# Detecting Wage Under-reporting using a Double Hurdle Model

Péter Elek[a] - János Köllő[b] – Balázs Reizer[c] - Péter A. Szabó[d]

January 2012[1]

**Abstract.** We estimate a double hurdle (DH) model of the Hungarian wage distribution assuming censoring at the minimum wage and wage under-reporting (i.e. compensation consisting of the minimum wage, subject to taxation, and an unreported cash supplement). We estimate the probability of under-reporting for minimum wage earners, simulate their genuine earnings and classify them and their employers as 'cheaters' and 'non-cheaters'. In the possession of the classification we check how cheaters and non-cheaters reacted to the introduction of a minimum social security contribution base, equal to 200 per cent of the minimum wage, in 2007. The findings suggest that cheaters were more likely to raise the wages of their minimum wage earners to 200 per cent of the minimum wage thereby reducing the risk of tax audit. Cheating firms also experienced faster average wage growth and slower output growth. The results suggest that the DH model is able to identify the loci of wage under-reporting with some precision.

a) Eötvös Loránd University, Budapest b) Institute of Economics, Budapest c) Central European University d) Reformed Presbyterian Church of Central and Eastern Europe. Corresponding author: János Köllő, kollo@econ.core.hu

# 1. Introduction

The evasion of payroll taxes has two main forms. One is unreported (black) employment, when the employee is not registered and neither she nor her employer pays any taxes. The other main form is the under-reporting of wages, or grey employment, when the compensation consists of an officially paid amount, subject to taxation, and an unreported supplement also known as an "envelope wage" or "under the counter payment". In order to maximize the total evaded tax, the officially paid wage is often (but not always) chosen as the minimum wage (MW).

In this paper we estimate the prevalence of disguised MW earners with the double hurdle (DH) model, first proposed by Cragg (1971), using linked employer-employee data. The DH is a potentially suitable method for disentangling genuine from 'fake' MW earners, relying on the assumption that MW payment is governed by two different processes: market imperfections implying censoring at the MW, on the one hand, and non-random selection to wage under-reporting, on the other. Our application of the DH for Hungary assumes that a spike at the MW was observed for two reasons (i) because of constraints and costs preventing firms from firing all low-productivity workers after a wave of exceptionally large hikes in the MW and (ii) because of tax fraud. That said, a worker's genuine wage is observed only if her productivity exceeds the MW *and* her wage is fully reported. The DH model simultaneously deals with the censoring problem and selection to tax fraud, and estimates the probability of cheating for each MW earner. In the possession of the parameters one can also simulate the 'genuine' wages of MW earners.

The DH model's reliance on distributional properties (as well as the difficulty in finding exclusion restrictions for the selection equation) warns us not to take the estimates at face value. Therefore, we test the validity of the DH results by exploiting a unique episode of Hungary's unconventional MW policies. The test examines the introduction of a minimum contribution base amounting to 200 per cent of the minimum wage (2MW), in 2007. After the introduction of the reform, firms paying wages lower than 2MW faced an increased probability of tax authority audit and a higher risk of being detected as cheaters. Firms were required to report that they paid wages below 2MW and provide evidence, upon request, that their low-wage workers were paid at the going market rate. The reform created incentives for cheating firms to raise the reported wages of MW earners to 2MW while non-cheaters (those paying genuine

minimum wages) had no interest to do so. We distinguish cheaters from non-cheaters on the basis of DH estimates for 2006 and check how the cheating proxies affected the probability that a worker earning the MW in 2006 earned 2MW in 2007. We also study how the wages of MW earners changed in 2006-2007. We find that suspected cheaters were more likely to shift their workers from MW to 2MW compared to non-cheating firms. Furthermore, we find that the sales revenues of cheating firms were adversely affected by the reform.

At least in the East and South-East of Europe, MW policies are strongly influenced by the conviction that nearly all MW workers earn untaxed side payments. Our results suggest that while the suspicions are not groundless they are overstated: we estimate the share of 'disguised' MW earners to be around 50 per cent and the share of cheating enterprises to fall short of 40 per cent. The high share of non-cheating firms and genuine MW earners warns against radical, fiscally motivated experiments with the MW, which may put unskilled jobs at risk. While the statistical profiles derived from the DH model may help the better targeting of tax authority inspection they can also facilitate more circumspect MW policies.

The paper is organized as follows. Section 2 gives a brief overview of the literature while Section 3 the MW regulations and the wage distribution in Hungary. Section 4 introduces the DH model, explains the estimation of its parameters, shows how the probability of cheating and 'genuine' wages are simulated and how we classify workers and firms on the basis of the DH estimates. Section 5 introduces the data. Section 6 presents the estimates of the DH model. Section 7 presents the methods, data and results of the test and Section 8 concludes.

## 2. Wage under-reporting and the minimum wage – An under-researched area

Compared to the vast literature on income under-reporting and MW regulations, respectively, the body of research on how these two areas relate to each other seems rather thin. Most of what we know empirically about this relationship comes from anecdotal evidence, inspection of aggregate data, scarce survey results and a few attempts to identify the incidence of envelope wages indirectly. Theoretical work is largely missing.

Although several mechanisms may cause a spike of the wage distribution at the MW, including the tacit collusion of employers (Shelkova 2008) or the extrusion of wages due to the effective MW (DiNardo et al. 1996), grey employment is certainly among the suspects. Cross-

country data suggest a positive correlation between the size of the spike and estimated size of the informal economy (Tonin 2007). Several accession countries including Hungary, Latvia, Lithuania and Romania have (or had) high shares of MW earners, while their Kaitz-indices are (were) in the middle range, suggesting that disguised MWs may be particularly wide-spread in these countries. Similar observations are interpreted in a similar way in World Bank (2005).

Erdogdu (2009) reports on the basis of several surveys that under-the-counter payments are prevalent in the wage policy of Turkish firms. There is a relatively extensive literature focus-ing on grey employment in the Baltic states. Relying on survey results, Masso and Krillo (2009) point out that 16-23 percent of the MW earners received envelope wages in Estonia and Latvia but only 8 percent in Lithuania in 1998. Meriküll and Staehr (2010) show that young employees and people working in construction and trade are most likely to get unre-ported cash supplement on top of their official salary in the three Baltic countries. Kriz et al. (2007) present similar results on the distribution of envelope wages using three different Es-tonian data sets. According to the Eurobarometer survey conducted by the European Commis-sion in 2007 (European Commission 2007), 5 per cent of employees in the EU receive part or all of their regular income untaxed and this ratio is over 10 per cent in some central and east-ern European countries (8 per cent in Hungary) but there is no information on how many of them are officially paid the MW.

Some studies obtain evidence on disguised MWs indirectly, by comparing the reported con-sumption-income profiles of households. Using household budget survey data from Hungary, Benedek et al (2006) looked at the winners and losers from the 2001-2002 MW hikes. They observed income loss *without* the loss of a wage earner in the high-income brackets where substantial under-reporting is most likely to occur. For these households the increasing MW may have implied higher taxes and lower net income. Based on the same data set, Tonin (2011) analyzed changes in the food consumption of households affected by the minimum wage hike compared to unaffected households of similar income. He found that food con-sumption fell in the treatment group relative to the controls – a fact potentially explained by a fall in their unreported income in response to the MW hike and growth of the associated tax burden.

The theories of wage under-reporting (Allingham and Sandmo 1972, Yaniv 1988) shed light on the incentives to engage in tax fraud under alternative penalty and withdrawal schemes but they do not explicitly discuss the case of reporting the MW to tax authorities. This is the cost-minimizing choice for the firm (unless MW payment provokes audits thereby decreasing the expected gain from cheating) but it also requires the cooperation of workers. As Madzharova (2010) notes: if the actual or perceived linkages between contribution payments and pensions or access to health services are weak and/or workers see that their payments feed corruption rather than are used to finance public services, they will be willing to accept the lowest possible reported wage.

Theoretical models explicitly addressing the issue of wage under-reporting *cum* MW regulations include Tonin (2011) and Shelkova (2008). Tonin argues that the MW induces some workers whose productivity is *above* the MW, but who would have declared less if there was no MW, to increase their declared earnings to the MW level. Workers with productivity below the MW either work in the black market or withdraw from the labor force while high-productivity workers are unaffected. This is a possible explanation of why a spike at the MW appears in the distribution of declared earnings. Shelkova assumes that low productivity labor is homogenous and easy to replace thanks to the low fixed costs of hiring. If a non-binding MW exists and employers act symmetrically then tacit collusion and offering the MW to low productivity workers is profit maximizing and dominant strategy for the companies. An increase in the minimum wage increases the probability of collusion since the incentive for deviation is weaker. This implies that a higher MW increases the spike there.

Our empirical work attributes the sudden nascence and decease of a huge spike at the MW to state intervention, on the one hand, and tax evasion, on the other. We look at a unique period in Hungary's MW history, which *quadrupled* the spike at the MW in only two years (when the MW was nearly doubled in 2001-2002) and decreased it by a *factor of 2.5* in only one year (when a double contribution base was introduced in 2007). We do not believe that these sudden and enormous changes could be explained by the established strategic behavior of enterprises underlying Shelkova's model. It is also hard to trust that Tonin's assumption, stating that the marginal products of those at the spike exceed the MW, was valid in the period we are looking at. When the plan of increasing the MW from Ft 25,500 to Ft 50,000 was announced, 32.7 per cent of the private sector employees earned less than that. When the idea of the minimum contribution base came up, 58 per cent had wages below 2MW. It is quite obvi-

ous that the vast majority of the affected workers remained in employment for a protracted period (or until recently) and many of them had productivity below the aforementioned thresholds after the hikes. It took time until mobility between jobs, changes of the product mix and technology, adult training and other forms of adjustment could restore (if at all) the optimum condition for mutually gainful employment without causing massive unemployment in between.[2] Therefore, we stick to the assumption that in the period under examination the spike at the MW was explained by under-reporting *and* the continuing employment of many low-productivity workers – two different processes that we try to model following the DH approach.

## 3. The minimum wage and the wage distribution in Hungary

MW regulations had minor impact on the Hungarian wage distribution until the millennium.[3] As shown in Figure 1, the MW-average wage ratio slightly decreased in 1992-2000 and fell short of Spain's, the laggard within the EU in that period. The fraction of workers paid 95-105 per cent of the minimum amounted to 5 per cent, a ratio similar to those reported by Dolado et al. (1996) for Austria, Belgium, the Netherlands, Denmark, and the US.

In 2001–2002 the MW was nearly doubled in nominal terms, resulting in a 14 percentage point rise in the Kaitz-index.[4] The fraction of private sector employees earning near the MW jumped to 11 per cent in 2001 and 18 per cent in 2002.

Figure 1

The wage distribution preserved its distorted shape until 2007, when a second spike appeared at 200 per cent of the MW, as shown in Figure 2.[5] That year, the Hungarian government introduced a minimum social security contribution base amounting to 2MW. Firms were allowed to pay wages lower than 2MW but in case they did so they faced an increased probabil-

---

[2] Independent studies by Halpern et al. (2004) and Kertesi and Köllő (2003) estimated the short-run aggregate disemployment effect of the first MW hike to fall to the range of 1-1.5 per cent in 2001.
[3] See Appendix 2 for further details of Hungary's MW regulations.
[4] The MW increased from Ft 25,500 in 2000 to Ft 40,000 on January 1, 2001 and Ft 50,000 on January 1, 2002. See Kertesi and Köllő (2003) on the motives and aftermaths of the large hikes.
[5] At the same time further minima were introduced for young and older skilled workers (1.2MW, 1.25MW) that flattened the spike near the MW.

ity of tax authority audit and a higher risk of being detected as cheaters (for paying disguised MW or for other reasons).[6]

Figure 2

The suspicion that the crowding of workers at the MW in 2001-2006 was partly explained by wage under-reporting is difficult to avert. In 2006, the fraction of MW earners amounted to 18 per cent among small firm managers, and close to 10 per cent among top managers in larger firms also earned the MW. High shares could be observed in a number of freelance occupations such as architects, lawyers, accountants, business and tax advisors, agents, brokers, artists, writers, film-makers, actors and musicians (13-17 per cent). The fraction was particularly high in those sectors, where cash transactions with customers frequently occur such as shops, hotels and restaurants (23 per cent), house building (21 per cent), personal services (18 per cent) and farming (21 per cent). In some low-wage occupations such as cleaners, porters and guards the fraction earning the MW fell short of the above-mentioned levels (Table 1).

Table 1

Further doubts arise if we look at the wage distribution within occupations (Figure 3). In 2006, the distribution for unskilled workers was strongly skewed at the MW with a small number of workers earning substantially more than that. By contrast, the wage distribution of managers, for instance, had a spike at the MW and another at 440 per cent of the MW, clearly pointing to a minority of managers under-reporting their earnings.

Figure 3

With the help of the double hurdle model we can utilize the information content of the different shapes of the wage distributions. In the next section we summarize how the estimation proceeds, how the probability of under-reporting and the MW earners' 'genuine' wages are derived, and how we classify workers and firms as cheaters or non-cheaters.

## 4. The double hurdle model

---

[6] Similar minimum contribution levels were introduced in Bulgaria and Croatia in 2003. The Hungarian regulations remained in effect until January 2010.

*4.1. The set-up of the model*

Let us use the notation $y$ for the (normalized) logarithm of the "true" wage, i.e. of the wage which would prevail in the absence of MW and under-reporting. (We normalize $y$ to be zero at the true MW.) The value of $y$ is determined by some characteristics $X$ of the employee and the firm, and we assume that its distribution is conditionally normal with expectation $X\beta$ and variance $\sigma^2$. (This is a standard assumption in the literature; see e.g. Meyer and Wise 1983a and 1983b.) In the presence of MW and under-reporting, a spike appears at the MW in the wage distribution. The observed wage (the logarithm of which – normalized again to be zero at the MW – will be denoted by $y*$) may be equal to the MW for two reasons: because of con-straints and costs preventing firms from firing low-productivity workers (in the simplest case those whose genuine wage would fall below the MW), or because of tax fraud (when the MW is reported to the authorities but an unobserved cash supplement is also given). The probabil-ity of cheating is determined by some characteristics $Z$ of the employee and the firm, and $X$ may be different from $Z$. Formally, omitting subscript $i$ for the individual, the following mod-el governs $y$ and $y*$:

(1)     $y = X\beta + u$,

and we observe the reported log-wage $y*$ according to the rule:

(2)     $y* = \begin{cases} y & \text{if } X\beta + u > 0 \text{ and } Z\gamma + v > 0 \\ 0 & \text{otherwise} \end{cases}$.

Under-reporting occurs when both $X\beta+u>0$ and $Z\gamma+v\leq0$ hold, and in this case the observed wage is equal to the MW. The residuals $u$ and $v$ are zero-mean normally distributed, possibly correlated ($\rho$) random variables. $\sigma^2$ stands for the variance of $u$ while the variance of $v$ is set equal to unity without loss of generality, hence the covariance matrix of $(u, v)$ is given by:

(3)     $S = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & 1 \end{pmatrix}$.

This is the double hurdle model first proposed by Cragg (1971), with the restriction $\rho=0$, to model the purchase of consumer goods in a setting where the decision to buy and the decision of how much to buy are governed by different processes. The name of the model comes from the fact that the spike of the distribution (in our case at the MW) is determined by two "hurdles": a standard tobit-type constraint (in our case following from the wage equation: $X\beta+u\leq0$) and a different second hurdle (following from the selection equation: $Z\gamma+v\leq0$). Note that the standard tobit model is obtained as a special case when the second hurdle is not effective, e.g. when $Z$ contains a sufficiently large constant and all other terms in $\gamma$ are zero, or when $X=Z$, $\beta=\gamma$ (apart from a constant), $\rho=1$ and $\sigma=1$. In our case, a second hurdle is needed because under-reporting and wage determination are governed by partly different processes.

Since the paper of Cragg the model and its extensions have been widely used to analyze consumer and producer behavior as well as problems in environmental and agricultural economics and banking (e.g. Labeaga 1999, Martinez-Espineira 2006, Moffatt 2005, Saz-Salazar and Rausell-Köster 2006, Teklewold et al. 2006). However, to our knowledge, only Shelkova (2008) used the model to analyze wage distributions, in a setting discussed earlier.

Figure 4

In our application, the baseline DH model (1)-(3) has to be slightly modified in order to better capture the features of the wage formation process. The first problem to be addressed is that the log wage distribution is *not* censored normal because of the crowding of wage earners just above the MW [7] (see Panel A in Figure 4). While at and above the median the distribution is close to the normal we have more workers on the left tail than expected under normality. This poses a problem because – as usual for nonlinear models – maximum likelihood estimation of the DH model yields consistent results only if the underlying distributions are well-specified. Therefore we apply a preliminary transformation that is roughly linear at higher wages and accounts for 'crowding' at lower wages. We assume that instead of $y^*$ we observe $g(y^*)$, where $r$ is a coefficient to be determined:

(4) $\qquad g(x) = x + r \cdot \exp(-x / r) - r \quad \text{if} \ \ x \geq 0.$

---

[7] This is explained by spillover effect as argued in Dickens et al. (1994) and elsewhere.

By the preliminary transformation $g^{-1}$ we can ensure that $y^*$ is close to a (censored) normal distribution and hence the DH model can be applied. Our approach is in line with the double hurdle literature, where a preliminary transformation is often needed to achieve normality: Martinez-Espineira (2006) and Moffatt (2005) use the Box-Cox, while Yen and Jones (1997) apply the inverse hyperbolic sine transformation.

The second possible problem concerns our assumption that cheating employers report the MW (and not a larger wage) to the authorities. This is a reasonable assumption for 2001-2006 because firms could maximize the evaded tax this way and the chance of tax audit was not increased for MW-reporting firms before 2007. The model can be extended to allow for cheating above the MW (see Elek et al. 2009) but external (e.g. survey-based) information is needed to identify its parameters. In this paper, we use the simpler formulation.

*4.2. Parameter estimation*

First, the parameter $r$ of the preliminary transformation (4) should be determined. Instead of a likelihood-based statistical procedure, we make use of the fact that the wage distribution was close to lognormal in 2000 (see Figure 2), changed substantially because of the MW increase and spillover effects in 2001-2002, and – in the absence of further drastic MW hikes – was practically unaltered in 2003-2006. Thus we create a quasi panel subsample of the LEED data for 2000-2002, and assign the median of the 2002 logarithmic wages (normalized to be zero at the MW) to the median of the 2000 logarithmic wages (again normalized) for each percentile of the wage distribution in 2000. (See section 5 for details of the LEED data set.) This graph hence shows the change in (normalized) logarithmic wages between 2000 and 2002 by percentiles. Finally, the function $g$ (with unknown parameter $r$) is fitted to the percentile graph with nonlinear least squares. This function gives a transformation for the log-wages in 2002 and – for the reasons mentioned above – for 2006 as well.

Our method yields $r=0.49$. Figure 5 displays the function and its appropriate fit to the 2000-2002 wage percentiles, while Panel B in Figure 4 shows that the transformed log wages ($g^{-1}(y^*)$) are approximately censored normal. For ease of notation, in what follows, we refer to $g^{-1}(y^*)$ as $y^*$.

Figure 5

Using the properties of the conditional distributions of the bivariate normal distribution, the likelihood function of the DH model (1)-(3) can be shown to have the following form (for the sake of clarity, here we use subscripts *i* for the individuals):

$$(5) \qquad L = \prod_{y_i^*=0} \left[1 - \Phi_{\rho,\sigma,1}\left(x_i\beta, z_i\gamma\right)\right] \cdot \prod_{y_i^*>0} \left[ \Phi\left( \frac{z_i\gamma + \frac{\rho}{\sigma}\left(y_i^* - x_i\beta\right)}{\sqrt{1-\rho^2}} \right) \frac{1}{\sigma} \phi\left( \frac{y_i^* - x_i\beta}{\sigma} \right) \right],$$

where $\Phi_{\rho,\sigma,1}$ denotes the bivariate normal distribution with covariance matrix given in (3), while $\Phi$ and $\phi$ stand for the univariate standard normal distribution and density, respectively. Parameter estimation can be carried out with maximum likelihood, where we use cluster-robust standard errors to tackle the potential within-firm correlation in the error terms.

If the DH model is correctly specified (including the distributional assumptions), then identification can be carried out even if *X=Z*, i.e. based merely on nonlinearities. However, to make the results more robust to deviations from the distributional assumptions, it is worth including variables that only influence the selection equation but not the wage equation (i.e. making valid exclusion restrictions). Therefore, in the wage equation we include the usual variables thought of as influencing the productivity of a worker such as her individual characteristics (experience, education, sex) and the characteristics of her firm (industry, productivity, fixed assets, location, size and ownership).[8] Since the majority of these variables affect cheating behavior as well, they are also present in the selection equation. (e.g. for larger firms it is more difficult to hide envelope wages from the tax authority thus they tend to be less involved in grey employment.)

We also include individual and firm-level proxies directly affecting the decision to evade taxes. In particular, we distinguish some occupational categories that are more prone to cheating than others, mainly due to the lower risk of being caught such as managerial and freelance occupations, occupations with frequent cash transactions or jobs in trade, hotels and restaurants (see Table 2 and Appendix Table A1 for definitions). We also choose proxies for tax

---

[8] For robustness check, in an alternative specification we use occupation dummies instead of industry dummies in the wage equation.

evasion from the corporate tax returns. It is expected that wage-underreporting firms tend to evade corporate taxation, thus tax liability correlates negatively with cheating. Another proxy is "other personnel related expenses" which contain fringe benefits: these are rather complementary to wage payments hence a high share of personnel related costs indicates compliance to the tax rules. The chosen indicators are indicative of compliance with the tax rules in fields other than wage payment. It is reasonable to assume that, after controlling for the usual factors in the wage equation, the firm-level instruments only influence the probability of cheating but not the genuine wages thus we have valid exclusion restrictions in the model.

*4.3. Under-reporting probabilities, 'genuine' wages and classification of workers and firms*

In the possession of the DH parameters the probability of cheating for each MW earner can be estimated as:

(6)
$$
\begin{aligned}
\text{P(underreporting)} &= P(X\beta + u > 0, Z\gamma + v \le 0 \mid y^* = 0) = \\
&= \frac{P(u > -X\beta) - P(u > -X\beta, v > -Z\gamma)}{1 - P(u > -X\beta, v > -Z\gamma)} = \\
&= \frac{\Phi(X\beta/\sigma) - \Phi_{\rho,\sigma,1}(X\beta, Z\gamma)}{1 - \Phi_{\rho,\sigma,1}(X\beta, Z\gamma)}
\end{aligned}
$$

Also, we can simulate the genuine wage of each MW earner as follows. We generate independent copies of bivariate normal random variables ($u$, $v$) with covariance matrix given in (3), and accept max ($X\beta+u$, 0) as the normalized genuine log-wage of an MW earner if $X\beta+u \le 0$ or $Z\gamma+v \le 0$. If none of these conditions hold, the person cannot earn MW according to the model. Technically, for each MW-earner, the ($u,v$) variables are simulated until at least one condition holds.

Let us denote the estimated probability of under-reporting by a MW earner with $P$ and the simulated wage with $w$ (i.e. $w=MW*\exp(g(y))$ ). As a benchmark definition cheating behavior is assumed in case of $P>0.5$, but $w>MW$ and $w>1.5MW$ will also be used for robustness checks.[9] If we find at least one MW earner classified as "cheater" in a firm we treat the firm as a cheater. Since the majority of cheating firms are small, the use of other, more advanced

---

[9] The definition $P>0.5$ is preferable to e.g. $w>MW$ because the latter includes some extra simulation uncertainties.

criteria such as a certain threshold for the ratio of cheaters would be of limited practical importance.

## 5. Data

Throughout the paper we rely on the Wage Survey (WS) of the National Employment Service. The WS is a linked employer-employee data set recently comprising observations on over 150,000 individuals in about 20,000 firms and budget institutions. The survey was carried out tri-annually until 1992 and annually since then. In the enterprise sector the WS covers businesses employing at least 5 workers. All Hungarian firms employing more than 20 workers are obliged to report data for the WS while smaller firms are randomly selected from the census of enterprises. In the years considered in our paper, firms employing 5-20 workers had to report individual data on each employee while larger ones reported data on a (roughly 10 per cent) random sample of their workers, selected on the basis of their day of birth. The observations are weighted by the Employment Service to correct for the selection of firms and individuals. The survey contains information on the wages and demographic and human capital variables of the workers and their job characteristics. The firm-level variables comprise industry, region, firm size, location, ownership, union coverage and financial variables including sales revenues, the net value of fixed assets, average wages, profits and several cost items. Our estimation sample covers the private sector and comprises 92,140 observations.

In section 7, we use panels of individual and firm-level observations. Firms in the WS can be directly linked and followed over time. Individuals cannot be linked directly but they can be identified across waves with acceptable precision using data on their firm identifier, location of their workplace, year of birth, gender, education and four-digit occupational code. The worker and firm panels are non-randomly selected from the base-period (2006) populations because of the survey design, on the one hand, and group-specific differences in firm survival, job destruction and quits, on the other. We control for selection on observables by estimating probit equations and using the inverse of the predicted probabilities of being in the panel as weights in those models, where weighting is allowed (for the method used see e.g. Moffit et al. 1999). The probits are presented in Tables A2 and A3 of Appendix 1. While the probability of making it to the panel was clearly non-random, weighting still had negligible effect on the estimated parameters.

## 6. Results of the double hurdle model

Table 2 presents the parameter estimates of the DH model. The parameters of the selection equation largely conform to intuition. 'Grey' occupations, male workers and employees in Budapest tend to under-report wages significantly, while foreign ownership, firm size, higher corporate tax liability and larger 'other personnel related expenses' of the firm are positively correlated with labor tax compliance. After controlling for other factors, education does not seem to have a direct effect on cheating. The correlation between the error terms ($\rho$) is significantly negative, implying that unobserved factors leading to higher genuine wages tend to increase the probability of cheating. Similar results are obtained in the alternative specification, when occupation dummies (defined in Table A1) are used instead of industry dummies in the wage equation.

Table 2

Using the estimated parameters, the probability of under-reporting among MW earners and their genuine wages was calculated. The results suggest that around half of all workers paid the MW hid part of their earnings from the tax authority. We estimate that the average „genuine" wage of the MW earners amounted to approximately 170 per cent of the MW and the average wage of cheating MW employees (using $w$>MW as the criterion for cheating) was around 250 per cent of MW. We should note that the exact share of cheaters and their simulated genuine wages are quite sensitive to the parameter $r$ of the preliminary transformation but – more importantly from a modeling point of view – the partial effects of the different factors (occupations etc.) are robust across different specifications.

Table 3

Table 3 displays the estimated probability of under-reporting among MW earners, their average genuine wage and a "cheating indicator" by occupation, industry and firm size for the two different specifications.[10] (The cheating indicator is defined as the share of cheating MW earners among *all* employees.) Looking at *occupations*, the estimated fraction of cheaters among MW earners is small for cleaners (10-20 per cent), unskilled laborers and agricultural

---

[10] One is the baseline specification containing industry dummies in the wage equation, while the other contains occupation dummies instead.

workers (20-30 per cent), while it is much larger than average e.g. for drivers and approaches 100 per cent for managers and professionals. It is also clear that the share of MW earners is not a good indicator of under-reporting because fraud is relatively frequent for some occupations with a high share of MW earners (e.g. in construction), while infrequent for others (e.g. among cleaners, unskilled laborers). The cheating indicator, which is the product of these two terms, is substantially higher than average in construction and trade professions and among drivers.

As far as *firm characteristics* are concerned, Table 3 also displays the relation of economic branch and firm size to under-reporting. The cheating indicator is higher than average in construction, trade and hotels and restaurants, while it is the lowest in financial services (where the share of MW earners is the smallest as well). Both the ratio of MW earners and cheating behavior are strongly negatively correlated with firm size: the cheating indicator is ten times higher for firms with 5-10 employees than for larger firms with more than 50 employees. Foreign-owned enterprises tend to employ much less workers at the MW than domestic and mixed ones but the ratio of under-reporting among them does not differ substantially.

The proportion of cheating *firms* (i.e. firms with at least one cheating employee) amounted to 17.3 per cent of all firms and 37.0 per cent among enterprises having at least one MW earner. While the estimates confirm that, in 2006, envelope wages existed at a large scale, they suggest that more than half of the MW earners did not receive cash supplement and the majority of firms paying MW did not cheat on taxes. However, for reasons discussed earlier the estimates should be treated with caution and the model's predictive power needs to be checked.

## 7. Testing the predictions of the DH: responses to the introduction of a minimum contribution base

As was briefly discussed earlier, the 2007 reform created incentives to raise the reported wages of disguised MW earners. Cheating firms could fully avert the risk of audit by officially paying 2MW or more to their grey employees instead of MW. Furthermore, the public debate preceding the reform gave a clear warning that the tax authority would treat MW payment as a signal of tax evasion. Therefore, cheating firms had stronger motivation to shift their grey employees away from the MW while non-cheating enterprises, in the position to demonstrate that they pay 'genuine' MWs, had less incentive to raise the wages of their MW earners.

The sudden shift of the spike of the wage distribution from MW to 2MW in 2007 (shown earlier by Figure 2) clearly indicated that firms – especially smaller ones – considered tax audit a credible threat. Before 2007, tax inspections were rather lax in Hungary. While firms employing more than 50 workers were checked by independent auditors and/or the tax authority annually and the monitoring activities of the tax authority concentrated on "accentuated tax payers" (companies having the largest tax liabilities), entities without legal personality were monitored only in every 7th year and individual entrepreneurs only in every 23rd year on average. Penalties were insignificant. Consistent with the reform's intentions, the new regulation changed the wage distribution of small firms dramatically while larger firms were weakly affected.[11]

We check how the wages of grey employees changed in response to the reform by estimating a probit equation (7) for a quasi-panel of individuals earning the MW in May 2006 and also observed in May 2007:

$$(7) \qquad P(w_1^* = 2MW_1 \big| w_0^* = MW_0) = \Phi(C\beta + Z\gamma)$$

In the equation, C denotes the dummy for cheating, *Z* comprises worker and firm characteristics, and $MW_0$ and $MW_1$ stand for the minimum wage in 2006 and 2007. Base period MW earners are defined as those earning the exact amount of the minimum and those earning 95-105 per cent of the minimum, alternatively. The expectation is that $\beta > 0$.

We use fraud indicators defined on the *individual* level since the reform affected only the fake minimum wage earners within firms: by shifting these particular employees away from the MW the enterprise could reduce the risk of audit.

The cheating proxies in equation (7) come from the DH model hence they are predicted regressors and the estimation of their effect by simple maximum likelihood would not yield valid results. Therefore, in calculating the standard errors in the equation we follow a two-step procedure. First, we simulate the parameter vector of the DH model from its asymptotic nor-

---

[11] The reform was initiated by a high (close to 10 per cent) budget deficit in 2006, and might be regarded as a simple form of presumptive taxation. For a discussion of the idea of presumptive taxation, practices in Italy, and an application to Bulgaria see Tanzi and Casanegra de Jantscher (1987), Arachi and Santoro (2007) and Pashev (2006), respectively.

mal distribution with its variance matrix, and create 100 simulated draws of firm-level cheating variables from the models. Second, using the different cheater classifications, we estimate equation (7) by bootstrap and finally take the sample mean and standard deviation of all simulated parameters. This way, the cumulated parameter uncertainty of the two stages is quantified – by using the asymptotic variance matrix in the first stage and direct bootstrap in the second. For simplicity, in the following text and tables we refer to this procedure as "two-step bootstrap". Note also that the resulting standard errors are only about 5 per cent larger than the ML standard errors of equation (7) because the error of the DH model, based on nearly 100 thousand observations, is negligible compared to the error of the test equation.

Table 4

The descriptive statistics in Table 4 yield preliminary support to our hypothesis: cheating enterprises were more likely to move their (apparent) low-wage workers away from the MW and shift them to 2MW than non-cheating firms. The MW earners (as of 2006) employed by fraudulent firms were 40 per cent less likely to earn the MW in 2007, 2.3 times more likely to earn 2MW, and 2.2 times more likely to earn 2MW or more.

The results from equation (7) are presented in Table 5. As shown in the first row, base-period MW earners classified as cheaters (victims of cheating) were significantly more likely to earn 2MW in 2007 than non-cheating MW earners. The estimated marginal effect of being a cheater amounts to 2.4 per cent when all controls are included – a remarkable impact if we take into account that the probability of earning $2MW_1$ in 2007 conditional on earning $MW_0$ in 2006 amounted to approximately 13 percent. In the second row of the table an alternative to equation (7) estimates the probability that a $MW_0$ earner was shifted to or beyond $2MW_1$ i.e. the worker was moved out of the 'danger zone'. The partial effects are positive and significant but lower.

Table 5

We may try to assess the magnitude of change induced by the 2007 reform and evaluate its economic significance in two ways. First, one can make back-on-the-envelope calculations relying on the results in Table 4, and taking into consideration that the share of MW earners amounted to 47.9 per cent in cheating firms and 5.7 per cent in non-cheating ones. This im-

plies that the reported wages of 6.4 per cent and 0.3 per cent of the employees were doubled in the two groups of firms, respectively.[12] Holding other wages constant these pay rises implied 6.3 and 0.3 per cent increase in the average reported wages, respectively.

Second, one may try to estimate the effect of cheating behavior on firm-level outcomes by estimating regressions of the form:

$$(8) \quad \Delta \ln x = \beta C + \mathbf{Z}\gamma + \varepsilon$$

where $\Delta \ln x$ stands for the log changes in average wages, sales revenues and employment, alternatively, while $C$ and $Z$ denote the firm-level cheater dummy and the controls, respectively. The equations are estimated for 5230 firms observed in 2006 and 2007, and the standard errors are estimated with the two-step bootstrap procedure described earlier. In the wage equation we expect $\beta > 0$ since raising the reported wages of grey employees must have increased the average reported wages of the cheating firms to some extent. The question of how actual costs and, therefore, output and employment were affected is more difficult to answer *a priori*. First, firms may have cut the cash payments of the affected workers, offsetting the impact of increased payroll taxes. Second, some of them may have increased the share of cash transactions in order to economize on VAT instead of payroll taxes.

Table 6

The results presented in Table 6 suggest that the firm-level cheating proxy had positive effect on the change of observed average wages. Reported wages grew faster by 12 percentage points after controlling for industry, region, firm size, ownership and skill composition. The estimated gap between honest and dishonest firms is larger than the 6 percentage points difference calculated beforehand. This may result from the effect of the reform on other reported wages, or from unobserved shocks, for which we can not effectively control with the firm-level variables at hand.

---

[12] Recall that cheaters shifted 13.4 per cent of their MW earners to 2MW while the respective share was only 5.9 per cent with non-cheaters.

In either case, the budgetary effect of the reform seems modest. According to the DH estimates, approximately 170,000 workers, or 11 percent of the labor force represented by the Wage Survey, were employed by cheating firms. The average wages of these firms equaled 1.3 times the MW. Starting from these data and considering that the combined (employer and employee) social security contribution rate was 49 per cent and the lowest personal income tax rate was 18 per cent, we can estimate that the excess increase of reported wages in fraudulent firms resulted in an extra revenue of 12 billion Ft, or about 0.05 per cent of GDP. If we accept the back-on-the-envelope calculations, the budgetary effect is proportionally lower (about 6 billion Ft).

The results indicate a significant negative effect on sales revenues and no effect on employment. A possible interpretation of this result is that the 2007 reform directed cheating enterprises to alternative forms of tax evasion and/or urged them to cut envelop wages.

The results presented in this section proved robust to changes in the definition of cheating and specification of the individual and firm level regressions. Weighting had practically no impact on the parameters. Using the exact amounts of the MWs rather than brackets around them left the qualitative results unchanged in the individual regressions. We also examined the sensitivity of results to alternative cheating indicators based on the simulated wage ($w$>MW, $w$ >1.1MW, $w$ >1.5MW and $w$ >2MW). Since there was no significant deviation from the presented results, the regressions using alternative indicators are not presented.

## 7. Conclusions

While grey employment and disguised MWs are widely debated issues in many emerging market economies, few attempts have been made to measure their magnitude and distribution. We applied a double hurdle model to this issue for Hungary in a period in which the presumptions of the model seemed to fit i.e. censoring at the MW and wage under-reporting (at the MW) occurred simultaneously. If these preconditions are met, a properly specified DH model can estimate the 'genuine' wage distribution, permits the calculation of cheating probabilities and allows the simulation of 'true' earnings.

The DH results for 2006 suggest that employers paid cash supplement to around half of all minimum wage employees, and hinted at a wide (150 per cent) gap between reported and ac-

tual wages in these cases. The estimated distribution of under-reporting across occupations, industries and firm size seem to be consistent with the anecdotal evidence and survey-based results. The DH model makes strong assumptions about the wage distribution, and finding variables, which affect selection to cheating without affecting wages, is also rather difficult.

Driven by the resulting uncertainty of the estimates, we conducted an experiment aimed at testing if the DH estimates have predictive power. It seems that the estimates worked well in the quasi-experimental setting analyzed in the paper: firms and workers suspected of tax evasion responded differently to the strong shock under investigation.

We obviously make both type 1 and type 2 errors in disentangling cheaters from non-cheaters but the results are encouraging for the analysis of 'grey employment' and, we believe, they also have practical importance. On the one hand, audits may be targeted by statistical profiles derived from the DH model, thereby improving compliance. However, by showing the loci of under-reporting the DH estimates also draw attention to the limits of tax enforcement. Disguised minimum wages have high shares in services provided to households and small businesses, freelance occupations, and small firm management – an attribute that limits the potential budgetary intakes from more stringent inspection. Cash transactions between households and the providers of personal services are difficult, if not impossible, to detect. Grey transactions of this kind can rather be whitened indirectly, by creating incentives to require receipts and making clear the link between reported income and access to publicly financed services and transfers such as pensions.

On the other hand, the DH results call for more cautious MW policies. The micro-data do not support the popular belief that in Hungary 'millions' are fraudulently paid the minimum wage – an assumption that served as a justification for regulations like the minimum contribution to be paid after 2MW. Reducing the under-reporting of wages by means of substantially increasing the MW and/or the tax burden on it is an undoubtedly cheap alternative to independent checks and carefully designed presumptive taxation. However, raising the costs of low-wage employment across the board is a poorly targeted policy, which can further reduce unskilled job opportunities: an undesirable outcome in a country, where six out of ten low-educated prime-age adults are out of work.

# References

Allingham, M. and A. Sandmo (1972): Income tax evasion: A theoretical analysis. *Journal of Public Economics* 1, 323-338.

Arachi, G. and Santoro, A. (2007): Tax enforcement for SMEs. Lessons from the Italian experience? *eJournal of Tax Research* 5(2), 225–243.

Benedek D., Rigó M., Scharle Á. and Szabó P. (2006): Increases in the minimum wage in Hungary, 2001-2006, *Working Papers in Public Finance* No. 16.

Cragg, J. (1971): Some statistical models for limited dependent variables with application to the demand for durable goods, *Econometrica* 39, 829-844.

Dickens, R., Machin, S. and Manning, A. (1994): The effects of minimum wages on employment: theory and evidence from the UK, *NBER Working Paper* 4742, Cambridge, MA.

DiNardo, J., N.M. Fortin, and T. Lemieux (1996), "Labor market institutions and the distribution of wages, 1973-1992", *Econometrica*, No. 5, Vol. 64: 1001-1044.

Dolado, J., Karamarz, F., Machin, S., Manning, A., Margolis, D. and Teulings, C. (1996): The economic impact of minimum wages in Europe, *Economic Policy* 23, 319-372

Elek, P., Scharle, Á., Szabó, B., Szabó P. A. (2009): Wage-related tax evasion in Hungary, (in Hungarian), *Közpénzügyi Füzetek,* No. 23.

Erdogdu, S. (2009): Turkey – Minimum wage in tension between economic and social concerns, in: Daniel Vaughan-Whitehead (ed.): *The minimum wage in the enlarged EU*, Edward Elgar 2009

European Commission (2007): Undeclared work in the European Union. Report, Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions.

Halpern, L., M. Koren, G. Kőrösi and J. Vincze (2004): A minimálbér költségvetési hatásai. (The budgetary effects of the minimum wage, in Hungarian), *Közgazdasági Szemle* 51(4), 325–345.

Kertesi G. and J. Köllő (2003): Fighting low equilibria by doubling the minimum wage – Hungary's experiment, *IZA Discussion Paper* No. 970, Bonn

Kriz K. A., Meriküll J., Paulus, A. and Staehr, K. (2007): Why do Individuals Evade Payroll and Income Taxation in Estonia? U. of Tartu Economics and Business Administration Working Paper No. 49-2007.

Labeaga, J. M. (1999): A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco, *Journal of Econometrics* 93, 49-72.

Madzharova, B. (2010): The effect of low corporate income tax on social security evasion, Cerge-EI Prague, EPRN Seminars, Univ. of Copenhagen, 2010 Nov 12, http://web.econ.ku.dk/eprn_epru/Seminar/EPRU%20Nov12.Madzharova.pdf

Martinez-Espineira, R. (2006): A Box-Cox double hurdle model of wildlife valuation, the citizen's perspective, *Ecological Economics* 58, 192-208.

Masso, J. and Krillo, K. (2009): Estonia, Latvia and Lithuania – Minimum wages in the context of migration and labour shortages, in: Daniel Vaughan-Whitehead (ed.): *The minimum wage in the enlarged EU*, Edward Elgar 2009.

Meriküll, J. and Staehr, K. (2010): Unreported Employment and Envelope Wages in Mid-Transition: Comparing Developments and Causes in the Baltic Countries, *Comparative Economic Studies* 52 (September), 637–670.

Meyer, R. H. and Wise, D. A. (1983a): Discontinuous distributions and missing persons: the minimum wage and unemployed youth, *Econometrica* 51, 1677-1698.

Meyer, R. H. and Wise, D. A. (1983b): The effects of minimum wage on employment and earnings of youth, *Journal of Labor Economics* 1, 66-100.

Moffatt, P. G. (2005): Hurdle models of loan default, *Journal of the Operational Research Society* 56, 1063-1071.

Moffit, R., Fitzgerald, J. and Gottschalk, P. (1999): Sample Attrition in Panel Data: The Role of Selection on Observables, *Annales d'Économie et de Statistique*. 55-56, 129-152.

Pashev, K. V. (2006): Presumptive taxation: lessons from Bulgaria, *Post-Communist Economies*, 18(4), 399–418.

Saz-Salazar, S. D. and Rausell-Köster, P. (2006): A double-hurdle model of urban green areas valuation: Dealing with zero responses, *Landscape and Urban Planning* 84, 241–251.

Shelkova, N. Y. (2008): Low wage labor markets and the power of suggestion, Working Paper 2008/33, *Department of Economics, University of Connecticut.*

Szabó, P. A. (2007): A 2000–2001. évi minimálbér-emelés hatása a jövedelemeloszlásra (The effect of the 2000-2001 minimum wage increase on income distribution, in Hungarian), *Közgazdasági Szemle* 54 (May), 397–414.

Tanzi, V., Casanegra de Jantscher, M. (1987): Presumptive income taxation: administrative, efficiency, and equity aspects, *IMF Working Paper* WP/87/54, Washington, D.C.

Teklewold, H., Dadi, L., Yami, A. and Dana, N. (2006): Determinants of adoption of poultry technology: a double-hurdle approach, *Livestock Research for Rural Development* 18.

Tonin, M. (2007): Minimum wage and tax evasion: theory and evidence, Institute of Economics, Budapest, DP 2007/1.

Tonin, M. (2011): Minimum wage and tax evasion: theory and evidence, *Journal of Public Economics* 95, 1635-1651.

World Bank (2005): Enhancing Job Opportunities: Eastern Europe and the Former Soviet Union, *World Bank*, Washington.

Yaniv, G. (1988): Withholding and non-withheld tax evasion. *Journal of Public Economics* 35, 183-204.

Yen, S. T. and Jones, A. M. (1997): Household consumption of cheese, an inverse hyperbolic sine double hurdle approach with dependent errors, *American Journal of Agricultural Economics* 79, 246-251.

**Table 1. Fraction paid the exact amount of the minimum wage in 2006**

|  | Per cent paid the MW | Composition All MW earners=100 |
|---|---|---|
| Top managers | 9.7 | 1.6 |
| Managers (heads of department, foremen, etc.) | 3.6 | 2.2 |
| Managers of small firms (5-20 employees) | 18.0 | 1.6 |
| Engineers | 2.4 | 0.6 |
| Architects and construction technicians | 9.5 | 0.3 |
| Professionals in health, education and social services (private) | 3.5 | 0.0 |
| Other professionals | 3.0 | 0.5 |
| Lawyers, business and tax advisors, accountants | 8.8 | 0.7 |
| Freelance cultural occupations (musicians, actors, writers etc.) | 16.5 | 0.5 |
| Technicians | 7.3 | 2.7 |
| Administrators | 8.3 | 6.8 |
| Agents, brokers | 12.6 | 0.8 |
| Office workers | 11.5 | 5.6 |
| Blue collars in retail trade and catering | 22.5 | 15.3 |
| Blue collars in transport | 7.7 | 0.1 |
| Services A (other than B and C) | 12.7 | 1.5 |
| Services B (health and social services, private) | 0.0 | 0.0 |
| Services C (personal services) | 17.7 | 0.7 |
| Farmers and farm workers | 20.9 | 5.1 |
| Blue collars in heavy industry and engineering | 8.9 | 6.6 |
| Blue collars in light industry | 14.6 | 9.2 |
| Blue collars in construction (house building) | 21.0 | 10.0 |
| Blue collars in civil engineering (roads, railways, bridges) | 20.0 | 0.6 |
| Assemblers and machine operators | 4.7 | 4.3 |
| Truck drivers | 20.5 | 3.8 |
| Porters, guards, cleaners | 18.2 | 7.6 |
| Unskilled laborers, casual workers | 37.6 | 11.2 |
| Total | 10.8 | 100.0 |

Source: Wage Survey, 2006, estimation sample of the DH model. Number of observations = 91,240
Note: For this table some occupations were divided into parts on the basis of industrial affiliation and firm size in order to capture differences in the scope for cash transactions with customers (personal versus other types of services, small firm versus large firm managers).

## Table 2. DH estimates of wages for 2006

|  | Coefficient | St. error[a] |
|---|---|---|
| **Wage equation for normalized log wages** (also includes industry controls) | | |
| Experience / 10 | 0.327*** | 0.013 |
| Exp squared / 100 | -0.049*** | 0.002 |
| Male | 0.205*** | 0.011 |
| Vocational edu. | 0.183*** | 0.013 |
| Secondary edu. | 0.485*** | 0.015 |
| Higher edu. | 1.191*** | 0.019 |
| Budapest | 0.135*** | 0.023 |
| Value added per worker log | 0.147*** | 0.010 |
| Fixed assets per worker log | 0.007 | 0.005 |
| Firm of foreign ownership | 0.255*** | 0.020 |
| Firm with 5-10 employees | -0.404*** | 0.037 |
| Firm with 11-20 employees | -0.371*** | 0.026 |
| Firm with 21-50 employees | -0.233*** | 0.024 |
| Firm with 51-300 employees | -0.112*** | 0.020 |
| Constant | 0.255*** | 0.020 |
| | | |
| **Selection equation** | | |
| Experience / 10 | -0.408*** | 0.070 |
| Exp squared / 100 | 0.108*** | 0.015 |
| Male | -0.214*** | 0.050 |
| Vocational edu. | 0.050 | 0.138 |
| Secondary edu. | -0.128 | 0.127 |
| Higher edu. | -0.012 | 0.136 |
| Managerial and freelance[b] | -0.392** | 0.161 |
| Cash transactions[c] | -0.226** | 0.113 |
| Retail trade[d] | -0.333*** | 0.100 |
| Budapest | -0.244** | 0.123 |
| Works in a city | 0.112 | 0.091 |
| Works in a village | -0.133 | 0.105 |
| Corporate tax payment / sales revenues | 10.03** | 4.00 |
| Other personnel related expenses /payroll | 2.261*** | 0.844 |
| Firm of foreign ownership | 0.778*** | 0.187 |
| Firm with 5-10 employees | -2.007*** | 0.272 |
| Firm with 11-20 employees | -1.709*** | 0.264 |
| Firm with 21-50 employees | -1.404*** | 0.270 |
| Firm with 51-300 employees | -0.930*** | 0.270 |
| Constant | 3.074*** | 0.315 |
| | | |
| **Rho** | -0.302*** | 0.047 |
| **Sigma** | 0.547*** | 0.008 |
| N of observations | 91,240 | |

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

a) Cluster robust standard errors, adjusted for firm-level clustering

b) Managerial and freelance occupations (the latter includes professionals in culture and arts, agents and brokers)

c) Occupations where cash transactions occur frequently. Includes car mechanics, electricians, plumbers, household employees, couriers, truck drivers and workers in personal services and house building (see Table A1).

d) Occupations in retail trade (see Table A1)

Data source: Wage Surveys 2006, private sector

## Table 3. Predictions of the DH model for 2006

| | Probability of under-reporting among MW earners (per cent) | | Share of MW earners (per cent) | Cheating indicator (per cent)[b] | | Simulated wage of cheaters (MW = 1.0) | |
|---|---|---|---|---|---|---|---|
| | (1)[a] | (2)[a] | | (1)[a] | (2)[a] | (1)[a] | (2)[a] |
| **Total** | 46 | 48 | 11.9 | 5.5 | 5.7 | 2.6 | 2.4 |
| **Occupations**[c] | | | | | | | |
| Agriculture | 31 | 29 | 27.5 | 8.6 | 7.9 | 2.3 | 1.8 |
| Construction | 45 | 56 | 23.4 | 10.6 | 13.0 | 1.9 | 1.8 |
| Services | 40 | 43 | 6.7 | 2.7 | 2.9 | 2.4 | 2.2 |
| Trade | 52 | 39 | 20.5 | 10.6 | 7.9 | 2.2 | 1.8 |
| Industry | 41 | 49 | 12.8 | 5.3 | 6.3 | 2.1 | 2.0 |
| *Other blue collar* | | | | | | | |
| Cleaners | 18 | 13 | 23.8 | 4.2 | 3.2 | 2.5 | 1.6 |
| Unskilled laborers | 30 | 22 | 33.3 | 10.1 | 7.5 | 2.0 | 1.6 |
| Machine operators | 35 | 45 | 5.7 | 2.0 | 2.5 | 2.1 | 2.1 |
| Porters and guards | 38 | 24 | 15.6 | 5.9 | 3.7 | 2.5 | 1.6 |
| Drivers | 59 | 72 | 15.8 | 9.3 | 11.4 | 2.2 | 2.2 |
| *White collar* | | | | | | | |
| Office clerks | 52 | 59 | 11.0 | 5.7 | 6.5 | 2.7 | 2.4 |
| Technicians. assistants | 72 | 84 | 5.3 | 3.8 | 4.4 | 3.2 | 2.8 |
| Administrators | 64 | 78 | 6.4 | 4.1 | 5.0 | 3.1 | 2.9 |
| Managers | 74 | 96 | 5.2 | 3.8 | 5.0 | 3.7 | 4.4 |
| Professionals | 94 | 97 | 2.5 | 2.4 | 2.4 | 5.1 | 5.1 |
| **Industries** | | | | | | | |
| Agricult., fishing | 34 | 37 | 15.9 | 5.4 | 5.9 | 2.4 | 2.5 |
| Manufacturing | 40 | 43 | 7.5 | 3.0 | 3.2 | 2.6 | 2.4 |
| Construction | 42 | 49 | 27.9 | 11.6 | 13.8 | 2.1 | 2.0 |
| Trade | 52 | 57 | 17.4 | 9.1 | 9.9 | 2.4 | 2.4 |
| Hotels. restaurants | 40 | 36 | 22.2 | 8.8 | 7.9 | 2.2 | 2.3 |
| Transport | 68 | 61 | 6.3 | 4.3 | 3.8 | 4.0 | 2.5 |
| Financial services | 72 | 35 | 2.4 | 1.7 | 0.8 | 3.6 | 3.6 |
| Real estate. business activities | 51 | 47 | 12.0 | 6.2 | 5.6 | 3.1 | 2.8 |
| Other | 49 | 42 | 8.1 | 4.0 | 3.4 | 2.8 | 2.3 |
| **Firm size** | | | | | | | |
| 5-10 employees | 58 | 60 | 32.3 | 18.9 | 19.5 | 2.2 | 2.1 |
| 11-20 employees | 50 | 52 | 23.3 | 11.6 | 12.2 | 2.3 | 2.2 |
| 21-50 employees | 44 | 46 | 14.1 | 6.2 | 6.6 | 2.7 | 2.5 |
| 51-300 employees | 30 | 36 | 6.9 | 2.1 | 2.5 | 3.1 | 3.1 |
| 300 + employees | 7 | 7 | 1.0 | 0.1 | 0.1 | 5.8 | 7.7 |

a) Models: wage equation with (1) industry dummies; (2) with occupation dummies
b) Cheating indicator: share of cheating MW earners among all employees
c) Occupations: see Table A1 in Appendix
Data source: Wage Survey 2006. Number of observations: 91,240

**Table 4: The wages of year 2006 MW earners in 2007**

| Wage in 2007 | Earned the MW in 2006 and estimated to be | |
| --- | --- | --- |
| | non-cheater (per cent) | cheater (per cent) |
| MW | 23.2 | 14.3 |
| Between MW and 2MW | 70.0 | 71.0 |
| 2MW | 5.9 | 13.4 |
| Above 2MW | 0.9 | 1.3 |
| Total | 100.0 | 100.0 |

Source: Wage Survey, MW earners in the worker panel of 2006-2007, Number of observations 3,940

**Table 5. The effect of estimated cheating behavior[a] on wage adjustment between May 2006 and May 2007**

*Probit marginal effects at the sample means*

| | Controls[b] | | | | | | Number of observations |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | No | | Education | | All | | |
| Model[c] | Partial effect | Z-value[d] | Partial effect | Z-value[d] | Partial effect | Z-value[d] | |
| probit1 | 0.072 | 9.511*** | 0.049 | 6.263*** | 0.024 | 3.580*** | 3940 |
| probit2 | 0.049 | 8.219*** | 0.026 | 4.262*** | 0.009 | 2.124** | 22996 |

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

a) Individuals suspected of cheating in 2006 on the basis of the DH model

b) Controls (all variables relate to 2006): dummies for education (college graduate, secondary school and vocational school), work experience in years, dummies for gender, municipality, and the logarithm of firm size.

c) At probit1, the dependent variable is $P(w_1^* = 2MW_1 \mid w_0^* = MW_0)$, at probit2 $P(w_1^* \geq 2MW_1 \mid w_0^* = MW_0)$.

d) Based on two-step bootstrap standard errors, adjusted for clustering by firms

Data source: Wage survey, MW earners in the worker panel of 2006-2007

**Table 6. The effects of estimated cheating behavior[a] on the changes of selected firm-level indicators in 2006-2007**

*OLS regressions*

| | Controls[b] | Partial effect | Z- value[c] | Number of obs. |
| --- | --- | --- | --- | --- |
| Change of average wage (log) | No | 0.1294 | 14.37*** | 5230 |
| | Yes | 0.1146 | 11.41*** | |
| Change of employ-ment (log) | No | 0.0073 | 1.02 | 5230 |
| | Yes | 0.0048 | 0.79 | |
| Change of sales reve-nues (log) | No | -0.0454 | -3.21** | 4824 |
| | Yes | -0.0352 | -2.33** | |

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

a) Firms suspected of cheating in 2006 on the basis of the DH model

b) Controls include skill shares, average wage, average age and dummies for sectors, regions, type of munici-pality and state ownership.

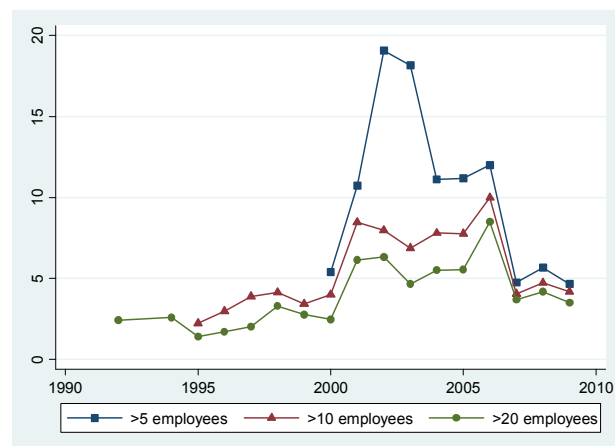c) Based on two-step bootstrap standard errors, adjusted for clustering by firms

Data source: panel of firms observed in the Wage Survey in 2006 and 2007.

## Figure 1: The minimum wage and minimum wage earners in Hungary 1992-2009

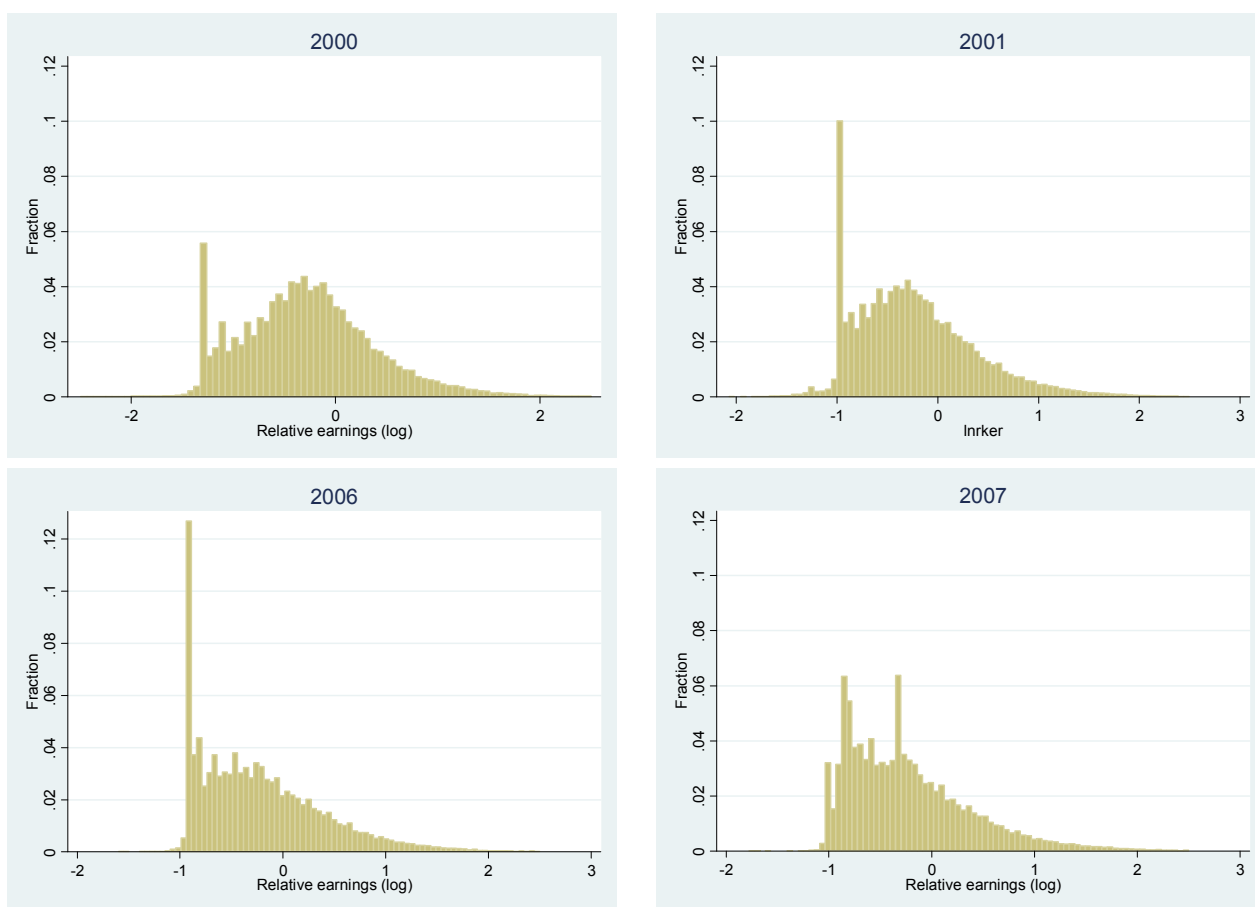(A) The MW compared to the average wage and the median wage 1992-2009
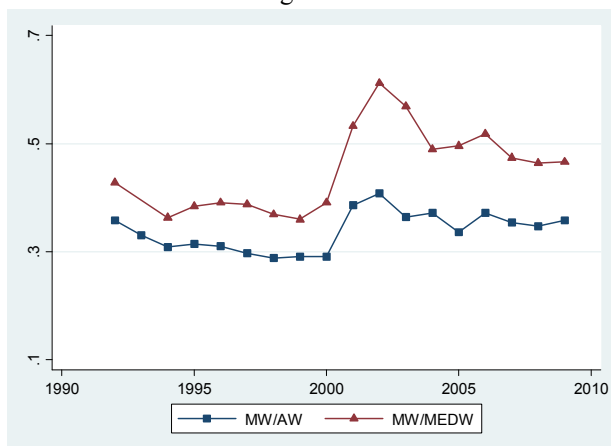
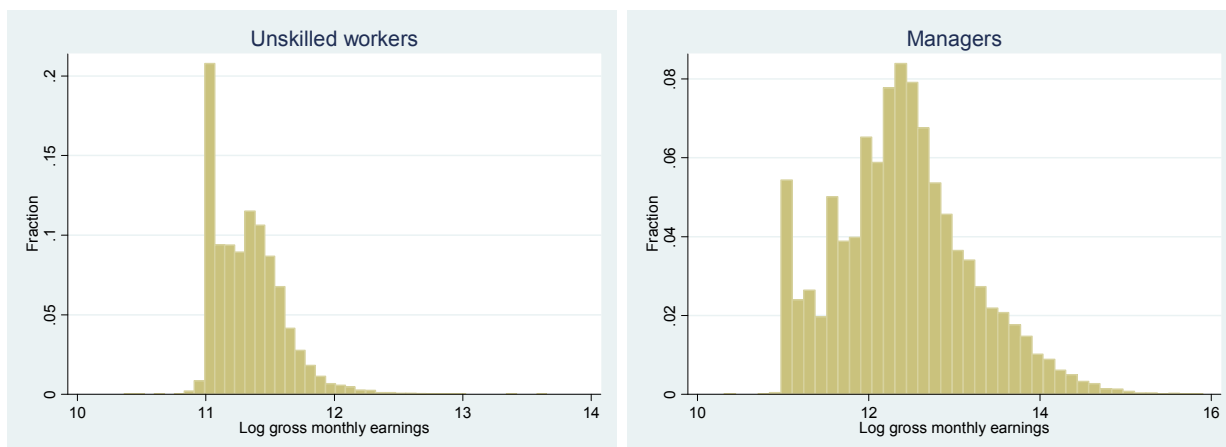(B) Fraction paid 95-105 per cent of the MW, 1992-2009



The data relate to gross monthly earnings in the private sector. Data source: Wage Surveys

## Figure 2: The wage distribution in selected years



Data: Wage Surveys. Samples: full-timers in the private sector

## Figure 3: The wage distribution in two occupations, 2006



Data: Wage Survey 2006, private sector. Occupational codes: managers 1311-1429, unskilled workers 9190

## Figure 4: Wage distribution before and after the transformation
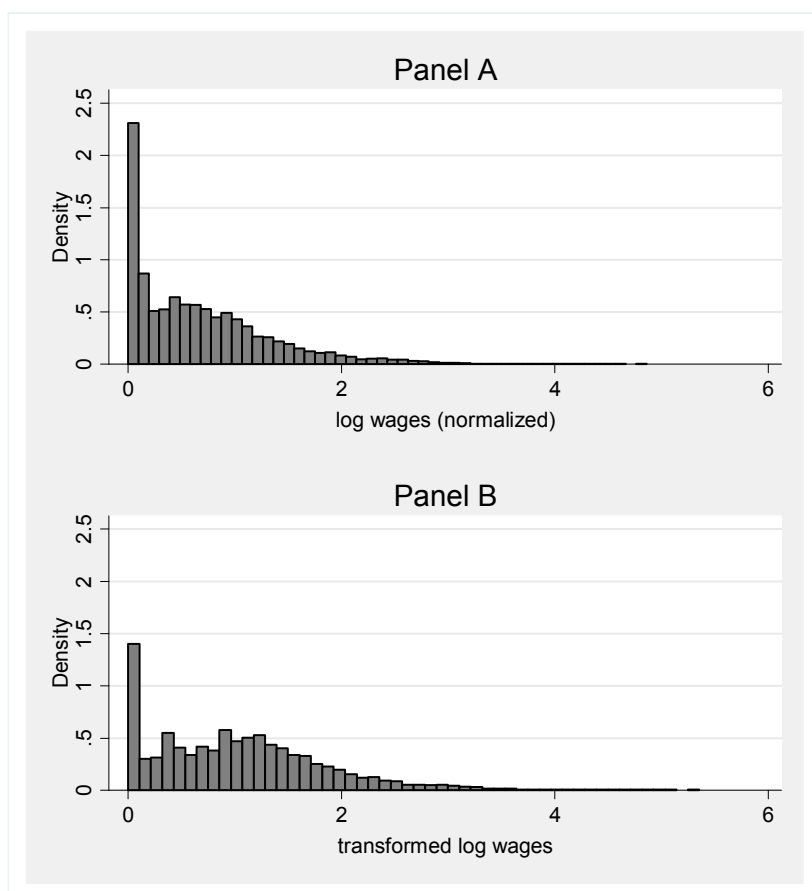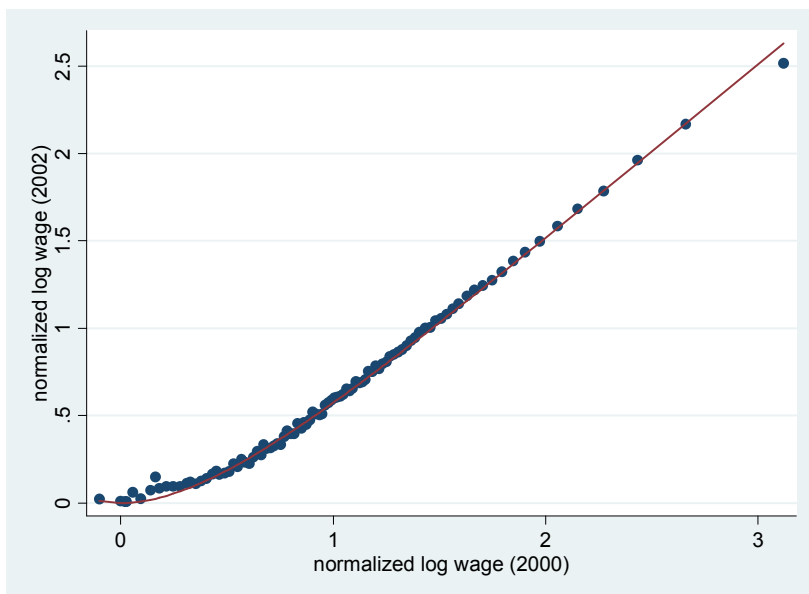Data: Wage Survey 2006, private sector, full-timers

**Figure 5: The function g(x) for r = 0.49 and its fit to the percentile graph 2000-2002**

# Appendix 1

**Table A1 Occupational classification used in the double-hurdle model**

| Occupations | Type* | Definition (based on standard classification of occupations) |
|---|---|---|
| **Agricultural** | E | Codes 61-64 and 92 comprising the drivers of agricultural vehicles |
| **Construction** | S | Code 76 |
| **Service** | S | Codes 52-53 except 532, 533 and 536. Includes transport, mail and telecommunication |
| **Trade** | S | Codes 51 and 421, 422 and 429 comprising cashiers |
| **Industrial** | S | Codes 71-75 |
| **Other blue-collar** | | |
| Cleaners | E | Code 911 |
| Unskilled laborers | E | Codes 913-919 |
| Machine operators | E | Codes 81-83. Includes the operators of mobile machines such as cranes, |
| Porters and guards | E | Codes 912 and 536 comprising porters and security guards, respectively |
| Drivers | S | Code 833, 835, 836 Car, truck and bus. Excludes the drivers of agricultural vehicles |
| **White-collar** | | |
| Office clerks | W | Codes 41-42 and 532-533 comprising office based jobs in health and social services |
| Technicians, assistants | W | Codes 31-34 |
| Administrators | W | Codes 35-39 |
| Managers | W | Codes 11-14 |
| Professionals | W | Codes 21-29 |

* E: elementary; S: secondary; W: white-collar

**Table A2.**
**Selection to the worker panel used in the test (probit)**

| Dependent variable: 1 if made it to the panel, 0 otherwise | Marginal effect | Z-value |
|---|---|---|
| Male | 0.013 | 6.05*** |
| Years in school | -0.000 | -1.27 |
| Experience | 0.009 | 25.68*** |
| Experience squared | -0.000 | -19.94*** |
| Earned more than the MW × log wage | 0.013 | 5.52*** |
| Earned the MW | 0.099 | 3.13*** |
| Firm size: 5-20 employees | -0.110 | -28.95*** |
| Firm size: 21-50 employees | -0.102 | 32.89*** |
| Firm size: 51-300 employees | -0.138 | -65.07** |
| Firm size: 301-1000 employees | -0.003 | -1.25 |
| Ownership: majority domestic private | -0.048 | -19.92*** |
| Ownership: majority foreign | -0.013 | -4.54*** |
| Ownership: mixed | 0.021 | 4.21*** |
| Sales revenues per worker (log) | -0.008 | 8.53*** |
| Negative value added | -0.009 | -0.64 |
| Micro-region unemployment rate (log) | -0.414 | -5.69*** |
| Western Hungary | 0.027 | 7.23*** |
| Northern Transdanubia | 0.064 | 16.20*** |
| Southern Transdanubia | 0.059 | 12.15*** |
| Southern Plain | -0.024 | -5.26 |
| Northern Plain | 0.091 | 19.69*** |
| Northern Hungary | 0.114 | 25.78*** |
| Agriculture, forestry, fishing | 0.131 | 24.92*** |
| Mining | 0.154 | 6.69*** |
| Construction | -0.013 | 3.41*** |
| Trade, tourism | 0.030 | 11.49*** |
| Transport | -0.051 | -8.17*** |
| Financial services | -0.032 | -8.50*** |
| Services | -0.164 | -4.46*** |
| Education and health (private establishments) | 0.035 | 6.22*** |
| Observations | 132115 | |
| LR chi2 (30), significance | 8473.98 | 0.0000 |

* significant at 10%; ** significant at 5%; *** significant at 1%

Reference categories: female, more than 1000 employees, majority state-owned, Central Hungary, manufacturing

Data: Wage Survey 2006, enterprise sector. All variables relate to May 2006.

**Table A3**
**Selection to the firm panel used in the test (probit)**

| Dependent variable: 1 if observed in the 2006 WS, 0 otherwise | Marginal effect | Z-value |
|---|---|---|
| Share of men | 0.000 | 0.01 |
| Average years in school | 0.000 | 0.01 |
| Average experience | 0.000 | 0.02 |
| Average wage | 0.000 | 0.00 |
| Share of workers affected by the 2001 MW hike | -0.000 | -0.000 |
| Firm size: 5-20 employees | -0.368 | -12.47*** |
| Firm size: 21-50 employees | -0.355 | -12.27*** |
| Firm size: 51-300 employees | -0.375 | -13.49*** |
| Firm size: 301-1000 employees | -0.053 | -2.17*** |
| Ownership: majority domestic private | -0.036 | -2.54** |
| Ownership: majority foreign | -0.058 | -3.08*** |
| Ownership: mixed | 0.059 | 1.58 |
| Sales revenues per worker (log) | 0.000 | 0.00 |
| Negative value added | -0.101 | -1.20 |
| Micro-region unemployment rate (log) | 0.008 | 0.02 |
| Western Hungary | -0.022 | -0.96 |
| Northern Transdanubia | -0.018 | -0.75 |
| Southern Transdanubia | 0.031 | 1.29 |
| Southern Plain | 0.019 | 0.83 |
| Northern Plain | 0.011 | 0.51 |
| Northern Hungary | 0.050 | 2.38 |
| Agriculture, forestry, fishing | 0.013 | 0.61 |
| Mining | 0.091 | 1.05 |
| Construction | -0.037 | -1.93* |
| Trade, tourism | -0.021 | -1.53 |
| Transport | -0.053 | -1.75* |
| Financial services | -0.020 | -1.11 |
| Services | -0.248 | --1.37 |
| Education and health (private establishments) | -0.030 | -1.19 |
| Firms in WS 2006 | 9574 | |
| Firms also observed in WS 2007 | 6348 | |

Absolute value of z-statistics in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
Data: Wage Survey 2006

## Appendix 2

### Minimum wage regulations in Hungary

*Target and coverage.* A single national monthly gross minimum wage was introduced by Hungary's last communist-led government in 1989. The minimum wage relates to monthly pre-tax base wages, that is, total monthly earnings net of overtime pay, shift pay and bonuses. Starting from 2007 weekly, daily and hourly levels are determined, too. The minimum wage is legally binding and covers all wages, including those paid to the self-employed by their own businesses. For part-timers, who account for about 5 per cent of total employment, the wage floor is proportionately lower. In 2006-2008 further minima applied to skilled workers (1.25MW) and young skilled workers (1.2MW). In 2009 the minimum for young skilled workers was abolished.

*MW setting.* The minimum wage is negotiated in a consultative body of employers and unions (Council of the Reconciliation of Interests). The government usually steps into the process at the end, by accepting the recommendations of the Council, but it is authorised to make a unilateral decision in case the negotiations fail, as it happened in 2001.

*Level of the MW.* At its introduction the MW amounted to 35 per cent of the average wage (AW), while in 2000 it stood at 29 per cent. Viktor Orbán's first government (1998–2002) nearly doubled the MW, by raising it from Ft 25,500 in December 2000 to Ft 40,000 in January 2001 and Ft 50,000 in January 2002. The two hikes raised the minimum wage–average wage ratio to 39 per cent and 43 per cent, respectively. Since 2003, the MW/AW ratio slightly fell but remained above its pre-hike level.[13]

*Compliance.* The Wage Survey's data suggest that sub-minimum wages accounted for less than 1 per cent of all wages in each year since 1989. Estimates based on personal income tax reports and pension contributions hint at higher rates, but these data do not allow proper adjustment for time out of work during the year.

*Fraction of employees affected.* The fraction of workers paid 95–105 per cent of the MW amounted to 5 per cent in 2000. It jumped to 19 per cent in May 2002 in firms employing five or more workers and increased substantially in larger firms, too. The ratio fell to 10–12 per cent by 2004 and fell further substantially after 2006, when the tax authority started to interpret MW payment as a signal of wage under-reporting.

*Taxing the MW.* In 1989-2001 the MW was subject to linear social security contribution and progressive personal income tax. In 2002 it became free of personal income tax. In 2007, a minimum social security contribution base amounting to 2MW was introduced, as discussed in Section 3 of the text. This measure was abandoned in 2010.

---

[13] All data quoted in Appendix 2 come from the Wage Survey.