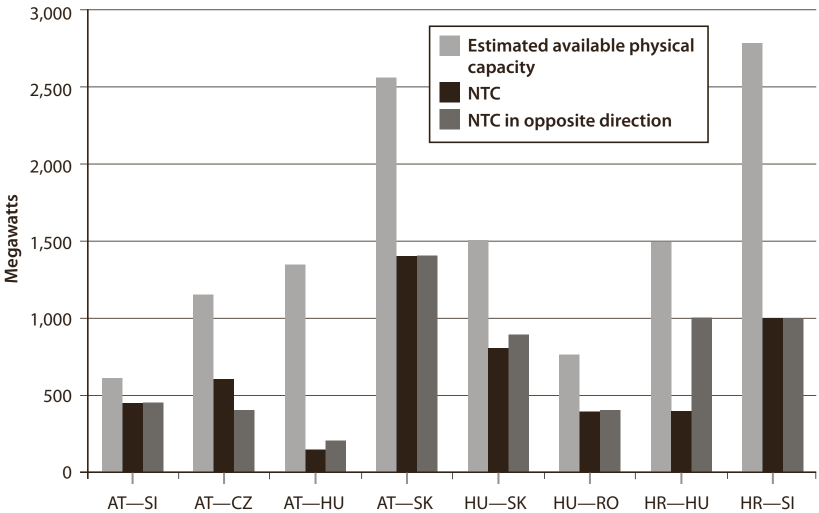
[6] PTDF stands for Power Transfer Distribution Factor. It shows the size and direction of physical flows generated by the transfer of one unit of electricity between two control zones connected directly and indirectly.

tribution mechanisms do not take into consideration the effect of non-direct (loop) flows, which has an adverse consequence: the bilateral transmission agreements generate negative external effects on lines that connect nodes indirectly. (They reduce available capacities elsewhere, which are not paid for.)

The existing continental capacity distribution mechanism (bilateral or coordinated auctions) „solves" the issues caused by externalities by reducing the actually available cross-border capacities by the amount of loop flows. This solves the problem of system security, but does not eliminate the basic welfare losses caused by external factors.

In tightly integrated systems, the effect of loop flows is explicitly taken into consideration during capacity allocation, and so called nodal pricing is applied.[7] In our model – in order to simplify the modelling of strategic behaviour – we assumed such an effective capacity allocation mechanism. However, we are aware that for the region under study this is far from the current practice. As a matter of fact, tighter regional integration could be interpreted as a shift to a more effective capacity allocation system. However, our model cannot be used to evaluate such a measure.

Figure 4 shows the size of cross-border capacities taken into account. As a starting point we can assume that the values of the so called NTC (net transfer capacity),[8]



Source: UCTE, ETSO, own calculations.

FIGURE 4 • **Estimated network capacity values available with regional nodal pricing and current net transfer capacity (NTC) values**

7 See for example, the PJM market on the Eastern coast of the United States (www.pjm.com).

8 NTC, or Net Transfer Capacity is the maximum capacity for exchange of power between two control zones.

known from the bilateral capacity allocation system, determine the maximum amount of electricity that can be transmitted between two nodes (countries). Since we use nodal pricing in the model, this is only a rough estimate. Therefore we also present an "integration" scenario in our model as well, where we determine the size of cross-border capacities available by subtracting the effect of average loop flows (and a 20 percent reserve margin) coming from countries that are outside the region from the physically available network capacity.[9]

As one can see, the available capacities estimated by us (based on the simplified network model) in each and every case exceed the NTC values actually published. The average difference is almost double.

Corporate behaviour

During the running of the model we distinguish two behavioural patterns by (the owners of) generators. The more basic assumption is price-taking behaviour (perfect competition). Every power plant assumes that their decision to generate electricity has no effect on market prices or the usage of cross-border capacities (and consequently their prices).

As a result, companies will keep increasing their electricity generation until the local market price exceeds their marginal costs (of course within the specific generating capacity constraints).

The first welfare theorem of economics states that perfect competition leads to efficient allocation in the market: competition maximises complete welfare attainable on the regional market given the existing constraints (including generating and transmission constraints). Of course, if we loosen these constraints – for instance, by assuming tighter integration, implying larger cross-border capacities – we can achieve a higher welfare level in the new equilibrium than previously.

The second possible assumption is that companies with large generating capacities recognize what effects their own decisions on their electricity output have on market prices. In extreme cases they may know perfectly well the demand curves as well as the reaction of the price-taking corporate sector (the „competitive fringe") and strategic competitors. To solve the model we are applying the so called Cournot-assumptions, meaning that when making their output decisions strategic companies assume that other large, strategic players do not react to the output changes of competitors, but the competitive fringe adapts to the new market price in a price-taking manner. In addition, strategic companies need to be able to forecast, which interconnectors will be congested. Equilibrium will occur where all

[9] It should be kept in mind that this calculation method ignores the effect the flows within a country have on cross-border lines.

these assumptions are in line with the generating decisions companies have made based on their forecasts.[10]

The ability to decide who the *strategic players* are, provides some decisive freedom in the course of modelling. Having examined several variations, we selected (non-state-owned) companies, which have strategic generating capacities that are significant both regionally and at a national level, but do not have full coverage of the sector. We have identified three such companies: CEZ (the Czech Republic), SE (Slovakia) and Verbund (Austria). Including such companies as MVM (Hungarian Electricity Ltd), AES-Tisza Power Plant Ltd, Electrabel Hungary Ltd or RWE Energy Hungary Ltd. which are regionally small (although significant in Hungary) in our study does not affect our findings greatly. Such a great part of the Romanian, Slovenian and Croatian generating capacities was state-owned at the time of modelling (and still are) that it would be more reasonable to expect a price-taking (or from a different perspective: optimally regulated) market behaviour from them than to think they would go for profit-maximisation.

OUTCOMES OF PERFECT COMPETITION

Having presented the model and input data, let us now look at the results. *Figure 5* shows the main scenario characterized by perfect competition and low level of regional integration.

There are two values corresponding to each country in the figure. The top box shows the equilibrium market price (EUR/MWh), while the box at the bottom displays the net export position of the country. A positive value in a white field means that the country is a net exporter (in MWh); while a negative value in a black field means that the country is a net importer.

Arrows crossing the borders indicate the direction and the strength of electricity flows (the stronger the flow the thicker the arrow). The tone of the arrow indicates whether the interconnector is congested (i.e. if the capacity of interconnectors is an effective constraint on further trade). Lines using 100 percent of their capacity are marked in black, while those marked in grey are not used to the full.

[10] For a detailed, formalized description of the model and the solution concept, see the paper of *Barquín–Vázquez* [2005]. Numerous studies have been conducted about the strategic modelling of electricity market competition, which for the most part differ in their assumed market mechanisms, the type of strategic game, the degree, to which they cling to the physical characteristics of electricity flow, or their equilibrium calculation methods (see *Cardell et al.* 1997], *Smeers* [1997], *Wei–Smeers* [1999], *Hobbs et al.* [2000], *Joskow–Tirole* [2000], *Day et al.* [2002], and *Metzler et al.* [2003]) A great overview of electricity market strategic competition modelling literature is given by *Neuhoff et al.* [2005] and *Ventosa et al.* [2005].

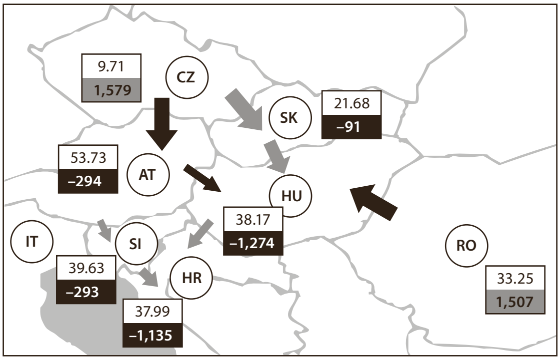


FIGURE 5 • **Competitive market, moderate integration**
(cross-border capacity = NTC value)

We can make the following observations about the diagram. There are large price differences between equilibrium prices in the countries as a result of severe congestion on three interconnectors (from the Czech Republic to Austria, from Austria to Hungary, from Romania to Hungary).

The lowest priced country is the Czech Republic, followed by Slovakia and Romania. The prices in Hungary, Croatia and Slovenia are higher, and they are relatively close to each other. The quite high price observed in Austria is probably the result of two effects: firstly, the limited import capacity existing from the direction of the Czech Republic, and secondly, the limitation of the capacity of storage power plants to average available capacity. (With less careful assumptions we would probably allow much a higher level of capacity usage for storage plants, which would lead to significant inexpensive extra capacity – and lower prices.)

Only the Czech Republic and Romania are in net exporting position. The most severe power deficit occurs in Croatia and Hungary.

Figure 6 shows what happens in a model assuming perfect competition, if, by increasing cross-border capacities, a tighter regional integration is achieved (for the degree of capacity increase see *Figure 4*).

The striking difference between *Figure* 5 and *Figure* 6 is the dramatic increase in the Czech export position towards both Austria and Hungary. The net exporting position of Romania has slightly decreased, but it is still very positive. Both Austria and Hungary have greatly increased their dependence on imports, while the export-import balance of all other countries has only slightly worsened.

The effect of tighter market integration is clearly visible in the equilibrium prices: market prices have converged considerably. In the case of the Czech Republic, this implies a two-fold increase in the system price; nevertheless, the Czech market remains the lowest priced country in the region. Slovakia comes in second place, followed by Romania, Hungary,

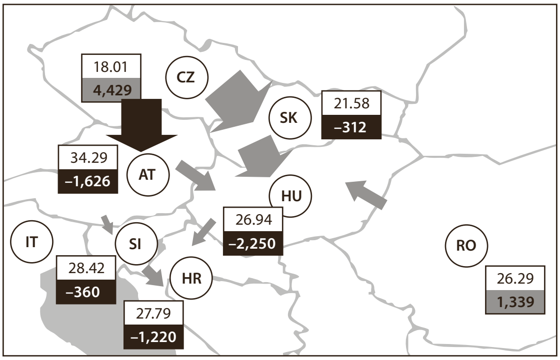


FIGURE 6 • **Competitive market, tight integration**
(cross-border capacity = estimated available physical capacity)

Croatia and Slovenia. Austria still has the highest electricity price, but proportionately it has had by far the largest price decrease of all countries.

Correspondingly, congestion still exists between the Czech Republic and Austria, but line limits from Austria to Hungary and Romania to Hungary are no longer binding.

THE EFFECT OF MARKET DOMINANCE IN THE REGION

We already covered the assumptions and effects that lie behind strategic behaviour, so now we will simply present and interpret the modelling results. *Figure 7* shows the market outcomes resulting from the strategic use of market dominance with moderate regional integration.

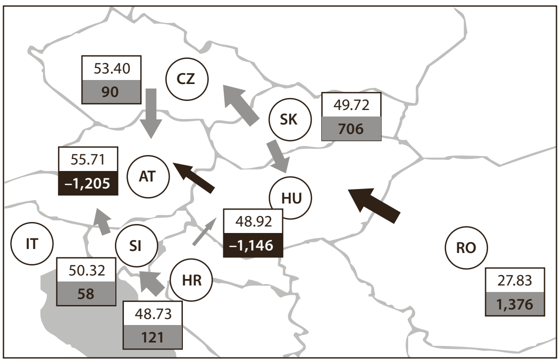


FIGURE 7 • **Oligopolistic market, moderate integration**
(cross-border capacity = NTC value)

To understand the effects of market dominance, compare *Figures 5* and *7*. Regarding the net export positions, the most striking is how much production in the Czech Republic was cut back. Numerical results show that as an oligopolistic company CEZ decreased generation by almost 3,400 MW (44 percent), which is only partly substituted by the 525 MW (14 percent) increase in Czech fringe production. There is also a sizeable decrease in power generation by Verbund (–1,019 MW) and SE (–809 MW). As a result, prices have gone up considerably throughout the whole region (except Romania, where competitive companies prevail).

Some countries (most notably Croatia and Slovenia) have turned from net importers to net exporters. Generally, the destination of electricity trade is still Hungary and Austria, but the Czech Republic is no longer the main source. Consequently, the direction of the flow of electricity has also changed at some interconnectors. The line from the Czech Republic to Austria is no longer congested, and the Austrian-Hungarian line is now congested in the other direction, towards Austria. Assumed price-taking behaviour in the Romanian power sector ensures that Romania remains a strong net exporter, as a result of which the flow on the Romania → Hungary interconnector has not decreased.

Finally, let us examine what effect tighter regional integration has in a market dominant position on the equilibrium *(Figure 8)*.

Relative to the corresponding scenario under the competitive setting *(Figure 6)*, we can make the following observations.

- Capacity withholding has decreased trade in the region and taken the load off the interconnectors. All congestions have been eliminated (although Romania-Hungary is still very close to being congested, using almost 100 percent of its capacity).
- As a result, all prices have converged to 41.23 EUR/MWh, which is much higher than any of the market prices under the perfect competitive market setting with tight integration.

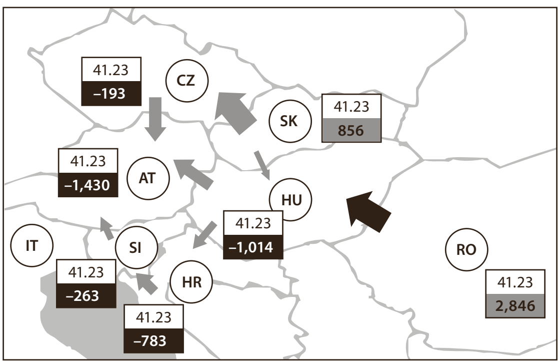


FIGURE 8 • **Oligopolistic market, tight integration**
(cross-border capacity = estimated available physical capacity)

- Slovakia has taken over the role of main regional exporter from the Czech Republic. Romania exports almost twice as much as in the competitive environment.
- If one takes advantage of market dominance, the average regional price level will still be higher under tight integration than with price-taking behaviour without integration.

Finally, let us see what effect regional market integration has on market dominance *(Figures 7 and 8)*.

- With the exception of Romania, prices in all of the countries dropped by an approx. average of 20 percent after the integration.
- The direction of flows remained the same.
- The Czech Republic, Croatia and Slovenia have turned from net exporters to net importers. The export-import balance of Austria has slightly worsened, while that of Hungary and Slovakia got better.
- The “missing” energy is provided by price-taking firms in Romania, which increases the local price there to regional levels as well.

SENSITIVITY ANALYSIS

Of the input data of our model – not taking into consideration structural assumptions – the information about the demand function are the most ad hoc. As a result, we have looked at some other scenarios regarding the level of demand and its price elasticity.

Looking at the results, we can say that the characteristics of the changes in demand and under competitive and strategic behaviour were qualitatively alike. So we only present the results about the oligopolistic market structure.

Figure 9 shows what happens if we reduce the demand to different extents on regional markets The specific cases (10–50 percent drop in demand) can be interpreted in two ways.

As we know, the demand for electricity fluctuates according to the time of the day, from week to week and seasonally. With the changes in the level of demand according to the first interpretation, we are examining how sensitive our results are to normal fluctuations in demand. (The ratio of peak period and off-peak period consumption can be as much as 2:1.)

In the second interpretation, sensitivity to changes in demand also affects one of our structural assumptions: the insularity of the region towards cheap and competitive import coming from outside (e.g. Poland or Ukraine). If some inexpensive and price-taking import electricity comes from over the borders of the region, we expect residual demand to drop. However, taking import capacities into consideration this decline in demand can be no more than 10-20 percent.

No matter which interpretation we chose, it becomes obvious that the decline in demand will result in lower equilibrium prices. The degree of price decrease is

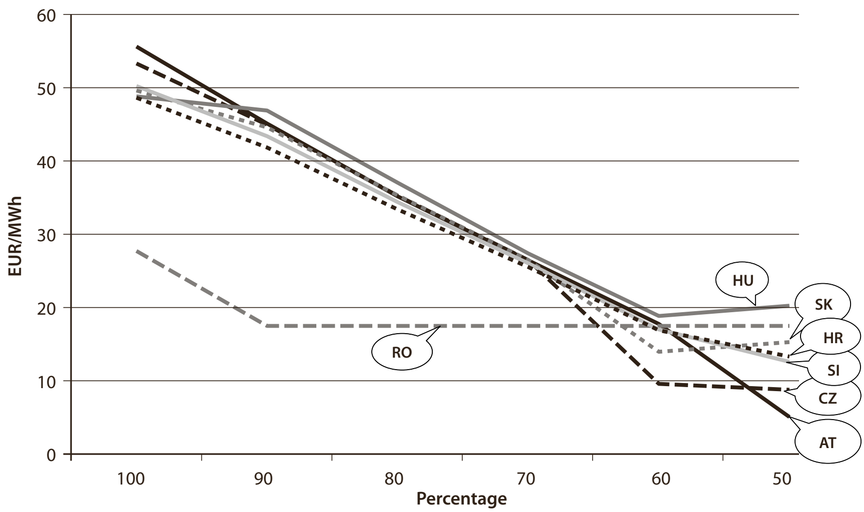


FIGURE 9 • **The effect decline in demand relative to peak demand (100 percent) has on equilibrium prices under an oligopolistic market structure and moderate integration**

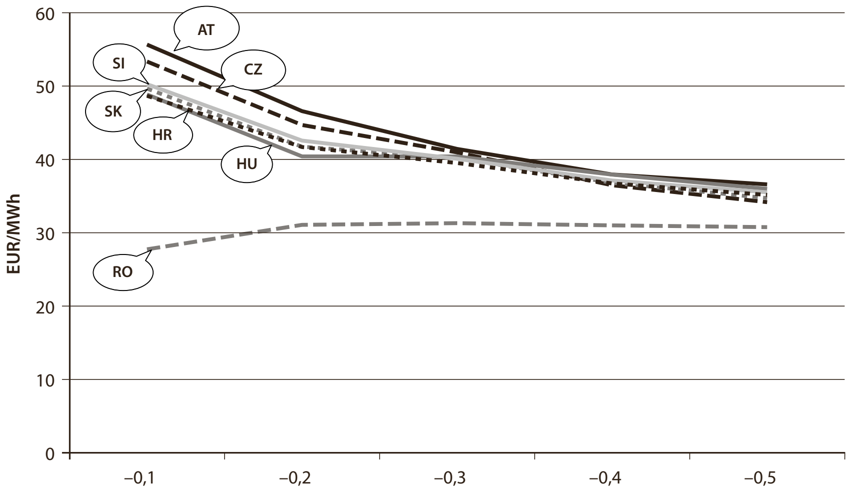


FIGURE 10 • **The effect greater demand elasticity has on equilibrium prices under oligopolistic market structure and moderate integration**

significant (it can drop to half of the market price); however it corresponds to the difference between peak and off-peak prices of power exchanges.

Figure 10 shows the effect the increase of demand elasticity has on equilibrium prices. Here our expectations inspired by economic theory are met: greater demand elasticity reduces price raising by strategic companies, since with the price increase companies can expect a sharper drop in demand, which makes the increasing of prices less profitable. However, the *Figure* also shows that even major changes in demand elasticity do not have such a strong reducing effect on prices as the daily fluctuation of consumption does.

CONCLUSIONS

We have now seen that large electricity generators have considerable market power in the modelled regional market environment, which raises market prices above competitive levels across the region. Using current NTC values as available cross-border capacity, this mark-up can range from 2 EUR/MWh (Austria) to 44 EUR/MWh (Czech Republic), with a typical value around 12-14 EUR/MWh. In percentages, the margin averages between 25-40 percent.

At the same time, two interconnectors are very heavily used. Congestion on the lines from Romania to Hungary and from Hungary to Austria reflects the effect of the competitive electricity supply coming from the Eastern end of the region and trying to reach the Western part of the region, where capacity withholding creates a shortage of supply.

We can observe that the modelled tighter regional integration does indeed reduce the price increasing power of dominant market players. The primary reason for this is also that the competitive supply coming from Romania is allowed to compete with strategic supply in Western markets. This result is of course not independent from the assumption (exogenous in our model) that electricity generators in Romania behave in a price-taking way.

On the other hand we have to note that large regional electricity generators have significant dominance even in a tightly integrated market. (Figure 8 shows the tightest integration that can be achieved, as there are no congestions at any borders in the region, and the same price applies to every market.) Even if we consider the effect of competitive supply coming from the east, we find that short-term equilibrium prices will be around 1.5-2 times the prices observed in the integrated competitive scenario.

Thus, it is fair to say that closer integration mitigates market power relative to a *more segmented market structure*, but it is not nearly sufficient to eliminate it altogether or to realize the potential welfare gains of market competition. (Indeed, the price mark-up of strategic players is barely dented by integration.) To draw our main conclusions we should not forget the original assumptions we worked with and their limitations.

The most important ones are: the static nature of the model, the idealized nature of the capacity allocation mechanism used, somewhat arbitrary drawing of the borders of the region under study, the isolation of the region, and the optimistic assumptions about the motivations of state controlled market players. It is hard to evaluate the practical importance of our modelling results until these limiting simplifications have been overcome.

REFERENCES

Barquín, J.–Vázquez, M. [2005]: Cournot equilibrium in power networks. Mimeo, Instituto de Investigación Tecnológica, Universidad Pontificia Comillas, Madrid, http://www.iit.upcomillas.es/docs/IIT-05-007A.pdf.

Cardell, J.–Hitt, C. C.–Hogan, W. W. [1997]: Market power and strategic interaction in electricity networks. Resource and Energy Economics. Vol. 19, No. 1–2, pp. 109–137.

Day, C. J.–Hobbs, B. F.–Pang, J.-S. [2002]: Oligopolistic competition in power networks: a conjectured supply function approach. IEEE Transactions on Power Systems, Vol. 17, No. 3, pp. 597–607.

Hobbs, B. F.–Metzler, C.–Pang, J.-S. [2000]: Calculating equilibria in imperfectly competitive power markets: an MPEC approach. IEEE Transactions on Power Systems, Vol. 15, No. 2, pp. 638–645.

Joskow, P.–Tirole, J. [2000]: Transmission rights and market power on electric power networks. RAND Journal of Economics, Vol. 31, No. 3, pp. 450–487.

Kiss, András–Barquín, J.–Vázquez, M. [2006]: Can closer integration mitigate market power? A numerical modeling exercise. In: Towards More Integration of Central and Eastern European Energy Markets. Regionális Energiagazdasági Kutatóközpont, Budapest, http://rekk.uni-corvinus.hu/c3em/pdf/modeling_study.pdf.

Metzler, C.–Hobbs, B. F.–Pang, J. S. [2003]: Nash–Cournot equilibria in power markets on a linearized DC network with arbitrage: formulations and properties, Networks and Spatial Economics, Vol. 3, No. 2, pp. 123–150.

Neuhoff, K.–Barquín, J.–Boots, M. G.–Ehrenmann, A.–Hobbs, B. F.–Rijkers, F. A. M.–Vázquez, M. [2005]: Network-constrained Cournot models of liberalized electricity markets: the devil is in the details. Energy Economics, 27, pp. 495–525.

Smeers, Y. [1997]: Computable equilibrium models and the restructuring of the European electricity and gas markets. Energy Journal, Vol. 18, No. 4, pp. 1 –31.

Ventosa, M.–Baíllo, Á.–Ramos, A.–Rivier, M. [2005]: Electricity markets modeling trends. Energy Policy, Vol. 33, No. 7, pp. 897–913.

Wei, J.-Y.–Smeers, Y. [1999]: Spatial oligopolistic electricity models with Cournot generators and regulated transmission prices. Operations Research, Vol. 47, No. 1, pp. 102–112.

• *Péter Kaderják–András Kiss–László Paizs–Adrienn Selei–Pálma Szolnoki–Borbála Tóth* •

NATURAL GAS MARKET INTEGRATION IN THE DANUBE REGION: THE ROLE OF INFRASTRUCTURE DEVELOPMENT

The paper introduces the Danube Region Gas Market Model, a network and contract constrained multi-country competitive equilibrium model and applies it to estimate the impacts of new gas infrastructure investments on market integration, social welfare and supply security in the countries of Central and South East Europe. Individual projects, project packages (e.g. the North-South gas corridor for Central and Eastern Europe) and international pipeline projects (like Nabucco West) are evaluated according to the Regional Cost Convergence Index. Estimates on price spill-over effects of new infrastructures are also presented. The model can support cost benefit analyses foreseen by the proposed European Infrastructure Package to identify EU projects of common interest.

INTRODUCTION

New EU member states and the wider Central and Southeast European region (in the following: the Danube Region or DR[1]) suffer from specific gas industry problems. The most serious one is the lack of sufficient interconnectivity of the region's countries prohibiting gas supply source diversification for the DR comparable to that of old member states.[2] The lack of interconnectivity also reduces the scope for gas market integration and supply security improvements at the regional level.

Since the shock of the 2009 January gas crisis, European energy policy has been seeking ways to address the above mentioned gas industry problems of the new EU member states, with consideration towards the Energy Community countries. A prominent example is EU gas supply security regulation 994/2010. More recently, the new European Infrastructure Package (EIP)[3] intends to identify and provide Union level support for gas infrastructure projects that might impact interconnectivity

[1] The 14 Danube Region countries are: Austria (AT), Bosnia and Herzegovina (BA), Bulgaria (BG), Croatia (HR), the Czech Republic (CZ), Germany (DE), Hungary (HU), Moldova (MD), Montenegro (MNE), Romania (RO), Serbia (SB), Slovakia (SK), Slovenia (SL) and Ukraine (UA).

[2] The price, supply security and political risks of a lock-in situation with dominant Russian import dependence for the DR are thoroughly discussed by *Kaderják* [2011a] and [2011b].

[3] COM(2011) 658 (in the followings: proposed Regulation), SEC(2011) 1233 and COM(2011) 665.

and market integration[4] in a positive way in this EU region. The EIP identifies certain priority corridors, which in the case of gas includes linking the Baltic, Black, Adriatic and Aegean Seas. The development of north-south interconnections in Central and Eastern Europe and Southeast Europe forms an important element of this corridor. In 2011 the EC commissioned a "High Level Group"[5] with the mandate to devise an action plan for the development of interconnections in gas, electricity and oil by the end of 2011. The High Level Group published its action plan in December 2011. In 2012 this work continues with a similar High Level Group activity for the Energy Community countries.

Finally, in October 2011, the EU approved the European Union Strategy for the DR that foresees a strengthened cooperation for the countries of the DR in a wide range of areas, including energy policy. Its Action Plan[6] states that for a secure and well-functioning natural gas market in the DR:

- "…the interconnections between national markets have to be improved and countries in the region need to gain access to new external sources. Reinforcing gas transmission infrastructure will be key for preventing potential supply disruption in the future. Well-functioning networks, interconnections and interoperability are needed for energy security, diversification and effective energy operation." (*EC* [2011] p. 18).

While an agreement seems to emerge that gas infrastructure development is the key to improving gas market integration and supply security for the DR, no solid methodology has yet been developed to assess the impacts of the proposed projects or project packages on regional gas market integration, security of supply, competition and sustainability. Moreover, while the proposed Regulation foresees the application of energy system-wide cost-benefit analysis for the evaluation of promoted projects, such a methodology is still to be developed – in the case of gas by the European Network of Transmission System Operators for Gas (ENTSO-G).

While, with the objective of supporting the work of the North-South gas working group, the study by *Kantoor Management Consultants* [2012] develops a methodology to set priorities for regional gas infrastructure developments, the proposed methodology still leaves many problems unsolved. Its basis is a physical flow model,

[4] Article 4 of the proposed infrastructure Regulation defines four criteria that will apply for the evaluation of gas projects of common interest. These are their impact on market integration, security of supply, competition and sustainability.

[5] The High Level Group on north-south interconnections is chaired by the EC and includes Bulgaria, the Czech Republic, Hungary, Poland, Romania and Slovakia as members, and Croatia as an observer. Austria, Germany and Slovenia also became members of this group. The High Level Group also established a "working group on natural gas" (GWG) consisting of representatives of the relevant ministries, regulatory authorities and transmission system operators (TSOs) in the participating countries, except for Austria and Germany.

[6] Com(2010) 715 and SEC(2010) 1489, respectively.

with country-level analysis, focusing mostly on security of supply issues. The gas market representation is rather simple, price formation modelling is neglected, therefore the understanding of how new infrastructures will impact market integration is incomplete. On the whole, the Kantoor study provides important insights on how changes in infrastructures affect the security of supply status of individual countries. However, the analyses do not answer what affects new infrastructures may have on prices, costs and benefits or social welfare.

This paper reports on an alternative approach to the evaluation and ranking of new gas infrastructure projects in a regional gas market context. We introduce the Danube Region Gas Market Model (DRGMM) and illustrate how model simulations can be used to assess the impacts of new infrastructures or infrastructure packages on regional gas market integration and for system-wide cost-benefit and security of supply analyses. If extended to include all the EU27 gas markets, the model could help the implementation process of the proposed infrastructure Regulation. First, it could serve as a potential component of the cost-benefit methodology envisioned by the proposed Regulation.[7] Second, model estimates on the distribution of consumer and producer benefits from new infrastructures across impacted countries could also support the Agency for the Cooperation of Energy Regulators (ACER) in elaborating its decisions on cross border cost allocation for Projects of Common Interest (PCI) when national regulatory authorities did not reach agreement on those matters.[8]

The structure of the paper is as follows. After a brief literature review on gas market modelling, we summarize the basic assumptions and characteristics of the DRGMM. Next we present several simulation results to illustrate the variety of analyses the model allows for, including market integration, cost-benefit and security of supply analyses. Finally, we reflect on the limitations of the model's present version and suggest areas for future research.

LITERATURE REVIEW

In the following, we give a short review of the most important complex, large-scale computational gas market models, which have been applied to analyze the security of gas supply and the impact of infrastructure developments in Europe.

The main focus of the EUGAS model (*Perner–Seeliger* [2004]) is to analyze the prospects of gas supplies to the European market in the coming decades. It assumes perfect competition among market players and contains an extensive infrastructure representation. The objective function and the constraints of this model are linear.

[7] See e.g. Article 12 of the proposed Regulation.
[8] See Article 13(6) on this matter.

The time periods include five years, and the annual gas consumption is split seasonally into three different load periods.

Contrary to the EUGAS model, most of the gas simulation models depict the strategic interaction between players. The GASTALE model (*Boots* [2004]) was the first attempt to apply successive oligopoly in natural gas production and trading in a large-scale simulation model. The model has a two-level structure, in which producers engage in competition *a la* Cournot, and each producer is a Stackelberg leader with respect to traders, who may be Cournot oligopolists or perfect competitors.

The extended, dynamic versions of the GASTALE model (*Lise–Hobbs* [2008], [2009]) include investments in scarce infrastructures (such as pipelines, storages and LNG infrastructure). However, this version assumes market power only for producers.

GASMOD (*Holz et al.* [2008] is similar in spirit to GASTALE. It also structures the European natural gas market as a two-stage-game of successive oligopolies: imports to Europe (first stage, upstream) and trade within Europe (second stage, downstream). As the model's main focus is to examine the possible effects of liberalization on trade, the geographical coverage of the model is wide. On the demand side it includes all European markets, and on the supply side it includes major exporters to Europe.

Egging et al. [2008] presented a more detailed complementary model of the European natural gas market which accounts for the issues of market power of exporters and of globalizing natural gas markets with LNG trade. The market structure that their model implements is different from that of GASMOD and the static GASTALE model as they assumed that only traders can exert market power by playing the Cournot game against each other. Other players are assumed to be price takers.

Based on their previous work (*Gabriel et al.* [2005a], [2005b]) *Egging et al.* [2010] presented the World Gas Model. It is a multi-period mixed complementarity model for the global natural gas market, which contains more than 80 countries and regions and covers 98% of worldwide natural gas production and consumption. It also includes a detailed representation of cross-border pipelines and constraints imposed by long-term contracts in the LNG market. The model operates with five year periods as well as two seasons (peak and off-peak). Similar to the previous models it includes market power in the upstream market, which lies with the traders, both representing pipelines and LNG deliveries. It allows for endogenous capacity expansions and seasonal arbitrage by storage operators.

The NATGAS model (*Mulder–Zwart* [2006]) assumes an oligopolistic producer market where a small number of strategic natural gas producers are facing price-taking traders in the downstream market. The main focus of the model is to compute long-term effects of policy measures on future gas production and gas prices in Europe. It contains long-run projections of supply, transport, storage and consumption patterns in the model region, aggregated in 5-year periods, distinguishing two seasons (winter and summer).

Abada et al. [2012] developed a dynamic Generalized Nash–Cournot gas market model (GaMMES model). In the applied oligopolistic market structure they take into account long-term contracts in an endogenous way, which makes the model a Generalized Nash Equilibrium problem. Their demand representation is specific because it captures the possible fuel substitution that can be made between the consumption of oil, coal, and natural gas in the overall fossil energy consumption.

THE DANUBE REGION GAS MARKET MODEL

The Danube Region Gas Market Model was developed by REKK to simulate the operation of an international wholesale natural gas market in the Central and South-East European (CSEE) region.[9] *Figure 1* shows the geographical scope of the model. Country codes denote the countries, for which we have explicitly included the demand and supply side of the local market, as well as gas storages. Large external markets, such as Germany, Italy or (indirectly) Russia, are represented by exogenously assumed market prices, long-term supply contracts and physical connections to the CSEE region.

All map outlines are based on the maps of Daniel Dalet, source: http://d-maps.com/m/europemax/europemax09.svg.

FIGURE 1 • **The geographical scope of the Danube Region Gas Market Model**

[9] For a first description and application of REKK's Regional Gas Market Model see *Kaderják* [2011a] pp. 121–147.

Given the input data, the model calculates a dynamic competitive market equilibrium, subject to constraints represented by the physical gas infrastructure and contractual arrangements specific to the region, for fifteen countries in the CSEE region, and returns the market clearing prices, along with the production, consumption and trading quantities, storage utilization decisions and long-term contract deliveries.

Model calculations refer to 12 consecutive months, with a default setting of April to March.[10] Dynamic connections between months are introduced by the operation of gas storages ("you can only withdraw what you have injected previously") and long-term take-or-pay (TOP) contract constraints (minimum and maximum deliveries are calculated over the entire 12-month period, enabling contractual "make-up").

The Danube Region Gas Market Model consists of the following building blocks: (1) local demand; (2) local supply; (3) gas storages; (4) external markets and supply sources; (5) cross-border pipeline connections; (6) TOP contracts; and (7) spot trading. We will describe each of them in detail below.

Local demand • We derive a set of local demand functions, each representing the aggregate demand of each modelled country for each month. Local demand functions are downward sloping, meaning that higher prices decrease the amount of gas that consumers want to use in a given period. For simplicity, we use a linear functional form, the consequence of which is that every time the market price increases by 0.1 €/MWh, local monthly consumption is reduced by an equal quantity (as opposed to equal percentages, for example).[11]

The linearity and price responsiveness of local demand ensures that market clearing prices will always exist in the model. Regardless of how little supply there is in a local market, there will be a high enough price so that the quantity demanded will fall back to the level of quantity supplied, achieving market equilibrium.

Local supply • Local supply shows the relationship between the local market price and the amount of gas that local producers are willing to pump into the system at that price.

In the model, each supply unit (company, field, or even well) has a constant marginal cost of production (measured in €/MWh). Supply units operate between minimum and maximum production constraints in each month, with the constraints

[10] The start of the modelling year can be set to any other month.

[11] For the construction of the demand functions we use 2011 consumption data taken from the Eurostat and EnC database. Given the lack of reliable data, the reference price is uniformly assumed to be 28 EUR/MWh across all countries and periods. We assume average retail price elasticity of -0.1 for gas consumers, to calibrate the linear demand curves on each local market.

being independent across months.[12] Therefore, production decisions in October, for example, have no direct effect on production possibilities in any other month.

Any number of supply units can be defined for each month and each local market. As a result, local supply will be represented by an increasing step-function, for which the number and size of steps can be chosen freely.

Gas storages • Gas storages are capable of storing natural gas from one period to another, arbitraging away large market price differences across periods. Their effect on the system's supply-demand balance can be positive or negative, depending on whether gas is withdrawn from, or injected into, the storage. Each local market can contain any number of storage units (companies or fields).

Storage units have a constant marginal cost of injection and (separately) of withdrawal. In each month, there are upper limits on total injections and total withdrawals. There is no specific working gas fee, but the model contains a real interest rate for discounting the periods, which automatically ensures that foregone interest costs on working gas inventories are taken into account.

There are three additional constraints on storage operation: (1) working gas capacity; (2) starting inventory level; and (3) year-end inventory level. Injections and withdrawals must be such during the year that the working gas capacity is never exceeded, intra-year inventory levels never drop below zero, and year-end inventory levels are met.

External markets and supply sources • Explicitly modelled local markets are limited to the countries of the CSEE region (including the DR), but their gas sectors are by no means closed to the outside world. There are comparatively large external markets and supply sources neighbouring the region, which can serve as import sources (e.g. Russia, LNG markets), export destinations, or both (e.g. Germany, Italy).

Prices for external markets and supply sources are set exogenously (i.e. as input data) for each month, and they are not assumed to be influenced by any supply-demand development in the local markets. As a consequence, the price levels set for outside markets are important determinants of their trading direction with the CSEE region. When prices are set relatively low, CSEE countries are more likely to import from the outside markets, and vice versa.

Cross-border pipeline connections • Any two markets (local or outside) can be connected by any number of pipelines, which allow the transportation of natural gas from one market to the other. Connections between geographically non-neighbouring countries are also possible, which corresponds to the presence of dedicated transit pipelines.

12 Minimum production levels can be set to zero. If minimum levels are set too high, a market clearing equilibrium may require negative prices, but this practically never happens with realistic input data.

Cross-border pipelines are unidirectional, but physical reverse flow can easily be allowed for by adding a parallel connection that "points" in the other direction. Each pipeline has a minimum and a maximum monthly transmission capacity, as well as a proportional transmission fee.

Virtual reverse flow ("backhaul") on unidirectional pipelines can also be allowed, or forbidden, separately for each connection and each month. The rationale for virtual reverse flow is the possibility to trade "against" the delivery of long-term TOP contracts, by exploiting the fact that reducing a pre-arranged gas flow in the physical direction is the same commercial transaction as selling gas in the reverse direction.

We disregard from modelling the internal gas transmission systems of local and external markets.

Long-term take-or-pay (TOP) contracts • A TOP contract is an agreement between an outside supply source and a local market concerning the delivery of natural gas into the latter. The structure of a TOP contract is the following.

Each contract has monthly and annual minimum and maximum quantities, a delivery price, and a monthly proportional TOP-violation penalty. Maximum quantities (monthly or annual) cannot be breached, and neither can the annual minimum quantity. Deliveries can be reduced below the monthly minimum, in which case the monthly proportional TOP-violation penalty must be paid for the gas that was not delivered.

Any number of TOP-contracts can be in force between any two source and destination markets. Monthly TOP-limits, prices, and penalties can be changed from one month to the next.

The delivery routes (the set of pipelines from source to destination) must be specified as input data for each contract. It is possible to divide the delivered quantities among several parallel routes in pre-determined proportions, and routes can also be changed from one month to the next.

Spot trading • The final building block, spot trade, serves to arbitrage price differences across markets that are connected with a pipeline. Typically, if the price on the source-side of the pipeline exceeds the price on the destination-side by more than the proportional transmission fee, then spot trading will occur towards the high-priced market. Spot trading continues until either (1) the price difference drops to the level of the transmission fee, or (2) the physical capacity of the pipeline is reached.

Physical flows on a pipeline equal the sum of long-term deliveries and spot trading. When virtual reverse flow is allowed, spot trading can become "negative" (backhaul), meaning that transactions go against the predominant contractual flow. Of course, backhaul can never exceed the contractual flow on a pipeline.

Equilibrium

The DRGMM algorithm reads the input data and searches for the simultaneous supply-demand equilibrium (including storage stock changes and net imports) of all fifteen local markets in all months, respecting all the constraints detailed above.

In short, the equilibrium state (the "result") of the model can be described by a simple no-arbitrage condition across space and time. However, it is instructive to spell out this condition in terms of the behaviour of market participants: consumers, producers and traders.[13]

Local consumers decide about gas utilization based on the market price. This decision is governed entirely by the local demand functions we introduced earlier.

Local producers decide about their gas production level in the following way: if market prices in their country of operation are higher than unit production costs, then they produce gas at full capacity. If prices fall below costs, then production is cut back to the minimum level (possibly zero). Finally, if prices and costs are exactly equal, then producers choose some amount between the minimum and maximum levels, which is actually determined in a way to match the local demand for gas in that month.

Traders in the model are the ones performing the most complex optimization procedures. First, they decide about long-term contract deliveries in each month, based on contractual constraints (prices, TOP quantities, penalties) and local supply-demand conditions.

Second, traders also utilize storages to arbitrage price differences across months. For example, if market prices in January are relatively high, then they withdraw gas from storage in January and inject it back in a later month in such a way as to maximize the difference between the selling and the buying price. As long as there is available withdrawal, injection and working gas capacity, as well as price differences between months exceeding the sum of injection costs, withdrawal costs, and the foregone interest, the arbitrage opportunity will be present and traders will exploit it.[14,15]

Finally, traders also perform spot transactions, based on prices in each local and outside market and the available cross-border transmission capacities to and from those markets, including countries such as Russia, Germany, Italy, Turkey, or LNG markets, which are not explicitly included in the supply-demand equalization.

[13] When assessing welfare effects, we omit storage operators, since injection and withdrawal fees are set exogenously, and stock changes are determined by traders.

[14] Traders also have to make sure that storages are filled up to their pre-specified closing level at the end of the year, since we do not allow for year-to-year stock changes in the model.

[15] A similar inter-temporal arbitrage can also be performed in markets without available storage capacity, as long as there are direct or indirect cross-border links to countries with gas storage capability. In this sense, flexibility services are truly international in the simulation.

SIMULATION RESULTS

This chapter presents an application of the DRGMM to assess the likely impact of all known gas infrastructure development project proposals on regional gas market integration in the DR. The types of projects we analyse are within-region pipelines (interconnectors, including reverse flow projects), underground storage sites, LNG terminals and new international long distance pipelines providing new sources of gas supply for the DR.

For this purpose we create and run a reference scenario with 2011 input data and additional assumptions discussed below. Next we add, one by one, the proposed projects to the reference case infrastructure *ceteris paribus* and compare model outcomes to the reference case. By this we ask how the outcome of regional gas trading and infrastructure operations would differ from the 2011 reference case if that piece of infrastructure was already in place. When adding new infrastructures to the reference case, we disregard from the cost and timing of infrastructure investment, so they get into the model '*overnight*' and do not change the tariffs paid by infrastructure users for transmission, storage or LNG terminal services. However, for the purpose of the cost-benefit analysis we collected available project related investment cost data.

After analysing individual projects one by one, we repeat the same procedure for project packages like the proposed project list of the north-south gas working group. Finally, the likely impact of new long distance pipelines on the regional gas market is assessed in the context of a 2020 reference scenario.

Input data

Table 1 contains the dimension and sources of technical input data used for the simulations. In order to create the 2011 reference scenario, we used estimated data when 2011 data was still not available (e.g. consumption data due to delayed publication).

For the 2020 annual consumption and production forecast we rely on a critical review of the forecasts of institutions listed in *Table 1.* The monthly distribution of gas consumption for the analysed countries was estimated using historic data (see *Figure 2*).

The pipeline infrastructure of the region for the 2011 reference scenario is depicted on *Figure 3.*

Finally, in order to run the model, we also have to assume TOP and spot prices for external markets and tariffs paid by infrastructure users for transmission and storage (injection and withdrawal).

Table 2 contains external gas product prices we use for simulation purposes in this paper. With regard to TOP contracts we assume a mixed pricing regime with a 20% weight for spot and 80% weight for oil indexed pricing. This is to reflect the

TABLE 1 • **Summary of input data structure and sources**

Data		Source	
Category	Unit	Actual data	Forecast / Planned
Consumption	Annual Quantity (bcm) Monthly distribution (% of annual quantity)	Eurostat, EnC data	N-S study, EnC data, Eurostat, ENTSO-G, own estimation
Production	Minimum and maximum production (mcm/day)	EUROSTAT, EnC data	N-S countries: N-S study, EnC data, ENTSO-G GRIPs, TYNDP
Infrastructure			
Pipeline	daily maximum flow	ENTSO-G, EnC	TSOs, N-S action plan, TYNDP, GRIPs, EnC
Storage	Injection (mcm/day), withdrawal (mcm/day), working gas capacity (mcm)	GSE	
LNG	Capacity (mcm/day)	GLE	
TOP contracts	Yearly minimum maximum quantity (mcm/year) Seasonal minimum and maximum quantity (mcm/day),	Gazprom, National Regulators Annual reports, Platts	

EnC: Energy Community Regional Energy Strategy Task Force data; N-S Study: *Kantor Management Consultants* [2012].

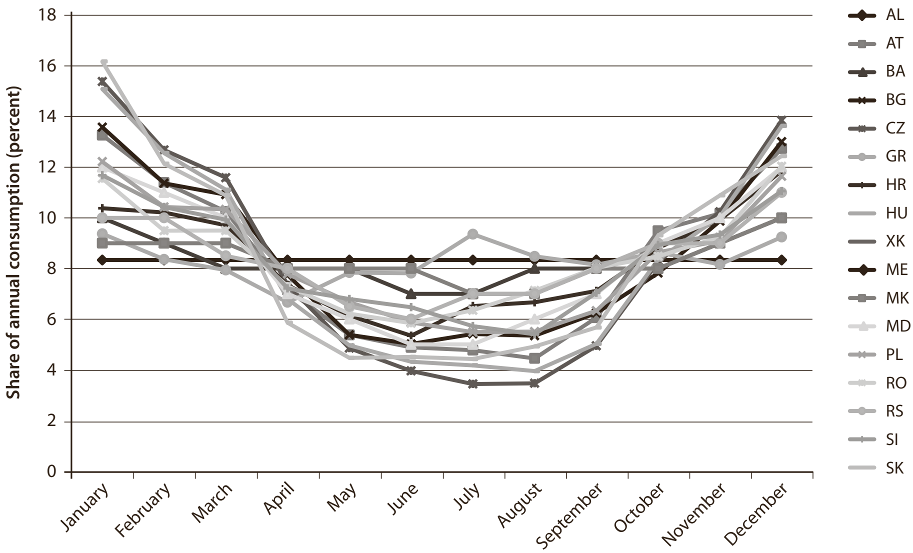
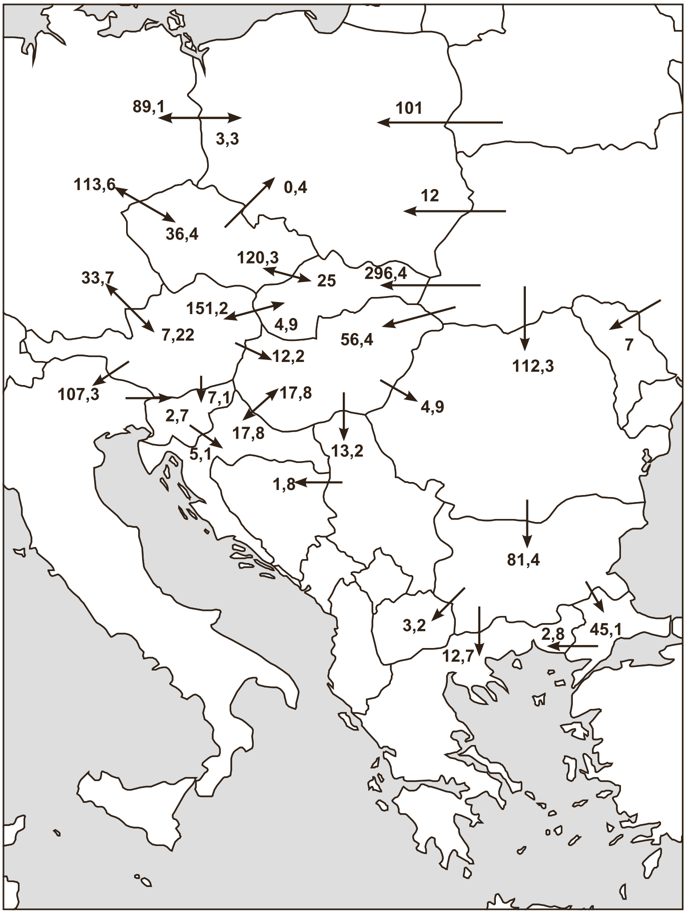


FIGURE 2 • **Estimated monthly distribution of consumption in the modelled countries** (% of annual consumption)

European gas industry experience of a significant renegotiation of e.g. Russian TOP contracts in recent years of economic crisis.[16] The assumed tolerance for TOP annual contracted quantity is ±15%. For the simulations with the 2020 reference scenario, we assume the renewal of the long term contracts expiring between 2011 and 2020, but also assume a 20% decrease in their annual contracted quantity.

[16] Note however that we assume no active *pricing behaviour* on external markets.



Arrows show the possible physical flow direction and the daily maximum capacity

FIGURE 3 • **Interconnector topology used for the 2011 reference scenario** (mcm/day)

TABLE 2 • **External market price assumptions (€/MWh)**

Market	Price in €/MWh
Western Europe (TTF spot)	24.2
Russia	34.2
Italy (PSV spot)	28.0
Turkey	31.6
LNG	24.2
LNG (Bulgaria, Romania)	31.6

Transit contracts are taken into consideration only if they use the infrastructure of the DR. In the case of Germany and France we assume 50% of their Russian imports will come through Nord Stream from 2013 on, thus 2020 flows are reduced accordingly. Furthermore, in case of Germany we assume that 50% of the transit requirements pass through the Yamal pipeline. For Turkey, we take into account only those Russian import contracts that are transmitted through Romania and Bulgaria. For Italy, Russian contracts go through Slovakia and Austria.

We do not have a realistic representation of local market transmission tariffs for the DRGMM, so we set them close to zero in this paper. We think that disregarding the transmission tariffs will not distort our conclusions because the unit transmission cost for a MWh of gas is negligible compared to its product price.[17] Another argument is that although significant differences in transmission tariffs across the region might distort cross-border arbitrage opportunities, including the utilization of gas storage assets, with the advancement of an EU wide gas market regulation and integration transmission tariffs are expected to level out for the region. Nevertheless, this is a point for further model development.

Data on gas storage tariffs (injection and withdrawal fees) were gathered from storage owners or national energy regulators. Besides direct storage costs, we also account for the foregone interest costs on holding working gas inventories. The real interest rate for calculating the interest costs of gas inventories is set at 5 percent.

Market integration measures

The first set of our project related analyses deals with the impact of having new interconnectors or LNG stations in the region on regional gas market integration.[18] Since market integration is a multi-dimension concept and difficult to measure *per se*, we have developed variations of a simple measure of market integration. Our *Regional Cost Convergence Index* (RCCI) is based on the assumption that an advance in market integration results in price convergence across the countries concerned *and* towards cheaper gas supply sources. In the 'Danube Region 2011' reference context this means that a new piece of gas infrastructure improves market integration when it results in reduced regional gas purchase costs through driving close-to-oil-indexed local prices down, closer to continental spot price levels.

[17] REKK has recently carried out a survey of gas transmission tariffs for a 80 MW gas fired power plant for 10 of the modelled countries and found a € 1.87/MWh average value for this group. This is 5.5% of the oil indexed and 7.7% of the German spot price we use in this study.

[18] A positive impact on gas market integration is the number one criteria a PCI should meet according to Article 4 of the proposed infrastructure Regulation.

Formally,

$$RCCI = \frac{\sum p_i \cdot q_i}{p_{\text{spot}} \cdot Q} - 1 \text{, where}$$

- i is an index for the DR countries, $i = 1, \ldots, k$;
- p_i is the annual weighted average gas price on local market i, calculated by the model;
- q_i is the annual gas consumption on local market i, calculated by the model;
- Q is the amount of DR gas consumption (sum of q_i over k), calculated by the model;
- p_{spot} is the continental spot price (TTF price).

The meaning of RCCI is the excess gas purchase cost (in per cent) the DR pays for its gas consumption over the case when it purchased the same amount at a continental spot price. The value of this excess cost for the 2011 reference scenario, measured by the RCCI, is 21.5%. *Figure 4* shows modelled 2011 reference scenario

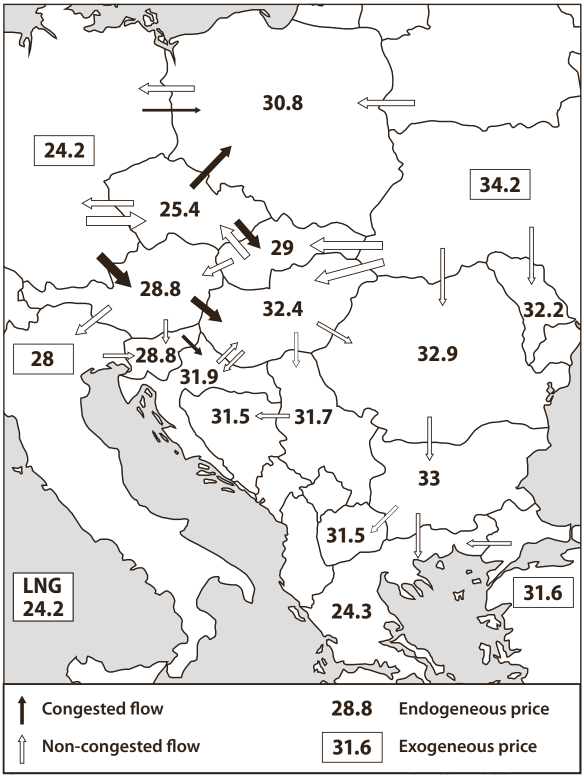


FIGURE 4 • **Reference scenario: 2011 current infrastructure**
($RCCI_{ref} = 21.5\%$)

local prices (€/MWh), and trade flows (arrows), assuming external market prices (included in the white boxes). White arrows stand for non-congested and grey for congested interconnections.

Analysis of individual projects case by case

First we added to the 2011 reference scenario all the proposed gas infrastructure projects (and then removed them) one by one (everything else unchanged) and calculated the RCCI for these simulations. No single gas storage project had a significant regional market integration impact. *Table 3* contains the investigated pipeline and *Table 4* the LNG projects in the order of increasing RCCI values. Those projects with lower RCCI save more gas purchase cost *for the region* – and not necessarily only for the countries directly involved in the project – than those with higher values. In the 2011 reference scenario, consumers of the Danube Region pay 4700 million € more than what they would pay for their consumption on a Western European spot market price.

TABLE 3 • **Individual pipeline project ranking by RCCI** ($RCCIr_{ef}$ = 21.51%)

Pipeline	*RCCI*	Pipeline	*RCCI*
CZ–PL2	17.10	PL–SK	21.51
SK–HU	18.35	BG–RO	21.51
GR–BG	21.13	PL–CZ	21.51
TR–BG	21.29	HR–IT	21.51
RS–BG	21.39	MK–GR	21.51
RS–RO	21.42	HR–HU2	21.51
RO–MD	21.47	RS–MK	21.51
BA–RS	21.50	RS–HR	21.52
MK–AL	21.51	BA–HR	21.52
HR–RS	21.51	MK–RS	21.55
HR–BA	21.51	RO–HU	21.56
HU–SK	21.51	BG–RS	21.56
MK–XK	21.51	RO–RS	21.56
AT–CZ	21.51	MK–BG	21.56
HR–SI	21.51	MD–RO	21.57
RS–BA2	21.51	SI–HU	21.67
HU–SI	21.51		

TABLE 4 • **Individual LNG project ranking by RCCI** ($RCCIr_{ef}$ = 21.51%)

	RCCI
LNG–PL	16.94
LNG–PL2	17.04
LNG–HR	20.03
LNG2–RO	20.40
LNG2–BG	21.29
LNG–GR2	21.51

We could identify seven pipeline and five LNG projects which, by themselves only, can have a significant and advantageous regional impact on gas prices and purchase costs. While the rest of the pipeline projects do not have a significant regional impact individually, we found some that result in increased RCCI values (that is, increasing gas purchase cost for the region). The latter results might seem counter-intuitive at first sight, but they are actually consistent with the workings of the market. The market equilibrium maximizes total welfare, i.e. the sum of welfare of *all* groups of market players (i.e. consumers, producers, storage and interconnector operators etc). Therefore while the addition of a new infrastructure element will never decrease short-term social welfare, it may well result in a welfare loss for one or more groups of market players.

According to RCCI, the best ranking pipeline project for the region is an upgrade of the Czech-Polish interconnector from its present 0.4 mcm/day to 8.6 mcm/day capacity. A new Slovak-Hungarian interconnector ranks second. Then come three projects helping to decrease very high Bulgarian prices and finally an interconnection from Romania to Moldova. The best ranking LNG projects are the Polish and the Croatian ones.

However, project ranking by RCCI alone might be misleading from a regional perspective, since it is neutral with regard to the distribution of price changes and cost savings across the countries. Impacts of some projects might be limited to the involved countries alone while others might provide a regionally more widespread effect. Our *Regional Spill-over Index* (RSoI) measures by how much the addition of a new piece of infrastructure will change the 2011 reference RCCI when we exclude the countries directly affected by the new project[19] from the RCCI calculation. *Table 5* contains the results for those interconnector projects that produce part of their cost reduction effects beyond the borders of the project countries.

TABLE 5 • **The reduction of regional gas purchase costs by individual pipeline projects in countries beyond the borders of the project countries (%)**

Pipeline project	Reduction (%)
SK–HU	1.59
GR–BG	0.51
RS–BG	0.11
MD–RO	0.02
TR–BG	0.01

We can see that the impacts of two of the top ranking projects by RCCI, the Czech-Poland and the Serbia-Romania interconnectors (see *Table 3*) are strictly limited to the involved countries. In contrast, the majority of benefits are generated beyond the borders of the project countries in the case of the Slovakia-Hungary and Greece-Bulgaria projects. We can identify similar differences in the case of

[19] One country in the case of LNG and two in the case of a new interconnection

LNG projects. The impact of a Polish LNG receiving terminal, without additional cross border pipelines put in place, is strictly limited to Poland itself by bringing gas wholesale prices sharply down at home. At the same time a Croatian LNG project could bring decreasing prices and purchase costs not only for Croatia but for Slovenia, Hungary, Serbia and Bosnia and Herzegovina without any additional infrastructure to be built.

Analysis of project packages: the North-South gas corridor

The DRGMM model can also be used to carry out similar analysis for project package impacts – instead of individual projects – on regional gas market integration. In recent years a number of proposals have been put forward to develop a set of infrastructure projects to improve gas market interconnectivity of the DR. The two prominent ones were the New Europe Transmission System (NETS) project (a European priority project under the EU's TEN-E program) and the recently developed North-South gas corridor for Central and Southeast Europe. Since the present status of the NETS project does not allow for tracking it down to a specific set of infrastructure projects, we remained with the North-South corridor project list that was published by the Commission in December 2011 (*EC* [2011).

Adding the 17 projects of the North-South corridor to the 2011 reference case brings down the RCCI index to 6.8% from the 25.1% reference number. This translates into an annual gas purchase cost saving of €2827 million for the DR (see *Figure 5*).

A few notes are worth to make here. First, all countries except for the Czech Republic seem to enjoy a significant drop in wholesale gas prices in the modelled countries. The implementation of the entire project seems to bring the Western part of the region very close to the German / Italian markets and the South-Eastern part to the Greek one. Four LNG terminals provide significant new supply sources for the region.

Second, the empty black circles on *Figure 5* stand for projects that are built but not utilized by market participants, according to the model. An interesting issue for future investigation is how the package could be reduced so that we preserve the rest of its benefits for the region. This needs a careful analysis since regional trading exhibits fairly strange patterns under the presence of significant TOP obligations and abundant spot trading opportunities supported by a robust infrastructure and new LNG supply sources. We can observe several trade flows from high to low priced countries (e.g. Bulgaria exporting to Greece or Hungary exporting to Serbia) or a lack of trade between countries with a meaningful price differential (e.g. an empty pipeline between Austria and the Czech Republic).

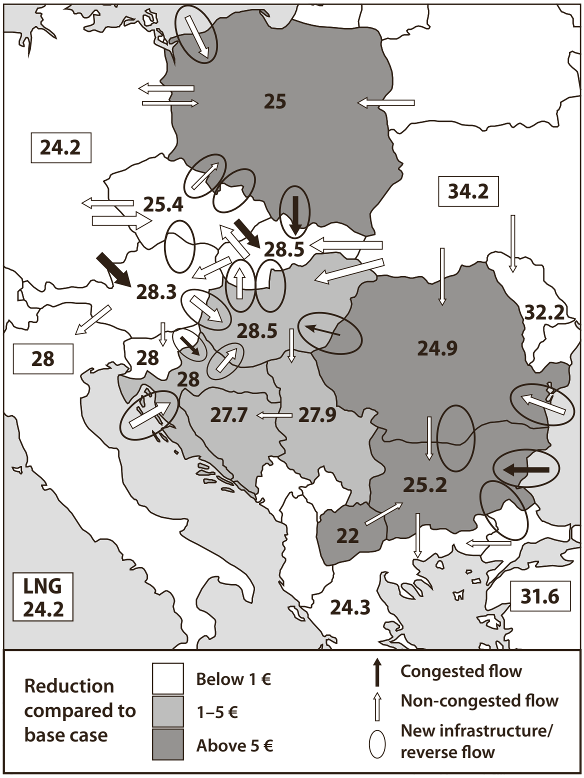


FIGURE 5 • **The impacts of the North-South corridor** ($RCCI_{ref}$ = 21,5%)

The impacts of new international gas pipelines entering the region

Up to now we have investigated the impacts of intra-regional projects and project packages on market integration. However, in recent years discussions about how to increase the gas supply source diversification of the DR have centred on the South Corridor gas pipeline project alternatives (Nabucco, Nabucco West, South Stream, TAP, etc). Now we have a look at the way we can analyse the potential impacts of new pipeline supply sources entering the DR by the model.

For this analysis we first create a 2020 reference scenario. Compared to the 2011 reference case, three major changes are made to the model. First, those – and only those – new infrastructures that are under construction in 2011 are added to the reference case. Second, load data is modified according to best available 2020 forecasts. Third, we assume that those TOP contracts expiring between 2011 and 2020 will all be signed again but at a reduced rate of annual contracted capacity (80% of the former contract). External price assumptions are unchanged compared to the 2011 reference scenario. The RCCI index for the 2020 reference case is up at 29.9%

New pipelines are represented in a schematic way. We assume new gas entering the region under a TOP regime. TOP is priced at Russian price minus 5%, the Russian TOP price being a mix of 20% spot and 80% oil indexed regime. We compare the impacts of two pipeline business models under two different intra-regional network configuration alternatives (four cases). The first pipeline brings 10 Bcm to the Turkish-Bulgarian border and then ships all of it to the Austrian hub of Baumgarten through Bulgaria, Romania and Hungary. Spot trading of this gas is then allowed (Project 1). This pipeline business model considers the DR as primarily a transit area. Alternatively, Project 2 brings 10 BCM to the Turkish-Bulgarian border again but then part of the gas is distributed along its paths more evenly: 1 Bcm for the Bulgarian and Romanian markets each, 2 Bcm for the Hungarian market and the remaining 6 Bcm ends up at Baumgarten. Sufficient additional pipeline capacities are assumed to bring these amounts to the affected markets. We estimate the impacts of Projects 1 and 2 on RCCI assuming either that the North-South corridor projects are completed or not completed. The corresponding RCCI figures are summarised in *Table 6.*

TABLE 6 • **The impacts of alternative 10 Bcm South Corridor projects on RCCI under alternative intra-regional network topology**

	With North-South package	Without North-South package
2020 base scenario	19.16%	29.86%
V1 (10 bcm TOP to AT)	16.89%	29.54%
V2 (10 bcm distributed along the route)	16.73%	27.38%

The conclusion of our analyses on this issue is that the bulk of the improvement in RCCI is due to improving intra-regional interconnectivity and adding LNG sources to the DR – that is implementing the North-South corridor projects. While a regionally more diversified pipeline business model performs slightly better for the region compared to the transit-like model, the positive impact due to the pipeline is of secondary importance.

Allowing virtual reverse flow (backhaul) transactions on EU–EU borders of major transit pipelines

Because of the apparent counter-incentives of transit pipeline owners, we have so far disregarded backhaul transactions on all transit pipelines, shipping Russian gas to Western and South Europe crossing the DR. However, one could with good reason argue[20] that, instead of building new infrastructures, a better and bi-directional commercial utilisation of major existing physical infrastructures could significantly help the integration of the DR and West European gas markets.

[20] The authors thank Pierre Noel for raising their attention to this point.

In order to estimate the potential impact of backhaul transactions on the Region's gas purchase costs we allowed for virtual reverse flow transactions to happen at all EU-EU borders – including Croatia[21] – along the transit pipelines. However, no backhaul transaction is allowed at EU-third country borders (EU–RU, EU–TR and EU–EnC[22]).

Table 7 contains the results of our simulations.

TABLE 7 • **The impacts of virtual reverse flow options on RCCI**

Backhaul option	Not allowed (base case)	Allowed on all EU–EU borders	Annual savings on regions gas bill compared to base case
2011	21.51%	17.20%	823 million €
2020 base scenario	29.86%	25.13%	1181 million €

We can translate the figures in *Table 7* so that allowing for the indicated backhaul transactions could help to save € 823 – 1181 million annually in gas purchase cost for the region with only changing the rules of commercial transactions along major transit pipelines.

Using the model for cost-benefit analysis: an illustration

Up to this point we have disregarded project related costs and concentrated merely on market integration and price impacts of projects and project packages. However, the combination of project related investment cost data and the various according benefit figures that the model produces allows us to develop more economically meaningful measures to evaluate and rank projects than the RCCI or the RSoI. Since the availability of investment cost data for future natural gas infrastructure projects are very limited,[23] we often used international benchmarks for this purpose. In this regard the following analyses can only be regarded as preliminary and illustrative.

First we calculate a *regional payback period* for the projects by dividing the project related investment cost with the estimated annual purchase cost decrease a project brings for the entire region. *Table 8* contains the results of the calculations and also compares project rankings by RCCI *versus* the payback period.

[21] Croatia will be a member of the EU from 01.07.2013

[22] Allowing backhaul transactions on the EU–EnC borders does not change the picture much, RCCI would be 25,01%

[23] Investment costs are from the project home pages and investors in case of pipelines, and are benchmarked in case of LNG.

TABLE 8 • **Individual project ranking by RCCI and regional payback period** ($RCCIr_{ef}$ = 21.51%)

Project	RCCI (%)	Annual saving on gas bill (million €)	Estimated investment cost (million €)	Pay-back period (year)
Individual pipeline				
CZ–PL2	17.10%	841.75	28	0.03
SK–HU	18.35%	598.51	150	0.25
GR–BG	21.13%	73.49	160	2.18
TR–BG	21.29%	41.77	75	1.80
RS–BG	21.39%	22.98	95	4.13
RO–MD	21.47%	7.73	50	6.46
LNG				
LNG–PL	16.94%	872.30	470	0.54
LNG–HR	20.03%	281.39	240	0.85
LNG2–RO	20.40%	205.51	470	2.29
LNG2–BG	21.29%	41.77	470	11.25

The results indicate that four of the best pipeline projects could break even *for the region* within just 3 years, with the two best (CZ–PL2 and SK–HU) in just a few months. The regional payback period for the Polish and Croatian LNG projects is also below one year. Interestingly, the ranking by regional payback period changes the RCCI ranking only slightly: the TR-BG project takes over the GR–BG pipeline.

The apparent question arises why these projects are not built – or proceed only very slowly – if their economics is so extremely good? Part of the answer to this question relates to the non-internalized positive network externalities of new interconnectors that are built and operated under the present regulated third party access remuneration scheme. The revenue of such a new interconnector is based on regulated transmission tariffs based on the investment and operation costs of the pipeline company. These costs are typically shared and paid by the consumers of those member states directly involved in the project through regulated transmission tariffs. However, a new pipeline might imply more dispersed additional costs and benefits for producers and consumers across a wider geographic area.[24]

We can illustrate this point by simulating the likely impacts of building one of the top ranking projects, the Greece-Bulgaria interconnector. This project ranks third by RCCI and its estimated cost is € 160 million. By adding this interconnector to the 2011 reference case, we can identify ten countries where the new line implies a measurable change in annual weighted average wholesale gas prices and thus induces changes in social welfare. *Table 9* summarizes the results of this simulation.

[24] Part of the benefits could be captured by tendering pipeline capacity through e.g. an open season procedure.

TABLE 9 • **Changes in welfare measures due to a new Greece–Bulgaria interconnector** (million €)

	Net consumer surplus	Producer surplus	Storage operation profit	Net profit from long-term contracts	TSO auction revenues	Total social welfare
GR	–76.8	41.0	0.0	43.9	114.9	122.9
BG	60.3	–8.2	0.0	–46.7	103.8	109.2
RO	94.8	–98.8	0.0	–24.5	–7.5	–35.9
HU	1.7	–0.4	0.0	–1.2	–7.6	–7.5
MK	3.2	0.0	0.0	–2.6	0.0	0.6
SI	0.0	0.0	0.0	0.0	–0.1	–0.1
AT	0.0	0.0	0.0	0.0	–0.1	–0.1
HR	0.6	–0.4	0.0	–0.2	–0.1	0.0
RS	0.3	–0.1	0.0	–0.2	0.0	0.0
BA	0.1	0.0	0.0	0.0	0.0	0.0

Aggregate welfare improvement is up at €190 million annually, Greece and Bulgaria being the most significant beneficiaries. In the meantime, Romania and Hungary suffer sizeable welfare losses. With regard to market players, TSOs and consumers are the beneficiaries of the project while, on aggregate, DR gas producers and TOP contract holders suffer losses from it. This is due to excess demand for the new pipeline capacity, resulting in significant congestion revenues for the participating TSOs and the gas price decrease that the pipeline investment carries for Bulgaria, Romania and Macedonia, respectively. Gas price decrease is, on the other hand, bad news for local producers and TOP gas holders. Clearly, TOP gas is crowded out by cheaper Greek LNG sources leaving TOP holders with a significant loss in all countries except for Greece. Since gas prices are to increase in Greece compared to the reference case (cheaper LNG flowing now to the North), consumers are suffering a significant welfare loss in this country, while producers and TOP traders are on the winning side.

We think that model simulations of this kind might help structure the disputes around new gas infrastructure projects for the DR by identifying the distributional impacts of them. Within the context of the EU, ACER could potentially make use of such results in preparing for its decisions on cross border investment cost allocation (see Article 13 of the proposed infrastructure Regulation).

Using the model for supply security analysis: another illustration

The DRGMM model can also support sophisticated gas supply security analyses at the regional level. In this section, we will outline the mechanisms behind such analyses with the help of a short exercise.

As we have noted before, the DRGMM model uses a fully dynamic solution algorithm over 12 consecutive months, in which we assume that traders optimizing the use of storage assets and the requested delivery of TOP contracts act with perfect foresight regarding the gas year. As a result, the model produces monthly forward prices for the entire year, which are "right on the spot" in the sense that if there are no subsequent changes in the input data, then all the outcomes (including prices) will turn out as predicted as the year unfolds.

Of course, in reality, supply and demand conditions do deviate from forecasts throughout the year. To capture this fact, the model allows for the possibility of intra-year runs, in which any input variable pertaining to the upcoming months can be changed. The following example might be illuminating.

Suppose that the gas year runs from April to March. In the initial model run, we have to include forecasts for supply-demand conditions in each of the 12 months, otherwise it would be impossible to take optimal storage and contract delivery decisions in the beginning of the year.[25] Taking the forecast as given, we can then calculate how each of the 12 months will "play out".

Now let us suppose that a supply disruption occurs in January. For the sake of the example, let it be another gas dispute between Russia and Ukraine, the consequence of which is zero transits through the latter for the whole of January. In the model, we would represent this incident by setting the maximum transport capacity of the pipelines through Ukraine (to Romania, Moldova, Hungary, and Slovakia) to zero for a month.

An important question is, in which month do market participants know that interconnectors crossing Ukraine will be unavailable in January? If they already know it in April, they will likely have enough time to stock up gas to weather the crisis better. But if it takes them by surprise, the price effects will be much more severe.[26] One can therefore imagine that the actual effects will be highly dependent on the length of time that is available for preparation.

Fortunately, the DRGMM model allows for a full exploration of these issues. Taking the start-of-year run as a reference of how market events occur naturally, it is possible to "stop" the year in any month (e.g. just before January), re-set the input parameters of the model for the rest of the year (e.g. interconnector capacities in January, and probably also the yearly TOP minimum constraints), and re-run the optimization procedure taking the outcomes of the past months (storage utilization from April to December, for example) as already given. The model results will then reflect the consequences of regional market-based responses to the supply shock, including the spillover effects on countries not directly affected by the shut-down of the pipeline (e.g. Serbia or Bulgaria, in this case).

[25] The key decision variables here are those with inter-temporal consequences.

[26] Since the model employs market mechanisms only, negative supply shocks will present themselves as price jumps in the affected areas.

Figure 6 shows the results of the crisis situation that we outlined above. The coloring of the markets indicates the extent of the price rise in January, or equivalently, the seriousness of the supply disruption if the market equilibrium is restored via mandatory consumption cuts.

Light grey colored markets experience a price rise of about 5-10 €/MWh for the crisis month, whereas the dark grey colors indicate a price rise beyond 100 €/MWh. As the actual numbers show, the supply disruption is quite severe in the Eastern part of the Balkans, whereas it seems to be more manageable in Hungary, Serbia, and Bosnia and Herzegovina. Interestingly, the Czech Republic is also affected through the decrease in SK → CZ pipeline flows.[27]

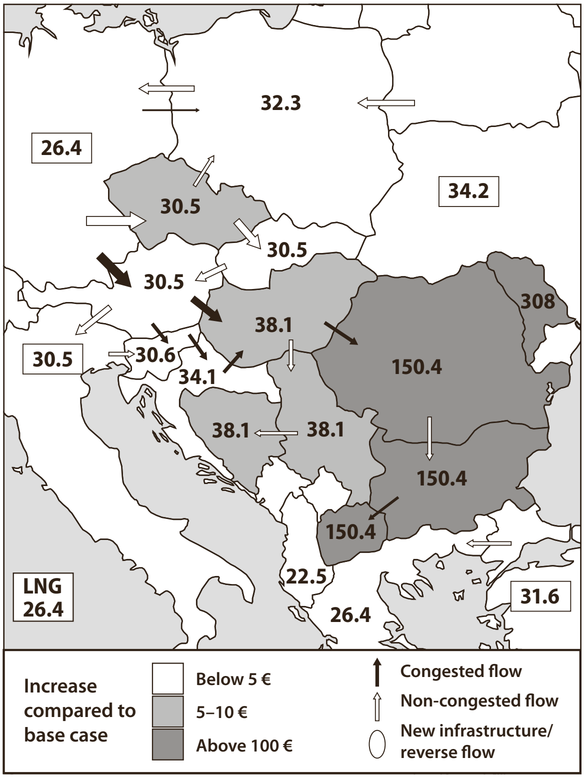


FIGURE 6 • **The effect of an unexpected supply disruption of all pipelines through Ukraine in January** (prices in €/MWh)

[27] The same crisis situation turns out to be almost fully manageable (except in Moldova) when market players start preparing for it in April, instead of only reacting to the events as they take place in January.

As a final point, we note that the regional (and country-level) supply security effects of various policies and new infrastructure elements can also be assessed using this methodology. One would simply carry out the supply shock analysis, such as the one above, with and without the policy or the new infrastructure, and compare the outcomes.

MODEL LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

The DRGMM is a unique analytical tool that represents the natural gas industries of Central and South East Europe in a detailed and consistent manner. In this paper we have described the assumptions and operation of the model and presented several simulation cases to illustrate the variety of analyses the model allows for, including market integration, cost-benefit and security of supply analyses. However, the model needs further development to cope with its present limitations.

The first of these limitations is the model geography. At present only those 15 countries indicated in *Figure 1* are represented in detail as 'local' markets in the model, while part of the EU has to be represented as 'external' markets. An extension of the model to present 'external' EU markets could result in a detailed representation of the entire interconnected EU natural gas wholesale market.

Second, the model lacks a sensible representation of the EU's outside suppliers' pricing behaviour. In its present form, the pricing of external markets to supply the DR is static: a static mix of oil product price and spot price indexation by Russia, related pricing by Turkey, spot pricing by Germany, Italy and LNG. Reality has provided several examples in the last four years when demand and supply shocks and the development of spot markets motivated a significant shift away from traditional oil indexation in natural gas pricing in Europe (*Stern–Rogers* [2011]). TOP contract pricing and quantitative characteristics have prohibited a remarkable flexibility under the pressure of market forces. We can also assume that a stronger internal and East-West integration of DR gas markets, promoted by a significant change in network topology in the DR could create a basis for more dynamic and market based gas pricing in the region compared to a present, very rigid oil indexation. Thus, developing a more realistic representation of outside supplier pricing behaviour is a key future model development task.

Third, the representation of gas transmission and storage access prices and pricing in the model requires refinement. This is made difficult by the lack of a consistent database and well-documented benchmarking of gas infrastructure access costs across Europe. Nevertheless, since the magnitude of transmission and storage access tariffs in comparison to product prices is marginal, we can argue that a more accurate and detailed representation of infrastructure access tariffs and rules are not likely to significantly change model results while they could overly complicate model algorithms.

Finally, one could argue that the representation of the DR gas market as being perfect competition under network and TOP contractual constraints is an unrealistic and, from the consumers' point of view, a too optimistic one. Beyond TOP constraints, national gas wholesale markets are often dominated by players with significant market power. The assumption of efficient utilization of cross border pipeline capacities can also be criticized on the basis of existing capacity allocation rules being fairly distant from market based mechanisms (see *REKK* [2011 on a Hungarian example). Nevertheless, the world represented by the model is the vision the European Union, including its south-eastern region, is following when restructuring its gas industry. The model thus provides for a normative reference case in a European spirit and allows for evaluating the impacts of changes as well as distortions of a different sort to this baseline.

REFERENCES

Abada, I.–Briat, V.–Gabriel, S. A.–Massol, O. [2012]: A Generalized Nash-Cournot Model for the European Natural Gas Markets with a Fuel Substitution Demand Function: the GaMMES Model. Networks and Spatial Economics, accepted, April.

Boots, M. G.–Rijkers, F. A. M.–Hobbs, B. F. [2004]: Trading in the Downstream European Gas Market: A Successive Oligopoly Approach. The Energy Journal, Vol. 25, 3, pp. 73–102.

EC [2011]: Action plan for North-South energy interconnections in Central-Eastern Europe. Final. http://ec.europa.eu/energy/infrastructure/doc/2011_north_south_east_action_plan.pdf

Egging, R.–Gabriel, S.A.–Holz, F. [2010]: The World Gas Model: A multi-period mixed complementarity model for the global natural gas market. Energy, 35, pp. 4016–4029.

Egging, R.–Gabriel, S.A.–Holz, F.–Zhuang, J. [2008]: A Complementarity Model for the European Natural Gas Market. Energy Policy, 36. No. 7, pp. 2385–2414.

Gabriel S.A.–Kiet, S.–Zhuang, J. [2005*a*]: A mixed complementarity-based equilibrium model of natural gas markets. Operation Research, Vol. 53. No. 5, pp. 799–818.

Gabriel S.A.–Kiet, S.–Zhuang, J. [2005*b*]: A Large-scale complementarity model of the North American gas market. Energy Economics. 27, 639–665.

Holz, F.–Hirschhausen, C.–Kemfert, C. [2008]: A strategic model of European gas supply (GASMOD). Energy Economics, Vol. 30, No. 3, 766-–788.

Kaderják, P. (ed.) [2011*a*]: Security of energy supply in Central and South-East Europe. Corvinus University of Budapest, Regional Centre for Energy Policy Research, Budapest.

Kaderják, P. [2011*b*]: The Lessons of the January 2009 Gas Crisis in Central and Eastern Europe. In: *Vinois, J.* A. (ed.): The Security of Energy Supply in the European Union, Claeys and Casteels, pp. 193–219.

Kantoor Management Consultants [2012]: Market analysis and priorities for future development of the gas market and infrastructure in Central-Eastern Europe under the North-South Energy Interconnections initiative (Lot 2). Final Report submitted to the Directorate-General for Energy, Unit B1: Security of supply and networks, European Commission. 19 January.

Lise, W.–Hobbs, B. F. [2008]: Future evolution of the liberalised European gas market: Simulation results with a dynamic model. Energy, Vol. 33, No. 7, pp. 989–1004.

Lise, W.–Hobbs, B. F. [2009]: A dynamic simulation of market power in the liberalised European natural gas market. Energy Journal. 30, 119–136.

Mulder, M.–Zwart, G. [2006]: NATGAS: a model of the European natural gas market. CPB Memoranda 144, CPB Netherlands Bureau for Economic Policy Analysis.

Perner, J.–Seeliger, A. [2004]: Prospects of gas supplies to the European market until 2030. Results from the simulation model EUGAS. Utilities Policy, Vol. 12, No. 4, pp. 291–302.

REKK [2011]: The Baumgarten saga. In: Hungarian Energy Market Report, Vol. 3, No. 3, pp. 9–12.

Stern, J.–Rogers. H. [2011]: The transition to hub based gas pricing in continental Europe. The Oxford Institute for Energy Studies. NG 49. March.

• *Balázs Édes–Vanda Bölcskei–László Lőrincz–Péter Nagy–Zoltán Pápai* •

SUBSTITUTABILITY BETWEEN FIXED AND MOBILE TELEPHONE SERVICES

The present paper discusses the results of an empirical study on substitution between fixed and mobile telephone services in the Hungarian retail market, carried out in 2008 and 2009. The study used a survey designed for examining fixed-mobile substitution; it was carried out using the stated preference method on a sample of 1000 people in September 2008. The effects of price changes were studied both with regard to access (telephone subscription) and usage substitution (the respondent's last five calls). In addition to prices, the effects of lifestyle and demographic factors on demand were also studied, as were alternative substitution options and the "telecommunications budget" hypothesis. The results indicate that mobile telephone access demand has a rather low price elasticity (a value of less than –0.3 for a hypothetical price raise). Fixed-line demand, on the other hand, is elastic (–1.4). The rate of usage substitution is significantly lower in the short term (in the presence of existing subscriptions) than it is for longer-term access decisions.

INTRODUCTION

The present paper discusses the results of a study on substitution between fixed and mobile telephone services, carried out in 2008 and 2009. In this period, the share and importance of mobile data services was significantly lower in Hungary than after 2010; the effects of the arrival of smartphones emerged only after the research was completed. Therefore, the data collected on voice service substitution are only valid with regard to the time period and market conditions in question; as a result, the results cannot be directly transferred to a smartphone dominated market environment. Nevertheless, the study was carried out at a time that was very interesting with regard to the development of telecommunications markets, and the effects of the events of this period are still felt to a significant extent today. Therefore, we feel that this analysis provides useful information regarding the development of markets and helps to understand the interactions of fixed-line and mobile services.

In September 2008 there were 3,145,000 fixed telephony connections on the Hungarian market. 91.4% of these were traditional PSTN or ISDN lines, while the remaining 8.6% were IP on broadband lines provided by cable operators. At the same time, there were 11,771,000 mobile phone connections in the country, 61.4% of which were prepaid SIM cards. According to market research commissioned by the National Communication Authority of Hungary and carried out in fall 2008, 55.8% of households had a fixed phone, and 81.7% of families had at least one mobile phone. The use of fixed phones was already declining by this stage: the share of households with a phone had fallen by 17 percentage points compared to the 73% in 2002. Up to this period, mobile phone penetration was increasing, although the rate of growth was gradually slowing down.

The number and minute volume of fixed telephone calls had essentially fallen by half. The number of mobile calls increased by 76% in the same period, and their aggregate length tripled.

These trends, indicative of changes in the consumption of communication services, clearly show that substitution is taking place to some extent, but in themselves they do not prove that this process constitutes substitution in the technical economic sense.

An analysis of *fixed-mobile telephone service substitution* is of great importance for understanding market changes and with a view to regulation. It also influences the definition[1] of the market of fixed telephone services, which is being carried out based on EU recommendations. Whether fixed and mobile telephone services are part of the same market depends, among other things, on the extent of demand substitution between the two services. Defining the market is the first step in any regulatory process based on the analysis of the market in question. Whether a regulator identifies market power in the affected markets defined by the regulator greatly depends on the definition of the market.

In the Hungarian regulatory framework, fixed and mobile services were not treated as part of the same market before 2009. The separation of these markets was not based on a market definition and analysis procedure carried out on economic grounds; it was primarily based on the differing characteristics of the products, and associated empirical studies also primarily targeted the characteristics of the services. Studies had previously been published on the issue, using theoretical-methodological approaches (e.g. *Infrapont* [2006]) or empirical research (e.g. *Tárki* [2004]) based on a subjective assessment of substitution, where respondents were asked if they felt that the services were substituted.

There are numerous theoretical and practical issues associated with measuring substitution between fixed and mobile telephone services. Our research attempted

[1] Commission Recommendation 2003/311/EC of 11 February 2003 on relevant product and services markets within the electronic communications sector susceptible to ex ante regulation in accordance with Directive 2002/21/EC of the European Parliament and of the Council of 7 March 2002 on a common regulatory framework for electronic communications networks and services.

to find a solution to these problems and to allow fixed-mobile substitution to be measured in practice and analyzed. The relevant data were subsequently collected.

The following section provides a short overview of the general economic framework of substitution between the two service types. After this section, we will briefly discuss the special characteristics of telephony services that need to be taken into account when analyzing substitution. Afterwards, we provide a very brief analysis of the methods and results reported in the international literature on fixed-mobile substitution, and then we discuss in detail the methodology of our research, the operational solutions for measuring substitution and the questions included in the questionnaire. When discussing the study results, we present the statistical distributions of the use of telecommunications services, consumers' views on such services and their subjective opinions on demand and substitutions. We also provide a brief analysis of the effects of demographic factors on the demand for such services. The next section presents our findings on the extent of substitution, also calculated as elasticity indicators. At the end of the paper, an interpretation of the elasticity indicators is offered, and conclusions are drawn.

MEASURING SUBSTITUTION

The simplest case of demand (consumption) substitution is substitution between two different goods (products or services). The present study examines a situation of this type: substitution in demand between fixed and mobile telephone services. Demand substitution between two goods exists when some external factor causes consumers to change their demand for both goods, and there is a causal relationship between the changes in demand. In such situations, some factor affecting demand for one product or service also affects demand for the other; therefore, changes in the demand for either of the goods cannot be separated from changes in the demand for the other.

Substitution is measured using the demand function. The demand function shows how much of a product or service a consumer consumes (how much their demand is) as a function of price (the price of the product or service in question and those of other goods that affect demand), the consumer's income and other relevant factors. The demand function used in demand analysis includes independent variables the changes of which can provoke substitution. However, the most important factor is the price change. The effects of price changes on demand are described by the price elasticity of demand. With regard to demand substitution between two goods, a positive cross-price elasticity indicates substitutability, and the value shows how extensive the substitution is. However, demand for telephone services differs from demand for other goods in many ways; the next chapter provides an overview of its characteristics.

Characteristics of the demand for telephone services

One of the unique characteristics of the demand for telecommunications services is that it is made up of two components. One is the demand for individual, specific calls. This is called *usage demand*. Demand for telephone subscriptions or *access* derives from the fact that the possibility (the option) of making phone calls is valuable to consumers. Probabilities of initiating or receiving individual calls with available subscribers vary between 0 and 1. These probabilities and the usefulness of the calls (instances of communication) determine the user's demand for a subscription. Therefore access demand depends on usage demand, i.e. the *derived demand* arising from expected usage (*Taylor* [1994]).

In the case of telephone services, the value of a connection to the network for a consumer also depends on how many other users have joined the network, i.e. the size of the network. How many people a consumer can reach through the telephone network and how many people can reach them affects how much they are willing to spend on the service. The larger the network – *ceteris paribus* – the more attractive it is to consumers.[2] This is called the *network effect*. *Liebowitz–Margolis* [2002] provided a good overview of the network effect in telecommunications services. Subsequently, numerous studies have examined the issue with regard to mobile telecommunications; the findings of *Kim–Kwon* [2003], *Fu* [2004], *Huang* [2006] and *Grajek* [2007] confirm the existence of a network effect.

This network effect has an impact on the substitution or complementarity of the two services under examination. As the utility of a service depends on the number of users on the network, the interconnection of fixed and mobile networks meant that the number of mobile phone subscribers also provided a utility increase to fixed subscribers and *vice versa*, creating a complementary relationship. According to *Liebowitz–Margolis* [2002], as long as the number of mobile subscribers is low, there is a positive externality influence, and the two services are complementary. As prices fall and the number of subscribers increases, however, substitution becomes the dominant phenomenon.

Due to the fact that telecommunications demand can be split up into two separate elements, consumer decisions are often modeled as a *two-phase decision process*. In this framework, the consumer first decides whether to subscribe to the service

[2] If the subscribers of all networks become accessible to users of all other networks due to the universal interconnection obligation, then the network size is the same for every user, but the price of access and accessibility still differs between networks. In such situations, rational consumers make decisions based on the average price, which is determined by the cost of calling the people they wish to communicate with, i.e. their community of interest. This in turn depends on which network the people in question are using. Therefore, when there are several networks, deciding which one to subscribe to requires solving a coordination problem in addition to taking externalities into account.

(this is a one-time discretionary decision). If the consumer has chosen to subscribe, they decide how much to use the service (this is a continuous decision).[3] This decision can be modeled as two successive decisions, or it can be seen as a simultaneous decision situation.[4] The model used in *Train–McFadden–Ben-Akiva* [1987] is based on a simultaneous approach, while the findings of *Miravete* [2002], *Miravete–Narayanan–Chintagunta* [2007] and *Iyengar* [2004] underline the importance of the temporal separation of the subscription decision and the usage decision.

Empirical analyses carried out over the last few years (*Lambrecht–Seim–Skiera* [2005], *Huang* [2006], *Goettler–Clay* [2006]) indicate that structural models must include the uncertainty of future usage, as the *faulty expectations* of consumers must also be taken into account when modeling demand.

Using telephone services requires the joint consumption decision of two parties, the calling party and the called party, which is to say that the call has utility not only for the calling party, who generates the demand, but also for the called party, who does not pay for the utility thus obtained (the incoming call). This is called *call externality*. Numerous studies examined call externality (see for instance *Littlechild* [1975]), but its impact is presumably not particularly great, and there is a high likelihood of a type of reciprocity emerging between parties that call each other regularly, exchanging roles as the calling party and the called party. As a result of this, call externality is generally internalized.[5]

Another particular feature of telephony demand is that it can self-generate, in the sense that a call may cause further communication to become necessary later on, thus *generating further calls* (e.g. calling back the original calling party).

Over the last few decades, numerous *new communication services* have emerged, such as e-mail, chat, or voice calls via the Internet (VoIP, of which Skype is a good example). These can replace telephone calls, as they can help people exchange information.

There are also *package offers* on the market, in which telephone subscriptions are sold at a discount bundled with television or Internet subscriptions. The system of discounts makes modeling more difficult, as it creates a sort of complementary relationship between the elements of the package.

Demand for *residential and business* services should be analyzed separately, as further special factors need to be taken into account when examining the latter. The present study is not concerned with the business market, focusing exclusively on demand on the residential market.

[3] Essentially, the concept of derived demand is also based on this characteristic.

[4] It should also be mentioned that the decision on access is a long-term decision (and there is a switching cost associated with it), while, once access exists, the other decision can change dynamically without switching cost in accordance with the consumer's current needs and external factors (such as pricing).

[5] As call externality emerges between a low number of parties (and is usually bilateral), it is likely to be internalized.

Fixed-mobile substitution

With regard to the modeling of fixed-mobile substitution, the fact that there are separate demands for access and for usage means that there are two types of substitution when it comes to telecommunications services. With regard to fixed and mobile telephone services, *access substitution* means that, when certain factors are present, consumers cancel their fixed telephone subscription and purchase a mobile subscription (or *vice versa*, but this is less interesting from a practical perspective). *Usage substitution or traffic substitution* refers to changes in the quantity of fixed and mobile calls.

The empirical literature contains the results of numerous measurements of usage substitution. Most of the models examining access substitution[6] and usage substitution[7] confirmed that substitution is taking place between mobile and fixed telephone services. However, a comparison of the models also shows that both the questions posed by various research projects and the methods employed in them varied considerably: there is no standard procedure for studying fixed-mobile substitution.

With regard to usage substitution, for instance, *Sung* [2003] examined fixed-mobile usage substitution in Korea by modeling point-to-point demand, in which the calls on the incumbent's network from region *A* to region *B* are affected by the price of long-distance calls and mobile calls from *A* to *B*. *Ahn–Lee–Kim* [2004] estimated the *ratio* of fixed and mobile call minutes, also in Korea. The model's explanatory factors include relative prices and the number of fixed and mobile subscribers. *Ward–Woroch* [2004] measured the effects of fixed and mobile prices on fixed and mobile call minutes. The authors used a special data collection method for their research: they asked households to submit their telecommunications bills for ten quarters (TNS "bill harvest").

With regard to access substitution, *Sung–Kim* [2002] did not model the number of fixed telephone connections, but rather the number of new subscriptions and cancellations. *Rodini–Ward–Woroch* [2003] also relied on telecommunications bills (TNS "bill harvest") for data. In this model, the authors measured the effects of fixed prices on mobile subscriptions. *Garbacz–Thompson* [2007] undertook the task of estimating the demand for residential fixed and mobile services (and penetration) in 53 developing countries in the period between 1996 and 2003.

There are various models designed to examine product consumption that separate one particular product group from other products. Demand for this product group is modeled using a conditional/limited demand function, where a separate budget is established within the consumer's income, designated by the consumer for this group of products (see for instance: *Eales–Unnevehr* [1988], *Baker–Blundell–Micklewright* [1989], *Hayes–Wahl–Williams* [1990], *Sellen–Goddard* [1997], *Edgerton* [1997]).

[6] *Sung–Lee* [2002], *Rodini–Ward–Woroch* [2003], *Horváth–Maldoom* [2002], *Garbacz–Thompson* [2007].

[7] *Sung* [2003], *Ahn–Lee–Kim* [2004], *Ward–Woroch* [2004], *Okada–Hatta* [1999].

This approach is based on the assumption that consumers make sequential decisions about how to spend their income. As a first step, the consumer divides their income between multiple product categories, and then proceeds to divide each sub-budget between the products of each category.

Hereinafter, the use of this two-step budgeting concept when analyzing the demand for telecommunications services will be referred to as the *telecommunications budget hypothesis*. According to this hypothesis, there is a sum that people devote to telecommunication, and if the price of one telecommunications service changes, they reallocate their consumption (spending) within this budget. When examining telecommunications service substitution, it is important to determine whether consumers apply the thinking posited by the telecommunications budget hypothesis, i.e. whether they tend to subscribe to new services if one of the services they use becomes cheaper.

RESEARCH METHODS

The goal of our research, funded by the Competition Culture Centre of the Hungarian Competition Authority and the National Communication Authority, was to measure economic substitution between fixed and mobile telephone services using a questionnaire-based survey. Demand for mobile and fixed telephone services is affected by price as well as by numerous lifestyle-related, sociological and psychological factors. Therefore, substitution was examined using demand models. Our research attempts to separate the effects of these two types of factors – prices on the one hand and other factors on the other. The effects of pricing were studied using a special type of preference analysis model that examines both *revealed* and *stated* preferences, while non-price factors were studied using a set of questions about facts and opinions. As the stated preference methodology is not commonly used in economics studies for measuring the effects of pricing, we first discuss this method and the specific subtype chosen for this study, the combined model. This is followed by a discussion of the survey questions regarding access substitution, usage substitution and non-price factors.

The stated preference methodology

The stated preference methodology has frequently been employed in certain areas of demand analysis, for example in the evaluation of environmental resources (see the overview by *Boxall et al.* [1996]), and certain tourism-related studies (see the overview by *Louviére–Timmermans* [1990]). The methodology is rarely employed for modeling telecommunications services, although there are precedents for such use (*Tseng–Tsiu* [2005], *Lee–Kim–Ahn* [2006]).

The methodology is built on *Lancaster's* [1966] and [1971] consumer decision models, conceptualizing goods as a linear combination of their main characteristics or attributes. This allows for the creation of a demand function that describes demand as dependent on the characteristics of the goods.

As part of the method, researchers compile a questionnaire in which respondents evaluate alternatives with differing features (in tourism surveys, for instance, features include distance, the type of activities available, accommodation and price). The method often relies on *conjoint* analysis when it comes to arranging and choosing alternatives and rankings to be included in the questionnaire. The next step is establishing utility levels (estimating the utility function) based on consumers' assessments, which is most often done using the ordinary least square (OLS) method. Then decisions are predicted based on a comparison of utility levels. Subsequently, expected market shares can be estimated on the basis of the distribution of individual decisions (*Louviére–Timmermans* [1990]).

A special kind of stated preference model is one that combines revealed preference and stated preference. This model takes into account the respondent's previous decisions when choosing the alternatives to be offered to them. First, respondents are asked about their previous experiences (e.g. where they spent their holidays the previous year), then they are asked to compare these with a hypothetical product, which is a modified version of their previous choice. Compared to the traditional *(conjoint)* analysis of stated preferences, this questioning method makes the decision more realistic for the respondent, as it is tied to a real experience, which may theoretically help to increase the reliability of responses. As an added benefit, the method allows researchers to offer more relevant choices. The method's drawback is that the proposed options are not independent of previous decisions, which in turn may depend on factors that are not being observed. This causes problems when it comes to estimating regression models; however, these econometrical issues can be managed by using mixed logit models (*Train–Wilson* [2008]).

The questionnaire

The objective when measuring substitution is to examine how changes in the pricing of fixed and mobile phone calls affect demand for the two services. Numerous other factors may also have an impact on demand and substitution. First, we present the method used for examining the effects of pricing, then we briefly discuss other factors. We start the examination of pricing effects by taking a closer look at consumers' decisions on *subscription* (access substitution), then we move on to issues associated with *telephone usage* (usage substitution).

Access substitution • By purchasing a *telephone subscription,* consumers purchase the option to make calls at the per-minute rate of the tariff plan in question, and they purchase accessibility (other people can call them). Accordingly, the choice between a fixed and a mobile phone is determined by the fixed monthly fee and the price of calls. Decisions also depend on what subscriptions (fixed, mobile or neither) the consumer already has. This information was used when generating alternatives to choose from.

People with only one type of subscription were asked about the effects of two price changes (for those who *have only a fixed phone* and those who *have only a mobile phone,* the questionnaire is perfectly symmetrical). *1.* First the effects of raising the price of the consumer's existing subscription were tested (fixed phone becoming more expensive). In this case, there are two possible reactions: simply purchasing the other subscription without making any other changes, and purchasing the other subscription while canceling the present one (i.e. either getting a mobile subscription in addition to the fixed phone, or getting a mobile subscription to replace the fixed subscription). *2.* The other scenario for those who only have one subscription is the reduction of the price of the other service (in our example, the mobile becoming cheaper), with the two possible reactions being purchasing the now cheaper other subscription in addition to the existing subscription or replacing it (either getting a mobile subscription in addition to the fixed phone, or getting a mobile subscription and canceling the fixed subscription).

For those respondents who have both types of subscription, obviously the only possible action with regard to access is the cancellation of services. Own price effects and cross-effects were included in the questionnaire in this case; that is, canceling mobile subscriptions in case of mobile price increases, and canceling fixed phones in case of mobile price decreases. Similarly, the cancellation of mobile subscriptions in case of the reduction of fixed tariffs was examined, as was the cancellation of the fixed phone in case of raised fixed tariffs.

The situation is simpler for households that have neither type of access. In their case, the possible reaction to mobile phone service price reductions is to get a mobile subscription, and the possible reaction to fixed phone service price reductions is to get a fixed subscription.

Thus, the alternatives offered include questions regarding own-price and cross-price elasticity, with regard both to price increases and to price reductions.

Questions designed to test the telecommunications budget hypothesis can also be asked, determining whether a reduction in the price of one service leads to an increase in the demand for the other as more money is left in the telecommunications budget (and *vice versa*). Estimating the effects of the telecommunications budget was not one of the principal aims of the research, but two options related to this issue were included in the questionnaire. These covered households with both types of subscription, using the above-described options (canceling the fixed phone if mobile prices rise and *vice versa*).

Access decisions merit a more detailed examination. Various types of subscription are available to consumers, and a price increase will not necessarily motivate them to cancel their subscription right away; they might only change tariff plans – or stick with the plan they have and use the phone less or more. The questionnaire included questions about these possibilities.

When framing the questions, we had to consider how much consumers know about their options and the associated prices. If we assume that they have a reasonably clear idea of the situation, the phrasing of the question may be based on price changes, such as: "Would you buy a fixed phone subscription if the prices were lowered by 20%?" On the other hand, if we assume that consumers lack this knowledge, we need to present hypothetical plans, such as: "Would you buy a fixed phone subscription if the monthly fee was 3200 HUF, half of which was usable as call credit and you could call fixed numbers for 10 HUF and mobile numbers for 60 HUF at any time of day?". The drawback of presenting hypothetical packages is that there is significant variation in multiple dimensions (call prices to various networks, monthly fee, crediting the monthly fee towards calls, free calls at certain times of day). Therefore, the following solution was chosen:

- Basing the questionnaire item on the current fee if the respondent has a subscription of the type in question;
- Presenting a hypothetical call plan if the respondent does not have a subscription of the type in question.

The advantage of the first option is that no detailed information is needed on the type of plan the respondent has (which the respondent might not be able to provide) in order to ask about the hypothetical plan they would find suitable. However, it presumes that the respondent has a reasonably good grasp of their call patterns and their total telephone bill, based on which they can decide what they would do in case of a price increase or a price decrease. Price changes were framed as an identical change to all elements of the bill; therefore, the price change does not depend on the call patterns of the respondent.

The advantage of the second option is that it does not presume that the respondent knows the price of services they do not currently use. Furthermore, in order to simplify things, we presumed that substitution is probably continuous, and therefore we proposed low-price retail phone plans similar to those that exist on the market to people who did not have a subscription. However, this may cause price elasticity to be somewhat underestimated. Presumably, there is a relatively small group of people who would buy a larger subscription right away when signing a fixed or mobile contract, or buy a small-value plan but choose a plan that is different in some way from the one we proposed.

Two questions were asked in order to examine the effects of insufficient information. First, respondents were asked about a plan costing the same as plans on the market at the time, then they were asked about a plan with a different price. Theoretically, this allows us to separate effects caused by a lack of information on prices (which we will call the information effect from now on) from those caused by price changes. By default, only the second, 'cleaned' effect was taken into account for calculating elasticities.

Usage substitution • Usage substitution may be defined as follows: presuming an existing subscription, how would changes to the cost of call minutes change consumption habits? This definition complements the previously discussed concept of *access substitution*, which is about the decision to change one's subscription. In this framework, usage substitution is only available to consumers who have both types of subscription. Access at the workplace (via fixed or mobile phone) plays a special role in this regard. On the one hand, the research covers the residential telephony market, and thus the examination focused on private telephone conversations, and only subjects with a private subscription were taken into account. Subscriptions belonging to the respondent's own business were considered private in this regard (and with regard to access substitution as well), as they can be considered private from the economic standpoint. However, use of a telephone at the workplace often exists as an option for people, and therefore it was included among the options offered to respondents.

When it comes to studying usage substitution, the fact that consumers are unsure about the volume of their usage constitutes an additional difficulty compared to studying access substitution. As the majority of consumers presumably do not know the various types of call traffic listed on their telephone bill (even if they receive such a list), it is difficult for them to tell what percentage of their calls they would transfer to the fixed network if the fixed call prices were reduced by 10 percent. In this situation, using the stated preference method involves drawing up hypothetical scenarios – as was done in a previous similar study by Tárki-NHH (*Tárki* [2004]). The primary drawback of this approach is that we do not know how often these hypothetical situations arise in the respondent's life, as respondents can probably only supply very imprecise estimates in this regard. Therefore, we used a method that combines stated and revealed preference in such situations. We asked respondents to think back to the last five calls they made and we asked them how they would have handled them if call costs had been different.

Decision options:
- no change,
- different call length,
- initiating the call from a different network,
- initiating the call from a workplace phone,
- making the call at a cheaper time of day,
- forgoing the call,
- using a different means of communication (e-mail, chat, sms).

Unfortunately, we can only study substitution from one angle: we can find out if the respondent would still have initiated a call they made even if the prices were higher, but we cannot find out if they would have called somebody they did not call if the prices were lower.

Information collected on various call characteristics was also useful for analysis. This information includes the following: call length, call type (family, friends or business), how urgent it was and where the respondent was when the call took place.

In addition to price, numerous *social and lifestyle factors* affect fixed-mobile substitution, and the impact of these factors may be greater than that of price. Our work also involved analyzing these factors. Although the primary purpose is to examine residential consumption, subscriptions at the workplace are an important factor, as work telephones are often used for private calls, which makes them function essentially as a substitute for residential consumption. Discounted tariff plans available to employees of a company – which are often cheaper than regular plans available to the public – can encourage people to use their corporate phone.[8] In such cases, we need to determine who actually pays the bill (the employee or the company) and whether there is a formal or informal limit to the use of the corporate phone. Software tools enabling voice calls or instant messaging over the Internet can offer another possible substitute for the private telephone. Additionally, mobile phone service providers offer options that can, to a certain extent, serve as an alternative to cheap fixed calls. One of these is a call plan option allowing the subscriber to call a limited set of numbers for free.

Complementary services may also affect demand. For fixed phone subscriptions, ADSL Internet subscriptions are one such complementary service; at the time of our research project, ADSL Internet was significantly cheaper for subscribers who also had a telephone subscription.

There is also the aspect of the economies of scale when we consider a fixed phone subscription: a larger household can use it more, lowering the per-call cost of the fixed monthly fee.

Examining elasticities

Substitution is generally measured in economic terms with the use of elasticity as a metric. The extent of substitution between two products is often measured by calculating the cross-price elasticity. Cross-price elasticity shows how much more is bought of one product when the other becomes more expensive, and how much less is bought of one when the other becomes cheaper.

[8] Market research indicates that this feature is clearly a relevant one in the Hungarian market.

In the hypothetical monopolistic test applied to market determinations for the purposes of competition law and telecommunications regulation, one product is considered to substitute another when both of the following conditions are met:

1. if the price of the product sold by the presumed monopoly is raised by a small but significant amount for a long period of time, its demand falls to such an extent that the price rise does not generate a profit, i.e. the own-price elasticity is high;
2. in the test carried out with the inclusion of an alternative product, which is functionally at least somewhat substitutive, and which is chosen based on cross-price elasticities, a price rise would be profitable.

The demand data collected using the survey method allowed us to estimate the own price and cross-price elasticities of the demand for various services. For our purposes, cross-product effects were most important, so estimating these was the main priority.

When drawing up the questionnaire, the primary goal was to investigate the background of substitution as deeply as possible, which also had a significant effect on elasticity estimations in many regards. A 20% price change was used in the survey to examine the effects of price changes. This rather substantial change was used on the assumption that significant and genuine reactions can only be expected from consumers if they are presented with a decision that really merits consideration.

When calculating elasticity based on such larger changes, the baseline chosen for examining reactions becomes an important factor. Therefore, we did not study one point on the demand curve (point-price elasticity) but a longer section (arc elasticity). The extent of change was calculated based on the arithmetic mean of the values before and after the change, i.e. a mean-based arc elasticity was calculated.

Sample

The questionnaire-based data collection was carried out by Szonda-Ipsos in September 2008. The target population included people over the age of 18 living in Hungary, and respondents were chosen on the basis of a sampling method that takes into account locality sizes. Responses were collected in person within the framework of an "omnibus" survey, in which various unrelated surveys are carried out at the same time by the same interviewers. The sample consisted of 952 people.

Differences between the sample and the target population were corrected by Szonda-Ipsos using four-dimensional weighting based on age, sex, residence (village, town, city etc.) and educational qualifications. All analyses were carried out on the weighted sample.

RESULTS

Telephone penetration

Our data indicates that 73.3% of the Hungarian population had at least one personal or corporate mobile phone subscription at the time of data collection, the share of fixed private subscriptions was 41.5% and a total of 51% of the population had access to some kind of fixed telephone network at home or at their workplace *(see Table 1)*.

TABLE 1 • **Distribution of access types and actual usage** (*N* = 952)

Access	Percentage
Mobile phone	
Private	63.7
Private and corporate	7.8
Corporate	1.8
None	26.7
Total	100.0
Telephone at home (mobile or fixed)	
Mobile only	44.8
Mobile and fixed	28.5
Fixed only	13.0
None	13.7
Total	100.0
Fixed	
Private	32.3
Private and corporate	9.2
Corporate	9.5
None	49.0
Total	100.0
Telephones total	
Mobile only	35.4
Mobile and fixed	37.9
Fixed only	13.1
Neither	13.6
Total	100.0

Taking into account all devices available at the consumer's home, only 13.6% of the population lacked a home telephone subscription, and the proportion of those without a telephone connection at work was essentially the same. The only cause for any difference between the two is that a significant number of people who only use a mobile phone had access to a fixed phone at their workplace.

Main characteristics of telephone use

The data collected using the questionnaire provide a good overview of how frequently people use the services available to them through various channels for their personal calls *(Table 2).* With regard to telephone access at the workplace or through corporate means, our analysis only examines private calls, as these are the calls regarding which consumers make independent decisions based on pricing and non-price factors. Generally, using fixed phones at the workplace is the rarest usage scenario. Only 6.5% of respondents use a fixed phone at their workplace more than five times a day.

TABLE 2 • **Main characteristics of private telephone usage** (percentages)

	Private mobile		Corporate mobile		Home fixed		Workplace fixed	
Usage frequency	Outgoing	Incoming	Outgoing	Incoming	Outgoing	Incoming	Outgoing	Incoming
More than 5 calls a day	20.4	25.4	33.6	34.8	5.9	7.8	6.5	6.1
1 to 5 calls a day	54.1	55.9	22.3	30.2	45.5	50.4	15.4	14.1
Weekly	15.7	14.2	9.8	7.5	28.0	25.5	11.2	6.2
Very rarely/never	9.7	4.4	34.2	27.5	20.6	16.3	66.8	73.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	681		92		395		177	

A significant number, about one third of those with a corporate mobile phone, use their phone for making private calls; at the same time, a similar number of people use their corporate phone for private calls only very rarely or never. The frequency of usage is by far the highest for private mobile phones. Private mobile phones are the most frequently used means of communication both for outgoing and incoming calls. Irrespective of whether a phone is private or corporate, consumers use mobile phones much more often than fixed phones.

Nevertheless, the ratio of *very rare usage* is surprisingly high for both private mobile and private fixed phones. 20% of people with a home fixed phone use it very rarely or not at all (either for outgoing, or for incoming calls). The lack of outgoing calls is explained by the fact that many consumers declared that they only have a fixed phone for receiving calls. However, the rarity of incoming calls still requires further explanation. One possible reason is that the fixed phone was only kept because it was bundled with an Internet subscription, or for security (many people feel that a fixed phone is worth having because of its reliability in emergency situations), or perhaps consumers simply did not get around to canceling it.

In the case of mobile subscriptions, the cause of the relatively large number of responses indicating rare or inexistent use (outgoing: 9.7 percent, incoming: 4.4 percent among those possessing a subscription) is likely to be security and aversion.

The monthly bill of most private consumers was between HUF 2000 and HUF 4000 for both types of telephone subscription; on average, mobile phone costs are not significantly higher than fixed telephone costs. It should be noted that these are not individual but household data, as home fixed phones are generally used by all members of the household *(Table 3).*

In the overwhelming majority of cases, the use of corporate fixed phones does not involve costs (and takes place at the workplace); even if there is a cost to the user, it is low (less than HUF 2000 per month). Corporate mobile phones are much more costly; 54.1% of users contribute to the costs, and 15% pay more than HUF 10,000 per month.

The goal of the questions regarding work-related issues was to identify the extent to which work-related situations occur that make it difficult or impossible to replace mobile connections with fixed connections *(Table 4).*

22.4% indicated working at home as a typical situation, 20% travel a lot and another 20% work at frequently changing locations; the latter two situations naturally make it impossible to rely on fixed phones. The most frequently indicated cause – marked by 47.8% of respondents – was that accessibility is important for work purposes.

TABLE 3 • **Costs of private telephone use paid for by the user, listed by access type** (percent)

	Mobile		Fixed	
Monthly expenditure (HUF)*	Private	Corporate	Private	Corporate
0	1.7	45.9	1.0	79.2
1–1,999	17.3	15.6	9.0	18.7
2,000–3,999	38.3	3.2	45.2	1.8
4,000–5,999	19.5	12.4	30.3	0.3
6,000–7,999	11.6	8.8	10.3	0.0
8,000–9,999	3.4	2.3	2.4	0.0
10,000–15,000	7.3	7.9	1.3	0.0
15,000–20,000	0.0	0.9	0.0	0.0
20,000+	1.1	3.1	0.4	0.0
Total	100.0	100.0	100.0	100.0
Average (HUF)	4104	2053	3530	111
N	693	95	406	191

* At the time of data collection cca. 275 HUF = 1 EUR

TABLE 4 • **Work characteristics** (percent, $N = 481$)

Statement	Yes	No	Total
Works at home often	22.43	77.57	100.00
It is important that colleagues or clients can reach them by phone at any time, and that they can reach colleagues or clients at any time.	47.82	52.18	100.00
Travels a lot for work	21.11	78.89	100.00
The location where they work changes often	20.57	79.43	100.00

Consumer attitudes and habits

In the section on attitudes, respondents were given statements related to the choice between fixed and mobile phones and asked to indicate to what extent they identify with the statements. Scores ranged from 1 to 4 (4 = fully agree). The penultimate column of *Table 5* shows the average of the scores.

According to the responses to the question on new technologies, only 4 percent of people are early adopters, while 51.3 percent do not see this as characteristic of them, with the average score being 1.7. However, most of the data indicate that this question has limited relevance to mobile phones, which cannot really be considered a new technology any longer.

Numerous questions were addressed to respondents who do not use a mobile phone. In many cases, the complicated nature of such phones caused a problem: for more than half of the respondents, difficulty of use was the reason for not using a mobile phone. The fear of adverse health effects – which have not been scientif-

TABLE 5 • **Attitudes and habits related to technology**

Statement	Fully agree	Mostly agree	Mostly not agree	Not agree at all	Average of scores (1–4)	*N*
	Percentage					
I tend to use new technologies much earlier than most people do	4.1	16.1	27.7	52.1	1.7	952
A fixed phone is important for me because it is always available in emergencies in case I can't use my mobile phone for some reason	12.7	16.6	15.0	55.7	1.9	406
It is important to have a shared fixed telephone in the household/family	33.2	39.7	14.1	13.0	2.9	103
I know the fixed number of many of my friends and acquaintances (perhaps in addition to their mobile number)	11.5	22.6	20.1	45.8	2.0	952
I prefer to call friends and acquaintances on their mobile phone because then I know it will not be someone else who answers the phone	34.6	26.4	12.6	26.3	2.7	952
It is important for me to be reachable at all times, wherever I am	33.3	28.1	17.4	21.2	2.7	952
Mobile phones are more complicated to use, so I use the fixed phone whenever I can	4.5	8.2	22.2	65.1	1.5	952
Mobile phone use causes damage to health, and therefore I often consider using the fixed phone instead	2.7	13.7	26.2	57.4	1.6	381
I have a fixed phone because it is important that my acquaintances and relatives can reach me at a cheap rate	34.8	38.4	11.6	15.2	2.9	381
I only have a fixed phone because of my Internet access (ADSL)	20.4	26.0	28.2	25.4	2.4	406
I don't use a mobile phone because they are complicated to use	34.7	16.9	15.8	32.5	2.5	243
I don't use a mobile phone because they cause damage to health	9.9	12.0	24.5	53.7	1.8	243
I don't use a mobile phone because I don't need one	57.2	16.0	10.0	16.8	3.1	243
I don't use a mobile phone because they are expensive	42.9	20.7	16.5	19.9	2.9	243
I don't use a fixed phone at home because I don't need one	58.8	17.6	8.4	15.2	3.2	546
I don't use a fixed phone at home because it is expensive	46.2	20.6	16.0	17.2	3.0	546

ically proven but are regularly discussed – received a score of 1.8. 55.1 percent of non-users firmly state that they do not need a mobile phone, and close to 40 percent clearly deem mobile phones too expensive. 43.3 percent of respondents who do not use a fixed phone say the same thing about the price of fixed subscriptions, and more than 50 percent state that they do not need one.

Interestingly, the statement about maintaining a fixed phone subscription for ADSL alone received an average score of 2.4 (strong agreement: 20.2%). The reason for maintaining a fixed connection may be that it allows friends and relatives to call the consumer cheaply (2.9 points), providing a common telephone number for the household/family (2.9 points), reliable access in case of emergency (1.9 points) and the fact that the respondent considers mobile phones complicated to use (1.5 points). Generally, only one-third of respondents know the fixed phone number of their friends (average: 2.0 points) and only 15.5 percent prefer to initiate calls from their fixed phone due to concerns about health risks.

For a rather high number of respondents (cca. 60%), two-way "immediate" accessibility (anyone can reach them and they can reach anyone) is important, which indicates the clear superiority of mobile phones. These are presumably the respondents who have a mobile phone in order to ensure accessibility (mobility).

Thus, those who use mobile phones mostly call mobile numbers and the primary consideration for them is to be accessible by phone at any time, anywhere (whatever the cost). Those who prefer fixed phones do so because of cheap rates and security.

Consumption characteristics affecting substitution

Whether a consumer replaces their fixed phone with a mobile phone depends – in addition to the price – on their habits and characteristics. This chapter examines the effects of three possible factors, first one by one, then as part of a unified model. These factors are: substitution options, household size and Internet subscription.

Fixed-mobile substitution can also be studied by examining whether a household is equipped with a fixed phone, or the respondent only uses a mobile phone at home. As previously shown, these are the two most typical arrangements. Accordingly, we compare these two groups, trying to identify the differences between mobile users who do not use a fixed phone and those who do. Among possible substitutes for a fixed phone, we consider Internet telephony and use of a corporate mobile phone *(Table 6)*.

The data show that respondents with a corporate mobile phone are more likely to have a fixed phone as well, which indicates that for them, substitution is less likely. In all probability, canceling the fixed phone is not important, as the two services do not significantly compete for a share of the family budget.

The role of Internet telephony was only studied among people with Internet access in order to eliminate any effect arising from the availability or absence of an Internet connection. That examination showed no difference: no significant substitution between calls over the Internet and calls over fixed phones was detected. The observation of any such effect is made more difficult by the behavior of service providers, who often offer discounted fixed phone services along with Internet subscriptions.

Next, the effects of Internet access were analyzed. In principle, as ADSL Internet subscriptions are cheaper for users who have a fixed phone, interest in an Internet connection in itself should increase the likelihood that the consumer will purchase a fixed phone connection. Similarly, Internet service providers often offer discounted telephone services for their Internet subscribers *(Table 7)*. The figures show that the proportion of respondents with a fixed phone is significantly higher among those who have an Internet connection. This indicates that the two services are complementary.

Finally, the impact of household size was studied. As all members of the household may use a fixed phone subscription, but the monthly fee is fixed, larger households in principle have lower per capita costs. To examine this issue, single-person households and multi-person households were compared. The results show that people who live alone are less likely to have a fixed phone subscription if they already have a mobile phone subscription *(Table 8)*.

TABLE 6 • **The effects of substitute services on fixed phone subscriptions** (percentages)

	Use of corporate mobile phone			Calls over the Internet		
	does not have	has		does	does not	
Has	a corporate mobile phone		Total	use the Internet for chatting and calls		Total
Only a mobile phone	61.8	56.8	61.1	48.6	50.5	50.0
Mobile + fixed	38.2	43.2	38.9	51.4	49.5	50.0
Total	100.0	100.0	100.0	100.0	100.0	100.0

TABLE 7 • **The effects of having an Internet connection on fixed phone subscriptions** (percentages)

Has	Does not have an Internet connection	Has an Internet connection	Total
Only a mobile phone	72.1	50.0	61.1
Mobile + fixed	27.9	50.0	38.9
Total	100.0	100.0	100.0

TABLE 8 • **The effects of household size on fixed phone subscriptions** (percentages)

Has	Single-person household	Household with multiple people	Total
Mobile only	70.5	59.8	61.1
Mobile + fixed	29.5	40.2	38.9
Total	100.0	100.0	100.0

Next, the combined effects of these three factors on respondent's decisions about whether or not to buy a fixed phone subscription will be examined *(Table 9).* Age can affect both household size and Internet subscription status, so it was included in the analysis as well. Additionally, having an Internet connection correlates with educational qualifications, so educational qualifications were also included as a control variable. Due to the nature of the issue, a logistic regression model was used.

TABLE 9 • **The combined effects of factors on fixed telephone subscription**

Explanatory factor	*P-value*	*exp(b) (odds ratio)*
Does not live alone	0.000***	0.561
Has a company-issued mobile phone	0.324	0.785
Has an Internet subscription	0.000*	2.140
Higher education qualifications (compared to secondary)	0.582	1.141
Secondary educational qualifications (compared to lower)	0.020·	1.533
Age below 40 (compared to ages 40 to 60)	0.000***	0.377
Age below 30 (compared to age below 40)	0.962	1.012
Age below 60 (compared to ages 40 to 60)	0.333	1.242

* Significant at the 10 percent level, ** significant at the 5 percent level, *** significant at the 1 percent level.

When the three factors are examined together, and the effects of age and educational qualifications are controlled for, two factors remain significant: Internet access and the size of the household. The marginal effect of these factors is rather significant. Having an Internet connection in itself more than doubles the chance of a respondent having a fixed phone line. Interestingly, people who live alone are more likely to have a fixed phone in addition to their mobile phone than people with a family. These effects are independent of age. An analysis of age distributions indicates that 40 years is a clear demarcation line with respect to fixed phone subscription. People above 40 are almost three times as likely to have a fixed phone as those under 40. There are no other major dividing lines, though: there is little difference between people above 60 and those between 40 and 60, or between those under 30 and those between 30 and 40. Additionally, people with secondary or higher level educational qualifications are 50 percent more likely to have a fixed phone than those with lower educational qualifications.

Among the examined hypotheses, the one that was confirmed was that having an Internet subscription increases the likelihood of also having a fixed phone subscription, and to a rather significant extent. The use of alternative technologies (Internet calls, corporate mobile phone) does not significantly reduce the likelihood of having a fixed phone subscription. The hypothesis about household size was not confirmed: among respondents with the same age and educational qualifications: single respondents were more likely to have a fixed phone in addition to their mobile phone. We also identified a cultural demarcation line at the age of 40, and found a smaller difference between respondents who had obtained a secondary school leaving certificate and those who had not.

Access substitution

Access substitution was examined by studying the responses of four consumer groups to different offers and price changes. The four groups of respondents were the following:

- has own mobile phone, but not fixed;
- has own fixed phone, but not mobile;
- has both own fixed and own mobile phone;
- has neither of the two.

The different groups were presented with different offers. The main types are the following:

- offer regarding a service type not currently used by the respondent at the current market price;
- offer regarding a service type not currently used by the respondent at 20% below the current market price;
- raising the price of a service currently used by the respondent by 20%.

We also asked respondents about other issues based on their responses whenever applicable. For instance, if a respondent said that they would buy a service they were not yet using, the next question covered effects on their preexisting other subscription, with the following options:

- cancellation;
- reduced usage;
- reduced usage and switching to a different tariff plan;
- continued usage without change.

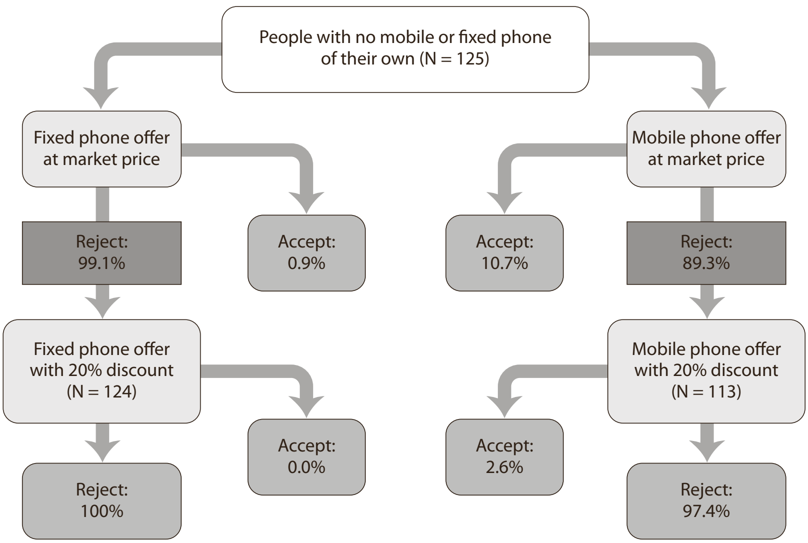
Similarly, if the questions proposed a change in the price of the existing subscription (price increase), we asked respondents for their reactions regarding both technologies in order to map crosswise influences as well.

Responses are presented in figures showing the entire decision tree. The figures indicate the number of people who provided a response to each question. In case of very low numbers, where showing the distribution would have been pointless, it was omitted. There were always three offers: one at the market price, one 20% lower and one 20% higher than the market price. Market prices were determined based on September 2008 pricing. Entry-level mobile or fixed call plan prices were chosen as the market price, assuming that if a respondent has no existing plan, they would most probably be interested in smaller plans. For fixed phones, this meant a hypothetical plan that largely matches the "Felező" plan offered by Magyar Telecom, and for mobile phones, the plan was the cheapest prepaid plan where pricing is not dependent on the time of day and the network called.

In addition to the responses on substitution, the results of estimates on the elasticity of telecommunications demand also allow us to draw conclusions regarding the relationship between fixed and mobile technologies.

When processing this part of the questionnaire, we did not follow the established procedure for analyzing opinion surveys: we included “do not know” responses among the percentage of valid responses. As this response is closer to “no” when the question is whether the respondent would buy a proposed service, these respondents were included in the “no” category. This was done in order to counteract the tendency of respondents to overestimate their willingness to make changes compared to their subsequent real behavior.

Figure 1 shows the results of the four offers proposed to those who have neither a mobile phone nor a fixed phone of their own. Only 0.9 percent would take the market-price fixed phone offer, while the rate is 10.7 percent for the market-price mobile phone offer. Of the remaining respondents, 0 percent would subscribe for a fixed phone line at the reduced price, and 2.6 percent would do so for a mobile phone connection. These data lead to two conclusions. First, demand for fixed phone connections is completely inelastic in this consumer group. Although the proportion is very low for mobile phones as well, there is a highly significant difference between the two. Second, compared to the effect of the price change, a comparatively large number of respondents said that they would subscribe at the market price. This is presumably (partly) due to the fact that people do not keep up with market prices.



Source: Infrapont.

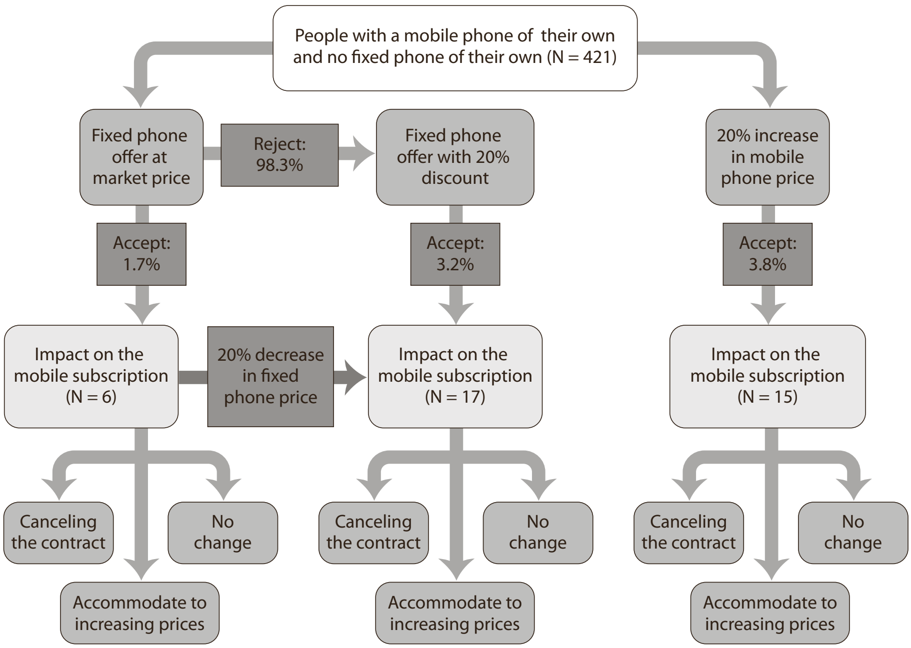
FIGURE 1 • **Responses of people with no mobile or fixed phone of their own**

The next group under examination is that of people with a mobile phone connection of their own and no fixed phone line of their own *(Figure 2)*. For them, the new offer was for a fixed phone line at the market price, which 1.7 percent of respondents would accept.

3.2 percent of the remaining respondents would subscribe for the fixed offers at a 20% price reduction. The crosswise effect is similar, i.e. if mobile telephony became significantly more expensive, 3.8 percent would purchase a fixed connection. The relatively low values indicate that the overwhelming majority of those with only a mobile phone would not wish to have a fixed phone even if the price was significantly lower.

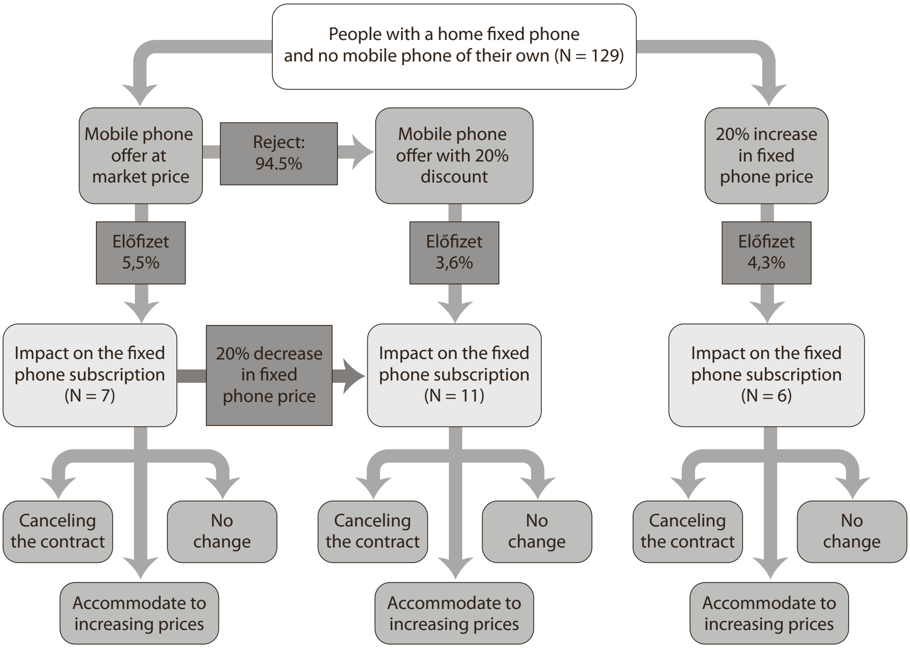
Further responses by people who gave a positive response regarding the subscription were omitted from the figure due to the very low sample size.

The third group under examination is that of people who do not have a mobile phone connection of their own but do have a fixed phone line of their own *(Figure 3)*. For this group, the structure of the questions was the same as for the previous group. The proportion of respondents who would accept market price mobile phone subscriptions is clearly higher than the number of those who would accept a fixed phone



Source: Infrapont.

FIGURE 2 • **Responses of people with a mobile phone of their own and no fixed phone of their own**



Source: Infrapont.

FIGURE 3 • **Responses of respondents with a home fixed phone and no mobile phone of their own**

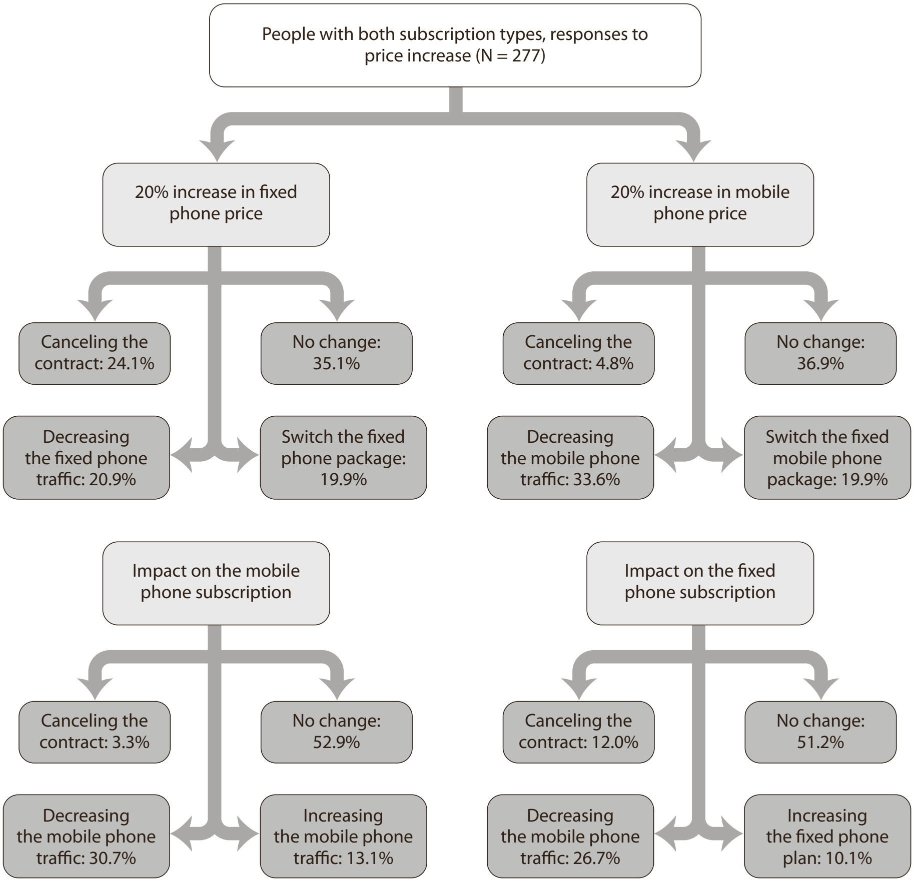
subscription in the previous group. Still, the sample size for the second response ("What would you do if you subscribed…") was still too low, as the size of the whole group is significantly lower than in the case of *Figure 2.* The reaction to the market price mobile subscription was 5.5% (information effect), and 3.6% of the remaining respondents stated that they would take the 20% cheaper offer. If fixed prices increased by 20%, 4.3% would buy a mobile phone subscription.

Overall, the results were similar to those of the second group: mobile phone price reductions did not lead to large amounts of substitution; at the same time, the willingness to make the change was almost twice as high as among those with a mobile phone only. This indicates that mobile phones are more of a substitute for fixed phones than the other way around.

With regard to substitution, the most interesting group is perhaps the one where both connection types are present (*Figure 4* and *Figure 5*). Here, crosswise relationships are more clearly observable, and the difference between the reactions to fixed and mobile price changes is easier to see as well. Additionally, the effect of price changes on the shift of usage between the two networks (usage substitution) can also be studied in this group. In *Figure 4* and *Figure 5*, consumers' reactions to price changes in the two services in the same direction (up or down) are summarized.

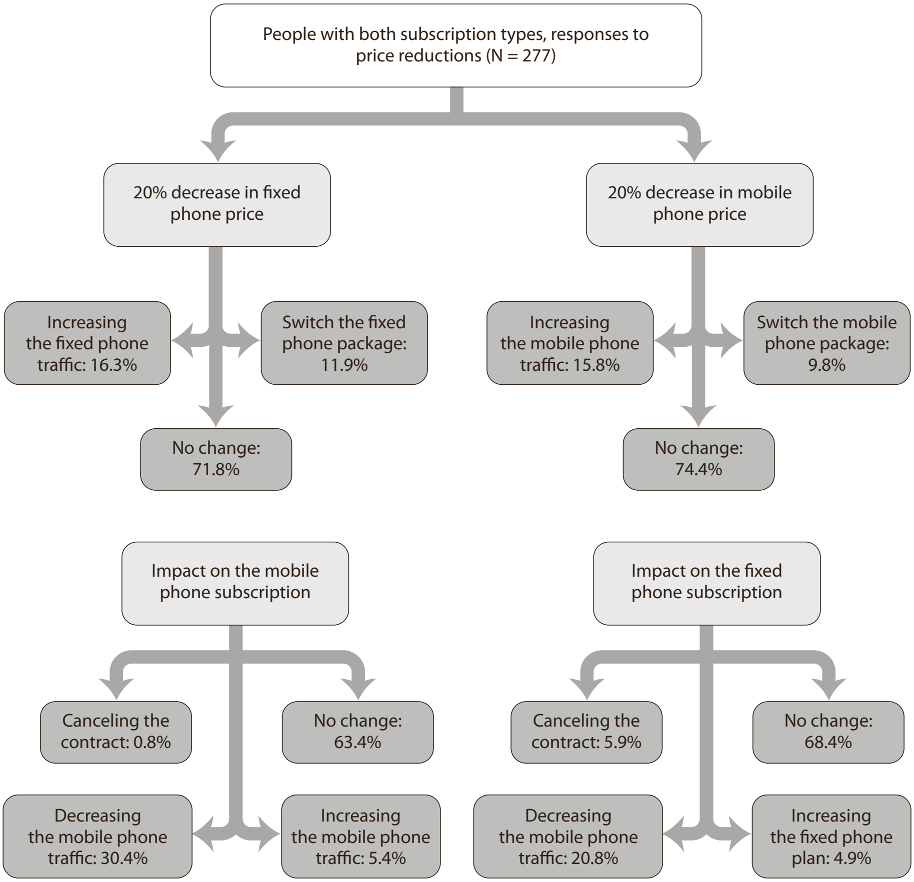
Although price rises are contrary to the dominant pricing trends, this hypothetical scenario tells us a great deal about how consumers value their two types of access: which of the two they consider more important. A significant group of consumers (about 35 percent) would not change their consumption if either service became 20 percent more expensive. However, while respondents tended to reduce their usage in response to rising mobile service prices, with only 4.8% canceling their contract, the same price increase would cause 24.1% to cancel their fixed phone contract. Thus, as opposed to the elastic reaction seen with fixed price rises, mobile access appears much more indispensable to consumers, making it less price sensitive.

The differences observed in cross-price influences are also worth noting. If fixed subscriptions became more expensive, only 3.3% would cancel their mobile



Source: Infrapont.

FIGURE 4 • **Responses of respondents with both subscription types to raising prices**



Source: Infrapont.

FIGURE 5 • **The responses to price reductions of users with both access types**

subscription. On the other hand, however, 12% of consumers would cancel their fixed line if mobile access became more expensive, which, in accordance with the presented results, indicates a rather significant difference between consumers' assessment of the two access types.

The examination of price increases also shows a non-negligible amount of usage substitution between the two services. If fixed prices were to rise by 20 percent, 13 percent of respondents would increase their mobile usage, while if mobile prices rose by 20 percent, 10 percent of respondents would use their fixed phone more. In other words, if a service becomes more expensive, a significant subset of consumers will reduce their usage (perhaps even cancel their subscription) and substitute it by using the other type of service more.

Among users who have both types of access, the number of consumers who, if faced with one service raising prices, would reduce their demand for the other service, the price of which is unchanged (by canceling the service or reducing their usage) is exceptionally high. This supports the previously discussed hypothesis of the telecommunications budget.

Only 0.8 percent of respondents would cancel their mobile subscription as a result of a reduction of fixed prices. On the other hand, 5.9 percent would cancel their fixed subscription as a result of a reduction of mobile prices. This indicates that there is a group of consumers who use a fixed phone as a secondary telephone in order to take advantage of cheaper rates, and members of this group would cancel their fixed service if mobile prices fell far enough. In the opposite direction, any such effect is negligible.

Canceling a service due to the reduction of the price of the other service is not widespread for either type of service, but adaptation by reducing usage is significant in both cases (30.4% and 20.8%). This clearly indicates usage substitution, i.e. if a consumer has both types of subscription and reduces their use of one service with an unchanged price when the price of the other service falls, the explanation is that the consumer is using the now cheaper service instead.

The difference between the strength of the phenomenon in the two directions is worthy of note. When mobile prices fall, consumers re-route their traffic from fixed to mobile less than in the opposite direction when fixed prices fall. The reason for this is that the demand for fixed telephone services is more elastic (more affected by price changes) than the demand for mobile services.

The telecommunications budget hypothesis

We need to examine to what extent the data support the telecommunications budget hypothesis. According to this hypothesis, changes to the price of one service should cause a similar response affecting both services. That is, if either fixed or mobile access becomes more expensive, we would expect both fixed and mobile demand to fall, and we would expect demand for both services to increase in response to price reductions. Data on this are presented in *Table 10.*

TABLE 10 • **Data on the telecommunications budget hypothesis (percentages)**

		Fixed	Mobile
		Response (reduction of use + cancellation)	
Price increase	Fixed	–	30.7 + 3.3
	Mobile	26.7 + 12	–
		Response (increase)	
Price reduction	Fixed	–	5.4
	Mobile	4.9	–

Table 10 shows that the proportion of people whose responses match the telecommunications budget hypothesis is higher – about 35% in both directions – in case of price increases. In case of price reductions, the effect is much smaller, but there is still a non-negligible proportion of 5% (i.e. when one service type becomes cheaper, about 5% of consumers increase their use of the other service).

This basically confirms that the effect outlined in the telecommunications budget hypothesis does exist, but it is present alongside other effects and is definitely not the only factor affecting behavior. In other words: the money lost or gained due to price changes is partly but not exclusively compensated (redirected) within the envelope of telecommunications expenses.

Usage substitution

Regarding the study of usage substitution, we previously discussed how many people stated that they would use a given service less or more as a result of changes in fixed or mobile pricing. In the next section, we examine the same issue using a different method. We asked respondents about the last five calls they initiated, inquiring about what they would have done if prices had been different. We also asked respondents about various characteristics of the calls *(Table 11).*

The results indicate that people use their own mobile phones most of the time (72.07%), with home fixed phones accounting for 23.47%, which means that when it comes to private calls, mobile phones are used more than three times as frequently as fixed phones. Telephones at work are used very rarely (4.5% in total).

Table 12 indicates that users generally call mobile phones from mobile phones and they call fixed phones from fixed phones: only 10.2% of calls crossed from one network type to the other.

Examining the phenomenon in greater detail, *Table 13* indicates that this characteristic is more powerful in the case of calls from mobile phones and to mobile phones (93.8 percent and 92.4 percent compared to 77.9 percent and 81.4 percent, respectively).

The data on the location of the calling party show that 91.4% of people use a fixed network call, unsurprisingly, from their homes. However, at close to 50%, the proportion of people calling from home is high even for mobile phone users *(Table 14).* Two-thirds of calls initiated from a fixed phone reach the called party at their home, 22% at their workplace or school; these proportions are lower for mobile phones as well.

With regard to the called parties, no significant differences were found between the two technologies: respondents called family members the most, 55.0% of the time, followed by friends and acquaintances (28.4%). The remaining calls were distributed fairly evenly between official and work-related calls *(Table 15).*

TABLE 11 • **Distribution of call initiation by connection type** (percentages)

What telephone did you initiate the call from?	Entire sample	Respondents with both
Home fixed phone	23.47	35.2
Workplace fixed phone	1.79	2.3
Corporate mobile phone	2.67	1.9
Own mobile phone	72.07	60.7
Total	100.0	100.0

TABLE 12 • **Relationship of initiation and termination technologies** (percentages)

	Called device	
Initiating device	Fixed	Mobile
Fixed	20.0	5.6
Mobile	4.6	69.8

TABLE 13 • **Relationship of initiation and termination technologies** (percentages)*

a) Distribution of call initiation by call termination

What network was called?	Call initiations		
	From fixed	From mobile	Total
Fixed	77.9	22.1	100.0
Mobile	6.2	93.8	100.0

b) Distribution of call reception by call initiation

	Received calls	
Initiating device	On fixed	On mobile
Fixed	81.4	7.6
Mobile	18.6	92.4
Total	100.0	100.0

* Part *a)* of the table shows the ratio of calls to fixed and mobile networks, respectively, among calls started from each network type, while part *b)* shows the ratio of calls from fixed and mobile networks, respectively, among calls received in each network type.

TABLE 14 • **The distribution of calls by location** (percentages)

	Fixed	Mobile	Total
Location when initiating the call			
Home	91.4	49.2	60.2
Work/school	7.1	22.5	18.6
Street, car, public transport	0.0	17.9	13.3
Other location	0.5	10.4	7.9
Total	100.0	100.0	100.0
Location of called party at the time of the call			
Home	67.9	39.6	47.1
Work/school	22.8	30.1	28.2
Street, car, public transport	6.3	17.1	14.2
Other location	2.9	13.2	10.5
Total	100.0	100.0	100.0

TABLE 15 • **The distribution of calls by called party**

Called party	Fixed	Mobile	Total
Direct relative, family	60.2	53.2	55.0
Friend, acquaintance	23.2	30.2	28.4
Official matters, private business	11.4	8.2	9.0
Work-related	5.1	8.4	7.6
Total	100.0	100.0	100.0

The distribution of calls by urgency level is somewhat surprising, as it does not indicate a very significant difference between the two technologies; the data do not show mobile phones being used more often for urgent calls from the perspective of either the calling party or the called party *(Table 16).* It is clear, however, that thanks to the available call plans, fixed phones are more often used for calls that are not urgent at all.

Table 17 sums up respondents' claims about how they would have reacted to higher service prices – i.e. the effects of price changes on usage. The answers to the question about the last five calls provide insight into the usage substitution taking place between the two networks.

Most respondents (about 60% for both services) would not change their last five calls in any way in case of a price rise. A smaller but still significant number of respondents (about 30%) would try to adapt to the higher prices in some way (shorter calls, calls made at discounted times, fewer calls). The rate of usage substitution is rather low: if fixed calls became more expensive, only 3.9% of consumers would switch to making calls from a mobile phone, and only 1.2% would switch in the other direction if mobiles became more expensive.

TABLE 16 • **Percentage distributions of call urgency and the technology chosen**

	Calling party		
How urgent was the call?	Fixed	Mobile	Total
Quite urgent	35.7	41.5	39.3
Not really urgent, but call needed to be made sooner or later	41.9	44.8	43.7
Not urgent at all	22.5	13.7	17.0
Total	100.0	100.0	100.0
	Called party		
How urgent was the call?	Fixed	Mobile	Total
Quite urgent	35.9	40.9	39.3
Not really urgent, but call needed to be made sooner or later	41.6	45.0	43.7
Not urgent at all	22.5	14.1	17.0
Total	100.0	100.0	100.0

TABLE 17 • **Effects of price rises** (percentages)

What would have been different if prices were 20% higher?*	Fixed	Mobile	Total
No change	56.4	63.0	60.5
Would have made shorter calls	33.0	25.6	28.3
Would have switched to mobile/fixed	3.9	1.2	2.2
Would have made the call at a cheaper time of day	2.6	4.6	3.8
Would not have made the call	3.2	4.2	3.8
Would have chosen other means of communication	0.9	1.5	1.3
Total	100.0	100.0	100.0

* With respect to the last five calls.

This shows that usage substitution with regard to specific calls is low – especially when compared to the results obtained regarding access (see *Figure 4*), where 13.1% of consumers (as opposed to the 3.9% listed in *Table 17*) reported that they would use their mobile phone more if fixed prices were to increase. The proportion is 10.1% for mobile price rises (compared with 1.2%).

This notable difference seems contradictory at first sight, but there is a fairly simple explanation for it. While questions regarding specific calls measure consumers' short-term reactions and adaptation, the data on usage substitution supplied with regard to subscriptions covers much more long-term adaptations. It is perfectly clear that in the short term, consumers adapt to price changes less than in the long term, in which case instead of changing decisions about individual calls, they change their habits.

Price elasticity assessments

As stated in the section on methodology, elasticity was examined using a relatively large 20% price change that provokes strong reactions. The methodological consequence of this choice is that we examine a section (arc) of the demand function, not a single point. Therefore, results obtained by calculating the change based on the original volume (before the price change) or the changed volume (after the price change) may be significantly different. In order to reduce any distortions as far as possible, we used the mean of the two quantities, calculating the arc elasticity.

Own- and cross-price elasticities were calculated with regard to both fixed and mobile telephone service and both price rises and price reductions. As described above, the information effect and the effects of prices were separated from each other, and the information effect was not taken into account when assessing elasticity. However, we also calculated elasticities using the assumption that the respondent's decision to accept a market offer and not just the 20% reduced offer is also the result of the price reduction. The results of this calculation, to be treated as a top-range estimate, are presented in the last row of *Table 18*.

TABLE 18 • **Access elasticity assessment results**

	The effects of mobile access price changes on		The effects of fixed access price changes on	
Price change	mobile demand*	fixed demand**	fixed demand*	mobile demand**
20 percent price increase	–0.27	–0.12	–1.39	–0.04
20 percent price reduction	–0.06	0.14	–0.07	0.022
20 percent price reduction (including the information effect)	–0.211	—	–0.10	—

* Own price elasticity. ** Cross-price elasticity.

CONCLUSIONS

Interpreting price elasticity assessments

With regard to own price elasticity, we can determine (primarily based on responses to price rises) that there is a significant difference between the two technologies. Mobile access is very inelastic (–0.27), but fixed phone price elasticity is much greater (–1.39). This indicates that *mobile access is much less readily substitutable than fixed access*. The obvious cause is the extra feature at the core of mobile telephony: mobile phones can be used anywhere. More precisely, the reduced substitutability expressed in low price elasticity is caused by the fact that mobility has become widespread and there is a very high demand for it; also, at the time of the study, the mobility surcharge was already low compared to other available options. On the other hand, mobile phones can provide a functionally almost perfect replacement for fixed phones, and, as the prices are relatively close to each other, even a small increase in fixed prices makes the switch to mobile worthwhile for many consumers.

It should also be noted that responses to price rises are stronger than the responses to price reductions: the price reduction response is less than –0.3% even if it is exaggerated by including the information effect.

Interpreting the results is more complicated with regard to cross-price elasticities. An examination of substitution shows that – as expected – these values are positive for price reductions. However, the values are negative for price increases. According to standard approaches used in economics, positive cross-price elasticity is an indicator of substitution, while negative values indicate a complementary relationship. Own price elasticities indicate that demand for mobile services is relatively inelastic, which means that mobile phones have no strong substitute (i.e. for regulatory purposes, they constitute a separate relevant market). Demand for fixed access is price-elastic, but its mobile cross-price elasticity is negative, which would appear to indicate that it is not being substituted by mobile access.

However, before leaping to the conclusion that mobile and fixed telephones do not substitute each other, a closer inspection of the issue is in order.

The first important methodological comment concerns cross-price elasticities. If a product's price is raised, the following effects may appear:

1. the consumer has less money left in total, and has to reduce consumption of the second product as well (the income effect);
2. if the two products are complementary, demand for the second product may also fall due to the fact that less of it is needed to complement the smaller quantities of the first product (this leads to negative cross-price elasticity if the income effect is filtered out);

3. if the second product is a substitute, then due to the relative price change, consumption of the first product needs to be substituted with the second product, thus increasing demand for the second product (which leads to positive cross-price elasticity.)

If the telecommunications budget hypothesis presented above is correct, i.e. if people devote a fixed (or more or less fixed) sum to telecommunications, then the first effect (the income effect) is strengthened. Thus, it is quite possible that the negative cross-price elasticity in case of price rises is due to the effects of the telecommunications budget, and the two products are not complementary but substitutes. A consumer who has both types of access may feel that as a result of a mobile price increase, their telecommunications costs rise beyond the amount they are willing to spend in this area, and therefore they need to reduce their consumption. If however the consumer still values mobile access more than fixed access, they will cancel the latter – despite the fact that it has just become relatively cheaper. In this case, cross-price elasticity may be negative, not as a result of complementarity but because of the income effect and the effects of the telecommunications budget.

It should be noted that the two access types are not in a complementary relationship in the traditional sense (hardware and software, for instance), as the consumption of one does not necessarily require or involve the consumption of the other. Furthermore, access demand is a special case in that there is no way to "consume less"; the consumer only has two options in case of a price rise: to cancel or to keep the subscription. Therefore, the classical complementarity effect described in paragraph 2 does not apply to fixed and mobile telephone access. The cross-price elasticity indicators presented here are created by a mixture of the income effect and the complementarity/substitution effect (as opposed to econometrics models for instance, where cross-price elasticity is measured using controls, separated from the income effect), and so the issue of complementarity and substitution cannot be resolved on the basis of these indicators alone. Therefore, the negative cross-price elasticity presented (if the income effect and the effect of the telecommunications budget are not filtered out) can certainly not be taken as evidence of a complementary relationship between the two products.

It should also be noted that the results presented here reflect the "average" responses of consumers, who are all in different situations. Consumers in different groups have different relevant options when faced with a price rise.

Canceling one connection if the price of the other is raised is an available option only for people who have both. The classic substitution effect could only be observed in people who have only one type of access, and are switching to the other service due to the price rise (and canceling the service that has become more expensive).

Taking into account own price elasticities and cross-price elasticities together, we can conclude that *fixed access is only a substitute for mobile access in a very limited way.* This is also confirmed by the low own price elasticity.

The own price elasticity of fixed telephone access is high, which suggests strong substitutability, but mobile cross-price elasticity is low. The question is whether this is due to fixed phones being substituted by some other service instead of mobile phones, such as VoIP calls. This is not likely, as indicated by the fact that other forms of communication (e.g. VoIP) provide a significantly poorer substitute for fixed *calls* than mobile calls do. (In case of a price rise on fixed calls, 3.9% would have substituted them with mobile calls, and only 0.9% would have chosen other means of communication; see *Table 17*). These results indicate that mobile phones still serve as the strongest candidate for substituting fixed phones.

The above explanations show why cross-price elasticity is low. Mobile and fixed access are by definition on/off options. If nearly everyone has a mobile phone, then there is limited scope for the effect that the traditional cross-price elasticity measures, i.e. consumers buying a mobile phone because of rising fixed prices. In an extreme case, where every single person has a mobile phone, this cross-price elasticity will not be positive. Therefore, the traditional concept of cross-price elasticity is not suitable for analyzing the type of substitution relationship where the consumer has both types of access and decides to cancel the fixed subscription and substitute it by using their mobile phone more. Yet this may be the most common consumer response today.

To examine these results in a wider context, it can be observed that the following phases followed each other in the development of Hungarian fixed and mobile telephony over the last twenty years:

1. an increasing number of consumers bought a mobile phone in addition to their fixed phone (complementary role, 1990s);
2. more and more of the traffic moved to mobile networks, owing to falling mobile prices, convenience and the fact that more and more people had mobile phones and mobile to mobile calls are cheaper than fixed to mobile (stronger usage substitution, around the turn of the millennium);
3. cancellation of fixed phones due to low usage, as maintaining the subscription was not worthwhile (after 2000) because significant monthly fees and declining usage meant that the total per-minute cost was getting progressively higher for consumers.

The cross-price elasticity of fixed access in case of increasing prices would be the best indicator of a substitution effect in the first of these phases, when most consumers still had only fixed access and might have switched to a mobile and canceled the fixed phone at the same time as a result of a price increase. However, at the time of the study, due to the widespread nature of mobile telephony, the indicators of phases 2 and 3 are already more suitable for examining the relationships between demand for fixed and mobile phones. These indicators suggest a significant element of substitution. Phase 2 is characterized by usage substitution. Examining specific calls shows that usage substitution using mobile phones is relatively low: 3.9% in case of a 20% price increase. This cross-price elasticity value of nearly 0.2 is not insignificant, especially considering that this is a realistic alternative only for people who have both access types (who make up less than 40% of subscribers). We also need to take into account that asking the question itself – by referring to substitution in a specific situation – measures short-term, non-planned effects, and therefore somewhat underestimates substitution. The most relevant indicator of phase 3 is the own price elasticity of fixed access, which our data show to be rather high.

It should be noted that access substitution, usage substitution and complementarity work differently in the three phases. In the first phase, when the penetration level of both services is generally increasing, the two connection types are complementary. Phase 1, the appearance of widespread mobile access (access complementarity) made it possible for usage substitution to emerge between the two services in phase 2, with fixed calls being substituted by mobile calls progressively more often. This process led to access substitution in phase 3, as fixed telephony (taking into account the average per-minute rate including the monthly fee) became more and more expensive compared to mobile services. It is also true that in this phase the cancellation of fixed subscriptions is associated more with increased mobile traffic than with increased mobile access.

Our interpretation of the data is that *fixed access is being substituted by mobile access*, despite the fact that cross-price elasticity is relatively low. With mobile phone penetration nearing 100%, access substitution can best be detected by examining the cross-price elasticity of usage and the own price elasticity of fixed access. In these circumstances, one cannot expect the number of mobile subscriptions to increase when the price of fixed access increases (since nearly all consumers already have a mobile subscription), but the cancellation of fixed lines and the parallel increase in mobile traffic clearly show that consumers are substituting fixed access with mobile.

From a regulatory perspective, too, the primary consideration is own price elasticity, not cross-price elasticity. What needs to be examined using the hypothetical monopoly (or SSNIP) test when determining relevant markets is whether raising the price of a product would be profitable. If own price elasticity is low, then the answer is yes and the product can be viewed as a separate market. If own price elasticity is high, then a price rise would generate a loss, and the market needs to be expanded by adding the closest substitute product. This is where cross-price elasticity has a role to play, in helping to identify the strongest of the candidate substitute products.

Thus, *the most important conclusion drawn on the basis of the substitution survey is that the own price elasticity of fixed access is high, and the closest substitute is mobile access; therefore, expanding the fixed access market by adding the mobile access market may be justified from a regulatory perspective. On the other hand, the own-price elasticity of mobile access is so low that even if there is a certain amount of substitution towards fixed telephones, it cannot be large enough to raise doubts regarding the definition of the mobile market as a separate relevant market.*

The results in the context of previous research

Above we identified three phases in the demand for fixed and mobile telephone services. Previous analyses of fixed-mobile access substitution generally only covered the first of these phases. For instance, *Horváth–Maldoom* [2002] and *Rodini–Ward–Woroch* [2003] analyzed the effects of fixed prices on purchases of new mobile subscriptions, while *Sung–Kim* [2002] studied the relationship between the cancellation of fixed subscriptions and the purchase of new mobile subscriptions. *Rodini–Ward–Woroch* [2003] measured the cross-price elasticity of mobile subscriptions and fixed prices in the US at 0.18 in 2001 and 0.13 in 2001. According to the authors' calculations, the own price elasticity of mobile subscriptions was –0.6 and –0.43 in the two years, respectively.

Our research generally found lower cross-price elasticities than this, which is entirely understandable in the light of the analysis presented in the previous section: the Hungarian market was for the most part already in phase 3 at the time of data collection. We also measured lower own price elasticity in mobile services than the values measured in the US. As previously explained, the cross-price elasticity of mobile phone services and rising fixed prices was low in Hungary in 2009 because of the high mobile phone penetration. The own price elasticity of mobile services measured with regard to price reductions was low for the same reason: even if a relatively high proportion of people with no mobile phone were to subscribe, this would still produce a low level of elasticity as there is only a very small pool from which to draw new subscribers.

However, this is not the reason why we measured mobile service demand as relatively inelastic with regard to price rises. This needs to be interpreted bearing in mind that along with the spread of mobile telephony, mobile phone prices have been falling continuously and quite significantly over the last decade,[9] thus changing the relative price of mobile and fixed telephone traffic. Calling a mobile phone number from a mobile phone now costs about half as much as making the same call from a fixed phone – and prices were almost equal ten years ago. Therefore, it is not surprising that people would generally not give up their mobile access even in case of a 20% price increase: they could call mobile numbers significantly more cheaply even after the increase. If the two prices were close to each other, more people would probably consider this option.

Thus, it is easy to see why, given the market's current characteristics, elasticities are entirely different than a few years earlier. The price ratios played a significant role in this. Therefore, even though they were not reached using econometrics methods, our results probably approximate the current Hungarian market situation much better than any estimate of elasticities based on the results of a thorough

[9] See the European Commission's annual implementation reports on this issue.

econometric analysis carried out in a country with a completely different market, or at a different time.

In addition to the papers known at the time of the 2009 data collection, a few papers have been written subsequently using newer data and data that overlap with our data collection period. The results presented in this paper need to be viewed in the light of these results as well, because they focus mostly on the phenomena in phases 2 and 3. In general, these papers use careful wording regarding the issue of fixed-mobile substitution, placing the emphasis not on specific numerical results but rather on the general trends they indicate. They identify a basic relationship of substitution between the two services, but at a relatively low intensity.

Suarez–Garcia-Marinoso [2013] examined the main driving forces behind fixed-mobile substitution on the Spanish residential market between 2004 and 2009 using household panel data, focusing specifically on situations where households substitute fixed access with mobile access by canceling their fixed subscription. The conclusion was that this type of substitution does happen, but only at a low rate (0.35%/quarter). Furthermore, the authors found that the likelihood of substitution is affected much less by the amount and composition of the sum spent on fixed telephony than by the lack of Internet access, the existence of mobile access and the demographic characteristics of the household (young age, single person household, sparsely populated area), all of which are factors that significantly increase the likelihood of substitution.

Grzybowski–Verboven [2013] studied fixed-mobile substitution in 27 EU countries between 2005 and 2011 based on Eurostat survey data. The authors found that mobile and fixed telephones are generally seen as potential substitutes for each other in households, and the parameter measuring substitution is significantly negative. At the same time, the level of substitution was much lower in the first three years (2006–2008) than in subsequent years. According to the authors' estimate, the existence of mobile phones only reduced fixed penetration by 6% in 2007. In 2012, however, fixed penetration would have been 14% higher if mobile phones had not existed. At the same time, significant differences were found between individual households and different regions: in many cases, households used the two services in a complementary manner. The rate of substitution was much higher in Central and Eastern Europe than in Western Europe. According to the conclusions of the study, households with different socioeconomic characteristics have different attitudes to fixed-mobile substitution.

Macher at al. [2013] analyzed the telecommunications decisions of US households between 2003 and 2010, using data from the annual National Health Interview Survey. The results of the empirical analysis of close to 200,000 observations supported the notion that mobile phones became close substitutes of fixed phones between 2003 and 2010. The study even attempted to divide "full substitution" into its constituent elements. Only a relatively low rate of substitution was observed between 2003 and 2006: only about half of the consumers who previously only had a fixed connection and bought a mobile connection switched to a "mobile-only" arrangement, with the other half subscribing to both services. After 2007, the dominant model became canceling the fixed service and using a mobile phone alone.

Barth and Heimeshoff [2012] estimated the effects of mobile substitution on fixed traffic based on panel data regarding 16 Western European countries between 2004 and 2010. According to their results, the own price elasticity of fixed calls varied between –0.1378 and –0.1661 in the short term, and between –0.4692 and –0.4867 in the long term. The effect of mobile prices on fixed traffic (cross-price elasticity) was between +0.1250 and +0.1256 in the short term, and between +0.4254 and +0.3680 in the long term.

The role of mobile telephony as a functional substitute for fixed telephony is obvious; however, substitution in the economic sense regarding a specific location can only established by analyzing that specific market situation.

REFERENCES

Ahn, H.–Lee, J.–Kim, Y. [2004]: Estimation of a Fixed-Mobile Substitution Model in Korean Voice Telephony Markets. Working Paper, Seoul.

Ahn, H.–Lee, M.-H. [1999]: An Econometric Analysis of Demand for Access to Mobile Telephone Networks. Information Economics and Policy, Vol. 11, No. 3, pp. 297–305.

Baker, P.–Blundell, R.–Micklewright, J. [1989]: Modelling Household Energy Expenditures Using Micro-Data. The Economic Journal, 99, pp. 720–738.

Barth, A. K.–Heimeshoff, U. [2012]: How Large is the Magnitude of Fixed-Mobile Call Substitution? Empirical Evidence from 16 European Countries, Düsseldorf Institute for Competition Economics, Discussion Paper

Boxall, P. C.–Adamowicz, W. L.–Swait, J.–Williams, M.–Louviere, J. [1996]: A Comparison of Stated Preference Methods for Environmental Valuation. Ecological Economics, 18, pp. 243–253.

Eales, S. E.–Unnevehr, L. J. [1988]: Demand for Beef and Chicken Products: Separability and Structural Change. American Journal of Agricultural Economics, 70, pp. 521–532.

Edgerton, D. L. [1997]: Weak separability and the Estimation of Elasticities in Multistage Demand Systems. American Journal of Agricultural Economics, 79, pp. 62–79.

Fu, W. W. [2004]: Termination-Discriminatory Pricing, Subscriber Bandwagons, and Network Traffic Patterns: The Taiwanese Mobile Phone Market. Telecommunications Policy, 28, pp. 5–22.

Garbacz, C.–Thompson, H. G. [2007]: Demand for telecommunication Services in Developing Countries. Telecommunications Policy, 31, pp. 276–289.

Goettler, R. L.–Clay, K. [2006]: Price Discrimination with Experience Goods: Sorting-Induced Biases and Illusive Surplus. Working Paper, Carnegie Mellon University.

Grajek, M. [2007]: Estimating Network Effects and Compatibility in Mobile Telecommunications. Working paper, Social Science Research Center, Berlin.

Grzybowski, L.–Verboven, F. [2013]: Substitution and Complementarity between Fixed-line and Mobile Access, Working Paper, 13–09.

Hayes, J. H.–Wahl, T. I.–Williams, G. W. [1990]: Testing Restrictions on a Model of Japanese Meat Demand. American Journal of Agricultural Economics, 72, pp. 556–566.

Horváth, R.–Maldoom, D. [2002]: Fixed-Mobile Substitution: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables. DotEcon Discussion Papers No. 02/02.

Huang C. [2006]: Estimating Demand for Cellular Phone Service under Nonlinear Pricing. Quantitative Marketing and Economics, 4, pp. 371–413.

Infrapont [2006]: A fix-mobil helyettesítés empirikus vizsgálatának módszerei. Tanulmány a Nemzeti Hírközlési Hatóság részére. [Empricical methods for istudying fixed- momobile substitution. Study for tha National Communication Authority of Hungary] Manuscript, Budapest.

Iyengar [2004]: A Structural Demand Analysis for Wireless Services under Nonlinear Pricing Schemes. Manuscript, University of Pennsylvania.

Kim, H.-S.–Kwon, N. [2003]: The advantage of Network Size in Acquiring New Subscribers: A Conditional Logit Analysis of the Korean Mobile Telephony Market. Information Economics and Policy, 15, pp. 17–33.

Lambrecht, A.–Seim, K.–Skiera, B. [2005]: Does Uncertainty Matter? Consumer Behavior under Three-Part Tariffs. Working Paper, University Of California, Los Angeles.

Lancaster, H. J. [1966]: A New Approach to Consumer Theory. The Journal of Political Economy, 74, pp. 132–157.

Lancaster, H. J. [1971]: Customer Demand: A New Approach. Columbia University Press.

Lee, J.–Kim, J.–Ahn, J. [2006]: A dynamic competition analysis on the Korean mobile phone market using competitive diffusion model. Computers & Industrial Engineering, 51, pp. 174–182.

Liebowitz, S. J.–Margolis, S. E. [2002]: Network Effects. In: *Cave, M.–Majumdar S.–Vogelsang, I.* (eds.): Handbook of Telecommunications Economics. North Holland Publishing Co., Amsterdam.

Littlechild, S. C. [1975]: Two-Part Tariffs and Consumption Externalities. The Bell Journal of Economics, 6, pp. 661–670.

Louviére, J.–Timmermans, H. [1990]: Stated Preference and Choice Models Applied to Recreation Research: A Review. Leisure Sciences, 12, pp. 9–32.

Macher. J. T.–Mayo, J. W.–Ukhaneva, O.–Woroch, G. A. [2013]: Demand in a Portfolio-Choice Environment: The Evolution of Telecommunications. Georgetown McDonough School of Business Research Paper, No. 2012-19.

Miravete, E. J. [2002]: Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans. Review of Economic Studies, 69, pp. 943–971.

Narayanan, S.–Chintagunta, P. K.–Miravete, E. J. [2007]: The Role of Self Selection, Usage Uncertainty and Learning in the Demand for Local Telephone Service. Quantitative Marketing and Economics, 5, pp. 1–34.

Okada, Y.–Hatta, K. [1999]: The interdependent Telecommunications Demand and Efficient Price Structure. Journal of Japanese and International Economies, 13. pp. 311–335.

Rodini, M.–Ward, M. R.–Woroch, G. A. [2003]: Going Mobile: Substitutability between Fixed and Mobile Access. Telecommunications Policy, 27, pp. 457–476.

Sellen, D.–Goddard, E. [1997]: Weak Separability in Coffee Demand Systems. European Review of Agricultural Economics, 24, pp. 133–144.

SUAREZ, D.–GARCIA-MARINOSO, B. [2013]: Which are the drivers of fixed to mobile telephone access substitution? An empirical study of the Spanish residential market, Telecommunication Policy, 37, pp. 282–291.

SUNG, N. [2003]: Fixed-Mobile Substitution and Regulatory Model in Telecommunications Markets. Telecommunications Market, 46, pp. 22–35.

SUNG, N.–KIM, M. [2002]: A Study on Call Substitution between Long-Distance and Mobile Services Using Estimation of Long-Distance Call Demand. Research on International Economics, 8, pp. 227–246.

TÁRKI [2004]: Mobiltelefon-szolgáltatások fogyasztói szokásainak felmérése a 14 éves és idősebb magyar lakosság körében. Elemző tanulmány, [A survey of consumer attittude in consumption of mobile telephony services in the fourteen plus population of Hungary. An emprical study] Tárki, Budapest.

TAYLOR, L. D. [1994]: Telecommunications Demand in Theory and Practice. Kluwer Academic Publishers, Dordrecht.

TRAIN, K. E.–MCFADDEN, D. L.–BEN-AKIVA, M. [1987]: The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices. Rand Journal of Economics, 18, pp. 109–123.

TRAIN, K.–WILSON, W. W. [2008]: Estimation on stated-preference experiments constructed from revealed preference choices. Transportation Research, Section B), 42, pp. 191–203.

TSENG, F. M.–CHIU, Y. J. [2005]: Hierarchical Fuzzy Integral Stated Preference Method for Taiwan's Broadband Service Market. Omega, 33, pp. 55–64.

WARD, M. R.–WOROCH, G. A. [2004]: Usage Substitution between Mobile Telephone and Fixed line in the U.S., Working Paper.

• *Károly Miklós Kiss* •

THE HUNGARIAN POSTAL SECTOR

This paper summarizes recent developments on the Hungarian postal markets and the manner, in which they are regulated. In the first part of this study, I describe the evolution and main features of the relevant Hungarian regulations. In the second section, I outline the process of market opening and the emergence of market competition, including key Hungarian legal cases concerning anti-competitive market conduct. In the third section, I assess the economic performance of the Hungarian postal sector and its various segments, examining changes in prices over time and price regulations in a separate section. Finally I provide a summary of service quality indicators.

THE DEVELOPMENT OF POSTAL REGULATIONS IN HUNGARY

Reform of Hungary's postal regulations began in the early '90s, drawing on the postal regulations of the European Union as they developed. Following institutional separation of postal, telecommunications and broadcasting services in 1990, separate legislation was introduced for the three sectors from 1992 to 1993. Act XLV of 1992 on Postal Services replaced the earlier Act II of 1964. The 1992 Act indicated, to a certain extent, the intention of opening the hitherto entirely monopolized postal markets and allowing other players to be present, subject to possession of the relevant license. The conditions set down in law did not enable actual market entry.

The concept of the convergence of the information and communications markets, which gained ground in the '90s, also impacted on Hungarian legislation: the need for common framework regulations for these sectors arose as part of the communications modernization program that got under way in the late '90s. Accordingly, Act XL of 2001 on Communications set the framework regulations for various information and communications activities. When regulating postal services under the Communications Act, the first EU postal services directive, namely Directive 97/67/EC of the European Parliament and of the Council, was taken into account, with the result that the fundamental principles of provision of universal postal services and market opening were adopted. The 2001 Act opened up those market segments involving non-universal services to new entrants. The impact was reflected in the increase in the number of market players – 18 registered service providers were present in these market segments in 2002 (the first year of the Communications Act), with that number rising to 43 in 2003.

However, the delegation of the European Commission that assessed implementation of the first postal services directive, criticized several aspects of the Hungarian Act on the grounds that it extended the privileges of the universal service provider. Criticisms were made in two important respects: first, regarding extension of the scope of reserved services (to include cash transfers and payment intermediary services) contrary to the directive, and, second, regarding *contributions to the subsidy fund for universal postal services*. In the meantime the second EU postal services directive (Directive 2002/39/EC) was adopted, with the result that there was both an internal and external need for reform of the Hungarian legislation. However, while preparing the amendment to the Act, it became increasingly clear that it was not expedient to set out the regulations in a joint law owing to the unique characteristics of the telecommunications and postal sectors. For that reason, these sectors were again regulated separately. Act CI of 2003 on Postal Services and Act C of 2003 on Electronic Communications were passed. The new Postal Services Act, which drew on the second EU postal services directive, was adopted, and within a short time the key government decrees and ministry decrees supplementing the Postal Services Act came into force.

As the regulations evolved, the scope of universal services was defined with increasing precision based on the directives. Although the 1992 Act used different terms (namely basic postal services and basic provision of services), it introduced numerous elements of universal services (for example, mail not weighing more than 2 kilograms and money orders). The provision of universal postal services as defined in the 2001 Communications Act was obliged to provide countrywide services with respect to domestic packages not weighing more than 10 kilograms, postal payment services and financial services, and the following extra services: registered mail, delivery with return receipt and insured mail. It also introduced a number of quality indicators.

The 2003 Postal Services Act extended the scope of universal services to include packages weighing up to 20 kilograms and set new, more exacting quality requirements. In addition, with respect to the manner of providing postal services, the obligations tied to universal services were defined with increasing precision. That manifested itself, for example, in provisions concerning household delivery to addresses beyond municipal boundaries. The Communications Act and the 2003 Postal Services Act also regulated access to universal services (minimum opening hours of permanent postal service points and minimum times that mobile services and mobile post offices are required to spend at the designated access points). These two Acts, unlike the 1992 Act, no longer allowed the universal service provider to not even attempt delivery of mail weighing over a specified amount to addresses outside municipal boundaries.

The 2003 Postal Services Act defined *universal* and *reserved* services as follows:

Scope of universal postal services:
- postal services involving domestic and international letters, addressed advertising mail and forms not weighing more than two kilograms;
- postal services involving domestic and international packages not weighing more than 20 kilograms;
- postal services involving domestic and international mail not weighing more than seven kilograms and containing Braille;
- extra services (registered mail, delivery with return receipt and insured mail) that can be ordered by the person posting the mail to supplement the services defined above.

Reserved postal services within the scope of universal services:
- services involving domestic or international letters or addressed advertising mail not weighing over 50 grams if the price of the service is less than two and a half times (earlier three and a half times) the price of a letter in the first weight category of the fastest service category of the universal services segment,
- postal services involving official documents, unless otherwise regulated by law or government decree.

With respect to market competition, another important question is how market entry is regulated in the various segments. The earlier postal services directive set out the possibility for the various Member States to apply a licensing procedure to non-reserved universal services, recommending two forms of licensing: the more liberal general license and the more stringent individual license. The practices of the Member States with regard to licensing can be categorized into four groups. In increasing order of stringency and difficulty of entry, these groups are as follows: *a)* no license is required; *b)* a general license is required; *c)* an individual license is required only for letter mail in the non-reserved universal segment or a limited scope of such letters; *d)* an individual license is required for all universal postal services (including packages). Hungary was among the countries in the last group, i.e. those countries that apply the strictest licensing procedures. The earlier 2003 Postal Services Act classified postal activities into three groups regarding entry requirements.

1. Non-universal postal services (express mail, courier mail, integrated mail and document exchange) can be performed subject to *registration* (which is a straightforward process, requiring only a formal declaration of intent).
2. Non-reserved universal postal services can be performed subject to an individual license (licensed service providers are not only entitled, but are also obliged to perform the universal service for the mail types and geographical area, to which the license applies).

3. The universal postal service provider (Magyar Posta Zrt. (Hungarian Post)), which has the exclusive right to provide reserved postal services, was *designated by the Act*.

To what extent the requirement for an individual license represents a barrier to entry also depends on the related conditions and obligations prescribed by the regulations and the national regulatory authority. One indicator of the extent, to which the requirement restricts entry may be the number of service providers present in the various universal postal service sectors. Even among the countries in the last group listed (countries where universal postal services can only be performed subject to an individual license) significant differences can be observed. (While for example, in 2005 more than 200 service providers operated in certain universal service segments in Italy, where an individual license is also widely prescribed, in Hungary there was not a single service provider aside from the designated universal service provider.)

Hungary was granted a moratorium until December 31, 2012 to transpose the third postal services directive into domestic law and to fully open its postal markets to competition. Following protracted negotiations and drafting, Parliament adopted the new Postal Services Act in fall 2012 (Act CLIX of 2012 on Postal Services) virtually at the last moment. The government decrees and ministry decrees for implementation of the Act then came into force in December 2012 (Government Decree no. 335/2012 (XII. 4.), Government Decree no. 336/2012. (XII. 4.), National Development Ministry Decree no. 67/2012. (XII. 15.).

The most important change in the new legislative environment was that almost all services were removed from the scope of reserved services, thereby opening up the hitherto monopolized postal service segments to competition. The only field, in which the universal postal service provider retains the exclusive right to provide postal services is the official documents segment.

The new Postal Services Act distinguishes between three types of postal services:

1. *Universal postal services:*
   a) non-registered domestic or international mail (differing from the mail defined in *b)–d)*) not weighing more than two kilograms;
   b) domestic or international packages not weighing more than 20 kilograms;
   c) domestic or international mail containing Braille;
   d) official documents.
   However, only services provided pursuant to the general terms and conditions constitute universal services. If the universal postal service provider provides the service according to terms and conditions set out in an individual agreement that differ from the terms and conditions for the universal postal service, according to an individually set price, then the service qualifies as a service replacing a universal service (hereinafter: replacement universal service).

2. Any postal service involving any mail type qualifies as a *replacement universal service* that does not come under the scope of the services specified in Section 8 of the Act (services that do not replace universal services; see the summarized group of the services in the section below) and is not provided by the postal service provider under its universal postal service obligation. In other words, replacement universal services are postal services involving all mail in the universal category if they are not provided by the universal service provider, or if they are provided by the universal service provider, but not under the general terms and conditions applicable to the universal service.
3. The below postal services that provide added value compared to universal postal services or replacement universal services constitute *postal services that do not replace universal postal services* (hereinafter: non-replacement universal postal services).
   a) courier service,
   b) express mail service,
   c) international EMS service
   d) mail service including at least one of the below supplementary services that represent considerable added value:
      - mail tracking,
      - service with delivery time guarantee,
      - cash-on-delivery service,
      - service whereby the mail is delivered at a time individually agreed with the recipient after the postal service provider has taken receipt of the mail,
      - delivery solely to the person designated as the recipient,
      - other supplementary services connected to mail delivery that are tailored to the needs of the sender and are provided on the basis of an individual agreement, provided they do not come under the scope of supplementary services that must be offered in connection with the universal postal service, and whose use means that from the perspective of the user the given service no longer qualifies as a replacement universal postal service.

   Non-replacement universal services may be performed by any business organization based on registration (services subject to registration). By contrast, a license is required for performance of replacement universal services (services subject to a license). Only the universal service provider may provide universal services. The Act designates Magyar Posta Zrt. as the universal service provider until December 31, 2020.

   The Act prescribes that the prices of universal services should be cost-based, but only the pricing of individually posted domestic letters not weighing more than 50 grams and domestic services involving official documents is regulated. The method for regulating these prices is set out in National Development Ministry Decree no. 67/2012. (XII. 15.). In place of the earlier regulatory prices, price cap regulations were introduced (which will be discussed later under “prices over time”).

The new Act creates the possibility for a *Subsidy Fund for Universal Postal Services* to finance the undue additional burden posed by the universal service obligation. Payments to and from that account are governed by a separate government decree (Government Decree no. 336/2012. (XII. 4.). All service providers, including the universal service provider that provides replacement universal services is obliged to make contributions to the subsidy fund from their revenue from replacement universal services. This contribution is payable according to market share on the relevant market (replacement universal services), according to the formula set out in the government decree. However, each postal service provider is not obliged to pay more than 10 percent of its net revenue from replacement universal services. The level of the contribution is determined by a formula set out in the decree according to the undue additional burden posed by the universal service. The amount of the undue additional burden is established by the National Media and Infocommunications Authority.

The Hungarian postal sector regulator

In Hungary state service provision and regulatory activities were separated in 1989. It was at that time that the separate, state-owned Hungarian Telecommunications Company, the Hungarian Broadcasting Company and the Hungarian Post Company were established. Simultaneously, the Ministry of Transport, Communications and Water Management assumed the role of regulatory authority. In 1993, based on the authorization of the Telecommunications Act (Act LXXII of 1992), a combined communications authority was created: an independent central authority (communications authority) operating under the control and supervision of the Transport, Communications and Water Management Minister, for the performance of postal, telecommunications and frequency management authority tasks. Its central body was the Communications Inspectorate General and its local bodies were the regional communications inspectorates. As successor to the Communications Inspectorate General the Communications Inspectorate was established by Government Decree 248/2001. (XII. 18), pursuant to which it became a central public administration body with legal personality and nationwide competence, operating under the control of the government and the supervision of the Minister in Charge of the Prime Minister's Office. In January 2004 the National Communications Authority (NHH) was established as legal successor to the Communications Inspectorate pursuant to the Act on Electronic Communications, and was vested with significantly greater powers than its predecessor (as it was granted regulatory, market-shaping and, under certain circumstances, legislative powers by the law).

The current regulatory authority, the National Media and Infocommunications Authority (NMHH) was established on August 11, 2010 by Parliament with Act LXXXII of 2010 on the amendment of certain laws governing media and communications. The NMHH was created by merging the National Radio and Television Authority (ORTT), which regulated the media, and the National Communications Authority (NHH), which regulated the communications sector.

The NMHH is a central budgetary body with independent financial management; it covers the costs of its operation from its own revenues. It is *tasked with* promoting the smooth, effective operation and development of the communications markets: it protects the interests of those performing communications activities and of communications users and it strives to ensure the development and maintenance of fair and effective competition and oversees legal compliance of entities performing communications activities. However, the independence of the NMHH is reduced by certain provisions of the new Act: 1) the head of the authority may be appointed by and removed from office by the prime minister, 2) the duties and powers of the regulatory authority may be prescribed not only by law, but also by government decree. Moreover, the new Act states that the NMHH is involved in the implementation of the government's communications policy.

Market structure and market size

In terms of competition, the Hungarian postal sector can be broken down into three categories up to the time of total market opening:

- absence of competition (because it was prohibited): reserved universal services: segment involving letters weighing below 50 grams,
- minimal competition: non-reserved universal services (until 2009 Magyar Posta did not have a competitor; in 2009 a competitor emerged with a small market share),
- strong competition: non-universal postal services, for which the service provider needs to be registered (courier services, express mail services, integrated postal services, postal intermediary activities, document exchange services).

International experiences show that, in addition to general reduction of reserved services, competition was primarily boosted by fully liberalizing segments of the addressed mail market (eliminating the reserved scope determined by weight limits in those segments), in which competitors were rapidly capable of offering competitive services. For example, the opening of the outbound international mail segment, the total liberalization of addressed advertising mail (e.g. in the Netherlands) or liberalization of the market involving (otherwise reserved) postal services that are supplemented by services with added value (e.g. in Germany the "D-license" or hybrid mail in Bulgaria). In Hungary's case, these market segments (within the prescribed weight limits) were classed among reserved services until the prescribed deadline for total market opening. The Hungarian regulations have consistently made full use of any flexibility allowed by the effective postal services directives to make the scope of *reserved* services as broad as possible. Throughout, Hungary has been in the group of countries that have made the least advances in terms of market opening.

In the case of non-reserved universal postal services, the earlier postal services Act (Act CI of 2003 on Postal Services) tied provision of such services to obligations that acted as a barrier to entry for a long time. As mentioned above, the provision of universal services was subject to a license in Hungary pursuant to the earlier Postal Services Act. Licensed service providers were not only entitled, but also obliged to provide universal services in the geographical region and for the mail types, to which the license applied. A license could be requested for provision of non-reserved universal postal services in one of the following public administrative regions:

- one or more settlements, with the exception of towns or cities;
- at least one county with the exception of Pest County;
- Pest County and at least one county;
- Budapest and – with the exception of Pest County – a combination of at least two counties;
- countrywide.

The obligations entailed by the license prevented potential market entrants from picking and choosing the more profitable market segments within universal services. However, they also hindered the development of competition in the non-reserved universal services segments. It was not until 2009 that a competitor emerged in those market segments, namely Feibra Kft. (a subsidiary of Österreichische Post (Austrian Post)). Feibra Kft. commenced operations in 1990, and for a long time focused on the unaddressed advertising mail market. In 2008, following a merger with Cont Média, it became the largest market player on the unaddressed advertising mail market, with a market share of 60 percent. Having established the necessary infrastructure, it entered the market for mail weighing more than 50 grams, which is classified as a universal service, in 2009. No reliable data are available with respect to the market share it achieved in that universal mail segment, but it is highly probably that, in line with the experiences of other countries with a liberalized postal market, it did not capture a significant share of Magyar Posta's market.

Following the complete market opening, in January 2013 Feibra Kft. submitted an application to provide replacement universal services (for which a license is required). However, the outcome of the authority assessment indicates that even the new Postal Services Act failed to lessen the entry barriers resulting from the regulations. Feibra Kft. was not granted a license by the regulatory authority for the postal sector (the authority established that the company's accounting procedures were not sufficiently transparent). Another applicant came on the scene in 2015. City Mail Hungary Kft.

submitted a license application to the authority to perform replacement universal services in Hungary, in addition to the postal services it was providing on the basis of registration. The NMHH also rejected that license application on the grounds that the necessary conditions for reliable delivery performance of the service provider in the whole public administrative area of Budapest and material conditions for providing the service were not in place. As a result, no competitor managed to enter the most recently opened market segments within the scope of universal services.

The anti-competitive tax conditions also act as a barrier to entry. Despite much criticism of the EU system of exemption from VAT, it is still the case that only a few Member States apply an equal level of VAT to the universal service provider and the other service providers on the market. In most countries either all activities of the universal service provider or its services involving mail that comes under the scope of universal services are exempted from VAT or receive a VAT discount, which significantly distorts competition. In Hungary, like in most Member States, universal postal services are exempt from VAT (on the grounds that they are in the public interest) [1], while replacement services are not, with the result that Magyar Posta's competitors are at a significant competitive disadvantage in those segments. In order for entrants to the universal services segment to make their services attractive (for instance, to the financial sector or public sector, which cannot reclaim VAT), they have to offer their services at below the universal service provider's prices, *less* the VAT amount, which significantly distorts price competition. That limits the range of potential customers, and thereby the commercial opportunities of competitors.

There is, however, fierce competition in the field of non-universal postal services. *Table 1* shows steady growth in the number of service providers in the various service categories (although the number of deleted service providers also rose sharply in 2009, the year following the financial crisis). Since 2010 the number of

TABLE 1 • **Number of non-universal postal service providers by service category**
(based on service categories designated at the time of registration, number of service providers)

	2005	2006	2007	2008	2009	2010	2015*
Total number of registered service providers in the various non-universal service categories	*96*	*111*	*137*	*178*	*237*	*265*	*179*
Deleted in the respective year	8	24	49	66	92	36	n/a
Other courier services	54	75	88	114	139	150	94
Only express mail	0	0	0	0	0	0	6
Integrated	0	10	13	12	12	15	10
Combined (courier, express, integrated, and intermediary services)	44	24	30	45	76	100	69
Postal intermediary services	0	2	6	7	10	20	2

* Data as of July 1, 2015 from the NMHH's searchable database.
Source: Data provided by the NMHH.

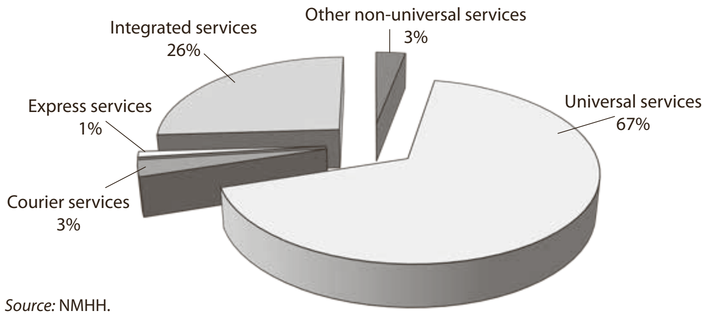
[1] Section 85 (1) of Act CXXVII of 2007 on Value Added Tax.

registered service providers has declined; according to the searchable database of the NMHH there were 179 postal service providers providing non-universal services at the start of July 2015.

Magyar Posta's main competitors in the non-universal packages segment and the express mail and courier segments are the large international integrator companies, namely DHL, TNT, FedEx, UPS, GLS, and Road Parcel Logistics. In those market segments, the role of international service providers has continuously grown, with Magyar Posta's market share falling from almost 30 percent in 2002 to around 15 percent in 2007. In 2007, the CR5 index indicating market concentration (the combined market share of the five companies with the largest market shares) in these segments was 75 percent (*ITA–WIK* [2009a] p. 60.). In addition to the large international service providers, there are also several small enterprises, which chiefly provide local courier and express mail services. In these competitive segments, Magyar Posta's market share (based on revenue) in 2006 was as follows in the various segments: 82 percent in the express mail segment, 1.5 percent in the courier segment (Magyar Posta is not present in the integrated mail services segment). (*Ecorys* [2008] p. 56)

Magyar Posta met with a considerable challenge in the newspaper distribution segment in 2007. At the end of 2006, three large newspaper publishers (Ringier, Népszabadság and Sanoma) and the Fiege Group established a new newspaper distribution enterprise called Médialog Fiege Zrt., which set up an independent distribution network in Budapest and another 104 towns and cities. As a result, on the newspaper distribution market, Magyar Posta had to scale down its activities to smaller settlements not affected by Médialog.

Figure 1 below shows the weighting of the various market segments in 2010. The market segments with actual competition accounted for approx. a quarter of the total postal industry in 2006. That figure increased to 32 percent in 2008, and was at roughly the same level in 2010 (33 percent). Between 2006 and 2008, revenue in the universal services segment grew to a lesser extent than that of the competitive segments, so the weighting of the competitive segment within the total revenue of the postal sector increased. However, between 2008 and 2010 the service providers in the competitive seg-



Source: NMHH.

FIGURE 1 • **Breakdown of postal revenues by service type in 2010** (based on revenue) ment were hit harder by the effects of the financial and economic crisis than the universal service provider (which held a monopoly or near monopoly in many of its markets).

Table 2 shows the volume of mail handled by the various competitive segments. We can see from the mail volume data that integrated postal services are the largest market segment within non-universal postal services, as is also shown by the revenue data set out in *Table 4.* Of the competitive segments, the combined revenue of the service providers is by far the greatest in this market segment (integrated mail services: HUF 37,585 billion, courier services: HUF 3,560 billion, express mail services HUF 1,411 billion).

From *Table 3* we can see that the number of letters handled by Magyar Posta unlike packages), contrary to international trends, has not yet fallen significantly (although, despite the minor fluctuations, a slight downward trend can be observed). In the case of packages, the decline can be attributed to the fact that some of the mail volume has shifted to services with greater added value (e.g. express mail services, courier services and integrated mail services). The majority of letters even in 2010 were in the reserved (i.e. monopolized) segment involving mail weighing below 50 grams (94 percent of letters and 73 percent of addressed advertising mail).

TABLE 2 • **Breakdown of mail in the competitive segments, 2010**

	Courier	Express	Integrated	*Total*
Domestic delivery	1,944,761	493,448	17,371,463	*19,809,672*
Posted abroad/inbound	0	79,725	2,229,808	*2,309,533*
Outbound international	0	31,913	1,544,715	*1,576,628*
Forwarded by another service provider	0	9,683	49,834	*59,517*
Forwarded to another service provider	1,921	36,329	34,123	*72,373*
Undelivered	148	17	33,677	*33,842*
Total	*1,946,830*	*651,115*	*21,263,620*	

Source: NMHH.

TABLE 3 • **Volume of domestic mail in the universal services category over time, 2005–2010**
(number of mail items)

	2005	2006	2007	2008	2009	2010
Standard letters	546,581,962	577,675,535	625,071,170	644,007,359	611,849,821	625,983,883
Registered	91,653,008	89,626,214	80,382,591	78,370,999	75,954,598	71,557,794
Value-declared letters	328,715	465,036	351,082	312,881	325,161	305,105
Official documents	38,307,238	39,640,710	43,774,374	45,799,263	46,712,382	50,398,334
Addressed advertising mail	75,490,746	89,579,418	31,721,887	27,788,101	27,509,099	26,930,883
Forms	13,138,625	10,330,540	5,766,055	3,338,148	5,031,472	2,314,206
Packages	4,559,849	2,673,383	1,760,958	1,541,523	1,443,324	1,379,066
Of the above: insured	1,881,753	1,033,462	1,178,690	1,303,731	252,201	200,743
Universal mail	*770,060,143*	809,990,836	788,828,117	801,158,274	768,825,857	778,869,271

Source: Data provided by the NMHH.

From the data in *Table 4* we can see that in recent years the net sales revenue increased in both the non-competitive universal services segment and in the competitive express mail services and integrated mail services segments, while the aggregate revenue from courier services decreased significantly between 2008 and 2010. If we examine the development of these market segments by looking at the volume of mail, we can see that during the three years the mail volume of the universal service provider showed a negative tendency, with some fluctuations, while the growth rate of the mail volume in the competitive segments was typically greater than the revenue increase in those same segments.

Payment services and financial intermediary activities play a highly significant role at Magyar Posta. If we look at Magyar Posta's 2011 Annual Report (*Magyar Posta* [2011]), we can see that revenue from payment services and financial intermediary services accounted for approx. 32 percent of its total revenue *(Table 5).*

TABLE 4 • **Net sales revenue, broken down by market segment** (HUF millions)

Segment	2006	2008	2010	Percentage change (2006–2008)	Percentage change (2008–2010)
Universal service	74,000	85,001	95,346	+14.87	+12.17
Courier service	2,455	7,237	3,560	+194.79	–50.81
Express mail service	1,484	349	1,411	–76.48	+304.30
Integrated mail service	22,406	31,227	37,585	+39.37	+20.36

Source: NMHH and *Magyar Posta* [2006], [2008], [2010].

TABLE 5 • **Revenue of Magyar Posta over time, 2009–2011**

	2009		2010		2011	
Services	HUF millions	%	HUF millions	%	HUF millions	%
Letter mail services	83,977	48.8	85,311	49.6	85,786	49.0
Newspaper services	6,173	3.6	5,489	3.2	5,364	3.1
Logistics services	7 ,497	4.4	7,608	4.4	8,549	4.9
Postal finance services	46,133	26.8	45,989	26.8	46,942	26.8
Banking and investment services	7,487	4.3	7,635	4.4	7,509	4.3
Insurance services	1,198	0.7	1,360	0.8	1,730	1.0
Retail and postal outlet activity	7,540	4.4	6,460	3.8	6,868	3.9
International services	12,038	7.0	12,058	7.0	12,432	7.1
Total revenue	172,043	100.0	171,910	100.0	175,180	100.0

Source: Magyar Posta [2011] p. 14.

Competition authority and regulatory authority proceedings

The proceedings of the regulatory authorities can be divided into three main categories. The cases in the first group are typical of the earlier period of regulation of the postal markets and resulted from the particular features of those regulations, namely cases concerning interpretation of the boundaries of the reserved scope of services. Since the universal service provider was granted exclusive rights to provide services in certain sectors, the exact boundaries of those sectors were a key question in terms of competition. It was naturally in the interest of the universal service providers for the reserved segment to be as broad as possible.

The second category of proceedings involves cases investigating conduct that restricts competition. Such cases involve abuse by universal service providers of their market strength (discriminative price discounts, cross financing that distorts competition, pricing that drives out competitors, and problems concerning accessibility). The earlier system of reserved services meant that there were exclusive rights in certain market segments. As a result, one purpose of the regulations was to prevent cross financing between services provided in a monopoly and services provided under competitive conditions in the case of economic players with exclusive rights. As the proportion of competitive markets has increased, the regulator also needs to examine whether companies extend their established dominant position on one reserved market to their activities in liberalized segments, or whether they abuse their monopoly by discriminating between the various consumer categories. Since the market strength of the incumbent national postal companies had not declined significantly after the market opening in the previously reserved segments, such regulatory cases remained important following the total market opening.

The third main group of competition authority proceedings involves merger inspections resulting from the international expansion of the large universal postal service providers. This last group does not differ from competition authority proceedings in other industries,[2] so I will not discuss it in detail, and will instead provide a summary of cases in the first two groups.

Proceedings concerning anti-competitive market conduct were launched against Magyar Posta in 2005. The NHH conducted an investigation into the discount system of the designated universal postal service provider. The authority chiefly investigated whether the pricing and discount system applied by Magyar Posta Zrt. to its contractual customers as per its *key partnership agreements,* and to other major consumers, were in line with the provisions of the Postal Services Act on the pricing of postal services and application of such prices, and whether the requirement for non-discriminatory treatment was met with respect to Magyar Posta's contractual partners. The Authority ruled that several aspects of the key partnership agreements gave cause for

[2] A summary of such cases can be found in the paper of *Geradin and Henry* [2004].

concern and ordered Magyar Posta to review its discount system and to develop a set of contractual conditions in line with the principles set out in the ruling and to render its active contracts legally compliant (ruling no. PS-19.710-4/2005 – *NHH* [2006]).

With respect to that same period, the Competition Authority also examined whether Magyar Posta had abused its dominant position (Vj-174/2005/55). In co-operation with the NHH, it examined in relation to Magyar Posta's discount system whether the conclusion of certain exclusive agreements had led to closure of the market. The Competition Authority terminated the proceedings since it established that Magyar Posta had only concluded an exclusivity agreement with Magyar Telekom Rt. and that Magyar Telekom Rt.' s share of the letter mail market was less than 10 percent. The Competition Council therefore concluded that the agreement did not lend itself to preventing competitors from entering the market or creating a disadvantageous market situation for such competitors (i.e. closure of the market).

The Competition Authority earlier examined whether Magyar Posta's discount system constituted abuse of its dominant position on the grounds that the discount system led to the bundling of products in a way that restricts competition. The 2011 proceedings (Vj-167/2001/52) focused on Magyar Posta's practice of giving greater than usual discounts to those customers that also had the delivered mail (typically invoices) produced by Magyar Posta. The investigation, although it was launched in relation to a specific order (Émasz public procurement), identified objectionable discounts in contracts between 1999 and 2001. According to the ruling, Magyar Posta used its dominant position on the basic postal services market to restrict and distort competition on the mail delivery market, which can be termed competitive. Thanks to the special discount it offered, it persuaded customers to commission it with mail production too. The Competition Authority fined Magyar Posta HUF 20 million and ordered it to desist from the practice in question.

It is also worth mentioning two other Hungarian legal cases concerning the anti-competitive conduct of Magyar Posta. The Competition Authority's Competition Council launched proceedings against Magyar Lapterjesztő Zrt. and Magyar Posta Zrt. on the grounds that they had *concluded an agreement restricting competition*, and established that their conduct had restricted competition in its ruling of November 8, 2007. Pursuant to an earlier agreement in effect between 1998 and 2001 in connection with the privatization of Magyar Posta's newspaper distribution companies, Magyar Posta would not compete for the delivery of newspapers from the printing presses to retailers. In exchange, Magyar Posta managed to have the commission paid by Lapker for newspapers sold in Magyar Posta's post offices increased from 13 to 23.5 percent in the cooperation agreement concluded between the two companies for the period between 2002 and 2007. In the amendment of the agreement effective from January 2003, the newspaper distributor agreed not to engage in subscription-based newspaper distribution, which was considered to be Magyar Posta's domain. In connection with the non-compete agreement, Magyar Posta received one-off market organization fee of HUF 260 millionfor not impeding Lapker's business for another five years. The Competition Council ordered Magyar Posta Zrt. and Lapker to pay a fine of HUF 468 million each. The appellate ruling of April 22, 2009 reduced that amount to HUF 250 million each (ruling no. Vj-140/2006/69).[3]

The other case worth mentioning involves postal services, but does not concern the letter and package markets. On November 8, 2007, the Competition Authority launched proceedings against Magyar Posta on the grounds of abuse of its *dominant economic position*. According to the investigation, Magyar Posta applied terms and conditions to the authorization or conclusion of contracts concerning the production of cash transfer orders that were liable to distort competition on the related market for the production of forms (on which Magyar Posta is itself present) or in certain segments of that market. The Competition Council, based on the evidence available, found that on at least five occasions between 2004 and 2007 Magyar Posta declined to authorize the production of forms required for cash transfer orders. In one case it had technical reasons for declining, but in the other cases its decision was based on commercial considerations. However, it was established during the proceedings that several enterprises are present on the market for production of the forms required for cash transfer orders, and that those market players are significant competitors of Magyar Posta, so, in the period examined, Magyar Posta's conduct would not have jeopardized competition on the market for production of forms for cash transfer orders or on other markets, and damage to consumers was also not considered likely. By its ruling of July 8, 2008, the Competition Council therefore terminated its competition supervision proceedings (ruling no. Vj-186/2007/36).[4]

MARKET PERFORMANCE

Economic performance

According to the estimates of the international study carried out by the ITA-WIK group, in Hungary the letter mail segment accounted for 59 percent of the mail volume of the postal industry in 2007 (with an estimated market value of EUR 405 million), while activities involving packages and express mail accounted for 41 percent (estimated market value of EUR 280 million) (*ITA–WIK* [2009*b*] pp. 59–60). By comparison, letters accounted for 56 percent of the EUR 94 billion revenue of the total EU mail sector in 2007, while packages and express mail accounted for 44 percent (*ITA–WIK* [2009*a*] p. 25). In general, the packages and express mail segment is growing more dynamically than the letter mail segment. Total internal

[3] http://www.gazdasagkifeheritese.uni-corvinus.hu/images/8/8b/GVH_-_lapterjesztokartell.pdf.
[4] http://www.concurrences.com/IMG/pdf/2008-09-18-Hungary03Doc_82_.pdf.

EU letter mail grew in the past 10 years by 0.4 percent on average, while package and express services rose by 6.1 percent on average (*ITA–WIK* [2009*a*] p. 36 and p. 140). First, that reflects the stimulating effect of competition, which developed far sooner internationally on those last two markets. Second, the fact that new forms of electronic communication are gaining ground primarily poses a challenge to the letter mail segment, since such information can be conveyed by other methods with new communications technology solutions. That is less true of package and express mail services involving tangible, physical mail.

Of course considerable deviation of the various Member States lies behind the EU averages. We can see from *Table 6* that the growth rate was considerably lower in the Western European Member States, which have more mature postal markets, than in the Southern and in particular Eastern Member States. The data, however, show that despite the more marked growth rate, the letter mail markets of the latter country groups are still significantly "less developed" (in the sense of saturation/maturity of the market): the number of letters per capita is significantly below the Western European average. *Table 6* shows that in Hungary the volume of mail per capita is below the EU average (and considerably lower than that of the Western Member States, which have more mature postal markets), but higher than that of the Eastern Member States that have joined the EU more recently. Understandably, the growth rate was lowest in the period between 2003 and 2007 in those Member States with a mature (and therefore largely saturated) postal market.

The negative Hungarian growth rate is somewhat misleading, however. If we look in more detail at the underlying data *(Table 7),* we can see that of the growth rates of the three separate groups of addressed domestic letters (numbers in italics), newspaper distribution is the only segment with a negative growth rate. That, however, is highly distorted, since for that segment the data only reflect the mail volume handled by Magyar Posta, which decreased considerably in 2007 as described above, owing to the arrival of a new competitor. The conclusion cannot be drawn from the data that the total volume of mail actually decreased to that extent on the newspaper distribution market. Further, the data concerning addressed advertising mail do not give an accurate impression of that segment, since Magyar Posta's classification of mail has since changed; in 2007 some mail types were reclassified from the earlier addressed advertising mail segment into the unaddressed advertising mail segment. If we eliminate all those distortions from the growth data of the cited international study, then the growth of the Hungarian letter mail market (dominated by the 0.7 percent growth rate of letters), would be around the modest EU growth average, below the 2.2 percent growth of the Eastern Member States, but, as we can see from *Table 6*, with a greater number of mail items per capita.

With respect to packages and express mail, the Hungarian 18.5 percent growth rate shown in *Table 6* is in line with the average growth trend of the Eastern European Member States.

TABLE 6 • **Domestic letter mail within the EU–27, 2007**

Member States	Number of mail items per capita*	Compound annual growth rate, 2003–2007 (percentage)	Share of revenues from mail within the EU (percentage)
Letters			
Western Member States	297	0.2	81.4
Southern Member States	94	1.1	14.5
Eastern Member States	68	2.2	4.1
Hungary	98	–1.5	–
Packages and express mail			
Western Member States	–	5.4	75.1
Southern Member States	–	9.3	21.7
Eastern Member States	–	18.8	3.2
Hungary	–	18.5	–

Western Member States: Austria, Belgium, Denmark, United Kingdom, Finland, France, Netherlands, Ireland, Luxembourg, Germany, Sweden; *Southern Member States:* Cyprus, Greece, Malta, Italy, Portugal, Spain; *Eastern Member States:* Bulgaria, Czech Republic, Estonia, Poland, Latvia, Lithuania, Hungary, Romania, Slovenia, Slovakia.
* Weighted average (weighted by population).
Source: based on *ITA–WIK* [2009*a*] p. 38, p. 143 and *ITA–WIK* [2009*b*] pp. 59–60.

TABLE 7 • **Size and growth of the Hungarian letter mail market**

Market segment	Items per capita, 2007	Compound annual growth rate, 2002–2007 (percentage)
Domestic letter post	98	–1.5
Letters	*90*	*0.7*
Addressed advertising mail	*3*	*20.6*
Newspaper distribution	*5*	*–23.1*
Unaddressed advertising mail	70	15.5
Outbound international mail	2	–1.3

Note: Data on the segments publications, unaddressed advertising, and cross-border outbound generally refer to incumbent postal operator / universal service provider.
Source: ITA–WIK [2009*b*] p. 59.

The vast majority of mail is sent by companies and public organizations. Eighty-eight percent of the total volume of letters in the EU is sent by such market players (B2X) and just 12 percent by private individuals (C2X) (including micro enterprises and sole traders). Of the letters posted by businesses, 35 percent are addressed to other businesses (B2B), while 65 percent are addressed to private customers (B2C). Unfortunately data of such depth are not available for Hungary, but the proportion of letters posted by businesses is estimated to be around 95 percent, i.e. higher than the EU average (correspondence by private individuals has steadily declined since 2004). *(ITA–WIK* [2009*b*] p. 60)

In *Table 8* more detailed data for the last three years have been compiled showing the performance of the postal sector and of the universal service provider.

TABLE 8 • **Data showing the economic performance of the Hungarian postal sector and the universal postal service provider (2006–2008)**

Designation	2006	2007	2008	Percentage change 2006–2007	Percentage change 2007–2008
*Net sales revenue (*HUF millions)					
Universal service[a]	74,000	79,465	85,001	+7.37	+6.98
Courier service	2,455	3,281	7,237	+33.65	+120.57
Express mail service	1,484	1,865	349	+25.67	−81.29
Integrated mail service	22,406	26,066	31,227	+16.33	+19.80
Number of mail items[b] (in thousands)					
Universal service	809,990.8	788,828.1	801,158.2	−2.61	+1.56
Courier service	1,192.1	1,789.6	2,599.5	+51.42	+45.26
Express mail service	411.0	501.5	483.6	+22.02	−3.57
Integrated mail service	8,167.9	10,072.8	13,133.7	+23.32	+30.39
Number of employees					
In the postal sector	39,706	37,648	n/a	−5.18	–
Universal service provider[c]	38,686	36,429	35,973	−5.83	−1.25
Productivity[d]					
Postal sector revenue (HUF million/employee)	2.53	2.94	n/a	+16.21	–
Number of mail items handled by the postal sector (in thousands/employee)	20.65	21.28	n/a	+3.05	–
Revenue of universal service provider (HUF millions/employee)	1.91	2.18	2.37	+14.13	+8.72
Number of mail items handled by the universal service provider (in thousands/employee)	20.93	21.65	22.27	+3.44	+2.86
Investment (at current prices, in HUF billions)					
Postal sector total	18.30	16.20	n/a	−11.47	
Postal sector, proportional to revenue[e]	0.18	0.15		−16.67	
Universal service provider, total[f]	13.63	11.29	7.97	−17.17	−29.41
Universal service provider, proportional to revenue[f]	0.18	0.14	0.09	−22.22	−35.71

[a] *Magyar Posta* [2006], [2007], [2008].
[b] Number of mail items with domestic delivery.
[c] *Magyar Posta* [2006], [2007], [2008].
[d] The productivity of the industry measured in output per employee and productivity of the incumbent service provider measured in output per employee (calculated on the basis of the other data in the table)
[e] Investment per forint of revenue (investment/sales revenue)
[f] *Magyar Posta* [2006], [2007], [2008].
Source: Unless otherwise indicated, the data are taken from the communications statistics database of the National Communications Authority (NHH).

We can see from the data in *Table 8* that the net sales revenue increased in both the non-competitive universal services segment and in the competitive market segments (courier, express and integrated mail services) in the period shown, but the competitive segments showed much stronger growth. If we examine the development of these market segments by looking at mail volume, we can see that in the case of universal services the volume of mail tended to fall, with some fluctuations, while the growth rate of the mail volume in the competitive segments was typically greater than the revenue increase in those same segments.

The output per employee was calculated as a measure of productivity, measuring output in revenue or volume of mail. The indicator calculated on the basis of the number of mail items is not suitable for the comparison of various industry segments (or for the comparison of the industry average and the universal service provider), because the mail types are very different in the various segments (Magyar Posta handles a significantly higher proportion of letters than its competitors, which chiefly operate in the packages and express mail segments). However, it can serve as a more useful indicator of development over time within the given market segment than the productivity indicator, which is calculated on the basis of revenue, since the latter can change as result of price increases, i.e. it does not necessarily reflect an actual change in efficiency. We can see that distinction, for instance, in the case of the indicators of the universal service provider, since the increase in mail volume per employee (3.44 and 2.86 percent in the periods of 2006–2007 and 2007–2008 respectively) was considerably more modest in the period examined than the revenue increase per employee (14.13 and 8.72 percent in the same periods respectively). The income per employee, however, is worth comparing directly for the various market segments. We can see from these data that the productivity of the sector as a whole was greater in that period than that of the universal service provider (the productivity of the sector as a whole was HUF 2.53 million/employee and HUF 2.94 million/employee in 2006 and 2007 respectively, while the productivity of the universal service provider was HUF 1.91 million/employee and HUF 2.19 million/employee) in 2006 and 2007 respectively).[5]

If we examine investments over time, we can see that there is not a considerable difference in investments proportional to revenue (investment per forint of revenue) of the sector as a whole and of the universal service provider. From the last rows of *Table 8* it can, however, be seen that in recent years the investments of the universal service provider decreased to a greater extent (by 17.17 percent between 2006 and 2007) than the investments of the sector as a whole (by 11.47 percent). Moreover, the investments of the universal service provider decreased even more sharply in the following year (by 29.41 percent between 2007 and 2008).

[5] Caution is advised when drawing conclusions since the number of employees refers to the entire workforce (those whose duties only include mail services do not constitute a separate group), while the revenues are from mail services only. In the case of Magyar Posta, the proportion of other, non-mail activities is greater (for example, in 2008 financial activities accounted for approx. 31.5 percent of Magyar Posta's revenue – *Magyar Posta* [2008] p. 13) than at its competitors, so the proportion of employees performing non-postal tasks within the total workforce is likely higher.

PRICING REGULATIONS AND PRICES OVER TIME

In most Member States price regulations refer to universal services as a whole, but in some countries they only apply to those universal services segments where there is no discernible competition. In Hungary until the complete market opening, only the prices of reserved services were regulated despite the fact that for a long time Magyar Posta did not have any competitors in the non-reserved universal services segment, so the universal service provider's prices in that segment were not restricted by either market competition or regulations. In Hungary, until the end of 2012 (until the time of total market opening), prices were regulated *ex ante* by ministry decree (by decree of the Ministry of National Development, and earlier by decree of the Ministry of Economic Affairs and Transport and of the Ministry of Informatics and Communications).

Price cap regulations came into force in Hungary too following the complete market opening. As of January 2013, National Development Decree no. 67/202 (XII.15.) sets out the method for determining the postal fees for individually posted domestic letters not weighing more than 50 grams that come under the scope of universal services (quantity and price in the respective year q_{ti} and p_{ti}, quantity and price in the preceding year p_{t-1i}) and of official documents (quantity and price in the respective year q_{ti} and p_{ti}, quantity and price in the preceding year p_{t-1i}). The price cap has to be specified per mail type, and the growth rate in the respective years is tied to the consumer index. If the latest forecast of the Hungarian National Bank (RPI_{t-1MNB}) indicates a rise in the consumer price index, the postal service provider is entitled to increase the price, and obliged to decrease it if a decrease in the consumer price index is forecast.

Given the possible inaccuracy of the forecast, the deviation from the actual change in prices (data published by the Hungarian Statistical Office: RPI_{t-1KSH}) needs to be taken into account when prices are changed in the following year as an inflation correction factor ($Z_t = RPI_{t-1MNB} - RPI_{t-1KSH}$). If the service provider is loss-making despite providing services effectively, then it is exempt from the obligation to reduce prices and does not need to apply the inflation correction rate (the loss needs to be demonstrated to the National Media and Infocommunications Authority). The percentage value of price changes of the respective year for mail of the postal service provider that comes under this category can be calculated as follows (using a Paasche-type price index):[6]

$$\Delta V_t = \left(\frac{\sum_{i=1}^{n} q_{ti} \times p_{ti} + \sum_{j=1}^{m} q_{tj} \times p_{tj}}{\sum_{i=1}^{n} q_{ti} \times p_{t-1i} + \sum_{j=1}^{m} q_{tj} \times p_{t-1j}} - 1 \right) \times 100.$$

[6] The Paasche-, Fisher- or Törnquist-type indexes can also be used for ex post regulations. The Törnquist index is widespread in telecommunications, whereas in the postal industry the Paasche index is frequently used, despite the fact that the Paasche indexes do not have Fisher-type properties and do not stand the time test (for there to be a multiplicative inverse relationship between the indexes calculated by inverting the time periods) or the factor test (i.e. the price index multiplied by the volume index does not give the value index).

Should the service provider fail to calculate the price change level in that manner, then an authority correction factor ($X_t = \Delta V_{t-2} - \Delta P_{t-2}$) may be applied. That authority correction factor is reflected in the maximum extent of the annual average price change of services to which regulated prices apply: $P_t = RPI_{tMNB} - Z_t - X_t$.

We can only directly examine price changes of those mail segments with regulated prices, since information for Hungary is only available for those segments. We can draw cautious conclusions about the other segments only by comparing the mail volume data and revenue data. In most Member States price regulations refer only to universal services, but in some countries they only extend to those segments of universal services where there is no perceptible competition. In Hungary until the complete market opening only the prices of reserved services were regulated, despite the fact that for a long time Magyar Posta did not have any competitors in the non-reserved universal services segment, so the universal service provider's prices in that segment were not restricted by either market competition or regulations. As mentioned earlier, it was not until 2009 that a competitor entered the non-reserved market segments involving addressed letters (primarily advertising mail) weighing more than 50 grams. From the data set out in *Tables 3–5* above, we can attempt to draw cautious conclusions about prices over time in the various market segments. If we compare the growth rate of revenue and the volume of mail, we can see that in the non-competitive universal services segment, the overall volume of mail declined slightly in recent years (by almost 3 percent from 2008 to 2010), while revenue increased (by 12 percent). The difference between the two indicates an increase in prices exceeding inflation in this market segment. By contrast, in the competitive segments the volume of mail increased to a greater degree than the resulting revenue, which suggests that prices did not increase significantly in those segments (we have to be cautious, however, when drawing conclusions, since we do not have information at our disposal concerning changes in the breakdown of mail types in the various price categories, so we cannot filter out the effect of changes to the proportions of the various mail types).

Table 9 shows regulated prices over time in the reserved segments. From the data it can be seen that prices increased for every mail type. The greatest price increase (for both letters and addressed advertising mail) was in the most frequent (non-priority standard) category (85 percent from 2007 to 2014 for letters and addressed advertising mail).

We can see therefore that in recent years the prices for letters weighing less than 50 grams, which are regulated, increased to a greater extent than inflation, in particular from 2011 to 2012, when the increase in price for letters and advertising were between 12.5 percent and 16.66 percent. We can also see from the data in the table, however, after the total market opening, which coincided with the introduction of price cap regulations, the increase in prices slowed. While in the previous years the increase in prices was considerably higher than the inflation rate; in 2013 and 2014 it only slightly exceeded inflation. The slowing down of the price increase is due to

TABLE 9 • **Regulated prices over time (in HUF)**

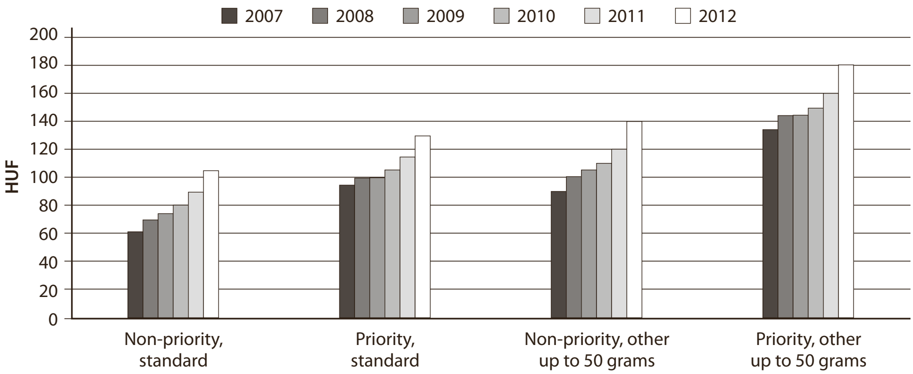
	2007	2008		2009		2010		2011		2012		2013		2014	
	price	price	%	price	%	price	%	price	%	price	%	price	%	price	%
Letters (letters, plain postcards, illustrated postcards)															
Non-priority, standard	62	70	+12.9	75	+7.14	80	+6.66	90	+12.5	105	+16.66	110	+4.76	115	+4.55
Priority, standard	95	100	+5.26	100	0.00	105	+5.00	115	+9.52	130	+13.04	140	+7.69	145	+3.57
Non-priority, other up to 50 grams	90	100	+11.11	105	+5.00	110	+4.76	120	+9.09	140	+16.66	145	+3.57	150	+3.45
Priority, other up to 50 grams	135	145	+7.40	145	0.00	150	+6.66	160	+6.66	180	+12.50	195	+8.33	200	+2.56
Addressed advertising mail															
Non-priority, standard	62	62	0.00	75	+20.96	80	+6.66	90	+12.50	105	+16.66	110	+4.76	115	+4.55
Priority, standard	95	89	−6.31	100	+12.35	105	+5.00	115	+9.52	130	+13.04	140	+7.69	145	+3.57
Non-priority, other up to 50 grams	90	89	−1.10	105	+17.97	110	+4.76	120	+9.09	140	+16.66	145	+3.57	150	+3.45
Priority, other up to 50 grams	135	128	−5.19	145	+13.28	150	+6.66	160	+6.66	180	+12.50	195	+8.33	200	+2.56
Domestic official documents															
Must be delivered to recipient in person	300	320	+6.66	330	+3.13	355	+7.57	370	+4.22	390	+5.40	410	+5.13	415	+1.22
Delivery to recipient in person not required	250	275	+1.00	280	+1.81	290	+3.57	300	+3.45	315	+5.00	325	+3.17	330	+1.54

Source: Economic Affairs and Transport Ministry Decree no. 85/2006. (XII. 15.), Economic Affairs and Transport Ministry Decree no. 100/2007. (XII. 19.), Transport, Communications and Energy Ministry Decree no. 39/2008. (XII. 20.), Transport, Communications and Energy Ministry Decree no. 83/2009. (XII. 30.), National Development Ministry Decree no. 21/2010. (XII. 16.), National Development Ministry Decree no. 75/2011. (XII. 15.) and the official price list published by Magyar Posta for 2013 and 2014 (from January 2013: http://www.posta.hu/static/internet/download/USZ_2013_tarifatabla_HU_BELFOLD_20121221.pdf, from 1 January, 2014: http://www.posta.hu/static/internet/download/2014_tarifatabla_HU_2014_01.pdf).

the stricter price cap regulations, rather than the total market opening directly, since new competitors did not manage to enter the universal services segment.

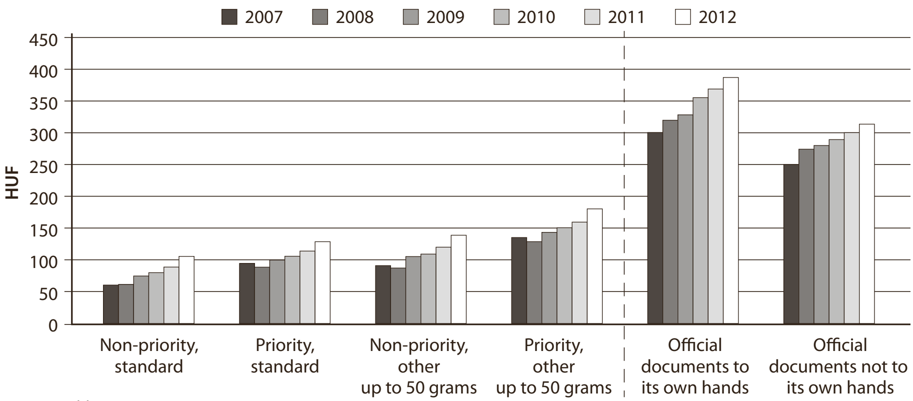
The figures below summarize price changes in the letter mail segment and the addressed advertising mail segment.

The prices of letters increased continuously *(Figure 2).* In the case of addressed advertising mail *(Figure 3*) following a minor decline in 2008 (there was no change in the most frequent category (non-priority standard) and a slight decline in the other categories), in 2009 prices went up again. In total, the combined price increase of those three years was at the same level as the price increase in the letter mail segment.



Source: Table 9.

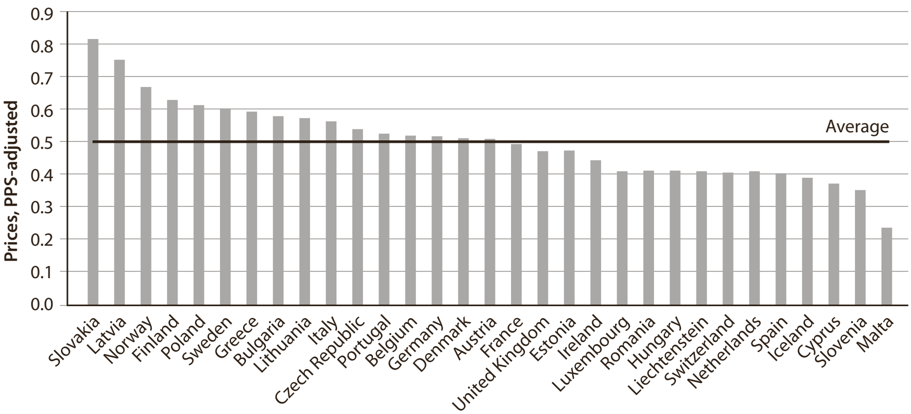
FIGURE 2 • **Regulated prices for postage of letters (letters, plain postcards, illustrated postcards) over time**



Source: Table 9.

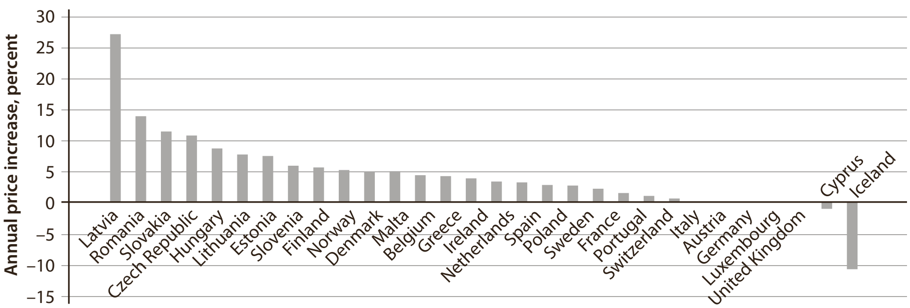
FIGURE 3 • **Regulated prices for postage of addressed advertising mail and official documents over time**

If we wish to assess the performance of the Hungarian postal sector and the regulations in terms of price changes, then it is worth comparing the prices of services with the data from other countries. For international comparisons, the price of 20 gram letters is a frequently used measure since the majority of mail is in that category. The prices of postal services differ considerably among European countries. *Figure 4* shows that the prices for 20 gram letters ranged between EUR 0.23 and EUR 0.8 in 2009 (adjusted for purchasing power). If the prices are simply converted to EUR and compared, then the Hungarian prices would seem relatively high (in the top third of the EU Member States). If, however, we compare the prices after adjusting for purchasing power, then the Hungarian prices were not high in 2009 and were below the EU average, despite the fact that the prices of services increased fairly significantly in Hungary between 2005 and 2009. A greater price increase than in Hungary was only experienced in four countries during that period *(Figure 5).*



Source: Copenhagen Economics [2010] p. 37.

FIGURE 4 • **Prices in EUR for 20 gram letters in 2009, PPS-adjusted**



Source: Copenhagen Economics [2010] p. 39.

FIGURE 5 • **Annual growth rate of nominal prices for 20 gram tariff letters, 2005–2009**

Service quality

With respect to universal services, the Hungarian Postal Services Act (similarly to the postal services directive), not only defines the scope of service obligations, but also contains provisions concerning regional accessibility and opening times, and quality requirements pertaining to those services. Below I will give a summary of the performance of the Hungarian universal postal service provider in light of those indicators.

Accessibility of services[7] • In 2010 no interruptions in the operation of postal service points, delivery districts or collection services were experienced. Services were provided in compliance with the legal requirements, and there was full household delivery in compliance with the law in all (3,154) settlements. Magyar Posta complied with the requirements for *opening times* (referring both to the times when mobile post offices are present and to the opening hours of permanent postal service points).

Compliance with the provisions concerning *geographical accessibility* in 2010 was as follows:

- Universal postal services were not available in three of the 1,250 settlements with a population of fewer than 600 residents.
- Three settlements with between 600 and 1,000 residents did not have the prescribed permanent postal service point.
- Currently 1,037 settlements only have a mobile post office, in addition to which services are provided by means of a combination of a mobile post office and a permanent service point in 85 settlements. An access point was designated and a postbox was installed in each of those settlements that only have a mobile post office.
- Of the 1,394 settlements with a population of greater than 1,000 residents, availability was not in compliance with the law in one case.

Act CI of 2003 also prescribed that when establishing and operating postal access points, it needs to be ensured that those using the services (including persons living with disabilities) can easily access the postal network without barriers.

- 1,743 of the 2,740 permanent postal service points did not comply with the legal requirements, despite the fact that the relevant grace period expired in 2007;
- 74 of the 8,988 postboxes available to the public did not comply with the statutory requirements;
- 17 of the 3,890 designated delivery/collection boxes outside municipality boundaries did not comply with the requirements for barrier-free access;
- the designated access points could be accessed barrier free in 1,036 of the settlements that only have a mobile post office; that was not possible in 70 settlements.

[7] Based on data provided by the NMHH.

Service quality indicators • In 2010 Magyar Posta's operations complied with all the quality indicators prescribed by law. There has been improvement in respect of almost all delivery time indicators in recent years. The indicators for reliability of mail forwarding showed either no change or a slight worsening, but the service provider's performance was well above the statutory requirements regardless. These quality indicators are summarized in *Table 10.*

TABLE 10 • **Postal service quality indicators, 2008–2010***

Type of quality indicator	Number of business days	Statutory requirement (%)	Performance (national average, %) 2008	Performance (national average, %) 2010
Delivery time: speed and reliability indicators				
Delivery time for priority letters	*D* + 1	85	92.69	93.68
	D + 3	97	99.73	99.76
Delivery time for non-priority letters	*D* + 3	85	95.07	95.59
	D + 5	97	99.39	99.58
Delivery time for packages	*D* + 1	85	92.89	92.27
	D + 3	97	99.73	99.89
Quality requirements for the date stamp				
Proportion of mail with an illegible date stamp		≤ 0.15	0.0017	0.0005
Requirements for reliability of mail forwarding				
Proportion of totally or partially lost registered domestic mail		≤ 0.06	0.0338	0.0338
Proportion of damaged registered domestic mail		≤ 0.05	0.0038	0.0080

* The specific quality requirements were established as a percentage of mail items arriving with a given number of business days following the day of being posted (*D*), i.e. the percentage of mail items that need to arrive by the given day. For each category there is a so-called speed indicator (*D* + 1 for priority letters and packages, *D* + 3 for other mail) and a reliability indicator (*D* + 3 for priority letters and packages, and *D* + 5 for other mail).
Source: NHH [2009] and the NMHH.

SUMMARY

Reform of the regulations of the Hungarian postal sector began in the early '90s. The postal, telecommunications and broadcasting services were had been previously combined into a single company with a state monopoly, and it was not until 1990 that the three fields were split into three separate companies. These fields were then also regulated separately by law. The Postal Services Act of 1992 opened up certain segments of the postal market (using current terms, certain non-universal segments) to competitors. Nevertheless, there were no actual market entrants owing to the strict conditions set out in the legal regulations. The Communications Act of 2001, which combined regulation of the information and communications markets, created an actual market opening in the field of non-universal services. After the Communications Act came into force, the number of registered service providers in the courier and express mail segments rapidly multiplied. The regulations set out in the Communications Act concerning the postal sector were intended to be in line with the first EU postal services directive (Directive 97/67/ EC). However, the Act was strongly criticized by the European Commission in its examination of Member State implementation of the postal services directive on the grounds that the privileges of the universal service provider had been extended (primarily with reference to extension of the scope of reserved services to cash transfers, payment intermediary services and postal order services contrary to the directive).

As there was both internal and external need to amend the legislation, a new Act on postal services was adopted in 2003, in which the regulation of the postal sector was again separated from the telecommunications and other communications services. The second postal services directive (Directive 2002/39/EC) was taken into account when drafting that Act. Hungary was granted a moratorium until December 31, 2012 for transposition into Hungarian law of the third postal services directive and total market opening. Owing to the lengthy negotiating and drafting procedures, at almost the last minute, in fall 2012, Parliament adopted the new Postal Services Act, which almost entirely eliminated reserved services. The only field, in which the universal postal service provider retains the exclusive right to provide postal services, is the official documents segment. Hungary consistently sought to define both universal services and reserved services as broadly as possibly (to the extent allowed by the EU directives), and prescribed conditions for entry to the opened universal services segments that barely allow for the emergence of new competitors. Hungary was in the last group of EU Member States with respect to the extent of market opening.

The anti-competitive tax conditions also act as a barrier to entry. In Hungary, similarly to most Member States, universal postal services (in view of being in the public interest) are exempt from VAT, while replacement universal services are not, with the result that Magyar Posta's competitors in those segments are at a considerable competitive disadvantage. In order for entrants to the universal services segment to make their services attractive (for instance, to the financial sector or public sector, which cannot reclaim VAT), they have to offer their services at below the universal service provider's prices less the VAT amount, which significantly distorts price competition.

In terms of competition, the Hungarian postal sector can be broken down into three categories up to the time of total market opening:

- absence of competition (because it was prohibited): reserved universal services: segment involving letters weighing below 50 grams,
- minimal competition: non-reserved universal services (until 2009 Magyar Posta did not have a competitor; in 2009 a competitor emerged with a small market share),
- strong competition: non-universal postal services, for which the service provider needs to be registered (courier services, express mail services, integrated mail services, postal intermediary services, document exchange services).

Nor did the picture change after total market opening, since competitors have not managed to enter the most recently opened reserved services segment (mail weighing less than 50 grams) since then. The two entry applications were rejected by the regulatory authority on the grounds that they did not fulfill the statutory requirements. There is, however, fierce competition in the field of non-universal postal services. In the non-universal packages segment and the express and courier segments. Magyar Posta's main competitors are the large international integrator companies. In addition to the large international service providers, there are also several small enterprises, which chiefly provide local courier and express mail services.

Based on revenue, the competitive segments account for just over a third of the total Hungarian postal sector, and the mail volume of the competitive segments grew faster in the last ten years than that of universal services. As a result, the total revenue of the competitive sectors as a proportion of the total industry revenue is increasing slightly from year to year. Integrated mail services are the largest market segment (26 percent in 2010) within competitive, non-universal postal services. Contrary to international trends, the number of letters handled by Magyar Posta (unlike packages) has not yet fallen significantly (though a slight negative trend can be observed despite minor fluctuations). The majority of letters sent in 2010 were in the reserved (i.e. monopolized) segment of mail weighing less than 50 grams: 94 percent of letters and 73 percent of advertising mail (in which Magyar Posta does not currently have a competitor). Payment services and financial intermediary services continue to play a major role within Magyar Posta's activities (accounting for roughly a third of its total revenue). The performance data from the period examined here show that the productivity of the total sector including the competitive segments was greater than the productivity of the universal service provider (measured in revenue per employee).

In Hungary until the complete market opening only the prices of reserved services were regulated, despite the fact that for a long time Magyar Posta did not have any competitors in the non-reserved universal services segment, so the universal service provider's prices in that segment were not restricted by either market competition or regulations. Until the end of 2012, prices were set by the relevant authority *ex ante* – the regulated prices of reserved services were stipulated by ministry decree. Price cap regulations came into force in Hungary too, following the complete market opening. Within universal services, the price cap regulations apply to the prices of individually posted domestic letters not weighing more than 50 grams and to official documents. The regulated prices have increased every year in the past ten years for every type of mail. The greatest price increase (with respect to both letters and addressed advertising mail) was in the most frequent (non-priority standard) category (by a total of 85 percent from 2007 to 2014 with respect to letters and addressed advertising mail). The increase in prices slowed following the complete market opening, which coincided with the introduction of price cap regulations.

While in the previous years the increase in prices was considerably higher than the inflation rate, in 2013 and 2014 it only slightly exceeded inflation. The slowing down of the price increase is due to the stricter price cap regulations, rather than the total market opening directly, since new competitors have not managed to enter the majority of the universal services segment.

REFERENCES

Copenhagen Economics [2010]: Main Developments in the Postal Sector (2008–2010). November 29, 2010 http://ec.europa.eu/internal_market/post/doc/studies/2010-main-developments_en.pdf.

Ecorys [2008]: Main Developments in the Postal Sector (2006–2008). Ecorys, September, http://ec.europa.eu/internal_market/post/doc/studies/2008-ecorys-final_en.pdf

Geradin, D.–Henry, D. [2004]: Regulatory and competition law remedies in the postal sector. In: *Geradin, D.* (ed.): Remedies in network industries: EC competition law versus sector-specific regulation. Intersentia, Antwerpen/Oxford, pp. 125–182.

ITA–WIK [2009*a*]: The Evolution of the European Postal Market since 1997. ITA Consulting–WIK Consult, August 2009, http://ec.europa.eu/internal_market/post/doc/studies/2009-wik-evolution_en.pdf.

ITA–WIK [2009*b*]: The Evolution of the European Postal Market since 1997 – Annex: Country Fiches. ITA Consulting–WIK Consult, August 2009, http://ec.europa.eu/internal_market/post/doc/studies/2009-wik-evolution-country-annex_en.pdf.

Magyar Posta [2006]: Annual Report Magyar Posta. Budapest, http://www.posta.hu/static/internet/download/MP_eves_jelentes_2006.pdf.

Magyar Posta [2007]: Annual Report Magyar Posta. Budapest, http://www.posta.hu/static/internet/download/MP_eves_jelentes_2007.pdf.

Magyar Posta [2008]: Annual Report Magyar Posta. Budapest, http://www.posta.hu/static/internet/download/MP_eves_jelentes_2008.pdf.

Magyar Posta [2010]: Annual Report Magyar Posta. Budapest, http://www.posta.hu/static/internet/download/MP_eves_jelentes_2010.pdf.

Magyar Posta [2011]: Annual Report Magyar Posta Zrt., Budapest, http://www.posta.hu/static/internet/download/Posta_EvesJelentes_2011.pdf

NHH [2009]: 2008 Quality Report on Universal Postal Services. National Communications Authority (NHH), Budapest.

Cited Hungarian legislation:

Act CLIX of 2012 on Postal Services

Act LXXXII of 2010 on the Amendment of Certain Acts Governing the Media and Communications

Act CLXXXV of 2010 on Media Services and Mass Communications

Act CI of 2003 on Postal Services

Act C of 2003 on Electronic Communications
Act XL of 2001 on Communications
Act XLV of 1992 on Postal Services
Act LXXII of 1992 on Telecommunications
Government Decree no. 336/2012. (XII. 4.) on the detailed rules of the Subsidy Fund for Universal Postal Services
Government Decree no. 335/2012. (XII. 4.) on the detailed rules of provision of postal services and of postal services involving official documents, on the general terms and conditions of postal services, and on mail excluded from postal services or mail that may only be delivered under certain conditions
National Development Ministry Decree no. 67/2012. (XII. 15.) on the method of determining the prices for postage of domestic letters not weighing more than 50 grams that are posted according to the per-letter price schedule and come under the scope of universal services and the prices for postage of official documents.
National Development Ministry Decree no. 75/2011. (XII. 15.) on amendment of Economic Affairs and Transport Ministry Decree no. 85/2006 (XII. 15.) on the maximum regulated prices for reserved domestic postal services
National Development Ministry Decree no. 21/2010. (XII. 16.) on amendment of Economic Affairs and Transport Ministry Decree no. 85/2006 (XII. 15.) on the maximum regulated prices for reserved domestic postal services
Transport, Communications and Energy Ministry Decree no. 83/2009. (XII. 30.) on amendment of Economic Affairs and Transport Ministry Decree no. 85/2006 (XII. 15.) on the maximum regulated prices for reserved domestic postal services
248/2001. (XII. 18.Government Decree no. 248/2001. (XII. 18.) on the Communications Inspectorate and on fines that may be imposed by the bodies of the Communications Inspectorate

APPENDIX

LIST OF ORIGINAL HUNGARIAN CHAPTERS

◆

LIST OF CONTRIBUTORS

◆

ABSTRACTS (2013, 2014)

◆

CONTENTS IN ENGLISH OF THE COMPETITION AND REGULATION SERIES 2007–2014

The chapters of this English-language volume were originally presented in Hungarian within the series published by the Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences, "Verseny és szabályozás" [Competition and Regulation], edited by *Pál Valentiny, Ferenc László Kiss, and Csongor István Nagy*. All Hungarian volumes, as well as the current compilation, are downloadable at the website of the Institute: MTA Közgazdaság- és Regionális Tudományi Kutatóközpont, Közgazdaság-tudományi Intézet, http://econ.core.hu/kiadvany/vesz.html.

LIST OF ORIGINAL HUNGARIAN CHAPTERS

András Kovács • Az üzleti titok védelméhez való jog gazdasági elemzése [Economic Analysis of the Right to Protect Commercial Secrets]. Verseny és szabályozás [Competition and Regulation], 2009, pp. 68–107.

Balázs Muraközy and Pál Valentiny • Az állami szabályozás alternatívái. Az ön- és együttszabályozás [Alternatives to State Regulation: Self- and Co-Regulation]. Verseny és szabályozás [Competition and Regulation], 2012, pp. 97–140.

Gergely Csorba • Magyarországi empirikus piacszerkezet-elemzések kutatási eredményeinek és gyakorlati alkalmazásainak áttekintése [Empirical Industrial Organization in Hungary: A Survey of Research and Applications]. Verseny és szabályozás [Competition and Regulation], 2012, pp. 13–31.

László Halpern and Balázs Muraközy • A verseny és a K+F összefüggései – elméleti megközelítések és számszerű eredmények [The Relationship between Competition and R&D: Theoretical Approaches and Quantitative Results]. Verseny és szabályozás [Competition and Regulation], 2011, pp. 13–37.

Ferenc László Kiss • A vállalati termelékenység elemzése vállalatközi összehasonlítás és dekompozíció segítségével [Inter-Firm Comparison and Decomposition of Productivity Gains for Regulatory Purposes]. Verseny és szabályozás [Competition and Regulation], 2008, pp. 206–244.

Csongor István Nagy • A tagállami árszabályozás keretei az Európai Unió jogában, különös tekintettel a liberalizált piacokra [The Framework for National Price Regulation in The European Union, with Special Emphasis on Liberalized Markets]. Verseny és szabályozás [Competition and Regulation], 2013, pp. 121–144.

Iván Major and Károly Miklós Kiss • Az összekapcsolási díjak ösztönző szabályozása rosszul informált szabályozó esetén [Interconnection and Incentive Regulation in Network Industries]. Verseny és szabályozás [Competition and Regulation], 2009, pp. 149–170. An English version was first published in *Acta Oeconomica*, Vol. 63, 1, pp. 1–21 (2013).

László Paizs • Ösztönzési problémák a kiegyenlítő energia hazai piacán [Incentive Problems in The Hungarian Energy-Balancing Mechanism]. Verseny és szabályozás [Competition and Regulation], 2007, pp. 179–196.

András Kiss • A regionális árampiaci integráció hatása az erőművek piaci erőfölényére [The Effect of the Regional Integration of Electricity Markets on the Market Power of Power Plants]. Verseny és szabályozás [Competition and Regulation], 2007, pp. 139–155.

Péter Kaderják, András Kiss, László Paizs, Adrienn Selei, Pálma Szolnoki and Borbála Tóth • Infrastrukturális fejlesztések szerepe a gázpiaci integrációban. Elemzések a Duna-régió gázpiaci modellel [Natural Gas Market Integration in The Danube Region: The Role of Infrastructure Development]. Verseny és szabályozás [Competition and Regulation], 2012, pp. 256–282.

Balázs Édes, Vanda Bölcskei, László Lőrincz, Péter Nagy and Zoltán Pápai • Vezetékes és mobiltávbeszélő-szolgáltatások közötti helyettesítés vizsgálata [Substitutability between Fixed and Mobile Telephone Services]. Verseny és szabályozás [Competition and Regulation], 2009, pp. 204–242.

Károly Miklós Kiss • Hungarian Postal Sector. The chapter based on two studies in the series: A postai piacok szabályozásának újabb fejleményei [New Developments in Postal Regulation]. Verseny és szabályozás [Competition and Regulation], 2009, pp. 270–298. and Piacnyitás a postai piacokon [Market Opening in Postal Sector]. Verseny és szabályozás [Competition and Regulation], 2012, pp. 321–354.

LIST OF CONTRIBUTORS

- VANDA BÖLCSKEI completed her Master studies at Budapest Corvinus University in 2006. She is an analyst at Infrapont Economic Consulting since 2006. Her main areas consist of telecommunication regulation and market analysis. She participated in projects regarding regulation of other network industries, e.g. post. She has publications on topics regarding telecommunication demand and behavioural economics in Közgazdasági Szemle. Email: bvanda2003@yahoo.com.

- GERGELY CSORBA has been working at the Institute of Economics since 2005. He finished his Master studies in Economics at the Budapest University of Economic Sciences and at the University of Toulouse, and received his PhD degree in Economics from the Central European University. His research focuses on theoretical and empirical Industrial Organization and Competition Policy, and published on this topic in referred international journals (Journal of Industrial Economics, International Journal of Industrial Organization, Information Economics and Policy, Journal of Industry, Competition and Trade). He also acted as the Chief Economist of the Hungarian Competition Authority (GVH) from 2006 till 2012. From 2012, he has been working together with Infrapont on various competition policy and regulatory projects. Email: gergely.csorba@gmail.com.

- BALÁZS ÉDES economist, graduated in 2006 at the Corvinus University of Budapest. Currently PhD student at the Department of Logistics and Supply Chain Management at the same university. He worked as a consultant at Infrapont between 2007 and 2010. His main area of work covered regulatory issues of network industries, especially the telecommunications and postal sector. Parallel to this he has been engaged with the field of transportation since 2005 – he participated in several research and advisory projects. Between 2011 and 2012 he acted as an advisor in the president DG's cabinet at MÁV Zrt. He worked as an expert at the Budapesti Közlekedési Központ, aiding the CEO between 2013 and 2014. Email: ebalazs@gmail.com.

- LÁSZLÓ HALPERN is director at the Institute of Economics, Centre for Economic and Regional Studies of the Hungarian Academy of Sciences in Budapest. He is research fellow in International Macroeconomics and International Trade research programmes of the Centre for Economic Policy Research (London) since 1996, and recurrent visiting professor at the Central European University (Budapest) since 1997. He is member of Scientific Committee of European Trade Study Group, and of Advisory Committee of Department of Economics at Eötvös Loránd University

(ELTECON, Budapest). He was member of the Research Advisory Committee of the Czech National Bank between 2004 and 2010. He has written widely on exchange rate and exchange rate policy in Central and Eastern Europe; on enterprise behaviour, on microeconomic environment and economic policy; on multinational enterprises, on foreign direct investment and economic development. He was given the Popovics Prize of the Magyar Nemzeti Bank in 2013. Email: halpern.laszlo@krtk.mta.hu.

- PÉTER KADERJÁK is the Director of the Regional Centre for Energy Policy Research at the Corvinus University of Budapest (www.rekk.eu). He received his MSc in economics from the Budapest University of Economic Sciences in 1987. In 1998 he was appointed as Chief of Cabinet of the Minister of Economic Affairs and started to work on the liberalisation of the electricity and gas sectors in Hungary. In January 2000 he became the President of the Hungarian Energy Office, the national energy regulator. Between 2000 and 2004 he also served as the Chairman of the Energy Regulators Regional Association (ERRA), an association of energy regulatory institutions of countries from Central and Eastern Europe, the CIS and South East Europe. Since 2004 he has been serving as Training Director for ERRA's in-house energy regulatory trainings. He is also directing a postgraduate program in Energy Economics at Corvinus University since 2010. He is research partner in the "European Energy Institute" at University of Leuven and a regular lecturer at the Florence School of Regulation. He has directed several recent research efforts with regional relevance. In 2011 he was appointed as alternate member of ACER's (Agency for the Cooperation of Energy Regulators) Board of Appeal. Email: pkaderjak@uni-corvinus.hu.

- ANDRÁS PÉTER KISS is an assistant professor at the Economics Department of the University of Amsterdam. His research interests include industrial organization, applied microeconomics, and behavioural economics. He currently works on estimating the role of inattention in consumer switching behaviour. Previously, András was a senior research associate at the Regional Centre for Energy Policy Research (REKK), specializing in the economic analysis and numerical equilibrium modelling of electricity and natural gas markets. He holds a PhD degree from Central European University (2015), an M.Sc. in Economics from University College London (2005), and a joint BA/MA degree from Corvinus University, Budapest (2004). Email: kiss.andras@gmail.com.

- FERENC LASZLO KISS, Ph.D., is an internationally well-known telecommunications economist, the author of nearly a hundred international publications. At Bell Canada he conducted productivity analysis and developed various simulation and econometric models during the 1970's. At Bell Communications Research in the United States, in addition to econometric productivity and demand analyses, his work also involved the investigation of the economic impact of newly developed digital technologies. He acted as occasional referee for The Journal of Industrial Economics, The International Journal of Forecasting, Review of Industrial Organization, Information Economics and Policy. For 16 years he was on the Board of Directors of the International Telecommunications Society (ITS). In 1985–6 he was a member of the organizing committee of the Telecommunications Policy Research Conference (TPRC). His involvement in Hungarian telecommunications began in 1988. He worked on a World Bank project to modernise the management information system of telecommunications operations. The success of this project led to further consulting assignments in Hungary, then in Austria and the Netherlands. He also did consulting work in Canada, the United States, New Zealand, Australia and South America. In 1992 Mr. Kiss was invited to become Chief Economic Advisor to Matáv, the Hungarian national carrier, newly divested from the Post Office. He provided complex and extensive help during the preparatory stages of privatization. Next he joined the newly formed communications authority as Chief Economic Advisor, and was instrumental in establishing the functions and organisation of a modern-day regulator. In the meantime he also consulted some large Hungarian corporations such as the electric utility (MVM), the Paks nuclear power station, the oil-gas utility (MOL). During the last 12 years he continuously taught various courses at the Budapest University of Technology, including the "info-communications" special program of the MBA School. In 2002 he joined Pannon University in Veszprém, Hungary and became head of the e-Economics Department. He retired from consulting in 2003 and teaching in 2012. In 2007 he was invited to be Editor-in-Chief for Competition and Regulation. Email: fkconsultingusa@yahoo.com.

- KÁROLY MIKLÓS KISS is a senior research fellow and the Head of Economics of Networks Research Unit at the Institute of Economics of the Hungarian Academy of Sciences, and associate professor at University of Pannonia. From 2009 until 2013 he was the Head of Department of Economics at University of Pannonia, Veszprém. His research is focused on Industrial Organization, Economics of regulation (mostly in network industries and public utilities), Economics of networks, Economics of Information (asymmetrical information). He has participated and coordinated in several research projects for Hungarian Infocommunications Authority, Hungarian Competition Authority, ministries. Email: kiss.karoly.miklos@krtk.mta.hu.

- ANDRÁS GYÖRGY KOVÁCS habil. Ph.D., is a lawyer-economist, working at the Supreme Court of Hungary as a judge and as a head of the III. judicial chamber. For the last 12 years he has been adjudicating competition law and sector-specific market-regulation cases, including cases of electronic communication and energy law. He is also the leader and coordinator of the administrative law section of the Network of Judicial Advisers in European Union Law. András is a professor of public, EU and international law at Kodolányi János University of Applied Sciences. András is also the author of the book "Market-Regulation and Legal Remedies" (2012, HVG-ORAC), covering all the areas of access to justice in the field of sector-specific competition law. Email: KovacsAndras@kuria.birosag.hu.

- LÁSZLÓ LŐRINCZ graduated at the Budapest University of Economic Sciences at 2002. He earned his PhD degree at the same institution (Corvinus University of Budapest, Doctoral School of Sociology) in 2009. He worked as a consultant at Infrapont Ltd. between 2002 and 2009 carrying out analysis and research primarily about economic regulation of telecommunications. Additionally, he took part in research projects about information and communication technologies at Institute of Sociology, Corvinus University of Budapest. From 2010 his primary research interest is social network analysis, as an external member of HAS-CSS RECENS. Between 2012 and 2013 he worked as head of unit and senior colleague at ECOSTAT Government Centre for Impact Assessment. He joined the "Economics of Networks" Research Unit of the Institute of Economics, HAS at 2013, where he participates in research about labour mobility networks between companies and about firm product diversification. Email: lorincz.laszlo@krtk.mta.hu.

- IVÁN MAJOR is professor of economics at the Budapest University of Technology and Economics, senior research fellow at the Institute of Economics, the Hungarian Academy of Sciences and recurrent visiting professor at the Departments of Economics, UCSD, USD and SDSU, California, U.S.A. He has done research in industry regulation – in telecommunications, transportation, the banking sector –, and privatization and restructuring in Central and East European countries after the system change. He is currently working on a research project that addresses the issues of performance and cost efficiency of the health care industry. Email: major.ivan@krtk.mta.hu.

- BALÁZS MURAKÖZY is the director of the Firms, Strategy and Performance Research Group at the Research Centre for Economic and Regional Studies, Hungarian Academy of Sciences. Balázs received his PhD from the Central European University in 2008. His research interest is centred on applied IO, international trade as well as regulation and competition policy. His work was published among others by the Journal of International Economics and the Canadian Economic Review. He also spent time as visiting researcher at UC Davis, IFW Kiel and the University of Nottingham. He has won the large "Lendület" grant from the Hungarian Academy of Sciences in 2013. Balázs is recently teaching an Econometrics course in Rajk László College for Advanced Studies. Email: murakozy.balazs@krtk.mta.hu.

- CSONGOR ISTVÁN NAGY is the head of the Federal Markets "Momentum" Research Group at the Hungarian Academy of Sciences and associate professor and the head of the Department of Private International Law at the University of Szeged, Faculty of Law; he is recurrent visiting professor at the Sapientia University of Transylvania (Romania) and at the Riga Graduate School of Law (Latvia) and visiting professor at the Central European University. He is admitted to the Budapest Bar. In 2015-2016, he is Fulbright visiting professor at Indiana University, Bloomington. Professor Nagy graduated at Eötvös Loránd University of Sciences (ELTE, dr. jur.) in Budapest, in

2003, where he also earned a PhD in 2009. He received master (LL.M, 2004) and S.J.D degrees (2010) from the Central European University (CEU) in Budapest/ New York. He earned habilitation at the University of Szeged in 2014. He pursued graduate and postgraduate studies in Rotterdam, Heidelberg and Ithaca, New York. He had visiting appointments in The Hague (Asser Institute), Munich (twice, Max Planck Institute), Brno (Masarykova University), Hamburg (Max Planck Institute), Edinburgh (University of Edinburgh) and London (British Institute of International and Comparative Law). Professor Nagy has more than 140 publications in English, French, German, Hungarian, Romanian and (in translation) in Croatian and Spanish. Email: csongor.nagy@gmail.com.

- PÉTER NAGY senior consultant, completed his MA studies at the Budapest University of Economic Studies. Between 1998 and 2000 he was senior fellow at the Strategy Directorate of Hungarian Telecom (MATÁV), where he was charged with developing the company level market model. Since 2000 he has been senior consultant at Infrapont participating in each major project of the company, such as providing expert support in the liberalization of Hungarian telecommunication markets and preparation of the electronic communication laws. He participated in several consultancy projects for the Hungarian Telecommunications Regulator and for telecommunications operators. He also has experience in other network industries, such as the electricity and the postal sector, and has been participating in projects related to competition cases. His main areas consist of market analysis, impact assessment, enhanced empirical research, price regulation, economic preparation of legal regulation, business planning, and market modelling. Email: peter.nagy@infrapont.hu.

- LÁSZLÓ PAIZS is a research fellow at the Institute of Economics, Research Centre for Economic and Regional Studies of the Hungarian Academy of Sciences and a founding partner of the Regional Centre for Energy Policy Research at the Corvinus University of Budapest, an independent energy policy think-tank. He completed his Master studies in Economics at the Budapest University of Economic Sciences and at the Central European University. He received his Ph.D. in 2007 from the Central European University. His research interest includes tax competition, the regulation of energy supply industries and energy market modelling. Email: paizs.laszlo@krtk.mta.hu.

- ZOLTÁN PÁPAI partner and CEO of Infrapont Economic Consulting, completed his Master and doctoral studies at the Budapest University of Economic Studies. He has been working as a regulatory and competition economist since the beginning of the nineties. Zoltán has regulatory and competition expertise in telecommunication, media, energy, postal services, transport and up to now has been involved in many project in these and also in other industries. He took part as an expert in the implementation of the liberalization network industries and setting up a new regulatory systems, preparation of mobile license auctions, completion of market analyses, impact assessments, and provided expert support in legal cases. Besides consultancy he has been involved in applied economic research concerning market power and competition issues in different markets. Email: zoltan.papai@infrapont.hu.

- ADRIENN SELEI has been working for Regional Centre for Energy Policy Research since 2011. Her work especially includes gas market modelling, but she has been also involved in different projects in the field of electricity markets. She has already finished her Phd studies in Economics (at Corvinus University of Budapest) and she is currently writing her thesis. Beyond the analyst and consultant activities she is teaching numerous courses in the field of economics at Eötvös Loránd University and Corvinus University of Budapest. Email: adri983@gmail.com.

- PÁLMA SZOLNOKI worked for the Regional Centre for Energy Policy Research between 2005 and 2012. From 2012 until 2013 she worked as a market surveillance analyst at CEEGEX Central Eastern European Gas Exchange Ltd. In 2013, she became a business development expert at MVM Partner's natural gas division. Currently she is doing her PhD studies in management and business administration at the Corvinus University of Budapest. Her research interests include electricity and natural gas markets analysis. Email: szolnokipalma@gmail.com.

- BORBÁLA TAKÁCSNÉ TÓTH has worked with REKK since its creation in 2004. With REKK she has been leading several international and national consultancy projects, with many using the European Gas Market Model as the primary analytic tool. Her main fields of expertise include: regional co-operations; security of supply issues; energy geopolitics; major infrastructure initiatives in the gas sector and incentives for investments; competition cases in the gas market; and the effect of gas release programs on competition in the gas market in Europe. In 2001 she received an M.A. in International Relations and European Studies at the Central European University in Budapest. Borbála was educated as an economist and received her degree from the Budapest University of Economic Sciences in 1998. She spent 5 years as a civil servant in government administration mostly in the field of energy regulation. Email: borbal.toth@gmail.com.

- PÁL VALENTINY is head of IO studies at the Institute of Economics, Centre for Economic and Regional Studies of the Hungarian Academy of Sciences in Budapest. He was invited by the University of Sussex (1980), University of Nottingham (1991) and University of Cambridge (1991) as visiting fellow. His principal interests lie in the theory of regulated markets, regulation and privatization of network utilities and economic aspects of competition policy. His research has also focused on public procurement issues, innovation policies, corporate governance, ICT and growth. His teaching experience covers several universities and colleges, he is visiting professor at the Department of Economics, Eötvös Loránd University (ELTECON, Budapest).

He was adviser to the Hungarian Communication Authority (1995–2003) and to the Hungarian Railways (1997–2002) on regulatory and restructuring issues. He has regularly provided assistance to the Hungarian Competition Authority in economic aspects of competition policy issues. He was member of the Board of Directors at the Hungarian Power Companies Ltd (1994–1998) and the Budapest Gas Company Ltd (1993–2007). He has advised the Budapest Municipality on privatization strategies. As expert witness he was involved in judicial reviews of regulatory decisions in telecommunication. Email: valentiny.pal@krtk.mta.hu.

ABSTRACTS OF THE COMPETITION AND REGULATION SERIES

2013–2014

COMPETITION AND REGULATION • 2013

I. OVERVIEW

Gergely Csorba • What Kinds of Standards are Applicable to Economic Analysis and its Judicial Review? A Competition Economist's Interpretation

The paper demonstrates the role of economic analysis in competition policy case, and discusses the best practices that could be expected from the parties building on such economic arguments. Based on the summary of European case law, it discusses the evaluation criteria applied by the European Courts when they reviewed the most important cases involving economic reasoning (which are mostly mergers). Finally, it discusses two recently decided cases by the European Courts (the Ryanair/Air Lingus merger and the abuse of dominance case against Tomra) to demonstrate whether these more settled evaluation criteria did indeed increase the quality of economic reasoning applied, but it finds mixed results.

László Lőrincz • The Role of Competition and Competition Policy in Promoting Economic Growth

Competition policy in addition to a number of other policy areas impacts the intensity of market competition. This literature review first identifies these areas. Afterwards, it discusses the theoretical relationships between innovation, productivity and growth, and also touches the necessary fundamentals of macroeconomic growth models. Next, the results of econometric analyses on the relationship between competition and innovation and between competition and productivity are presented, which is an area that has shown significant progress the recent years. When discussing the econometric results, it also presents the policy measures, which were used for identification. Based on the presented studies, liberalization essentially can be associated with improvements in productivity, but between innovation and competition an inverted U-shaped relationship can be assumed. Considering the direct relationship between competition policy and growth, suspending competition policy measures in the United States in the thirties was shown to cause substantial negative impact on the GDP. In the final sections, the study briefly discusses the results considering the effect of trade liberalization on economic growth, the consequences of bankruptcy regulations, and the effects of reallocation policies on productivity and growth.

Judit Szabó • State Aid and Public Service Obligations on the European Market: The European Regulation of Services of General Economic Interest

The paper reviews the recent European regulation of services of general economic interest (SGEI) from an economic point of view. The emphasis is on the significant changes put forth by the European Union rules on state aid to services of general economic interest („State aid SGEI package") during the years 2005-2006 and 2011–2012. The paper outlines also important previous changes in some relevant European rules and that of some relevant judgements of the European Court of Justice (ECJ) and the Court of First Instance (CFI). The paper puts the emphasis on the birth of the compensation approach of state aid and gives an economic interpretation of this approach and its application. The paper doesn't evaluate the performance of the compensation approach in practice, but gives a detailed survey of its forming since the 1970's until being part of the European rules in the 2000's. This review might make some grounding for the future empirical analyses.

II. THE TRANSFORMATION OF PRICE REGULATION

Csongor István Nagy • The EU Law Framework of National Price Regulation, with Special Emphasis on Liberalized Markets

The paper, in the first step, examines the applicability of EU competition rules (antitrust and state aid) to national price regulation. Afterwards, it analyses the requirements against regulated prices erected by free movement (internal market) law. The paper also examines how market-opening and liberalization rules (and the judgments of the European Court of Justice in Federutility and ENEL) affect the Member States' playing field as to price regulation. It demonstrates the differences between electricity and natural gas sectoral regulation entailed by the fact that the former qualifies as an EU universal service, while the latter does not. Finally, it synthetizes the EU law requirements against national price regulation and assesses the jurisprudence of the European Court of Justice from a critical perspective.

Balázs Felsmann • The Impacts of the Price Regulation on the Level of Retail Electricity Prices in Europe

There is a strengthening debate within the European Union in recent years about the impact of the affordable industrial and household electricity prices on the general competitiveness of European economies. While the European Institutions argues for the further liberalization of the energy retail sector, there are others who believe in centralization and price control to achieve lower energy prices. Current paper reviews the regulatory models of the European countries and examines the connection between the regulatory regime and consumer price trends. Although the current regulatory practice is heterogeneous within the EU member states, there is a clear trend to decrease the role of regulated tariffs in the end-user prices. The study did not find a general causal relationship between the regulatory regime and the level of consumer electricity prices in a country concerned. However, the quantitative analysis of the industrial and household energy prices by various segments detected significant differences between the regulated and free-market countries. The first group of member states tends to decrease the prices in the low-consuming household segments through cross-financing technics, including increased network tariffs and/or taxes for the high-consuming segments and for industrial consumers.

Pál Belényesi • Pricing of Water Services: The Hungarian Market

Pricing of water supply and related services has been, and continue to be, a principal concern in Hungary – and not purely as part of the conundrum around the government's price minimizing propaganda of utility payments. The exclusion of the water services from European-wide liberalization, their restriction for concession-related activities in the directive designed to open up markets in the EU are only some of the worrying elements. Renationalization of network elements and forced appropriation of the service provision are attempts to consolidate water service provision in Hungary. The paper examines the influential European trends of water supply provision in the light of the 2011 market reform in Hungary. It is examined whether benchmarking, competition for the market, regulatory competition, regulated monopolist, free market competition would qualify as good examples for the Hungarian market. The importance of environmental externalities in water pricing is introduced in detail.

Károly Miklós Kiss and Zsolt Stenger • Price Regulation in Postal Markets

The first part of the study summarizes the purpose and causes of price regulation and its relations with other regulatory targets. Not only pure efficiency considerations show up in prices, since they are limited by some welfare objectives, such as the universal service obligation. Then it presents the development of the EU regulatory framework of postal services as well as the implementation of main requirements (cost-based, transparent, non-discriminatory and affordable pricing) defined in the guidelines at Member State level. In the next section the specifications of cost-accounting are reviewed as the basis for price regulation. Then it discusses the features, advantages and disadvantages of the applied pricing principles and methods (ex-ante, price cap or ex-post). In a separate section it analysed the characteristics of the Hungarian postal regulation. Closing the study some critical remarks were drawn up about the price regulation of postal services in the EU.

III. MARKETS AND REGULATIONS IN NETWORK INDUSTRIES

Gábor Koós • **Market Opening in Rail Transport: The Fourth Railway Package**

The aim of this study is to shortly present the fourth railway package announced by the European Commission as the newest stage of the European rail market opening process focusing on its differences to the earlier stages. Although the objective of the Commission’s proposal was quite clear and progressive already the amendments adopted by the European Parliament in the end of February 2014 has shown that it will be very difficult to get through them especially in the so-called market pillar. To bring the gradual market opening process to the end is highly questionable therefore it is worth to detect from time to time the problems of this sector and to discuss the attempts made by the Commission to fulfil its mission to make this sector much more effective. The study contributes to this task through systematically presenting the main content of the fourth railway package, the main results of the recast and the main consequences of the judgements of the ECJ in this sector.

Pál Valentiny • **Institutional Changes at Regulatory Authorities**

The paper assesses the development of institutional settings at regulatory and competition authorities. It follows their path from municipal concession contracts to state and federal level regulation. Technological change, competitive pressures and convergence between services are emphasized among the prime movers. The study discusses the experiences in the US and in the EU in the context of changing relations between regulation and competition policy.

András Mezősi and László Szabó • **Cost-Benefit Evaluation of Electricity Network Investments in Central Eastern Europe**

The paper analyses the complex welfare impacts of proposed transmission investments in the Central Eastern Europe (CEE) region with the application of the EEMM electricity model. The applied model is bottom-up model, having 36 European countries, 85 aggregated transmission connections and detailed technological breakdown of the European power systems (over 5000 generation units are modelled). The assessment is made at regional level, as new transmission lines have significant spillover effects over third countries. The cost-benefit assessment (CBA) focused on the CEE region identifies those transmission lines that increase the regional welfare the most. In addition, the paper also identifies those methodological and policy issues, that have significant impact on the results. The results indicate, that many projects are interlinked, so an individual project assessment would only bring partial results.

COMPETITION AND REGULATION 2014

I. OVERVIEW

Gergely Csorba • Competition Policy Assessment of Parallel Events in Merger Control

This article discusses the questions concerning the competition policy assessment of parallel events. First, it reviews the statements of economic theory on sequential mergers and optimal merger policy. Then it presents the implications derived from the European Commission's general methodology on merger assessment, and the arguments for and against the priority principle used in practice. Finally, it presents three groups of mergers to discuss the various problems that can arise and how the scarce case practice dealt with them.

Pál Belényesi • How Do I Get My Money? The Economics behind the Design and Application of Contingency Fee Arrangements in Antitrust Private Damages Actions

The present study examines the more significant country examples of contingency fee arrangements in Europe, in particular for competition law related private damage claims. Furthermore, it analyses the economics behind such agreements, the reasons why such agreements are on the borderline of competition economics and law. The final section concludes on what lies behind efficiency claims of both parties. Second part of the study observes the financial and the economic arrangements that urge or discourage potential private claimants from initiating such actions. Finally, this section details the specific arrangements, which may have direct impact on the right to access to justice. In its conclusions, the paper argues for a parallel application of private and public tools in order to draw the right balance between frivolous suits of private parties and the under-deterrence of administrative fines when it comes to mandating contingency fee arrangements in EU Member States. The paper argues that the current framework – in spite of the recent changes in EU legislation – is still imperfect and significant improvements are needed. A well-designed, promoted and efficiency-based contingency fee coupled with class action arrangements could substantially expand the possibilities for the parties involved.

Csongor István Nagy • The Right to Fair Trial and Judicial Review in Competition Proceedings

The paper examines the right to fair trial as to the judicial review of competition authority decisions. It presents the general administrative law models of judicial review and analyses the recent judgments of the European Court of Human Rights, the European Court of Justice and the Hungarian Supreme Court (Kúria); it also addresses the structural questions of Hungarian competition procedure.

II. ECONOMICS IN ANTITRUST AND REGULATION

Pál Valentiny • Does Forensic Economics Really Exist?

In the last couple of years forensic economics obtained a JEL code, as a tribute to its scientific performance. This paper is about the emergence of this sub-discipline and the different interpretations of its notion. First the study provides a brief insight into the development of the discipline of industrial organization and its use in legal enforcement. Then mapping the procedures in antitrust and regulation it examines the changing relationship between law and economics and lawyers and economists. Finally it delivers an answer for the question in the title.

Balázs Muraközy • Application of Quantitative Methods in Selected Competition Law Cases, 2009–2013

This study discusses the role of quantitative methods in some important cases of four competition authorities between 2009 and 2013. The most complex quantitative methods, like merger simulation, can only be applied in ‘textbook’ cases in reality. When data are less suitable, market structure or the competitive concern is less standard, often less sophisticated empirical methods are applied. The cases we focus on suggest that recently the role of measuring the positive effects of mergers, non-horizontal fusions and specialized models for some markets has increased and these issues also present important empirical challenges.

Krisztina Antal-Pomázi • On the Choice of Court-Appointed Experts

This paper examines the use of court-appointed experts under Federal Rule of Evidence 706. The use of economic expert witnesses is common in civil litigation. In an adversarial system, expert opinions are expected to be contradicting, or even slanted according to the interests of the parties. The more disputed the field of science concerned, the harder a judge or jury may find it to decide on the expert testimonies' credibility. It often happens that all expert opinions are excluded, and the case is decided on the basis of the non-expert evidence. To avoid the welfare loss caused by the waste of resources spent on this way of proof, several authors recommend the more extensive use of court-appointed experts. Taking in concern that it might be hard for a judge to pick a 'neutral' on his own, the paper first examines the proposal that the parties should agree upon a 'neutral' expert to be appointed by the court. The more informed the parties are, the less likely are they to reach an agreement. Second, an alternative mechanism is proposed for the uninformed judge to choose a 'neutral' expert on the basis of information acquired from the well-informed parties. The mechanism implements in Nash-equilibrium the social choice rule that the plaintiff should win the lawsuit if and only if she is right.

László Paizs • Market Power in the Electricity Generation Market: A Survey of Game-Theoretic Electricity Market Models

A large volume of research on strategic behaviour on deregulated electricity markets has accumulated over the last 25 years. Among the different approaches, the study surveys the game-theoretic models of power generator behaviour. It provides a review of both theoretical and empirical findings in the oligopoly and auction literature.

Balázs Muraközy and Pál Valentiny • On the Measurement of Welfare Effect of Competition Policy

In parallel with the increasing role of competition policy and data availability, assessing the decisions of Competition Authorities becomes more and more important. After discussing the possible assessment frameworks, it focuses on one specific method, the ex ante impact assessment of consumer welfare conducted by the Competition Authorities, and shows how it is applied in different countries, what are its main challenges and their possible solutions.

III. MARKETS AND REGULATIONS IN NETWORK INDUSTRIES

Zoltán Pápai and Bertalan Papp ◆ **Competition Issues on the Liberalized Postal Markets and the Evaluation of the Hungarian Experience**

The study examines the competition problems that emerged in the postal markets throughout the multi-step liberalization process of the European Union. Following a review of the prevailing barriers to entry, the paper investigates 1) the deterrence arising from incumbent strategies, 2) the legal and administrative hindrances from state regulation, and 3) the regulatory discrimination of the incumbent. Finally, the Hungarian liberalization act of postal markets is scrutinized, including the assessment of the 2012 Postal Law's effect on regulatory barriers. The study comes to the conclusion that, instead of allowing for a meaningful competition on the postal letter market, the Hungarian state policy and regulation is still protective of the incumbent postal operator.

Adrienn Selei and Borbála Takácsné Tóth ◆ **Short-term Effects of the Ukrainian Crisis on the Security of Gas Supply in Central-Eastern Europe and Hungary**

In this paper the vulnerability and the short term resilience of the European gas market to supply side shocks are modelled. The paper analyses the effect of regulatory interventions on the wholesale gas price and welfare of the most vulnerable Central and Eastern European gas markets. A short term security of supply situation is modelled through a one month disruption of Russian gas supplies through the Ukraine in January, and a longer term crisis is modelled through a six month disruption on the same route. Three short term measures were analysed that aim to increase the resilience of gas markets to supply disruptions in the short run through the better use of existing infrastructure: by allowing spot trade on all interconnection points; allowing virtual trade against the physical flow; allowing bidirectional flows on all EU-EU borders. It was found that the first two measures have rather local effect on the markets, but bidirectional physical flows increase the security of supply of the whole region - mostly because of better connecting Central and Eastern Europe to the Western gas markets. It concludes that Hungarian consumers realize significant part of the monetized benefits that are measured; hence Hungary has a profound interest in supporting the European gas market integration. It was also found that the Hungarian strategic storage is a useful tool to mitigate the damages of a potential crisis not only within the Hungarian borders but also on a regional basis.

CONTENTS OF THE COMPETITION AND REGULATION BOOK SERIES 2007–2014

COMPETITION AND REGULATION 2007

COMPETITION AND REGULATION 2008

COMPETITION AND REGULATION 2009

COMPETITION AND REGULATION 2010

COMPETITION AND REGULATION 2011

COMPETITION AND REGULATION 2012

COMPETITION AND REGULATION 2013

COMPETITION AND REGULATION 2014

INSTITUTE OF ECONOMICS • CERS • HUNGARIAN ACADEMY OF SCIENCES

IE CERS HAS

Hungarian
Competition
Authority

The Institute of Economic Sciences at the Hungarian Academy of Sciences launched a new series of publications entitled "Verseny és szabályozás" (Competition and Regulation) in 2007. Eight annual volumes have been published so far, all in Hungarian. The current volume is the first one in English, and it contains 12 selected translations from the crop of the first seven years. It offers the reader a glimpse into the current state of research in its chosen field in Hungary. The published studies covered a very broad range of topics. Some articles of general theoretical and methodological nature dealt with the background in the law and economics of regulated markets. Others investigated current legal, economic and policy issues and cases. Others again dealt with regulation and the regulators themselves. The functions, methods, analytical tools, the institutions and the impact of regulation were discussed in those articles. Special attention was paid to regulation by the European Union, and also to recently de-monopolized key industries such as communications, energy, media, the postal sector or water and sewage. The publications were designed to provide a meeting place for economists and lawyers to work together on the economic background of legal problems and the legal solutions to economic problems. Five of the 12 articles selected for publication in English in this tome deal with broad economic and legal issues of competition and regulation, while the remaining 7 discuss the state and specific problems of key industries in Hungary and, in some cases, in the surrounding region.

ISSN 1789 9702
9 771789 970051 20150