INSTITUTE OF ECONOMICS
Centre for Economic and Regional Studies

Közgazdaság- és Regionális Tudományi Kutatóközpont
KÖZGAZDASÁG-TUDOMÁNYI INTÉZET

KRTK-KTI WORKING PAPERS | KRTK-KTI MŰHELYTANULMÁNYOK

Digitalization against the shadow economy: evidence on the role of company size

BÁLINT VÁN – GÁBOR LOVICS – CSABA G. TÓTH – KATALIN SZŐKE

KRTK-KTI WP – 2022/24

December 2022

https://kti.krtk.hu/wp-content/uploads/2022/12/CERSIEWP202224.pdf

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES
EÖTVÖS LORÁND RESEARCH NETWORK (ELKH)
BUDAPEST, 2022

ABSTRACT
Online cash registers (OCRs) are important tools for reducing the size of the shadow economy. This paper analyzes the impact on reported turnover and tax liability of introducing OCRs in Hungary using a fixed-effects panel and event study model. We identify strong size-related heterogeneity in the retail and the accommodation and food services sectors: smaller companies increased their reported turnover more than larger ones. Since large companies pay the dominant part of value-added tax, the effects on the payment of this tax were mitigated. We find significant spillover effects in both sectors, which are slightly stronger among larger companies.

JEL: E26, H25, H26
Keywords: Value-Added Tax, Tax Evasion, Shadow Economy

BÁLINT VÁN
Ministry of Finance
e-mail: balint.van@pm.gov.hu

CSABA G. TÓTH
KRTK KTI and CIAS
e-mail: toth.gcsaba@krtk.hu

GÁBOR LOVICS
Hungarian Central Statistical Office
e-mail: gabor.lovics@ksh.hu

KATALIN SZŐKE
Central Bank of Hungary
e-mail: szokek@mnb.hu

Digitalizációval a rejtett gazdaság ellen: a vállalati méret szerepe

VÁN BÁLINT – LOVICS GÁBOR – TÓTH G. CSABA – SZŐKE KATALIN

ÖSSZEFOGLALÓ

Az online pénztárgépeknek fontos szerepük lehet a rejtett gazdaság visszaszorításában. Tanulmányunkban vállalati szintű adatbázisok, illetve különböző ökonometriai módszerek felhasználásával arra keressük a választ, hogy az online pénztárgépek magyarországi bevezetése miként hatott a bejelentett értékesítési forgalomra, illetve az adófizetési kötelezettségre. Azt találtuk, hogy a kiskereskedelmi szektorban kisebb mértékben, a szálláshely-szolgáltatás és vendéglátás szektorban nagyobb mértékben növelte az online pénztárgépek bevezetése a bejelentett forgalmat. Eredményeink azt jelzik, hogy a fehéredési hatás fordítottan arányos a vállalat méretével: az online pénztárgépek bevezetése nagyobb mértékben növelte a kisebb cégek bejelentett forgalmát, mint a nagyobb vállalatokét. Mivel az általános forgalmi adó domináns részét éppen a legnagyobb vállalatok fizetik, ezért az online pénztárgépek bevezetésének mérsékelt hatása volt az áfa-bevételek alakulására. Mindkét ágazatban jelentős tovagyűrűző hatásokat találtunk, amelyek a nagyobb vállalatok körében kissé erősebbek voltak.

JEL: E26, H25, H26

Kulcsszavak: általános forgalmi adó, adókerülés, rejtett gazdaság

I. INTRODUCTION

Since computers, smartphones, credit cards, and many other devices now generate far more data on our habits and activities than ever before, the current period can easily be referred to as the age of big data (Mayer-Schönberger and Cukier, 2013). One aspect of the current information revolution is that authorities have a wealth of new data sources to collect information on individuals, firms, and other market participants. The use of new technologies and the resulting new data sources can change the behavior of market participants and promote their compliance behavior, thus leading to efficiency increases in several ways (de Mello and Ter-Minassian, 2020). Nevertheless, we still have limited knowledge about their real effect.

Electronic Fiscal Device (EFD) is a general term used for a wide variety of technological devices that can help tax authorities monitor business transactions (OECD, 2019). At the core of these fiscal tools is the automated and systemic sharing of sales data with the tax administration, and one of the most advanced types is the online cash register (OCR). Although the first EFDs were introduced in Italy in 1983, their origin can be traced to the 19th century (Varian, 2010). In the last two decades of the 20th century, EFDs were introduced in various other countries in the Mediterranean (OECD, 2013), Eastern Europe, and Latin America. At the beginning of the 21st century, various East African and some Asian countries also began using these devices (Eilu, 2018; OECD, 2013). As EFDs spread globally, they also became more and more advanced (Chacaltana et al., 2018; OECD, 2019); they could record more data, and some could connect and send information to the national tax authority.

As EFDs become increasingly popular worldwide (OECD, 2017), policymakers need empirical evidence of their effectiveness to guide the further introduction and development of such devices. The effect of EFDs has been examined in some countries, but the results, especially those based on macro data, are controversial. The experiences of European tax administrators suggest that the introduction of EFDs has not been associated with noticeable increases in value-added tax (VAT) revenues. However, together with other, simultaneously implemented reforms, it can increase tax revenues (Casey and Castro, 2015). In contrast, EFDs had a positive effect on VAT revenues in Tasmania, but it was smaller than expected (Fjeldstad et al., 2018). According to Mandari et al. (2017), awareness of the introduction of an EFD system is a key element of taxpayers' acceptance of that system and its increased impact.

Studies using microdata and methodology similar to ours show more favorable results (Awasthi and Engelschalk, 2018; Fan et al., 2018;). Eissa et al. (2014) find that the average effect of EFD introduction on VAT payments in Rwanda was 6.5 percent. In the paper most closely related to our own, Ali et al. (2015) estimate the EFD introduction effect in Ethiopia and find that the short-term effect (at a 1-quarter time horizon) on VAT revenues was 15 percent, while the long-term effect (at a horizon of 6 quarters) was 30 percent. They also find different effects for firms with institutional or personal ownership. The main difference between their paper and ours is that while they estimated the effect over time, we focus on the heterogeneity of the effects across different size categories of firms. In Sweden, the estimated effect of EFD introduction on reported turnover was 5.2 percent (Skatteverket, 2013); when they included smaller companies that report VAT on a lower frequency in the research, they obtained a slightly higher result of 7 percent. To the best of our knowledge, this is currently the only indirect evidence related to the relationship between company size and the impact of installing OCRs on VAT payments.

The novelty of our research is that we reveal the detailed role played by company size in the effect of OCR introduction on turnover and VAT payment by examining the Hungarian experience. Furthermore, we assess the spillover effect through the supply chain and the contribution to reducing the VAT gap. In addition to heterogeneity in company size, we investigate the difference in OCR effects between the retail and the accommodation and food services (AFS) sectors. Previous results are mixed in this respect: in Sweden, the effect of OCRs on turnover is higher in restaurants than in the retail sector (Skatteverket, 2013), while in Rwanda, the effect in restaurants is smaller than the one in retail and hotels (Eissa et al. 2014).

The paper is organized as follows. In Section II, we summarize the context of the OCR's introduction. Next, we present the data set and then describe the methodology. In Section V, we present our main findings on the effect of OCR introduction on reported turnover estimated by a fixed-effects panel model and summarize the related event study. In Section VI, we set out the impact on taxation and the size of spillover effects. Then, we present several robustness tests and conclude in the last section.

II. LEGAL, TECHNICAL AND ECONOMIC BACKGROUND

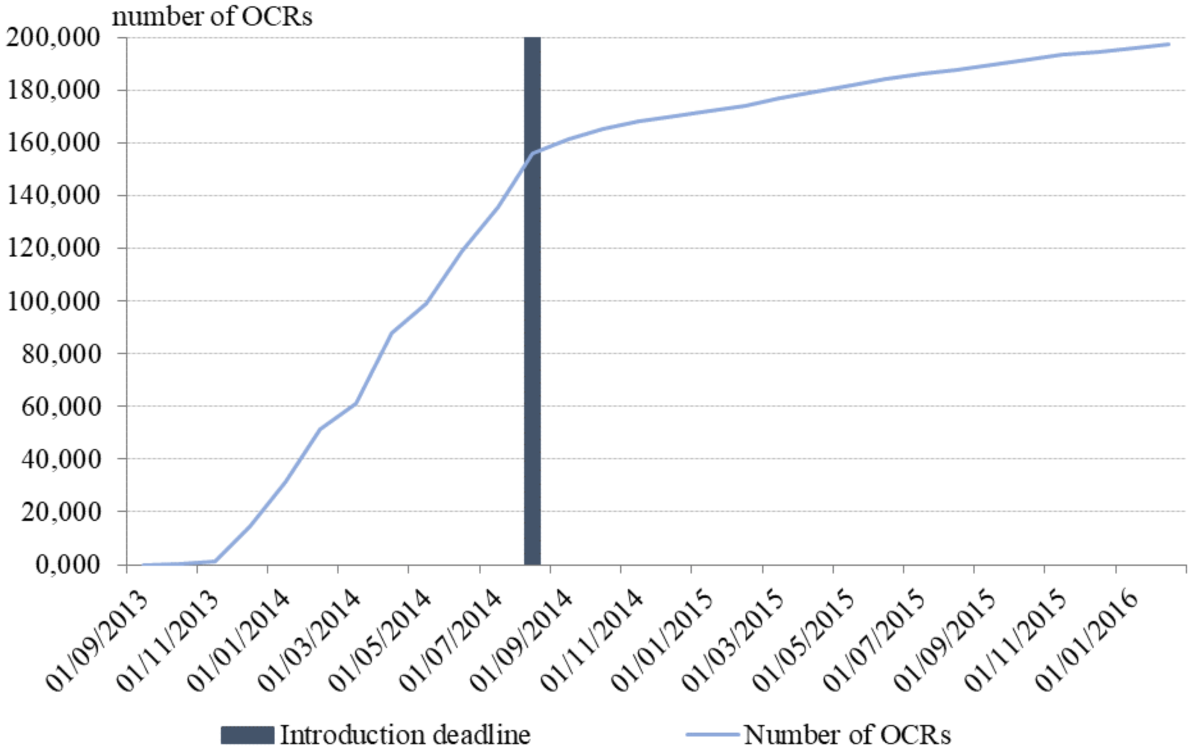
The introduction of OCRs was a part of a multi-year program (Ritzlné, 2021) aiming to reduce the size of the shadow economy. The latter includes all market-based legal production of goods and services deliberately concealed from public authorities to avoid paying taxes, legal labor market standards, and administrative procedures (Schneider et al. 2010). In addition to the OCRs, the core of this program in this period was the introduction of reverse taxation (from 2012) in several sectors (agriculture, construction) and the implementation of the Electronic Trade and Transport Control System (2015). In addition, the government made it compulsory to submit domestic itemized VAT summary statements (from 2013) and limited cash payments above HUF 1.5 million between persons regularly engaging in economic activities (from 2013).

The authorities' original aims in introducing OCRs were the following: (1) Enhancing market competition by reducing the number of sales without invoices, (2) Increasing the government's tax revenues by reducing the size of the shadow economy, and (3) Strongly supporting the control and selection processes of the National Tax and Customs Administration (NTCA).

The first draft of the relevant decree by the Finance Ministry was presented in December 2012, with an original deadline for implementation of April 1, 2013.[1] This deadline was postponed several times, and finally, the affected firms had to introduce OCRs before August 31, 2014. The number of OCRs increased continuously until the final deadline (see Figure 1). After August 2014, we continue to observe stable and moderate growth, but this is due to the opening of new firms that were required to introduce OCRs.

[1] For further use of the data of OCRs, see Illyés and Varga (2017).

Figure 1: Number of introduced OCRs over time



Source: National Tax and Customs Administration of Hungary

By 2016, almost 200,000 OCRs had been introduced by around 100,000 companies. In 2015, 75 percent of the total turnover documented by these newly introduced OCRs was in the retail sector,[2] and another 8 percent was generated in the AFS sector.[3] In this paper, we focus on these two sectors, and thus our turnover-based data coverage is 83 percent.[4] The retail sector's share was 11.7 percent, and the share of the AFS was 1.6 percent in total VAT income in 2013, the year preceding the introduction of the OCRs. It is worth pointing out that not all companies in the two surveyed sectors were obligated to introduce an OCR. Retail firms that sold through stalls, markets, the internet, or mail-order houses were exceptions. Companies providing event catering also were not obliged to introduce OCR in the AFS sector. In these two sector categories, the contribution to the total VAT of companies that installed OCRs was 10.2 and 1.4 percent, respectively.

The most important part of the introduction of OCRs, as applied in Hungary, is that these devices serve as a fiscal memory that collects all relevant tax information (opening and closing time of the register, blackouts, value and tax rate of the sold items, etc.) and saves it indefinitely. No information is allowed to be deleted. The memory is part of the register and must be placed inside its casing. No hidden software can run on the register, and this must be certified by three independent experts. After saving the information, the device transfers it to the NTCA, typically every 30 minutes. A special mobile internet connection is used to send the information, which is encrypted before it is sent, and a special encoding technique

[2] NACE G.47.

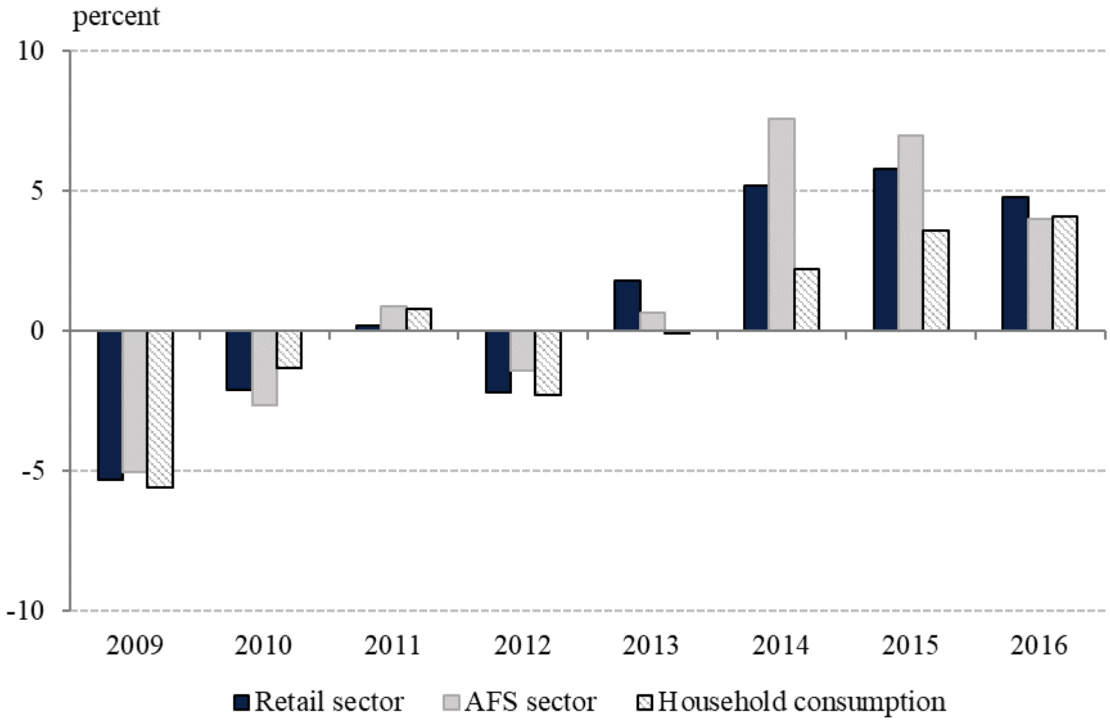
[3] NACE I.

[4] We show and discuss results for companies in other sectors in Appendix C.

prevents ex-post modification of the data. The NTCA collects the information on a server and can thereafter monitor the sales and the current content of the register.

A proper assessment of the effect of the introduction of OCRs requires an understanding of the macroeconomic context of the introduction and spread of OCRs, which occurred in the middle of the recovery of the Hungarian economy. The Great Recession was preceded by procyclical and expansionary fiscal policy, and the consequences of the global economic crisis proved to be more severe in Hungary than in most other countries in the region (Darvas, 2010; Tóth, 2011). The fall in household consumption was 5.7 percent in 2009, and the average annual change in the next 4 years was −0.7 percent, indicating the particularly lengthy nature of the crisis. The recovery of domestic consumption started in 2014 (the year of OCR introduction), and the growth rate climbed over 4 percent in the following years. Household consumption reached its pre-crisis level in 2015. This slow recovery is captured in the evolution of the two sectors where the introduction of OCRs is concentrated.

Figure 2: Growth rate of real turnover in retail and AFS activities and actual final consumption of households (YoY)



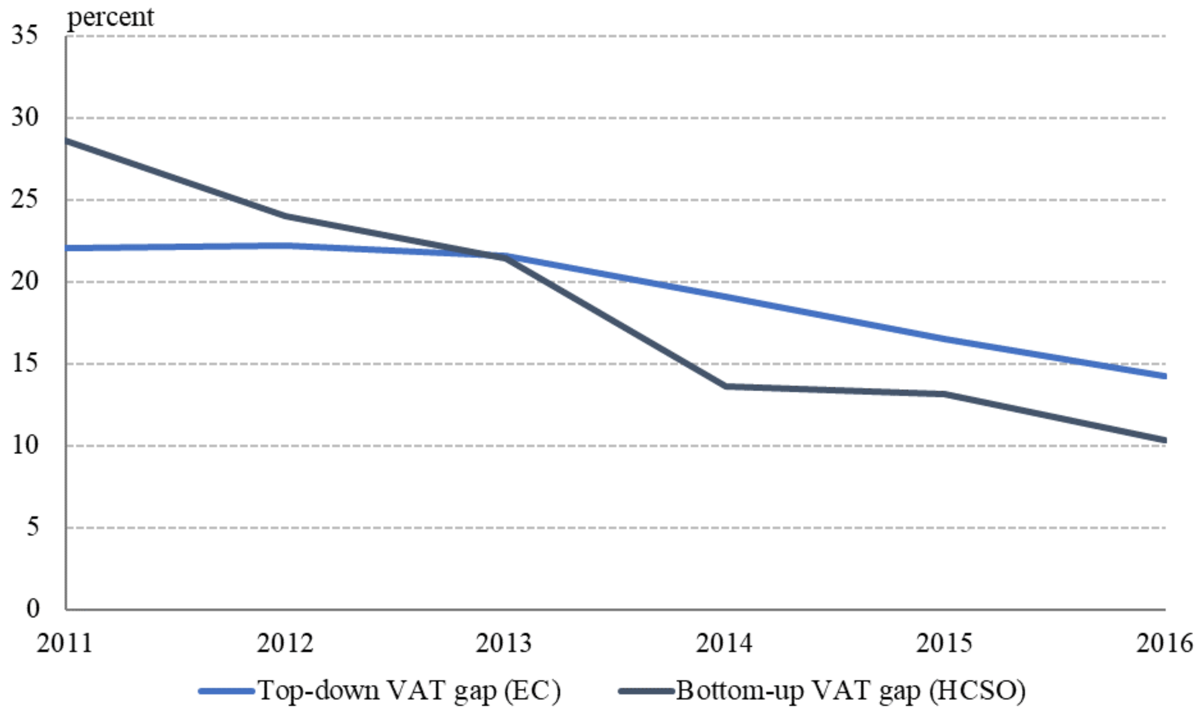
Source: Hungarian Central Statistical Office

The turnover of the retail and AFS sectors suffered huge losses in 2009. The sharp decline was followed by 4 years of weak performance with more declines than increases (Figure 2). After this near-stagnation, in the year OCRs were introduced, the annual increase in retail turnover jumped to more than 5 percent, and the AFS growth in turnover jumped to almost 8 percent. This dynamic growth rate continued into the following year, after which a moderate slowdown set in.

If we compare these changes in the two sectors with household consumption, we observe a temporary divergence around the years of OCR introduction. Since household consumption statistics include an estimation of the hidden economy, and sectoral statistics do not, the larger increase in the latter may indicate, among other things, some reduction in tax evasion.

One of the main goals of introducing OCRs was to reduce the size of the shadow economy, and thus it is worthwhile to analyze the total VAT gap (Bostan et al., 2017). The VAT gap or VAT compliance gap is the estimated difference between what taxpayers would pay if they complied with tax regulations and what they actually pay. This value is particularly important from a fiscal point of view since, of the available taxes, VAT makes the largest contribution to the revenue side, accounting for almost one-quarter of total tax revenues.

Figure 3: Total VAT gap (percent of potential VAT)



Source: European Commission, 2021 and Ritzlné and Máténé, 2020

There are two main published figures on the VAT gap in Hungary, using different methods of estimation. The European Commission publishes its estimates yearly (European Commission, 2021). They use national accounts data to estimate potential VAT and compare this with the official VAT revenue. Researchers of the Hungarian Central Statistical Office (HCSO) use tax returns and audit microdata to estimate the VAT gap with a bottom-up approach (Ritzlné and Máténé, 2020). Both methods show a decrease in the VAT gap between 2012 and 2015: from 22.2 percent to 16.5 percent in the top-down and from 24 percent to 13.1 percent in the bottom-up estimation (Figure 3).

In summary, there was a clear increase in the turnover of both analyzed sectors during the introduction of OCRs, especially in AFS. Nevertheless, this period coincides with a recovery in the economy and the introduction of other fiscal reforms aiming to reduce the shadow economy. This makes it difficult to distinguish the effect of the OCRs from the effects of a favorable macroeconomic environment and the other fiscal tools. However, there is a remarkable decrease in the VAT gap, possibly indicating that tax-related fiscal measures, including the introduction of OCRs, had an impact on reducing the size of the shadow economy. Since the analysis of aggregate numbers is not decisive in judging the effect of the use of OCRs, analysis of microdata is required.

III. DATA AND FILTERING

The empirical analysis focuses on reported net turnover and net purchases. Reported net turnover can be calculated from VAT returns, but different institutions use different definitions. Our definition includes all turnover that the OCRs should record. We add four lines in the return: net domestic sales with 27 percent VAT, net domestic sales with 18 percent VAT, net domestic sales with 5 percent VAT, and net domestic sales with other tax rates.[5] Using alternative definitions (e.g., adding sales free of VAT in the public interest) does not affect the main results. Our dependent variable in the estimation of responses in purchases includes three lines in the VAT return: net domestic purchases with 27 percent VAT, net domestic purchases with 18 percent VAT, and net domestic purchases with 5 percent VAT.[6]

The main data sources are individual VAT returns linked with the individual OCR and corporate income tax (CIT) returns databases. To create a dataset in which the observation frequency is homogeneous, we aggregate the monthly data to the quarterly level and exclude companies with only annual VAT returns available.[7] We do not use data from companies installing their first OCR after the deadline (August 31, 2014) since most correspond to new companies and those starting a new economic activity in which OCR use is obligatory; these companies are not sufficiently informative in considering the effect of OCR installation. Our information about firms' economic activity is drawn from the NACE (*Nomenclature des Activités Économiques dans la Communauté Européenne*) codes declared in their VAT returns.

However, these do not always reflect the firm's real activity; they are usually based on what was declared when the firm was founded, which is not necessarily its actual area of activity, or the real activity could have changed. A total of 48,474 firms in the retail and AFS sectors introduced at least one OCR during the period analyzed (Table 1). We do not use data from annual VAT returns as our estimation method requires more frequent observations. Of our initial target group in the two sectors, we exclude 822 firms (1.7 percent of the total) whose turnover going through OCRs was around 0.3 percent of the total. We do not use data with zero turnovers. Due to the model's specification, these data had to be excluded.

[5] For some regressions, we use variables valid before 2012: net domestic sales with 25 percent VAT, net domestic sales with 12 percent VAT, net domestic sales with 15 percent VAT, and net domestic sales with 20 percent VAT.

[6] For some regressions, we use variables valid before 2012: net domestic purchases with 25 percent VAT, net domestic purchases with 12 percent VAT, net domestic purchases with 15 percent VAT, and net domestic purchases with 20 percent VAT.

[7] Taxpayers shall submit a monthly tax return if the tax balance of two years prior is positive and this amount reached HUF 1 million (~EUR 2,820). Taxpayers shall submit a quarterly tax return, if the tax balance of the 2nd previous year did not reach HUF 1 million (~EUR 2,820). Taxpayers shall submit an annual tax return, if the tax balance of the 2nd previous year did not reach HUF 250,000 (negative or positive) and the NET value of the supply of goods and the provided services did not reach HUF 50 million (~EUR 140,000), and they do not have EU VAT ID number.

Table 1: Number of observations and percentage of gross OCR turnover after data filtering

	Retail sector		AFS	
	N-cumulative	Percentage of gross OCR turnover (Initial population=100%)	N-cumulative	Percentage of gross OCR turnover (Initial population=100 %)
Initial population	34,533	100	13,941	100
Without annual tax returns	34,019	100	13,633	99
VAT returns around OCR introduction	28,296	97	9,904	87
Outliers with >100* growth	22,298	86	7,376	68
OCR turnover ratio (sample)	15,046	67	5,588	50

Source: National Tax and Customs Administration

In the first step, we include VAT returns from the time of the OCR introduction, and it was a criterion for inclusion in our sample that the company has data available for the quarter in which OCR was introduced and for two quarters prior to and after that date. In the next step, we remove companies that, in at least 1 quarter, had turnover growth so high that their new turnover was 100-times higher than that of the previous quarter; we consider these companies to be outliers. In our baseline specification, we keep companies whose OCR turnover ratio is less than 1.5. This means that if the turnover in VAT returns is more than 50 percent higher than the OCR turnover, then activities unrelated to OCRs are significant, and we wanted to avoid possible estimation bias. The analysis was performed for the year 2015, as we only have data on OCR turnover for that year. Furthermore, by that time, initial issues with the system had been resolved, so the relevant data were cleaner than in the first year of operation. The robustness tests (Section VII) contains results with alternative turnover-ratio cutoffs. Our final panel dataset for the estimation includes 20,634 firms, accounting for 65 percent of the gross OCR turnover.

Since some companies installed more than one OCR, we had to provide an exact definition of the most important variable in our econometric specification, the installation date. In our baseline specification, we use the installation date of the first OCR. The robustness section of this paper (Section VII) shows the results with two alternative definitions: the quarter with the largest number of new OCRs introduced and the quarter of the last introductory-period OCR installation. The main results are robust to these modifications.

IV. MODEL SPECIFICATION

Our main goal is to identify the effect on turnover of introducing OCRs, distinguishing this from the impact on turnover of other factors. We do this by exploiting the heterogeneity in installation dates and estimating the following panel econometric model with company- and time-fixed effects:

$$y_{i,t} = \beta_0 + \beta_1 o_{i,t} + \gamma_t + \delta_i + \varepsilon_{i,t} \tag{1}$$

where $y_{i,t}$ is the log reported turnover of the ith company at quarter t; $o_{i,t}$ is the OCR dummy, which takes the value of 1 if company i has at least 1 operating OCR in quarter t, and 0 otherwise; γ_t is the time-fixed effect at quarter t; δ_i is the company-fixed effect for the ith company; and $\varepsilon_{i,t}$ is the residual of the ith company at quarter t.

Our parameter of interest is β_1, which shows the relative change in reported turnover (measured in log points) after the introduction of the first OCR. We avoid cross-sectional comparison of companies that have never introduced OCRs and are thus possibly different from those that did. It is important to highlight that, with this specification, we only measure the direct effect of OCR introduction in specific companies. Further increases in reported turnover could have been caused by OCRs if they occurred at a time outside of the introduction period.

The key condition for parameter β_1 as a measure of the causal effect of OCR introduction is the exogeneity of the time of introduction. As mentioned above, the introduction deadline was postponed several times due to technical difficulties with installation: thousands of firms ordered OCRs at the same time, and distributors were only able to deliver the devices gradually. As a result, actual installations took place several months, and sometimes even half a year after the order was submitted, and the delivery delay could not be influenced by the company. This assumption of the randomness of installation dates is further supported by the descriptive statistics set out in Table 2.

Table 2: Descriptive statistics - introduction dates in the sample

Date of OCR installation	2013-Q4	2014-Q1	2014-Q2	2014-Q3
Number of companies	211	6,056	4,500	5,283
Share of retail companies	67	70	70	68
Share of companies with one OCR	60	54	53	60
Median size [1]	23,390	20,496	18,916	19,038

Notes: [1] average yearly turnover before introduction in thousand HUF

Source: Own calculation

The sectoral distribution of companies that installed their first OCRs is very similar across quarters. The share of companies with one OCR is somewhat larger for the first and the last quarter, but not decisively so. The median size of the companies, measured by average yearly turnover prior to the first installation, is almost the same for the last three introduction quarters (2014 Q1–Q3). A few companies introducing OCRs in 2013-Q4 have, on average, a 20 percent larger turnover than companies introducing OCRs at the other times, but they do not distort the results. As the size distribution of firms is extremely skewed, we report the median sizes for the different installation dates.

Differences in seasonality in the sub-groups could distort our estimations, especially if seasonality were correlated with the OCR introduction dates. Therefore, we estimate the effect of introduction using seasonally adjusted and non-adjusted turnover data and obtain almost identical results. We use the non-seasonally adjusted data in our baseline specification.

We estimate Equation (1) on the effects of OCR introduction separately for the two sectors and different firm sizes. This allows for the use of sector- or size-specific time-fixed effects, which would not be possible using interaction variables between the OCR introduction date and sectors or firm sizes. These estimated time-fixed effects are indeed different between sectors and firm sizes (e.g., the general rate of growth in turnover of smaller companies was lower than that of large companies). Firm size is measured by annual average turnover (from VAT returns) prior to OCR introduction.

When deciding the number of firm size and sector categories, we face the following trade-off. We can better observe the category-specific heterogeneity with more categories, but the estimated coefficients have larger standard errors as the number of observations decreases. We run regressions for categories with different granularity, the most detailed versions having 20 size categories and several sectors, making estimations separately for AFS and dividing retail into five subcategories. We select a level of granularity where the confidence intervals are not too big to see the differences between estimates. In our final specification, companies are categorized into two main sectors: retail and AFS.

The retail and AFS sectors have very different cost structures that can affect how they behave in response to OCRs. Most of the costs in the retail sector are related to product purchases, while for accommodation and restaurants, the cost of labor is more important. In terms of size categories, we divide both sectors into quintiles. This means that we estimate the model for ten subsamples. The division into size quintiles is conducted separately for the two sectors, and, as a result, the size limits (that separate the different size categories) are somewhat different in the two sectors. This decision was motivated by having sector-specific quintiles with the same number of companies and thus similar standard errors. The bounds are very close in the retail and AFS sectors (Columns 3–4 of Table 3), so the different size boundaries do not cause any serious comparability concerns.

Table 3: Composition of the final sample

Sector	Turnover Quintile	Number of observations	Number of companies	Lower bound, thousand HUF [1]	Upper bound, thousand HUF [1]	Analyzed turnover, thousand HUF [1]
Retail	1	52,762	3,010	0,093	9,256	6,236
	2	55,870	3,009	9,257	16,971	12,853
	3	57,215	3,010	16,971	29,609	22,641
	4	58,331	3,008	29,623	63,182	43,056
	5	59,188	3,009	63,184	589,515,118	964,700
AFS	1	18,425	1,118	0,132	7,397	5,456
	2	20,481	1,117	7,413	11,599	9,327
	3	20,959	1,118	11,604	19,938	15,183
	4	21,356	1,117	19,939	39,160	28,019
	5	21,868	1,118	39,188	7,196,442	126,483

Notes: [1]: average yearly turnover before introduction in thousand HUF
Source: Own calculation

We use an event study design regression to investigate the potential delayed effects of the OCR introduction. The following regression is estimated:

$$y_{i,t} = \beta_0 + \sum_{r=-10}^{11} \theta_r e_{r,i,t} + \gamma_t + \delta_i + \varepsilon_{i,t} \tag{2}$$

Where $e_{r,i,t}$ are dummy variables that take the value of 1 in the quarter that is *r* quarters away from the quarter of OCR introduction for company *i*. For example, $e_{0,i,t}$ is 1 in the quarter of the introduction and 0 otherwise. The exception is $e_{11,i,t}$, which takes the value of 1 for every quarter where more than quarters have passed since the first OCR introduction. This is done to estimate the total effect of the introduction after 2.5 years have passed. The θ_r values estimate the OCR effect for each quarter.

In addition to estimating delayed effects, the event study design serves as a test of our identification design. If our estimation strategy is correct, the OCR effect should be zero before the introduction. The disadvantage of this design is that we estimate many more coefficients, which makes their standard errors higher. This requires a trade-off in respect of category granularity; we cannot use categories that are as fine as those in our main specification. Therefore, in this case, we use only two categories, retail and AFS, and do not separate firms within those by size.

V. TURNOVER RESULTS

According to our estimation, the introduction of OCRs increased turnover in most sectors and size categories. However, the sector- and size-specific OCR effects are significantly different. We find an inverse relationship between the impact of OCR introduction and company size. The effect of OCR is definitely greater for smaller than for large firms. In addition, we find sizable sectoral differences since the impact proved to be more severe in the AFS sector than among retail companies.

Table 4: OCR effects on turnover: main results

	Retail					AFS				
Turnover Quintile	1	2	3	4	5	1	2	3	4	5
Estimate	0.104**	0.024*	0.018	0.013	−0.008	0.134**	0.040*	0.077**	0.042*	0.051**
T-statistic	7.395	2.150	1.600	1.283	−0.790	5.710	1.987	3.746	2.130	2.729
R^2	0.596	0.391	0.391	0.451	0.900	0.451	0.334	0.379	0.421	0.784
Number of obs.	52,762	55,870	57,215	58,331	59,188	18,425	20,481	20,959	21,356	21,868

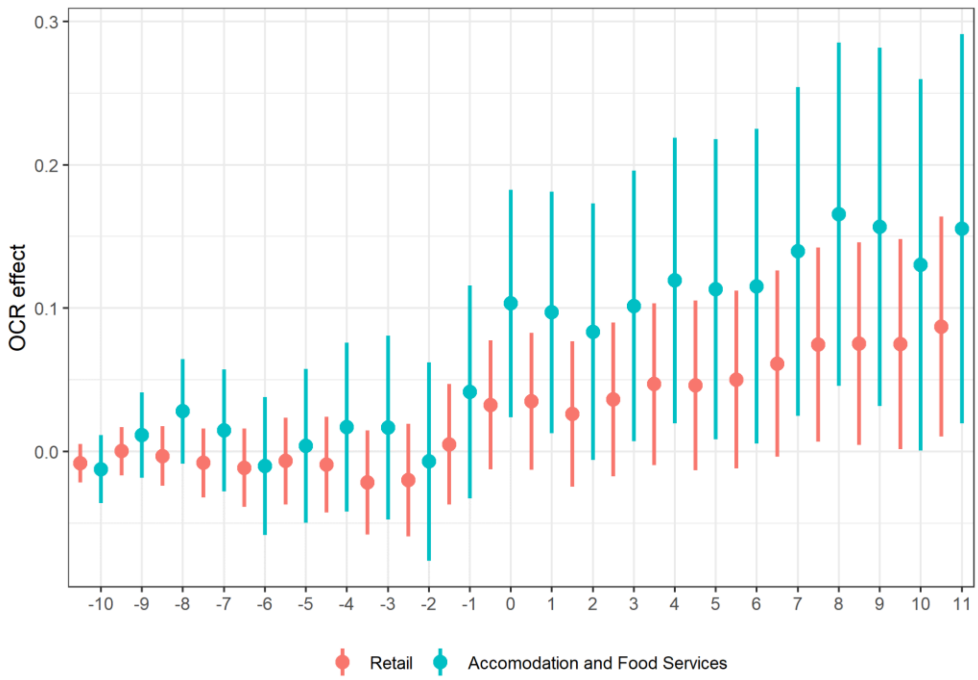
Source: Own calculation

In detail, in the retail sector, we find significant effects among the smallest companies in the 1st quintile (10,4 percent) and in the 2nd quintile (2,4 percent). The estimated OCR effect does not significantly differ from zero among the larger companies (3rd, 4th and 5th quintile). In the AFS sector, the effects are significant in all quintiles. We find the largest impact (13,4 percent) for the group of the smallest firms, but the effect was sizable, varying from 4,0 to 7,7 percent in all other quintiles (Table 4). Only the effect in the 1st quintile is significantly different from the others;[8] in this category, there is greater variance as there are fewer companies than in the retail sector. Results for the companies introducing OCRs in sectors other than retail and AFS can be found in Appendix C.

These results (which we refer to as the main results) resonate with the literature and, we believe, expand common knowledge on the efficiency of OCRs. Our findings on the general and sector-specific impact of OCRs on turnover fit with those of other studies (ADB, 2020; HCSO, 2016; Jakubowska, 2019; Skatteverket, 2013). Although some other papers emphasize the role of company size, their focus is different from ours. For example, Naritomi (2019) measures the impact of consumer monitoring, not OCRs, while Bachas et al. (2019) assesses the impact of OCRs on taxation, not turnover. Our results pertaining to the size-dependency effect of OCRs on turnover can be considered new evidence.

We now present results from the event study design. Our finding that the OCR's effects before the introduction date are not significantly different from zero supports our assumption that the introduction date is not related to some unobserved firm characteristics that would undermine our estimation strategy. In both sectors, the effects start to increase around the OCR introduction date (0). In the AFS, the effect first shift to around 10 percent and then increases to around 15 percent; however, this difference is not significant (Figure 4). The retail sector runs on a similar trajectory, but with more limited effects of around 5 percent, and they only are significantly different from zero on a few occasions. This is reasonable since our main model indicates the OCR effect in the retail sector is significantly different from zero in only two of the five size categories. In the AFS sector, we also observe a slight increase in the impact of OCRs effect in 1 quarter before their introduction. This might be due to firms with introduction dates in the last month of the previous quarter or firms anticipating the OCR introduction.

Figure 4: OCR effects on turnover: event study design



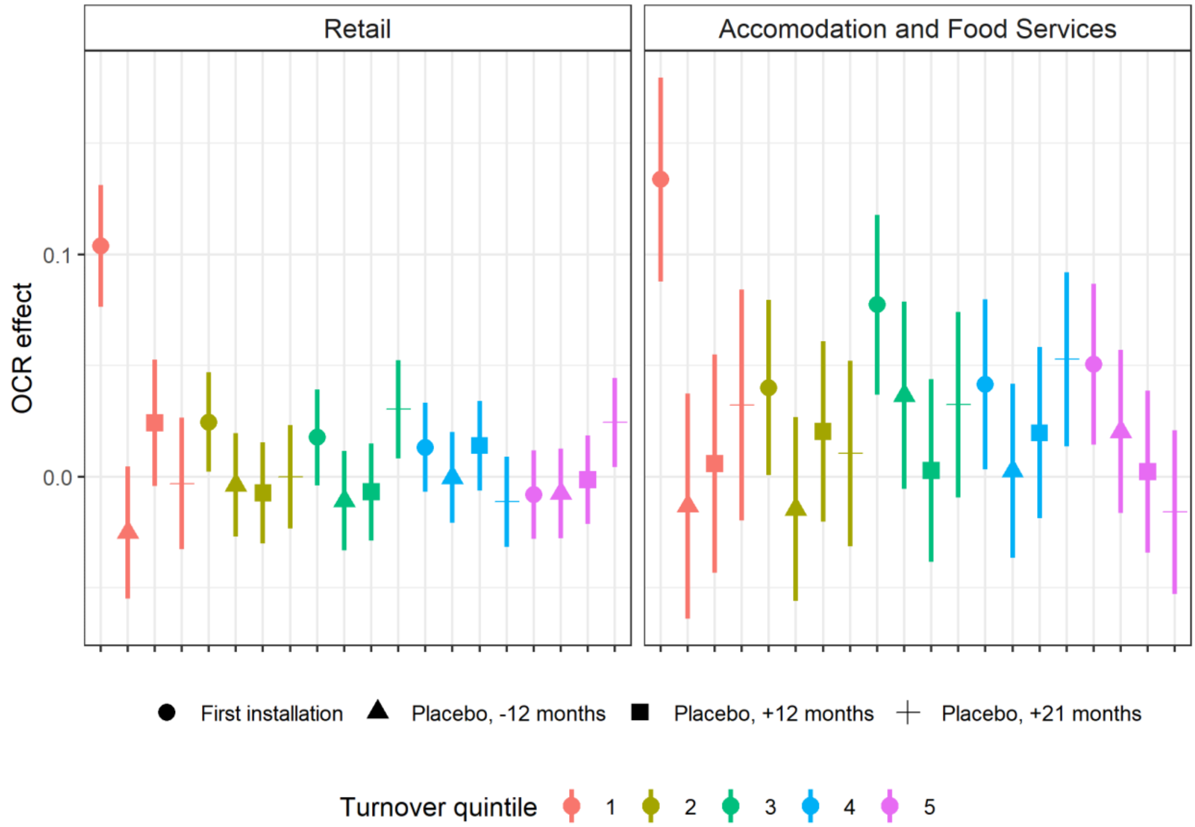
Source: Own calculation

[8] This result also has strong implications in respect of policy recommendations. Prior to the introduction of OCR, classical investigations and monitoring activities of the NTCA focused mainly on large companies. The emergence of OCRs can support the work of the tax authority against tax evasions by smaller companies.

The main question that arises in respect of our results is whether or not the time of the introduction could be manipulated by companies to hide the effect. In Section II, we described the legal context of the introduction and noted that companies' decisions had small and barely predictable effects on the introduction time. Nevertheless, in this section, we present relevant econometric evidence to ensure that the effects detected were caused by the introduction of OCRs.

We re-estimated our baseline model with placebo OCR introduction dates. We expect that the estimated OCR effects at these false dates are not significantly different from zero, and, therefore, there is no systematic distortion in our dataset that would lead us to falsely detect an OCR effect. We use three different placebo dates: 1 year before the real introduction date, 1 year after the introduction date, and 7 quarters after the introduction.

Figure 5: Results with placebo introduction dates



Source: Own calculation

The results depicted in Figure 5 show that the OCR effects that we find in our baseline model do not exist in the models with placebo introduction dates. Estimated coefficients are almost always closer to zero than our true estimates. From the thirty placebo tests, the placebo is significantly different from zero twice. This ratio shows that the model is reliable; the effects are not due to a more general change in the data but are related to the OCR introduction.

VI. SPILLOVER EFFECTS AND TAXATION

For a comprehensive description of the effects of OCR introduction, it is necessary to consider the macro-level impacts, especially on VAT. Doing so allows us to assess whether this reform has met its original aims, such as reducing the number of sales without invoices and increasing tax revenues by reducing the size of the shadow economy. Before presenting our results, it is worth summarizing the main features of VAT and its implications for the macro-level estimation.

The core feature of VAT is that all participants in the supply chain incur a payable tax that is based on their sales, which can be reduced by the tax already paid on purchases (Naritomi, 2019). One important consequence of this structure is that an increase in a participant's sales does not necessarily lead to an increase in their net tax liability. Rather, this depends on changes in the company's purchases compared to its increase in sales. If sales and purchases increase by the same amount, payable and deductible tax change correspondingly, leaving the firm's overall tax liability unchanged. It is worth emphasizing that if the two variables change in the same proportion, then their differences (and hence the tax liability) will change in the same proportion. However, this leads to the final customers paying the full tax. If the sales by the last actor in the supply chain increase, this will increase the state's VAT revenue regardless of the change in purchases by the company; such changes determine only which actor in the supply chain pays the extra VAT.

For this reason, the tax impact of introducing OCRs can be calculated in two ways. First, we focus on the entire turnover. We take the increase in reported turnover implied by introducing the OCRs and calculate the extra VAT paid after the extra sales. This method identifies the total impact of the introduction of OCRs on tax revenues but does not generate any details about the sectoral contribution to the state's extra income. Next, we calculate the increase of the tax liability of the different sectors by considering not only the increase of the turnover (and hence the payable tax) but also the increase of the purchase (and hence the deductible tax). These numbers cannot capture the full impact of the OCR introduction on tax revenue. Nevertheless, they provide useful information about the size of the spillover effects and the share of the analyzed sectors in the payment of extra tax revenue as a result of the introduction of OCRs.

First, we assess the extra VAT paid after the extra turnover due to the introduction of OCRs. In this case, we do not care about changes in purchases and calculate the tax effect on the whole supply chain. It is assumed that all companies that have introduced one or more OCRs responded similarly to companies of the same size. Based on this, the introduction of OCRs increased the annual turnover of the retail sector by 0.1 percent and the AFS sector by 2.5 percent. Together the turnover growth in the two analyzed sectors increased the annual VAT revenue through the whole supply chain by HUF 6 billion (at 2012 prices), which is 0.2 percent of the annual VAT revenue.

We can compare this to the level and the change of the VAT gap. According to the European Commission estimate, the VAT gap was HUF 743 billion in 2012 (European Commission, 2018), and by 2015 it had decreased to HUF 652 billion (European Commission, 2021). This shows that the introduction of OCRs decreased the VAT gap by less than 1 percent. As a rough estimate, approximately 7 percent of the decrease in the VAT gap between 2012 and 2015 can be attributed to the OCR introduction.

The specification guarantees that the measured OCR effect is only due to the introduction of OCRs and not other phenomena or measures introduced to reduce the size of the shadow economy. The specification captures the total long-term effect of OCR introduction (as opposed to the short-term effect in the quarter of the introduction). However, this estimate should be considered a lower bound for the effect of OCRs on tax revenues. The reason for this is that we do not consider effects that were unrelated to the introduction of the first devices. For example, it may be the case that some firms started to change their behavior when the legislation was adopted or when the media wrote about it. This “announcement effect” cannot be estimated in our specification as there is no observable firm-specific heterogeneity in this respect; thus, its impact is captured by the time-fixed effects. In the analyzed period, the tax agency and the government communicated thoroughly regarding the various measures to reduce tax fraud and the new opportunities for the NTCA to monitor taxpayers. This could have caused a general sense of threat among fraudulent taxpayers, which could have increased reported turnover significantly, but this would be unrelated to the introduction of specific machines and is thus not measured. The potentially important role of tax-agency communication is underlined by the results of Bíró et al. (2021), who find effects on reported income from an increase in the audit threat in Hungary that is much larger and more immediate than the experience of actual audits would suggest. We also do not measure possible further effects on taxes of types other than VAT.

Our results resonate with Casey and Castro (2015) but are relatively small compared to those of other studies providing estimates on macro-level turnover (Awasthi and Engelschalk, 2018) and tax payments (Bachas et al., 2019; Cardoza, 2012; Eissa et al., 2014;). Besides, there is an undeniable tension between the obtained sizable coefficients from the micro-level analyses and the relatively small effect on turnover and tax revenues at the macro-level, especially in the retail sector. But this is only an apparent contradiction that can be easily resolved. The most important factor is the concentration in the retail sector. While the estimated impact of OCR’s introduction on turnover proves insignificant for the largest retail companies, they are responsible for most of the sector’s total turnover and tax payments. The top 1 percent of companies accounted for 64 percent of the total turnover of the retail sector and paid 76 percent of the total VAT. The exceptional weight of the largest companies can easily be captured if we assume that the effect of the OCR’s introduction on the largest companies is equal to that for the same group in the AFS sector (5.1 percent). In this hypothetical case, the introduction of OCRs would have increased the annual turnover of the retail sector by 3.1 percent, and the turnover growth in the two analyzed sectors would have increased the annual VAT revenue through the whole supply chain by HUF 49.1 billion, which is 1.8 percent of the annual VAT revenue.

Another factor of importance is that, according to VAT returns, only around 60 percent of the companies in the sectors considered were obliged to install OCRs. These companies accounted for 84 percent of the total turnover in retail and 76 percent in AFS. Companies declaring retail as their main activity do not have to install OCR if they sell products with prices above HUF 100,000, if they operate in a sub-sector that was not obliged to install the new devices, or if their actual economic activity was not in retail. In addition to the above, the difference in the sizes of the two analyzed sectors is also worth emphasizing. The total turnover of the AFS sectors is about one-tenth of the total turnover of the retail sector.

In sum, the relatively small size of the macro-level impact comes from three factors. While the largest companies in the retail sector are responsible for most of the total turnover and tax payment, we do not detect a significant effect in their case. We find a significant impact among the largest companies in the AFS sector, but this sector is minor compared to the size of the retail sector. Even in the two sectors surveyed, a significant proportion of firms could legally avoid the introduction of OCRs, which would also reduce the size of the macro impact.

In addition to calculating the tax impact through the whole supply chain, it is important to reveal the sectoral contribution by assessing the spillover effects, which we capture by estimating the effect of OCR introduction on domestic purchases. The results indicate that the response in purchases to the introduction of the OCRs is similar to that in respect of turnover in most categories (Table 5). This suggests that, in most cases, the companies passed through most of their increase in payable taxes to their purchases. The exception to this overall observation is the smallest quintile in the AFS sector (and, to a lesser extent, the smallest quintile of the retail sector and the 3rd quintile of the AFS sector). In those cases, the pass-throughs are significant but not complete since these companies increased their sales more than their purchases as a result of the OCR introduction. This result indicates that the spillover effects increase with company size.

Table 5: Results for turnover and purchases

Sector	Turnover Quintile		Turnover (Main estimate)	Purchases
Retail	1	Estimate	0.104**	0.082**
		T-statistic	7.395	5.075
		R2	0.596	0.626
		Number of observations	52,762	52,222
	2	Estimate	0.024*	0.033**
		T-statistic	2.150	2.636
		R2	0.391	0.544
		Number of observations	55,870	55,692
	3	Estimate	0.018	0.019
		T-statistic	1.600	1.560
		R2	0.391	0.524
		Number of observations	57,215	57,087
	4	Estimate	0.013	0.020
		T-statistic	1.283	1.867
		R2	0.451	0.573
		Number of observations	58,331	58,224
	5	Estimate	−0.008	−0.003
		T-statistic	−0.790	−0.242
		R2	0.9	0.887
		Number of observations	59,188	59,013
AFS	1	Estimate	0.134**	0.082**
		T-statistic	5.710	3.651
		R2	0.451	0.503
		Number of observations	18,425	18,052
	2	Estimate	0.040*	0.055**
		T-statistic	1.987	2.685
		R2	0.334	0.497
		Number of observations	20,481	20,278

	3	Estimate	0.077**	0.046*
		T-statistic	3.746	2.380
		R2	0.379	0.479
		Number of observations	20,959	20,804
	4	Estimate	0.042*	0.045*
		T-statistic	2.130	2.302
		R2	0.421	0.515
		Number of observations	21,356	21,164
	5	Estimate	0.051**	0.051**
		T-statistic	2.729	2.754
		R2	0.784	0.790
		Number of observations	21,868	21,777

Source: Own calculation

The size of the spillover effects suggests that most of the tax increases from the introduction of OCRs were paid in sectors other than the two we analyze here. To check this, we estimate the direct effect of OCR's introduction on a net tax base, which is the difference between the payable and the deductible tax bases. The net tax base can be negative due to various reasons: different timing of sales and purchases or if the tax rate of the sales is lower than the tax rate of the purchases, for example, in the case of pharmaceutical trade. Therefore, we estimate the impact on the net tax base without logarithmization. The results fit our findings related to the spillover effects. We find a significant positive effect in the 1st and 3rd category of AFS (where the spillover effect is smaller); in all other cases, however, the coefficients were not different from zero (See Appendix A).

Using the micro data, in addition to the response on turnover, we calculate the sectoral net tax impact by considering the impact on the purchase (hence the impact on the deductible tax). Based on this, we find that extra tax revenue paid by the retail sector was only 0.1 percent of the annual tax liability of the sector, while the corresponding share in the AFS was 2.6 percent. The total extra tax coming from the OCR's introduction and paid by the two surveyed sectors was HUF 1.4 billion (at 2012 prices), which is 0.05 percent of the annual VAT revenue. Our results on the sizable spillover effect resonate with the related literature. Pomeranz (2015) analyzes the impact of audit announcements, Carillo et al. (2017) measure the effect of firm notification concerning detected revenue discrepancies, and Naritomi (2019) focuses on the impact of consumer monitoring. In addition to the primary responses in turnover, these studies also identify strong spillover effects.

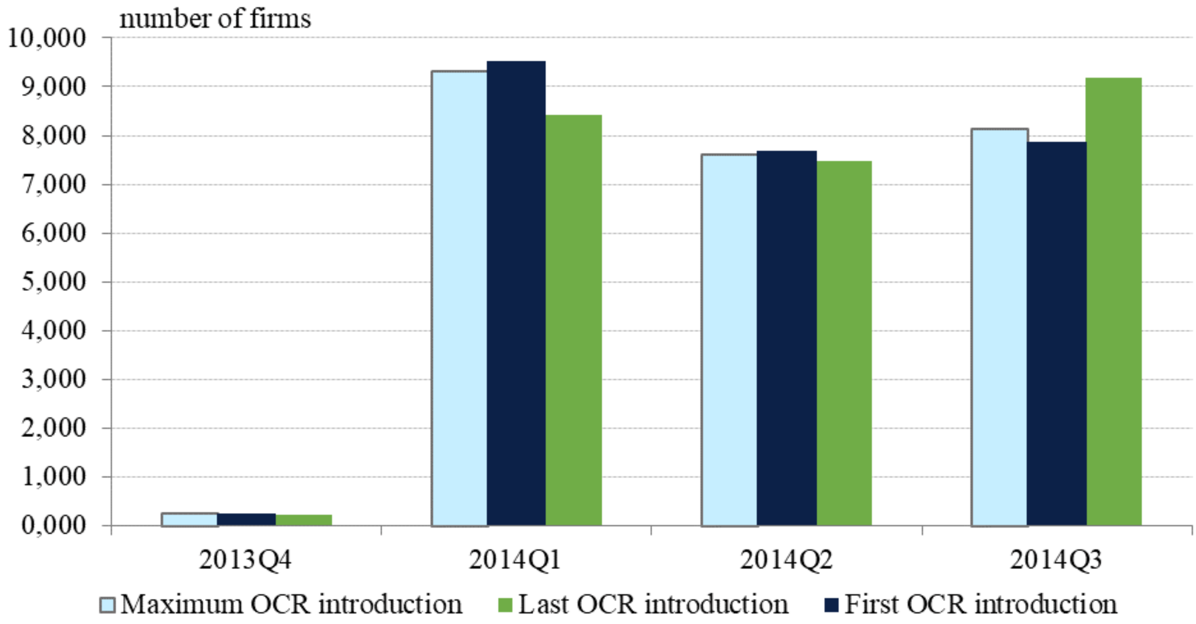
Being aware of all the reasons our estimations should be considered the lower bound of the total effect of OCRs on tax revenues, we can conduct an approximate cost-benefit analysis of the obligatory online cash register regulation from the perspective of the state. What was higher, the effects on VAT or the cost of the introduction? The government supported the purchase of OCR with HUF 50,000 per cash register for businesses with less than HUF 500 million of annual turnover, for up to five cash registers per company. The total cost of this support scheme amounts to HUF 6.2 billion since August 2014 (the deadline for the introductions). The tax administration incurred additional costs with the installment of IT systems prepared to receive the data sent by OCRs and has additional operational costs, but we do not have exact information on these. However, it is safe to assume that these costs were relatively low. The HUF 6.2 billion spent is of the same magnitude as the estimated increase in annual VAT revenue through the whole supply chain. As the positive effects last several years, it is safe to say that the policy had a net benefit from the perspective of the Hungarian state. This is without considering the further benefits of reducing the extent of the shadow economy and enhancing market competition.

VII. ROBUSTNESS TESTS

We conduct various robustness tests to see if our results change with different specifications. In this section, we present the robustness tests using regressions where the dependent variable is turnover. Results from similar robustness tests on regressions with purchases as the dependent variable are included in Appendix B.

In our first robustness test, we re-estimate our baseline model with alternative definitions of OCR installation dates. In the baseline specification, the OCR dummy variable changes to one when the company introduces the first OCR (and remains one thereafter). We test two alternative definitions of the OCR introduction dates: the quarter of the last OCR installation and the quarter with the highest number of new installations. This latter definition is reasonable as larger companies often installed only a few OCRs in the beginning as a test and then installed most of the new machines at a later date (and may have continued installations at a slower rate). The quarter of the last installation is for the last installation within the introduction period; later installations are not included in the sample.

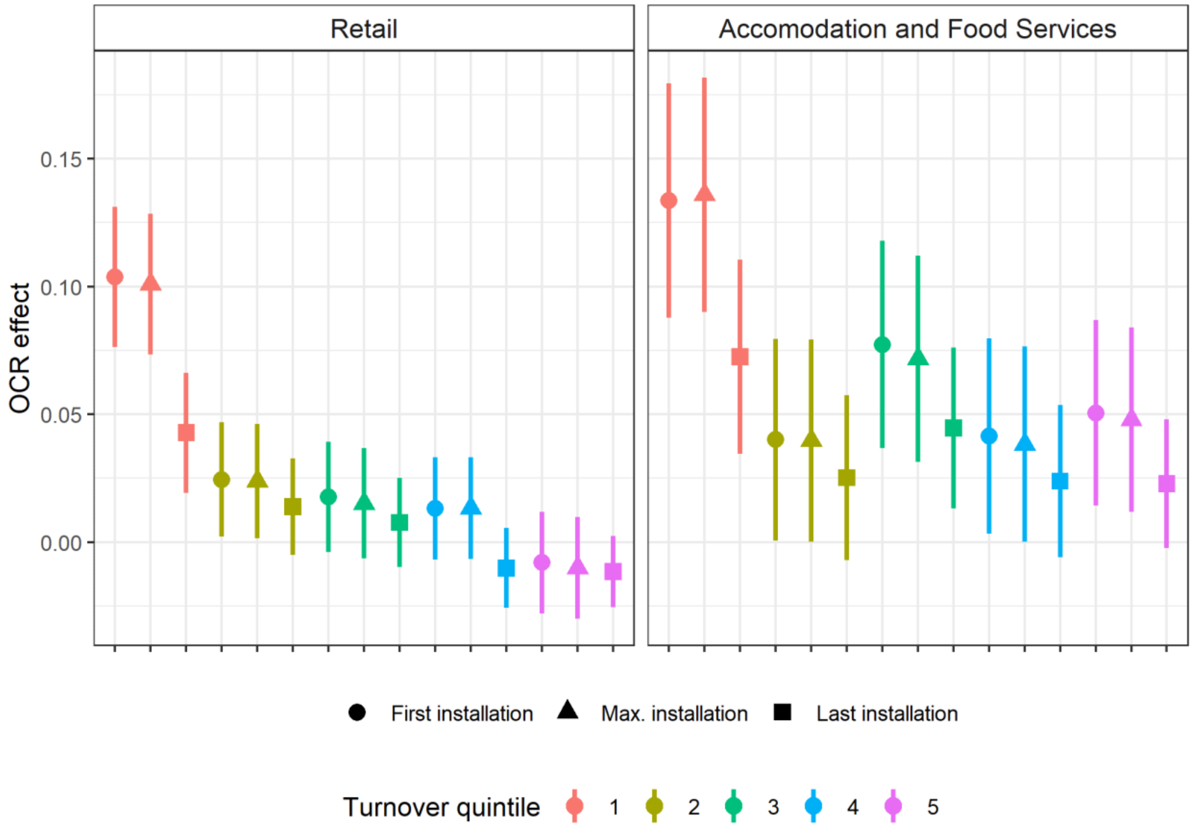
Figure 6: OCR introductions by quarter



Source: Own calculation

Figure 6 shows the number of installations by quarter for all three alternative definitions of the introduction date. In some cases, the quarter of the first, last, and maximum number of installations differ. However, as most companies have only one or two OCRs, the differences in the distribution of introduction dates are not that large. In fact, for 93 percent of the companies in our sample, the OCRs installations occurred in one quarter, and thus, all three definitions are the same. Therefore, it is perhaps not surprising that the estimated coefficients are very similar to our baseline estimates and are almost never significantly different from these (see Figure 7). The main exception is for the last installation in the category of the smallest firms, where the estimates are much lower than the main result; this can also be observed in some other categories, albeit to a lesser degree. It is plausible that the OCRs had the biggest effect at the time of the first or maximal number of installations and not when the last was installed since companies previously engaging in fraudulent activities would be aware that their already installed cash registers were sending data to the tax agency and could adjust their behavior accordingly.

Figure 7: Results with alternative definitions of introduction date



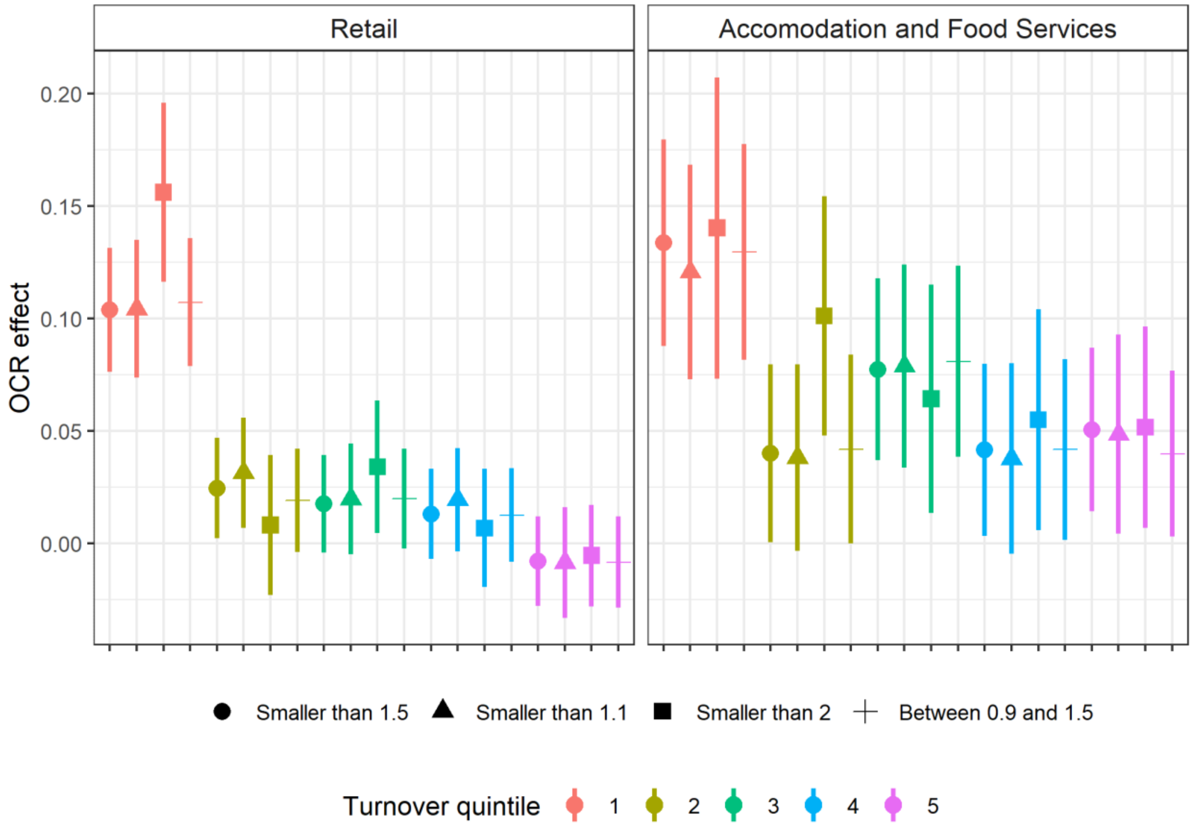
Source: Own calculation; detailed statistics in Appendix D. Table D2

During the data selection process, we excluded companies that had relatively high revenues that did not have to go through OCRs. The reported turnover in VAT returns, the variable from which we identify the OCR effect, is not divided into OCR and non-OCR turnover. We expect that OCRs affect only the turnover that they register, and in our baseline specification, we thus exclude firms where the VAT/OCR ratio was greater than 1.5.

In a second robustness test, we re-estimate our model with alternative VAT/OCR cutoff numbers in the data selection process: 1.1 and 2. In another run, we keep 1.5 as the upper limit for the VAT/OCR ratio but introduce 0.9 as a lower threshold. Ratios much lower than 1 are theoretically impossible, and they usually reflect frequent data errors, especially in the first years of OCR operations. We keep the baseline size categories constant while re-estimating the model for different VAT/OCR sales ratios. The companies belong to the same size categories across the different specifications, but the number of observations (in each size category) varies, rendering the standard errors not entirely comparable.

The estimated coefficients for the alternative VAT/OCR cutoffs are very similar to our baseline estimates, and the differences are significant only twice out of the 30 observations (see Figure 8). In these two cases, the smaller than two cutoff version has a somewhat higher OCR effect than in the main estimate.

Figure 8: Results with alternative OCR turnover-ratio cutoffs



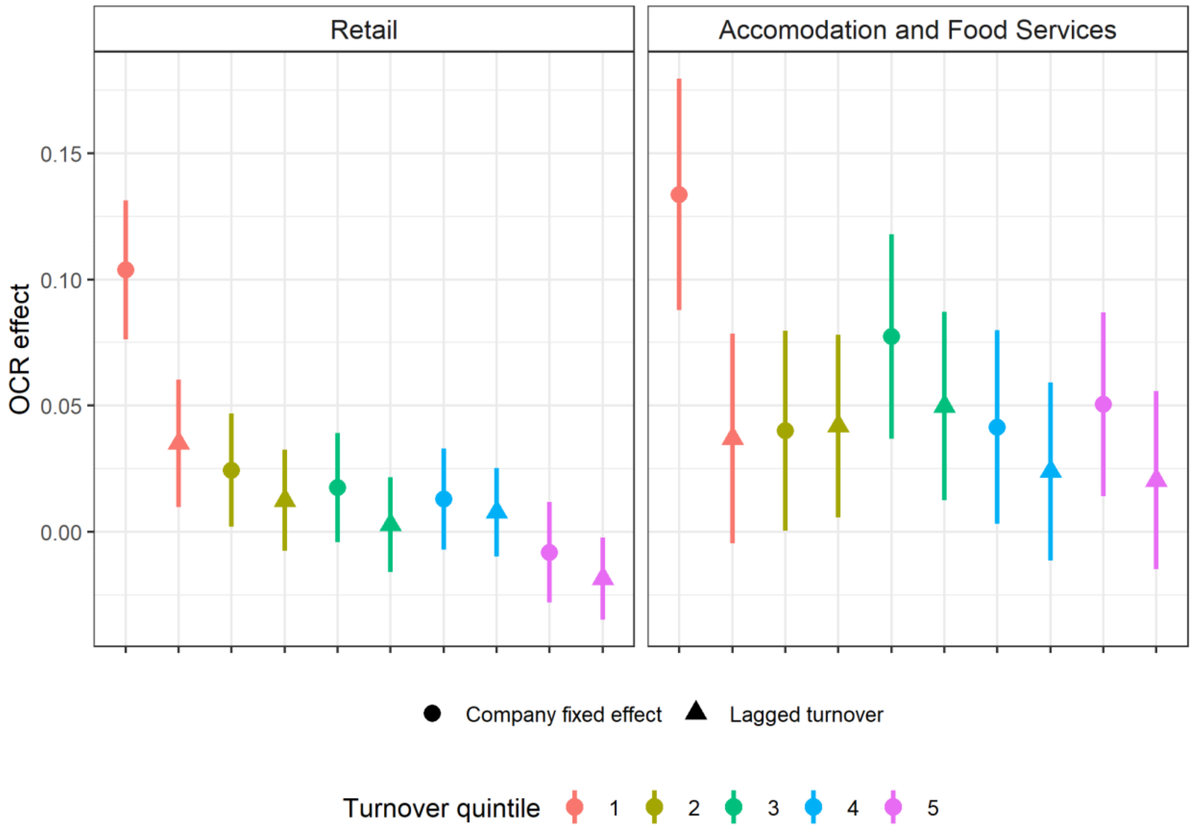
Source: Own calculation; detailed statistics in Appendix D. Table D4

It may be that the invariant company fixed δ_i was unsuitable to represent the unobserved company characteristics in our model (Angrist and Pischke 2009). To ensure that this is not the case, we estimate the following alternative specification of the model.

$$y_{i,t} = \beta_0 + \beta_1 o_{i,t} + \beta_2 y_{i,t-1} + \gamma_t + \varepsilon_{i,t} \quad (3)$$

The results from the model with lagged dependent variable, compared with the main, fixed-effects model, can be seen in Figure 9. The results of these two models are not the same since the estimate using lagged turnover is sometimes significantly smaller than our main result. At other points, the difference in the results is not significant. Based on these estimations alone, it is not possible to decide which is biased; thus, more analysis is needed.

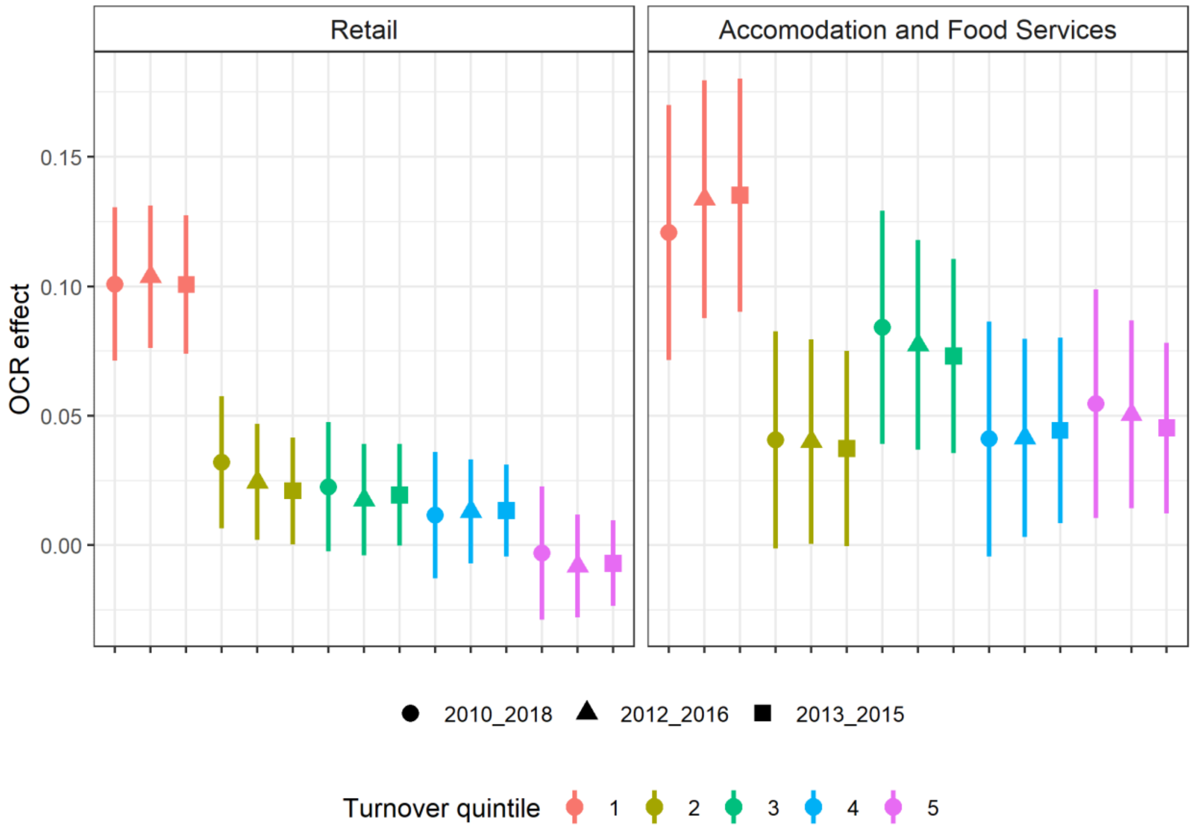
Figure 9: Results with lagged dependent variable



Source: Own calculation; detailed statistics in Appendix D. Table D5

It is known (Nickell 1981) that if there is bias in our main result, the inconsistency depends on the length of the time period. Thus, we re-estimate the model changing the length of the time period of the data. In our main specification, we use data between 2012 and 2016. We estimate Equation (1) with alternative time frames 2010–2018 and 2013–2015. As OCR introductions started in 2013-Q4 and had to be carried out by 2014-Q3, the variation that we exploit to identify the OCR effect is unaffected by the addition or removal of further years. As a consequence, estimates are not significantly different from our baseline estimates (Figure 10). This result suggests no bias in our main result and ensures that the measured effect was caused by OCR introduction and is not influenced by later developments in turnover or by some error in the estimation method.

Figure 10: Results with alternative time frames



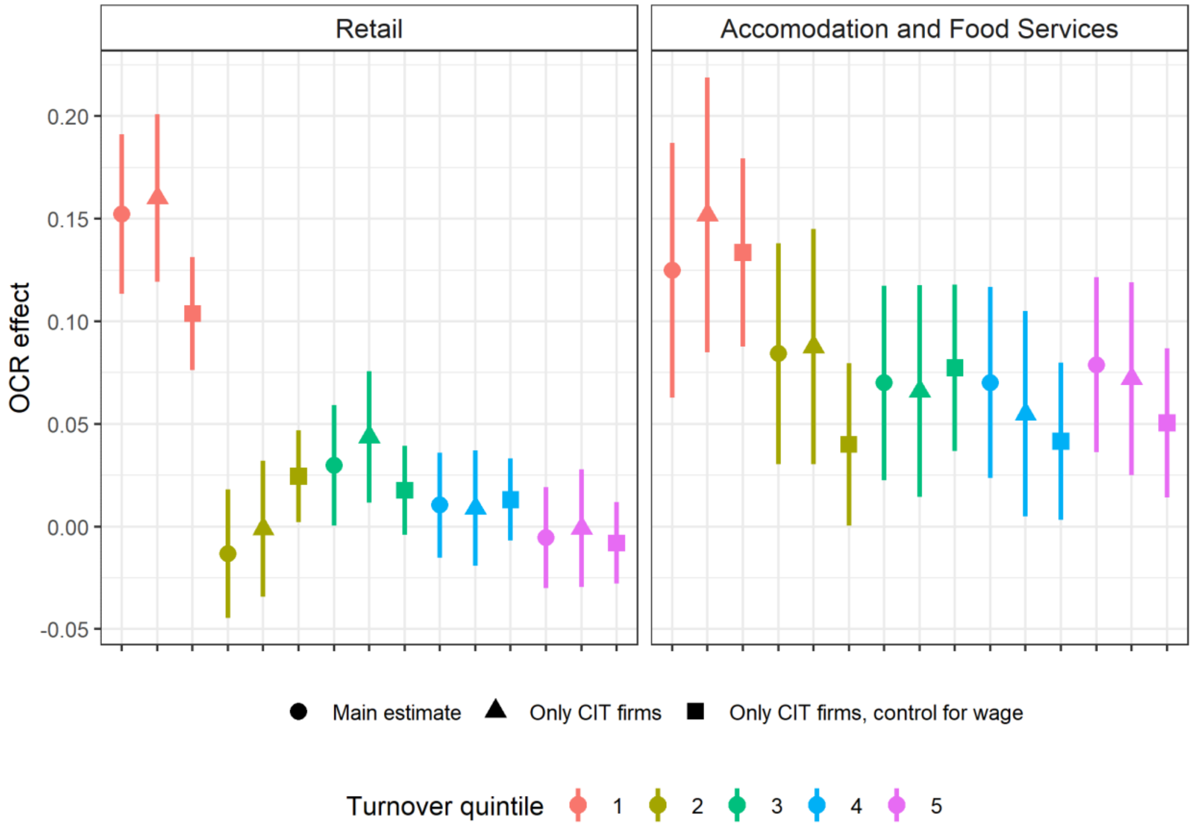
Source: Own calculation; detailed statistics in Appendix D. Table D5

An option to increase the explanatory power of our estimation is to add further explanatory variables which describe changes in the size of the companies independently of turnover. Such a variable is the amount spent on personal allowances, mostly wages. This data is available in the annual corporate income tax (CIT) return, but not every company in our sample is paying CIT. So we created a restricted sample with only CIT firms. The number of companies decreases from 15,046 to 7,965 in the retail sector and from 5,588 to 3,156 in the AFS sector. The separating thresholds of the size quintiles are also somewhat different in the restricted sample. As CIT is declared yearly, wage data is available yearly, not quarterly, as turnover from VAT returns. We estimate the following equation. The only difference from the main equation is the addition of the wage variable.

$$y_{i,t} = \beta_0 + \beta_1 o_{i,t} + \beta_2 w_{i,t} + \gamma_t + \delta_i + \varepsilon_{i,t}, \tag{4}$$

where $w_{i,t}$ is the log reported total wage cost for the company i in quarter t. To separate the effect of using a restricted sample from the addition of a control variable, we show the OCR effects for the restricted sample with the main specification as well. Estimates in the main and restricted sample are broadly the same, with a few exceptions, such as the smallest retail firms where the only CIT companies have a higher average OCR effect. This can be explained by the sizable difference in the number of firms between the two samples. Results with or without controlling wage are almost equal; there is no significant difference in either category (Figure 11). This shows that adding an additional control variable about changes in the size of firms is not necessary to have reliable estimates of the OCR effect.

Figure 11: Results using wages as a control variable



Source: Own calculation; detailed statistics in Appendix D. Table D6

VIII. CONCLUSIONS

With the spread of the internet and digitalization, many countries have introduced OCRs to reduce the extent of their shadow economy. The Hungarian government made it compulsory to install OCRs in some sectors in 2014. The switch from the old to the new cash registers took place gradually and mostly affected the retail and AFS sectors. During this process, almost 200,000 OCRs were installed by approximately 100,000 companies. In this paper, we used econometric panel techniques to identify the effect of this measure on the reported turnover and tax liability. We assume that the introduction of the OCRs did not in itself change the company's operations, and the additional turnover – after controlling for other factors – can thus be considered the extent of the reduction of the shadow economy.

To quantify this effect, we used a linked firm-level dataset of Hungarian firms that included their VAT and CIT returns and OCR data. We found that the introduction of OCRs had a remarkable effect on the reported turnover of the companies in the surveyed sectors. However, the sector- and size-specific OCR effects were heterogeneous. The introduction of OCRs had a significant impact on reported turnover among smaller companies in the retail sector, with 10.4 percent growth in the 1st quintile and 2.4 percent in the 2nd quintile. The impact did not significantly differ from zero among the larger companies. In the AFS sector, we found the larger impact (13.4 percent) among the smallest group, but the effect was sizable in all other quintiles, varying from 4.0 to 7.7 percent.

The size-dependency result and the fact that most VAT is paid by large companies resulted in the mitigation of the impact on VAT payments of introducing OCRs. In both analyzed sectors, the introduction of OCRs increased annual VAT revenue by 0.2 percent throughout the whole supply chain. Focusing on the contribution of the surveyed sectors to the VAT payment, we observed significant spillover effects, which were slightly stronger among larger companies.

Acknowledgment: We are grateful to Ádám Reiff and Mária Pécs for their strong support and are also indebted to Danila Pankov and Róbert Lieli for their valuable comments. We are also grateful for the comments we received at the 15th Conference of Economic Modeling, at the workshop conducted by the Central Bank of Hungary, and at the 2019 Conference of the Hungarian Society of Economics.

References

Ali, Merima, Abdulaziz Shifa, Abebe Shimeles, and Firew B Woldeyes, 2015. "Information Technology and Fiscal Capacity in a Developing Country: Evidence from Ethiopia." *SSRN Electronic Journal 31*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2587857

Angrist, Joshua D., and Jörn-Steffen Pischke, 2008. "Mostly Harmless Econometrics" Princeton University Press. https://press.princeton.edu/books/paperback/9780691120355/mostly-harmless-econometrics

Asian Development Bank, 2020. "A Comparative Analysis of Tax Administration in Asia and the Pacific 2020 Edition." Metro Manila, Philippines http://dx.doi.org/10.22617/TCS190240

Awasthi, Rajul, and Michael Engelschalk, 2018. "Taxation and the Shadow Economy." World Bank Policy Research Working Paper 8391. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3153229

Bachas, Pierre, Roberto N. Fattal Jaef, and Anders Jensen, 2019. "Size-Dependent Tax Enforcement and Compliance: Global Evidence and Aggregate Implications." *Journal of Development Economics* 140, 203–22. https://doi.org/10.1016/j.jdeveco.2019.05.001.

Bíró, Anikó, Daniel Prinz, and László Sándor, 2021. "The minimum wage, informal pay and tax enforcement." Working Paper 21/41. Institute for Fiscal Studies https://doi.org/10.1920/wp.ifs.2021.4121

Bostan, Ionel, Cristian Popescu, Costel Istrate, and Ioan-Bogdan Robu, 2017. "The Impact of Taxation of the Domestic Economic Transactions on the Vat Collection Through Electronic Fiscal Devices." *Amfiteatru Economic Journal* 19 (45), 581–94. https://www.econstor.eu/handle/10419/169090

Cardoza, Marvin, 2012. "Fiscal Printers Dominican Experience." *Tax Administration Review* 33, 16–33. https://www.ciat.org/Biblioteca/Revista/Revista_33/Ingles/rev33-e-book_ing.pdf#page=20

Carrillo, Paul, Dina Pomeranz, and Monica Singhal, 2017. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement." *American Economic Journal: Applied Economics* 9 (2), 144–64. https://doi.org/10.1257/app.20140495

Casey, Peter, and Patricio Castro, 2015. "Electronic Fiscal Devices (EFDs) An Empirical Study of Their Impact on Taxpayer Compliance and Administrative Efficiency." IMF Working Papers 15/73 https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Electronic-Fiscal-Devices-EFDs-An-Empirical-Study-of-their-Impact-on-Taxpayer-Compliance-and-42820#:~:text=IMF%20Working%20Papers-,Electronic%20Fiscal%20Devices%20(EFDs)%20An%20Empirical%20Study%20of%20their%20Impact,Taxpayer%20Compliance%20and%20Administrative%20Efficiency&text=Summary%3A,(VAT)%20payable%20on%20sales..

Chacaltana, Juan, Vicky Leung, and Miso Lee, 2018. "New technologies and the transition to formality: The trend towards e-formality." ILO Employment Policy Department Working Paper 247. https://www.bollettinoadapt.it/wp-content/uploads/2018/08/wcms_635996.pdf

Darvas, Zsolt, 2010. "The Impact of the Crisis on Budget Policy in Central and Eastern Europe." *OECD Journal on Budgeting* 10 (1), 1–42. https://doi.org/10.1787/16812336

de Mello, Luiz, and Teresa Ter-Minassian, 2020. "Digitalisation Challenges and Opportunities for Subnational Governments." OECD Working Papers on Fiscal Federalism 31, 1–24. https://doi.org/10.1787/22265848

Eilu, Emmanuel. 2018, "Adoption of Electronic Fiscal Devices (EFDs) for Value-Added Tax (VAT) Collection in Kenya and Tanzania: A Systematic Review." *The African Journal of Information and Communication* 22, 111–134. http://dx.doi.org/10.23962/10539/26169

Eissa, Nada, Andrew Zeitlin, Saahil Karpe, and Sally Murray, 2014. "The Adoption and Impacts of Electronic Billing Machines for VAT in Rwanda." Report submitted by the International Growth Centre http://www.salesdatacontroller.com/wp-content/uploads/2015/09/Eissa-Incidence-and-Impact-Rwanda-VAT.pdf

European Commission, Directorate-General for Taxation and Customs Union, 2018. "Study and reports on the VAT gap in the EU-28 Member States: report 2018" https://data.europa.eu/doi/10.2778/013500

European Commission, Directorate-General for Taxation and Customs Union, Poniatowski, G., Bonch-Osmolovskiy, M., Śmietanka, A., 2021. "VAT gap in the EU: report 2021" Publications Office, 2021 https://data.europa.eu/doi/10.2778/447556

Fan, Haichao, Yu Liu, Nancy Qian, and Jaya Wen, 2018.”The Dynamic Effects of Computerized VAT Invoices on Chinese Manufacturing Firms.” CEPR Discussion Paper No. DP12786 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3138848

Fjeldstad, Odd Helge, Cecilia Kagoma, Ephraim Mdee, Ingrid Hoem Sjursenc and Vincent Somville, 2020. “The Customer Is King: Evidence on VAT Compliance in Tanzania.” World Development 128. https://doi.org/10.1016/j.worlddev.2019.104841.

Hungarian Central Statistical Office. 2013. “A Hálózatba kötött pénztárgépek gazdaságfehérítő hatásának vizsgálata.” HCSO, Budapest, Hungary http://www.ksh.hu/docs/hun/xftp/idoszaki/pdf/opg.pdf

Ilyés, Tamás, and Lóránt Varga, 2018. “Acceptance of Payment Cards by Retailers in Hungary Based on Data of Online Cash Registers.” *Financial and Economic Review* 17 (1), 83–109. http://doi.org/10.25201/FER.17.1.83109

Jakubowska, Marta. 2019, “Tradiotional Fiscal Devices in Polnad and the Concept of Their Modernization.” *Copernican Journal of Finance & Accounting* 8 (1), 125-138. https://doi.org/10.12775/CJFA.2019.005

Mandari, Herman, Daniel Koloseni, and Jerome Nguridada, 2017. “Electronic Fiscal Device (EFD) Acceptance for Tax Compliance among Trading Business Community in Tanzania: The Role of Awareness and Trust.*” International Journal of Economics, Commerce and Management* 5 (3), 142–58. http://154.72.94.133:8080/xmlui/handle/123456789/159

Mayer-Schönberger, Victor, and Kenneth Cukier, 2013. “Big Data: A Revolution That Will Transform How We Live, Work, and Think”. Houghton Mifflin Harcourt. https://www.amazon.com/Big-Data-Revolution-Transform-Think/dp/0544227751

Naritomi, Joana, 2019. “Consumers as Tax Auditors†.” *American Economic Review* 109 (9), 3031–3072. https://doi.org/10.1257/aer.20160658

OECD, 2013. “Electronic Sales Suppression : A Threat to Tax Revenues.” OECD Publishing. Paris. https://www.oecd.org/ctp/crime/ElectronicSalesSupression.pdf

OECD, 2017. “Technology Tools to Tackle Tax Evasion and Tax Fraud.” OECD Publishing. Paris.

http://www.oecd.org/tax/crime/technology-tools-to-tackle-tax-evasion-and-tax-fraud.htm.

OECD, 2019. "Implementing Online Cash Registers: Benefits, Considerations and Guidance." OECD Publishing, Paris https://www.oecd.org/tax/forum-on-tax-administration/publications-and-products/implementing-online-cash-registers-benefits-considerations-and-guidance.htm

Pomeranz, Dina, 2015. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." *American Economic Review* 105 (8), 2539–2569. https://doi.org/10.1257/aer.20130393

Ritzlné Kazimir, Ildikó, 2021. "Az áfacsalás alakulása Magyarországon 2006 és 2016 között." Budapest Corvinus Egyetem. http://phd.lib.uni-corvinus.hu/1149/1/ritzlne_kazimir_ildiko_dhu.pdf

Schneider, Friedrich, Andreas Buehn, and Claudio E. Montenegro, 2010. "New Estimates for the Shadow Economies All over the World." *International Economic Journal* 24 (4), 443-461. https://doi.org/10.1080/10168737.2010.525974

Skatteverket, 2013. "Requirement of Cash Registers Impact Evaluation." English Summary of the Swedish Tax Agency Report. https://www.skatteverket.se/download/18.46ae6b26141980f1e2d11b2/1382084687220/Summary+of+Report+2013_2+The+Requirement+of+Cash+Registers.+An+impact+evaluation.pdf

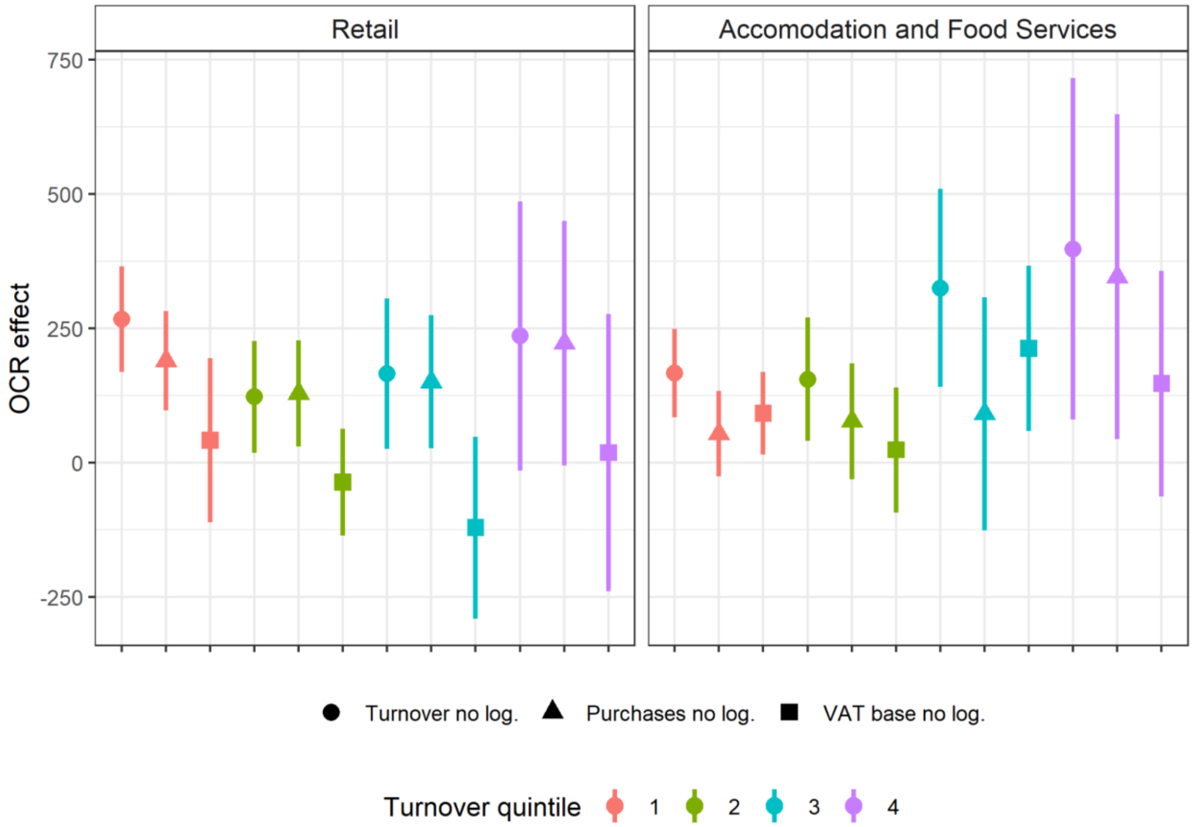
Tóth, Csaba G, 2011. "Debt Dynamics and Fiscal Sustainability." *Hungarian Statistical Review Special* 16, 123–49. https://ssrn.com/abstract=2352836

Varian, Hal R, 2010. "Computer Mediated Transactions." *American Economic Review* 100 (2), 1–10. https://doi.org/10.1257/aer.100.2.1

Appendix A: Results without log transformation

Results on turnover and purchases without log transformation were similar to our main results with log transformation, with the exception of those in the 5th size quintile. This supports the view that valid results are possible without log transformation if the size difference within the category is not too large. In the largest size categories, the difference between the smallest and largest companies is substantial, making the estimation unreliable.

Figure A1: Results using variables without log transformation



Source: Own calculation

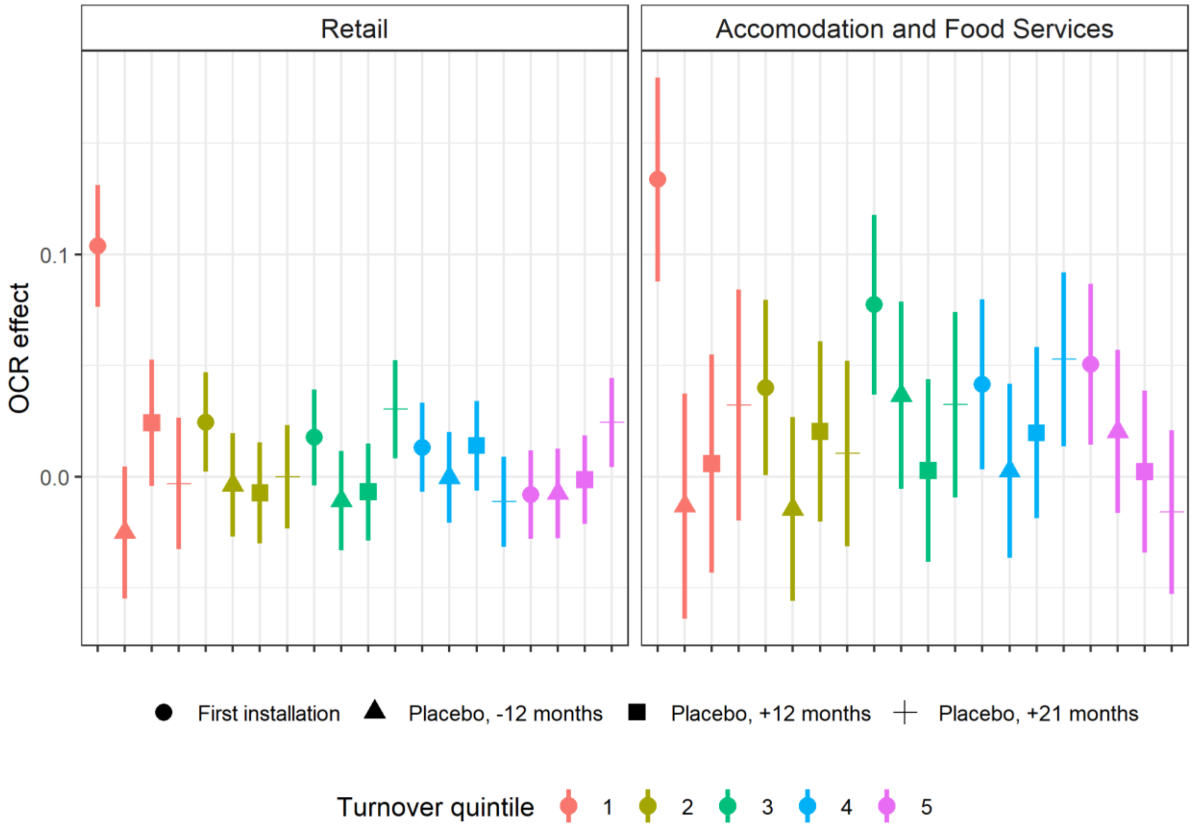
Table A1: Results using variables without log transformation

Sector	Turnover Quintile		Turnover no log.	Purchases no log.	VAT base no log.
Retail	1	Estimate	267.820**	190.308**	42.247
		T-statistic	5.342	4.034	0.543
		R2	0.594	0.610	0.366
		Number of obs.	52,762	52,762	52,762
	2	Estimate	122.989*	129.231*	−36.084
		T-statistic	2.322	2.575	−0.713
		R2	0.434	0.540	0.560
		Number of obs.	55,870	55,870	55,870
	3	Estimate	166.142*	151.022*	−120.416
		T-statistic	2.329	2.389	−1.395
		R2	0.452	0.505	0.208
		Number of obs.	57,215	57,215	57,215
	4	Estimate	236.184	222.707	19.016
		T-statistic	1.849	1.921	0.145
		R2	0.493	0.519	0.233
		Number of obs.	58,331	58,331	58,331
	5	Estimate	−15,659.666	−9,865.420	−6,657.154
		T-statistic	−1.194	−1.292	−0.734
		R2	0.983	0.987	0.935
		Number of obs.	59,188	59,188	59,188
AFS	1	Estimate	167.160**	54.331	92.360*
		T-statistic	3.981	1.345	2.363
		R2	0.482	0.447	0.280
		Number of obs.	18,425	18,425	18,425
	2	Estimate	155.871**	77.479	24.350
		T-statistic	2.662	1.412	0.410
		R2	0.399	0.560	0.610
		Number of obs.	20,481	20,481	20,481
	3	Estimate	325.519**	91.282	213.093**
		T-statistic	3.460	0.823	2.708
		R2	0.428	0.347	0.298
		Number of obs.	20,959	20,959	20,959
	4	Estimate	398.290*	346.367*	147.707
		T-statistic	2.456	2.244	1.379
		R2	0.442	0.417	0.339
		Number of obs.	21,356	21,356	21,356
	5	Estimate	2,998.803*	2,346.623*	1,017.198
		T-statistic	2.308	2.068	1.279
		R2	0.930	0.876	0.867
		Number of obs.	21,868	21,868	21,868

Source: Own calculation
Notes: ** significant at 1% level, * significant at 5% level

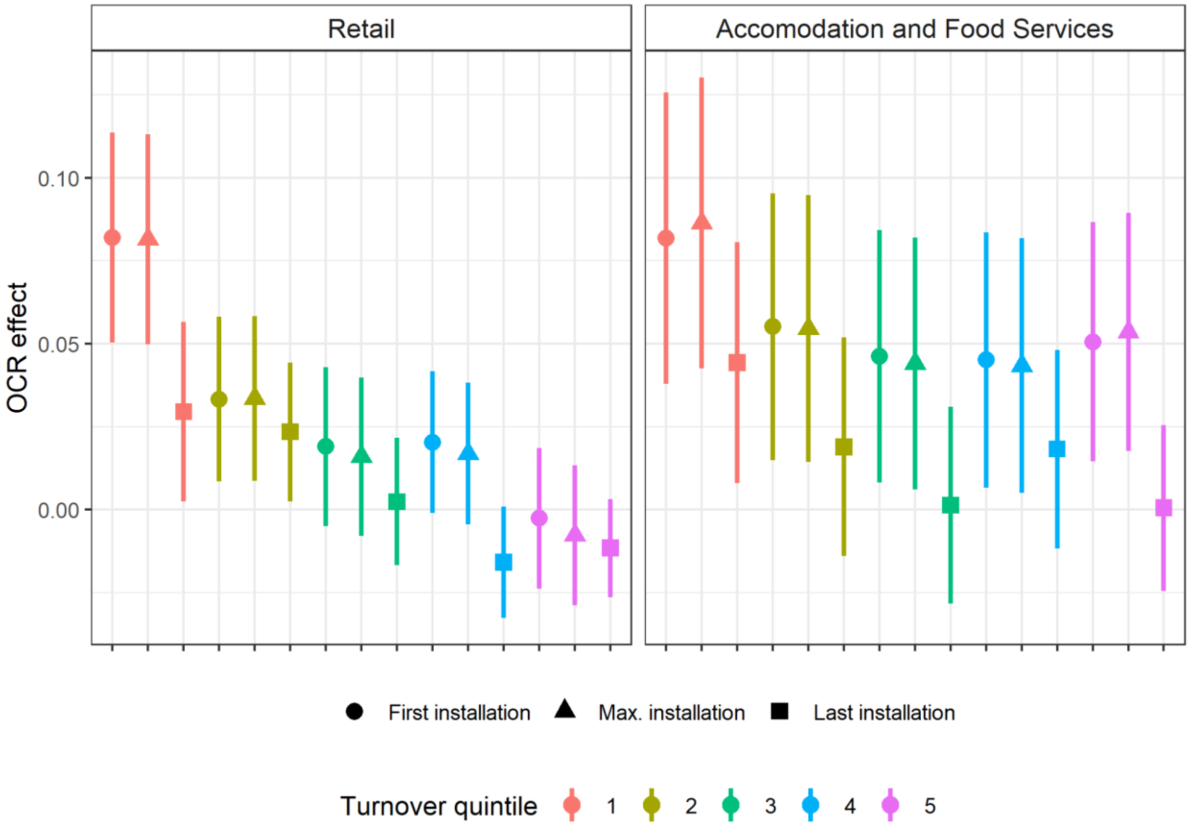
Appendix B: Robustness tests for purchases

Figure B1: Results with placebo introduction dates for purchases



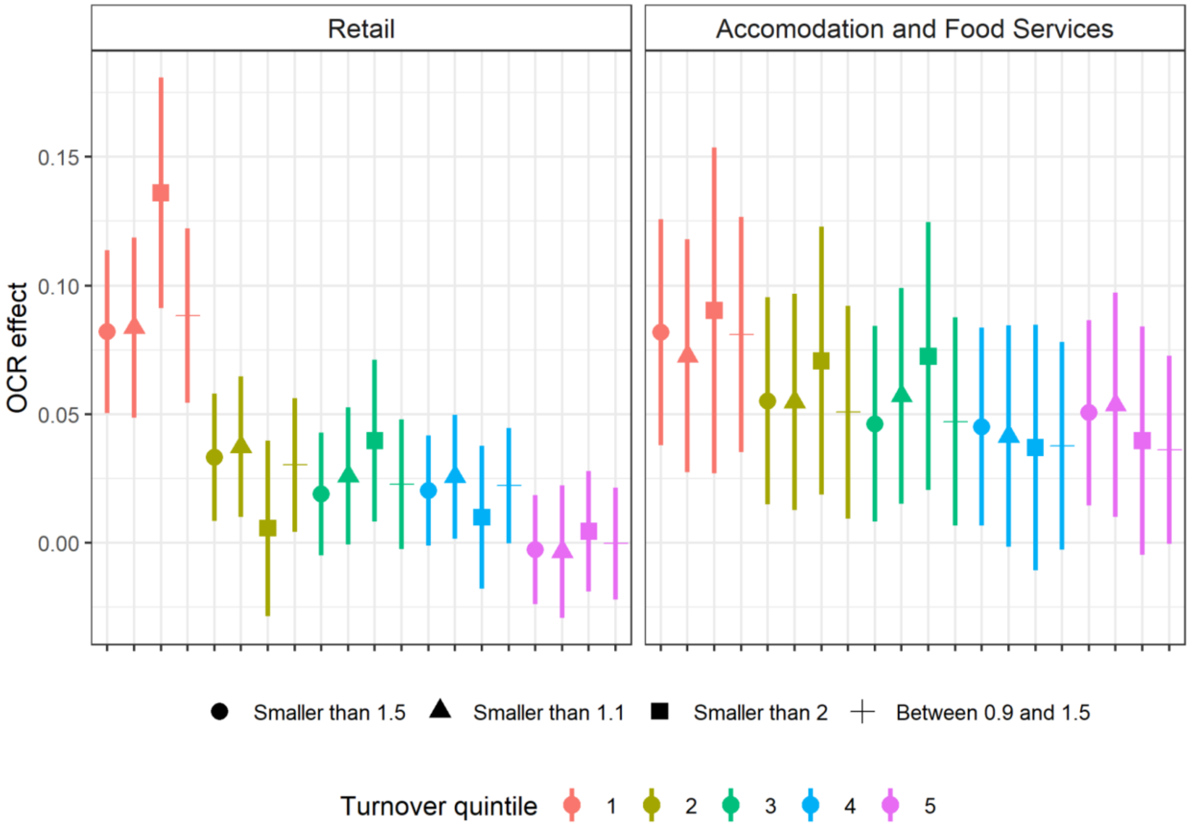
Source: Own calculation

Figure B2: Results with different definitions of the introduction date for purchases



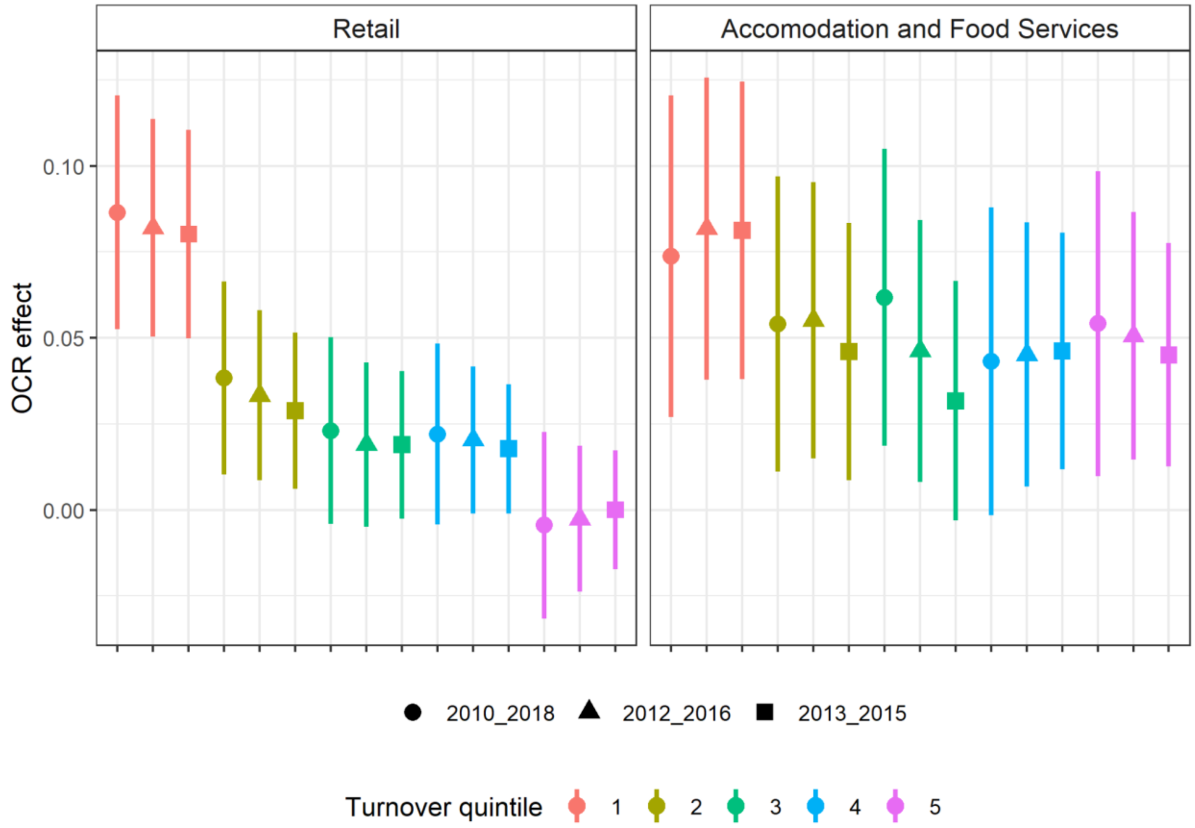
Source: Own calculation

Figure B3: Results with alternative OCR turnover-ratio cutoffs for purchases



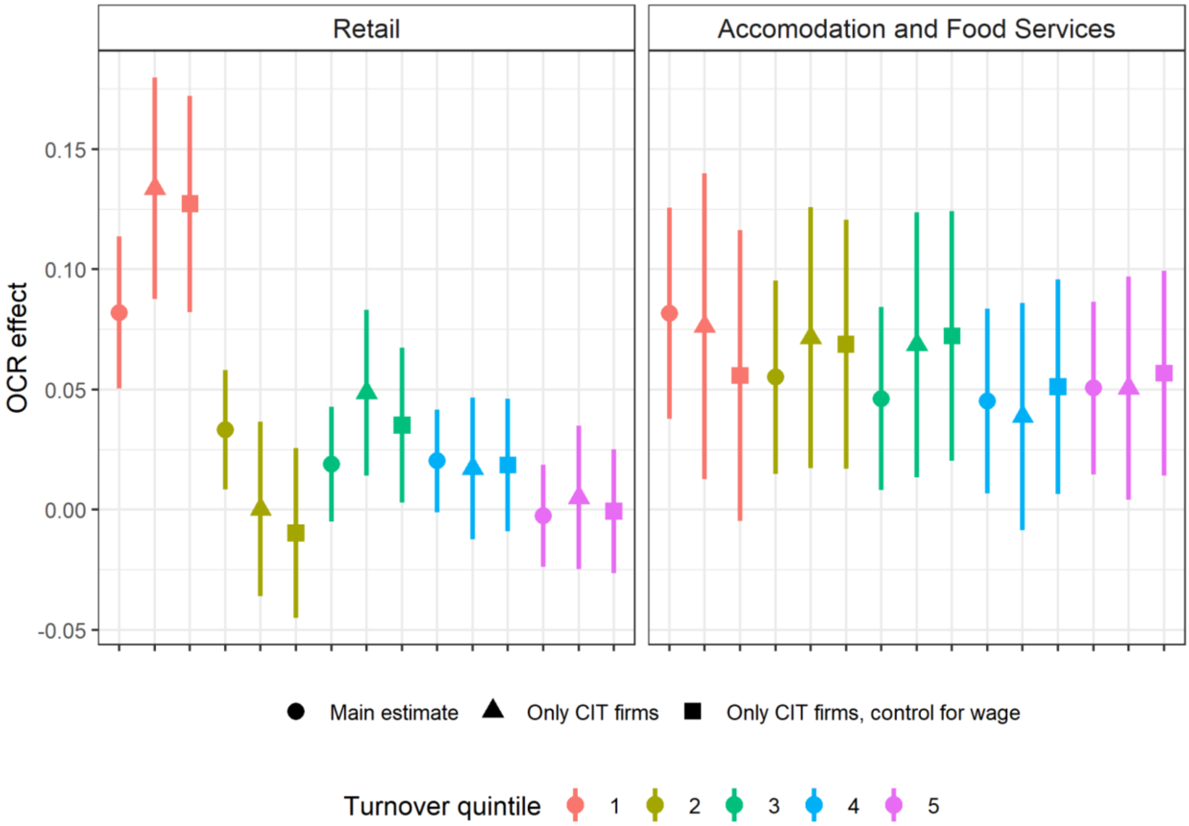
Source: Own calculation

Figure B4: Results with alternative time frames for purchases



Source: Own calculation

Figure B5: Results using wages as a control variable for purchases



Source: Own calculation

Appendix C: Other sectors

In addition to the firms in the two main sectors – retail and AFS – there are firms that could not be categorized into these based on their NACE codes. They are a very mixed group. It is most likely that these firms were conducting economic activity for which OCRs were obligatory but which, based on their NACE code, was not their main activity, making it difficult to categorize them into the retail or AFS sectors. It is also possible that some companies bought OCRs even though they were not obliged to do so. Including these in the main estimation would be a mistake as there could be selection bias. There are 4,679 companies in this category, and we also develop estimations for these. Compared to Table 1, the data-filtering step using the OCR turnover ratio in Table C1 shows that companies having only a smaller portion of their turnover processed through OCRs were much more important in sectors other than retail and AFS – the “other” sector.

Table C1: Number of observations and percentage of gross OCR turnover after data filtering in “other” sector

	Other	
	N - cumulative	Percentage of gross OCR turnover (Initial population=100%)
Initial population	20,089	100
Without annual tax returns	19,975	100
VAT returns around OCR introduction	17,302	97
Outliers with >100* growth	14,005	88
OCR turnover ratio (sample)	4,679	23

Source: National Tax and Customs Administration

The OCR effects also decrease with size in this “sector.”. The effect is only significant in the smallest quintile as standard errors are quite large due to the relatively small sample size. The point estimate in the smallest group is almost 7 percent, in the 3 middle quintiles, it moves to around 4 percent, and in the biggest quintile, it is −2 percent.

Table C2: OCR effects on turnover: main results for "other" sector

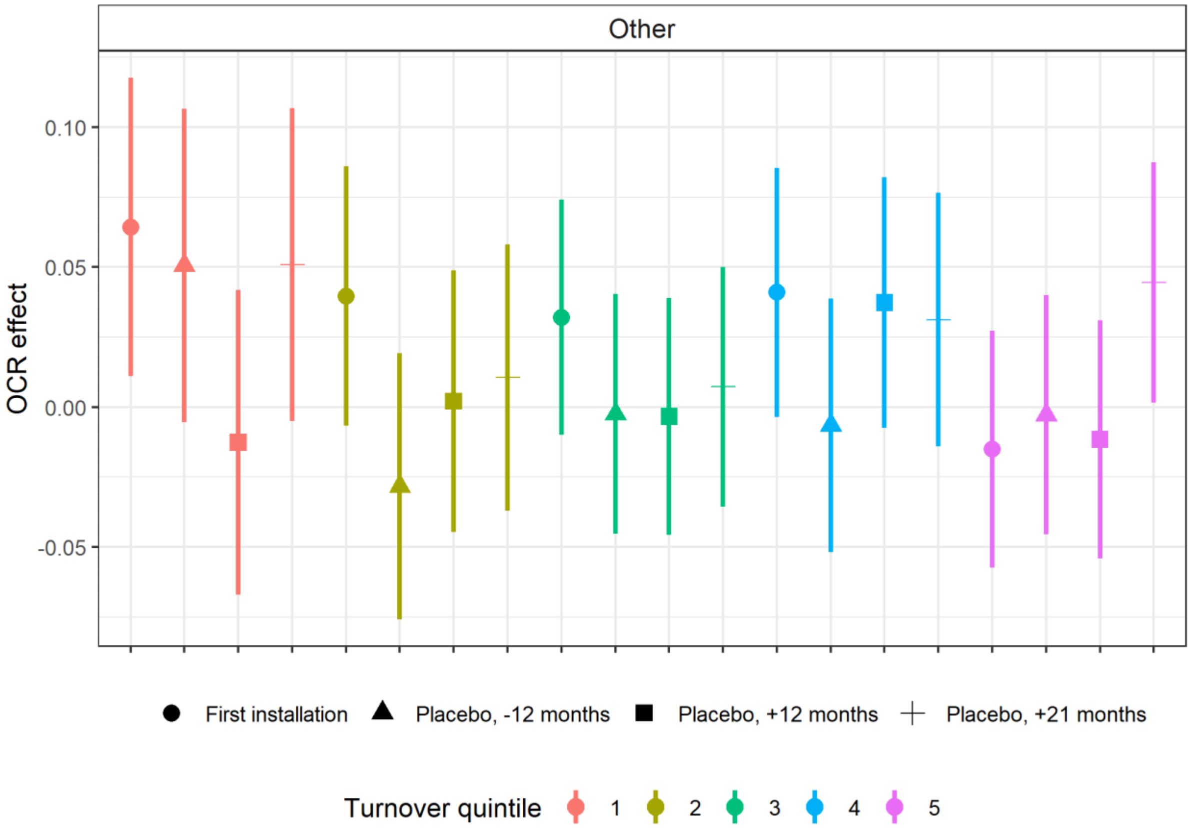
Sector	Turnover Quintile		First Installation
"Other" sector	1	Estimate	0.064*
		T-statistic	2.364
		R2	0.588
		Number of observations	17,194
	2	Estimate	0.040
		T-statistic	1.676
		R2	0.359
		Number of observations	17,881
	3	Estimate	0.032
		T-statistic	1.495
		R2	0.387
		Number of observations	18,058
	4	Estimate	0.041
		T-statistic	1.804
		R2	0.423
		Number of observations	18,020
	5	Estimate	−0.015
		T-statistic	−0.696
		R2	0.800
		Number of observations	18,381

Source: Own calculation

Notes: ** significant at 1% level, * significant at 5% level

Results with placebo introduction dates are never significantly different from zero, with one exception in the largest quintile.

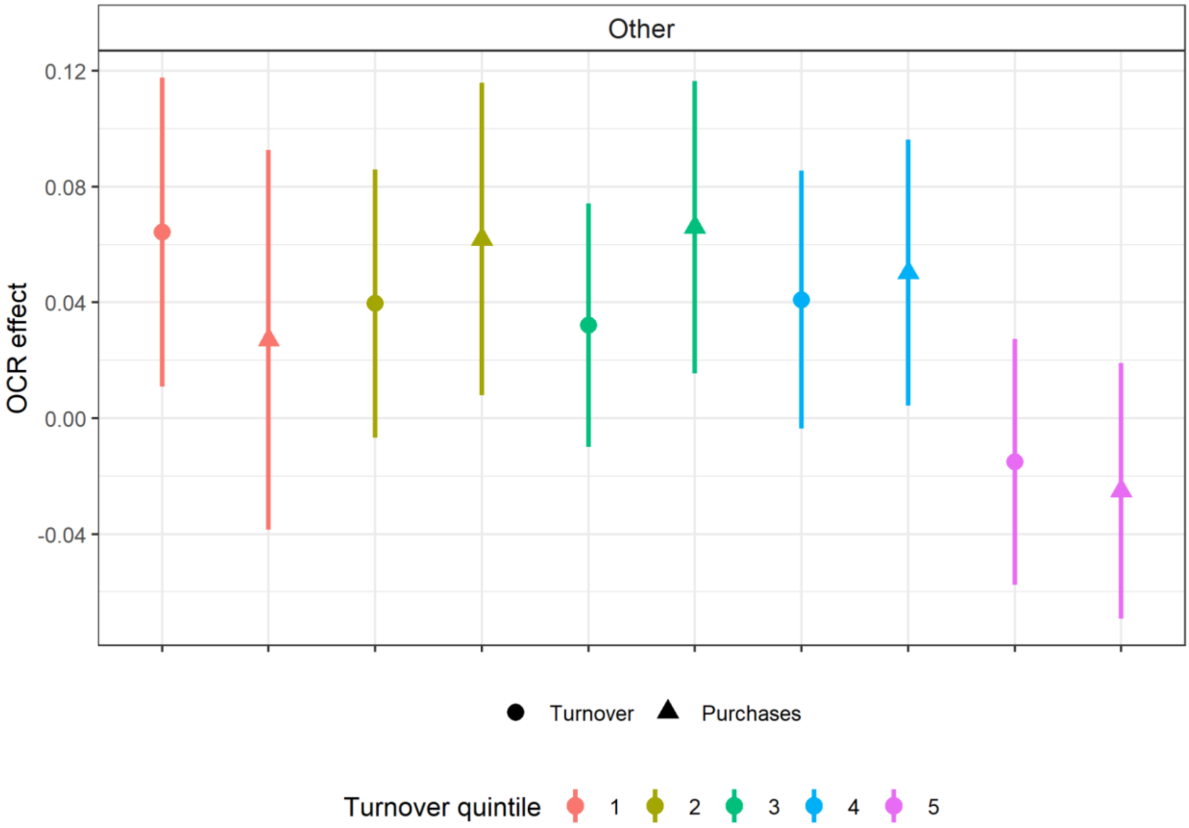
Figure C1: Results with placebo introduction dates for “other” sector



Source: Own calculation

The OCR effect on purchases is not significantly different from the effect on turnover in either size quintile. The difference is biggest in the smallest quintile, where the greater turnover effect implies less spillover than in bigger size categories.

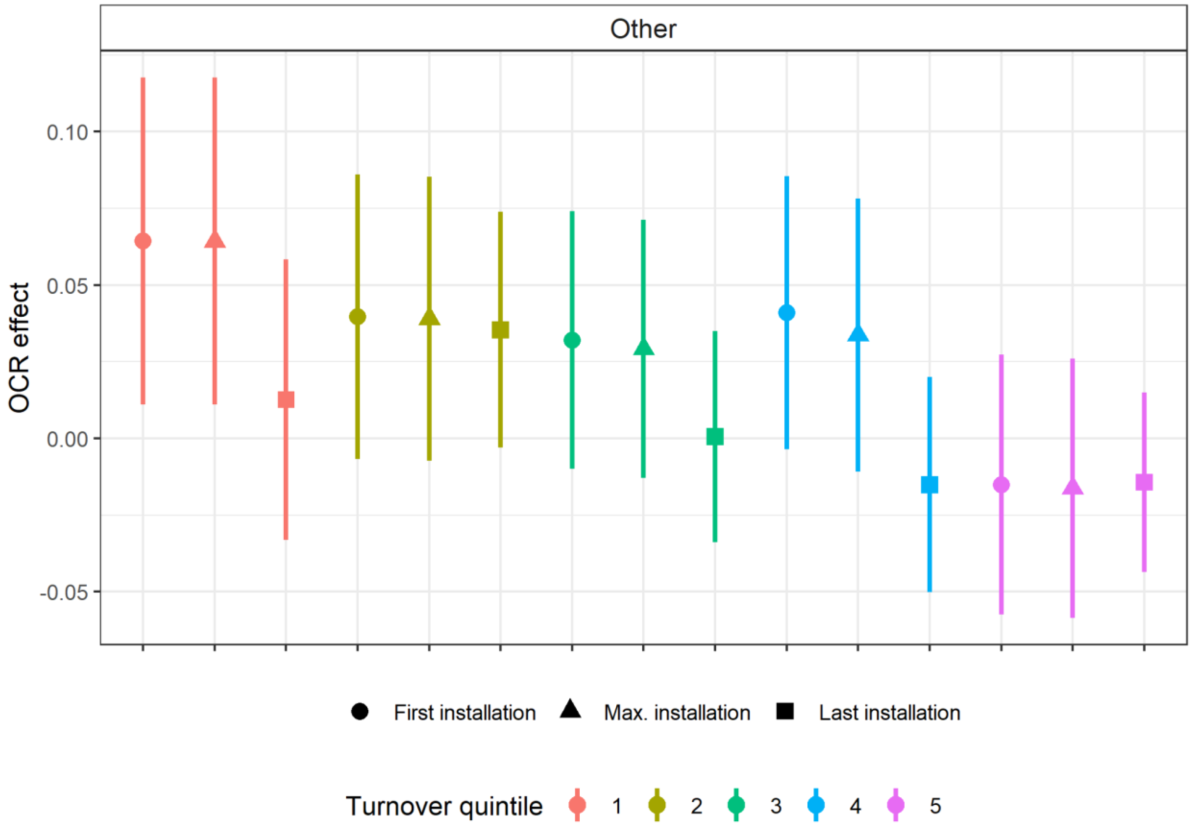
Figure C2: Results for turnover and purchases for "other" sector



Source: Own calculation

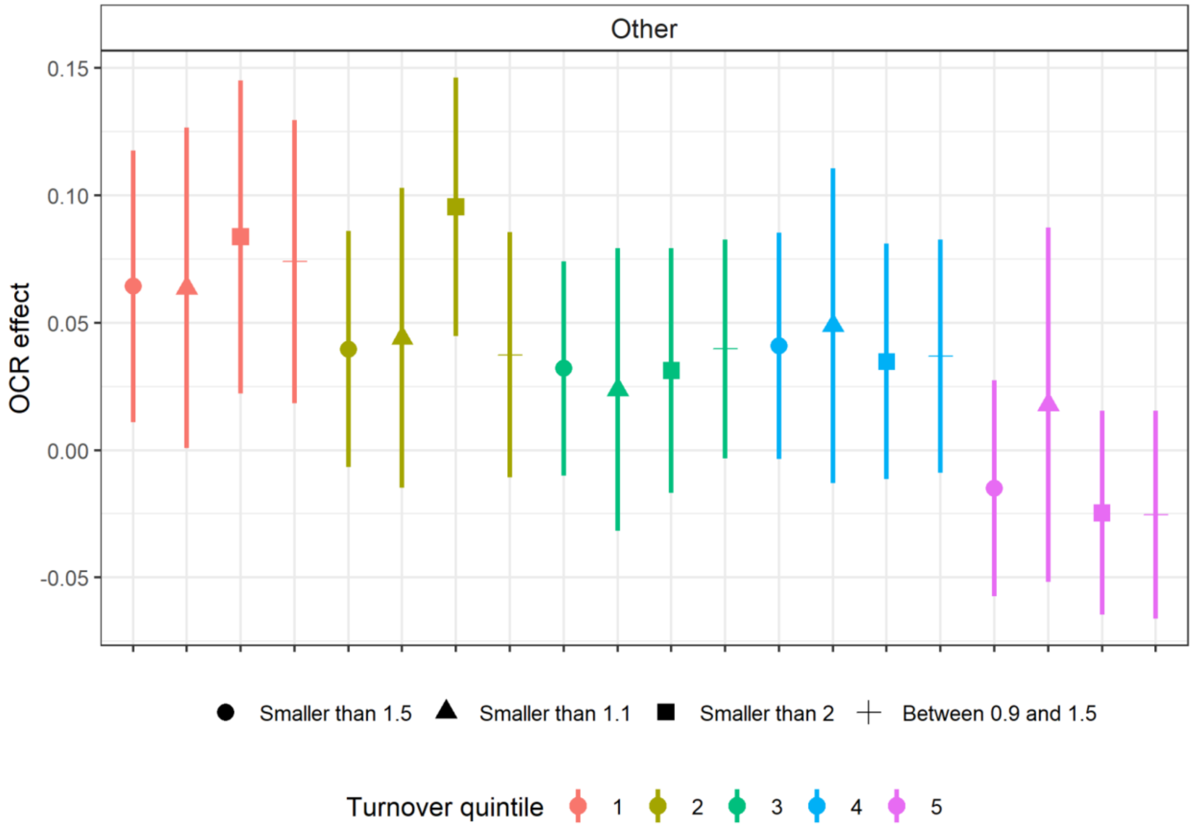
We present further robustness tests for the "other" sector, the results of which are very similar to those in the retail and AFS sectors.

Figure C3: Results with alternative definitions of the introduction date for "other" sector



Source: Own calculation

Figure C4: Results with alternative OCR turnover ratio cutoffs for "other" sector



Source: Own calculation

Appendix D: Tables of graphs

Table D1: Results with placebo introduction dates

Sector	Turnover Quintile		First installation (Main estimate)	Placebo, -12 months	Placebo, +12 months	Placebo, +21 months
Retail	1	Estimate	0.104**	−0.025	0.024	−0.003
		T-statistic	7.395	−1.660	1.667	−0.209
		R2	0.596	0.596	0.596	0.596
		Number of obs.	52,762	52,762	52,762	52,762
	2	Estimate	0.024*	−0.004	−0.007	0,000
		T-statistic	2.150	−0.322	−0.636	−0.005
		R2	0.391	0.390	0.390	0.390
		Number of obs.	55,870	55,870	55,870	55,870
	3	Estimate	0.018	−0.011	−0.007	0.030**
		T-statistic	1.600	−0.960	−0.619	2.687
		R2	0.391	0.391	0.391	0.391
		Number of obs.	57,215	57,215	57,215	57,215
	4	Estimate	0.013	0,000	0.014	−0.011
		T-statistic	1.283	−0.042	1.356	−1.090
		R2	0.451	0.451	0.451	0.451
		Number of obs.	58,331	58,331	58,331	58,331
	5	Estimate	−0.008	−0.008	−0.001	0.024*
		T-statistic	−0.790	−0.740	−0.142	2.377
		R2	0.900	0.900	0.900	0.900
		Number of obs.	59,188	59,188	59,188	59,188
AFS	1	Estimate	0.134**	−0.013	0.006	0.032
		T-statistic	5.710	−0.511	0.234	1.217
		R2	0.451	0.450	0.450	0.450
		Number of obs.	18,425	18,425	18,425	18,425
	2	Estimate	0.040**	−0.015	0.020	0.010
		T-statistic	1.987	−0.692	0.977	0.487
		R2	0.334	0.334	0.334	0.334
		Number of obs.	20,481	20,481	20,481	20,481
	3	Estimate	0.077*	0.037	0.003	0.032
		T-statistic	3.746	1.697	0.131	1.520
		R2	0.379	0.378	0.378	0.378
		Number of obs.	20,959	20,959	20,959	20,959
	4	Estimate	0.042**	0.003	0.020	0.053
		T-statistic	2.130	0.127	1.002	2.640
		R2	0.421	0.421	0.421	0.422
		Number of obs.	21,356	21,356	21,356	21,356
	5	Estimate	0.051*	0.020	0.002	−0.016**
		T-statistic	2.729	1.083	0.116	−0.849
		R2	0.784	0.784	0.784	0.784
		Number of obs.	21,868	21,868	21,868	21,868

Source: Own calculation

Notes: ** significant at 1% level, * significant at 5% level

Table D2: Results with alternative definitions of introduction date

Sector	Turnover Quintile		First installation (Main estimate)	Max. installation	Last installation
Retail	1	Estimate	0.104**	0.101**	0.043**
		T-statistic	7.395	7.200	3.575
		R2	0.596	0.596	0.596
		Number of obs.	52,762	52,762	52,762
	2	Estimate	0.024*	0.024*	0.014
		T-statistic	2.150	2.097	1.437
		R2	0.391	0.391	0.391
		Number of obs.	55,870	55,870	55,870
	3	Estimate	0.018	0.015	0.008
		T-statistic	1.600	1.374	0.866
		R2	0.391	0.391	0.391
		Number of obs.	57,215	57,215	57,215
	4	Estimate	0.013	0.013	−0.010
		T-statistic	1.283	1.303	−1.272
		R2	0.451	0.451	0.451
		Number of obs.	58,331	58,331	58,331
	5	Estimate	−0.008	−0.010	−0.012
		T-statistic	−0.790	−0.987	−1.626
		R2	0.900	0.900	0.900
		Number of obs.	59,188	59,188	59,188
AFS	1	Estimate	0.134**	0.136**	0.072**
		T-statistic	5.710	5.814	3.739
		R2	0.451	0.451	0.450
		Number of obs.	18,425	18,425	18,425
	2	Estimate	0.040*	0.040*	0.025
		T-statistic	1.987	1.972	1.528
		R2	0.334	0.334	0.334
		Number of obs.	20,481	20,481	20,481
	3	Estimate	0.077**	0.072**	0.045**
		T-statistic	3.746	3.480	2.779
		R2	0.379	0.378	0.378
		Number of obs.	20,959	20,959	20,959
	4	Estimate	0.042*	0.038*	0.024
		T-statistic	2.130	1.966	1.568
		R2	0.421	0.421	0.421
		Number of obs.	21,356	21,356	21,356
	5	Estimate	0.051**	0.048**	0.023
		T-statistic	2.729	2.602	1.779
		R2	0.784	0.784	0.784
		Number of obs.	21,868	21,868	21,868

Source: Own calculation
Notes: ** significant at 1% level, * significant at 5% level

Table D3: Results with alternative OCR turnover-ratio cutoffs

Sector	Turnover Quintile		Smaller than 1.5 (Main estimate)	Smaller than 1.1	Between 0.9 and 1.5	Smaller than 2
Retail	1	Estimate	0.104**	0.104**	0.107**	0.156**
		T-statistic	7.395	6.673	7.388	7.689
		R2	0.596	0.601	0.582	0.631
		Number of observations	52,762	41,674	46,170	30,111
	2	Estimate	0.024*	0.031*	0.019	0.008
		T-statistic	2.150	2.507	1.637	0.515
		R2	0.391	0.395	0.395	0.483
		Number of observations	55,870	43,608	50,549	34,315
	3	Estimate	0.018	0.020	0.020	0.034*
		T-statistic	1.600	1.583	1.757	2.262
		R2	0.391	0.401	0.387	0.474
		Number of observations	57,215	42,557	51,625	34,162
	4	Estimate	0.013	0.019	0.013	0.007
		T-statistic	1.283	1.653	1.183	0.506
		R2	0.451	0.466	0.451	0.538
		Number of observations	58,331	40,540	52,868	36,975
	5	Estimate	−0.008	−0.008	−0.008	−0.005
		T-statistic	−0.790	−0.673	−0.807	−0.471
		R2	0.900	0.906	0.907	0.909
		Number of observations	59,188	36,625	55,057	45,702
AFS	1	Estimate	0.134**	0.121**	0.129**	0.140**
		T-statistic	5.710	4.961	5.290	4.107
		R2	0.451	0.474	0.461	0.482
		Number of observations	18,425	15,281	16,272	10,833
	2	Estimate	0.040*	0.038	0.042	0.101**
		T-statistic	1.987	1.807	1.959	3.725
		R2	0.334	0.344	0.332	0.432
		Number of observations	20,481	16,802	17,846	13,721
	3	Estimate	0.077**	0.079**	0.081**	0.064*
		T-statistic	3.746	3.416	3.733	2.484
		R2	0.379	0.383	0.381	0.447
		Number of observations	20,959	16,133	18,761	13,176
	4	Estimate	0.042*	0.038	0.042*	0.055*
		T-statistic	2.130	1.752	2.037	2.195
		R2	0.421	0.432	0.422	0.490
		Number of observations	21,356	15,326	19,865	13,060
	5	Estimate	0.051**	0.049*	0.040*	0.052*
		T-statistic	2.729	2.149	2.123	2.260
		R2	0.784	0.778	0.789	0.818
		Number of observations	21,868	13,972	20,840	13,119

Source: Own calculation

Notes: ** significant at 1% level, * significant at 5% level

Table D4: Results with lagged dependent variable

Sector	Turnover Quintile		Company fixed effect	Lagged turnover
Retail	1	Estimate	0.104**	0.035**
		T-statistic	7.395	2.721
		R2	0.596	0.578
		Number of observations	52,762	49,486
	2	Estimate	0.024*	0.013
		T-statistic	2.150	1.224
		R2	0.391	0.408
		Number of observations	55,870	52,604
	3	Estimate	0.018	0.003
		T-statistic	1.600	0.299
		R2	0.391	0.451
		Number of observations	57,215	53,998
	4	Estimate	0.013	0.008
		T-statistic	1.283	0.880
		R2	0.451	0.503
		Number of observations	58,331	55,191
	5	Estimate	−0.008	−0.018*
		T-statistic	−0.790	−2.237
		R2	0.900	0.924
		Number of observations	59,188	56,147
AFS	1	Estimate	0.134**	0.037
		T-statistic	5.710	1.752
		R2	0.451	0.409
		Number of observations	18,425	17,162
	2	Estimate	0.040*	0.042*
		T-statistic	1.987	2.267
		R2	0.334	0.308
		Number of observations	20,481	19,241
	3	Estimate	0.077**	0.050**
		T-statistic	3.746	2.614
		R2	0.379	0.346
		Number of observations	20,959	19,759
	4	Estimate	0.042*	0.024
		T-statistic	2.130	1.331
		R2	0.421	0.410
		Number of observations	21,356	20,184
	5	Estimate	0.051**	0.020
		T-statistic	2.729	1.141
		R2	0.784	0.760
		Number of observations	21,868	20,741

Source: Own calculation
Notes: ** significant at 1% level, * significant at 5% level

Table D5: Results with alternative time frames

Sector	Turnover Quintile		2012-2016 (Main estimate)	2010-2018	2013-2015
Retail	1	Estimate	0.104**	0.101**	0.101**
		T-statistic	7.395	6.692	7.410
		R2	0.596	0.543	0.647
		Number of observations	52,762	84,272	33,963
	2	Estimate	0.024*	0.032*	0.021*
		T-statistic	2.150	2.466	2.001
		R2	0.391	0.349	0.459
		Number of observations	55,870	93,709	34,827
	3	Estimate	0.018	0.023	0.019
		T-statistic	1.600	1.772	1.943
		R2	0.391	0.342	0.478
		Number of observations	57,215	98,027	35,281
	4	Estimate	0.013	0.012	0.013
		T-statistic	1.283	0.932	1.489
		R2	0.451	0.380	0.551
		Number of observations	58,331	101,167	35,611
	5	Estimate	−0.008	−0.003	−0.007
		T-statistic	−0.790	−0.220	−0.824
		R2	0.900	0.840	0.935
		Number of observations	59,188	104,150	35,863
AFS	1	Estimate	0.134**	0.121**	0.135**
		T-statistic	5.710	4.812	5.882
		R2	0.451	0.417	0.490
		Number of observations	18,425	28,474	12,115
	2	Estimate	0.040*	0.041	0.037
		T-statistic	1.987	1.901	1.939
		R2	0.334	0.343	0.384
		Number of observations	20,481	34,041	12,837
	3	Estimate	0.077**	0.084**	0.073**
		T-statistic	3.746	3.657	3.821
		R2	0.379	0.363	0.441
		Number of observations	20,959	35,553	13,031
	4	Estimate	0.042*	0.041	0.044*
		T-statistic	2.130	1.776	2.434
		R2	0.421	0.405	0.480
		Number of observations	21,356	36,541	13,181
	5	Estimate	0.051**	0.055*	0.045**
		T-statistic	2.729	2.423	2.700
		R2	0.784	0.718	0.827
		Number of observations	21,868	38,177	13,321

Source: Own calculation

Notes: ** significant at 1% level, * significant at 5% level

Table D6: Results using wages as a control variable

Sector	Turnover Quintile		Only CIT firms, control for wage	Main estimate	Only CIT firms
Retail	1	Estimate	0.152**	0.104**	0.160**
		T-statistic	7.677	7.395	7.693
		R2	0.656	0.596	0.621
		Number of observations	27,885	52,762	27,885
	2	Estimate	−0.013	0.024*	−0.001
		T-statistic	−0.829	2.150	−0.061
		R2	0.486	0.391	0.417
		Number of observations	29,562	55,870	29,562
	3	Estimate	0.030*	0.018	0.044**
		T-statistic	1.993	1.600	2.669
		R2	0.497	0.391	0.400
		Number of observations	30,190	57,215	30,190
	4	Estimate	0.010	0.013	0.009
		T-statistic	0.799	1.283	0.623
		R2	0.561	0.451	0.473
		Number of observations	30,515	58,331	30,515
	5	Estimate	−0.005	−0.008	−0.001
		T-statistic	−0.430	−0.790	−0.060
		R2	0.934	0.900	0.911
		Number of observations	30,960	59,188	30,960
AFS	1	Estimate	0.125**	0.134**	0.152**
		T-statistic	3.946	5.710	4.447
		R2	0.542	0.451	0.467
		Number of observations	10,472	18,425	10,472
	2	Estimate	0.084**	0.040*	0.088**
		T-statistic	3.073	1.987	2.996
		R2	0.448	0.334	0.371
		Number of observations	11,483	20,481	11,483
	3	Estimate	0.07**	0.077**	0.066*
		T-statistic	2.898	3.746	2.508
		R2	0.470	0.379	0.372
		Number of observations	11,692	20,959	11,692
	4	Estimate	0.070**	0.042*	0.055*
		T-statistic	2.961	2.130	2.154
		R2	0.521	0.421	0.446
		Number of observations	11,925	21,356	11,925
	5	Estimate	0.079**	0.051**	0.072**
		T-statistic	3.625	2.729	3.011
		R2	0.837	0.784	0.803
		Number of observations	12,195	21,868	12,195

Source: Own calculation

Notes: ** significant at 1% level, * significant at 5% level