

CASE Working Papers

Informal employment and wages in Poland

Jacek Liwiński

No. 14 (138) / 2020



WARSAW

BISHKEK

KYIV

TBILISI

CHISINAU

MINSK

This publication is a continuation of a “CASE Network Report” series. Previous works in this series are available on CASE website: www.case-research.eu

The opinions expressed in this publication are those of the author and do not necessarily reflect the views of CASE.

Keywords:

informal workers, undeclared employment, wages, wage penalty, PSM

JEL Codes:

J24, J31, J46

Graphic Design:

Katarzyna Godyń-Skoczylas | grafo-mania

© CASE – Center for Social and Economic Research, Warsaw, 2020

ISBN: 978-83-7178-719-5

Publisher:

CASE – Center for Social and Economic Research
al. Jana Pawła II 61, lokal 212, 01-031 Warsaw, Poland
tel.: (48 22) 206 29 00, fax: (48 22) 206 29 01
e-mail: case@case-research.eu
www.case-research.eu

Table of Contents

Author	4
Abstract	5
Introduction	6
1. Review of literature	8
2. Data	12
3. Method	13
4. Results	16
5. Conclusions	24
References	27
Appendix	32

Author

Jacek Liwiński – assistant professor of economics at the University of Warsaw, Faculty of Economic Sciences, senior economist at CASE - Center for Social and Economic Research, research fellow of the Global Labor Organization (GLO). His research has focused on labor economics, education economics, and evaluation of public education programs. Address: ul. Długa 44/50, 00-241 Warsaw, Poland. E-mail: jliwinski@wne.uw.edu.pl.

Abstract

Purpose: This paper tries to identify the wage gap between informal and formal workers and tests for the two-tier structure of the informal labour market in Poland.

Design/methodology/approach: I employ the propensity score matching (PSM) technique and use data from the Polish Labour Force Survey (LFS) for the period 2009–2017 to estimate the wage gap between informal and formal workers, both at the means and along the wage distribution. I use two definitions of informal employment: a) employment without a written agreement and b) employment while officially registered as unemployed at a labour office. In order to reduce the bias resulting from the non-random selection of individuals into informal employment, I use a rich set of control variables representing several individual characteristics.

Findings: After controlling for observed heterogeneity, I find that on average informal workers earn less than formal workers, both in terms of monthly earnings and hourly wage. This result is not sensitive to the definition of informal employment used and is stable over the analysed time period (2009–2017). However, the wage penalty to informal employment is substantially higher for individuals at the bottom of the wage distribution, which supports the hypothesis of the two-tier structure of the informal labour market in Poland.

Originality/value: The main contribution of this study is that it identifies the two-tier structure of the informal labour market in Poland: informal workers in the first quartile of the wage distribution and those above the first quartile appear to be in two partially different segments of the labour market.

Introduction

Informal or undeclared work is usually understood as performing a job without paying taxes and social security contributions. But when it comes to the details, there are many different definitions and methods of measurement used to examine this phenomenon. Correspondingly, the level of undeclared employment is assessed differently depending on the definition used. Williams *et al.* (2017) used the Labour Input Method (LIM) and estimated the average level of undeclared employment among the European Union (EU) member states at 7.7% of the total labour input. Interestingly, they found that the country most affected by undeclared employment among the 27 EU economies was Poland (14.4%). Another source of data is the Eurobarometer survey, where undeclared employment is measured as the share of individuals who report to work in the grey sector. The EU average of this share amounted to 4% in 2013 and 3% in 2019 (European Commission, 2014, 2020). Poland, with shares of 3% and 1%, respectively, was below the EU averages in both survey years. The undeclared employment level is also monitored in Poland by Statistics Poland (GUS) on the basis of a special module added to the Labour Force Survey (LFS). The results of this survey show that undeclared employment amounted to 4.5-4.9% of total employment in the period 2009–2017.

In fact, three parties are involved in undeclared work – the customers willing to buy goods and services produced in the grey economy because they are cheaper, the employers willing to hire workers informally in order to reduce labour costs, and the employees willing to work without a formal contract. As for the employees, they generally have two reasons to work informally. First, they may believe they will earn more when employed informally because they will not have to pay income tax. Second, they may agree to work without a formal contract if they are not able to find a job in the formal sector. These two reasons correspond with the “exit” and “exclusion” hypotheses that explain the determinants of informal employment. The “exit” hypothesis is based on the assumption that some individuals leave the formal sector of employment in order to increase their net earnings by avoiding taxes (Maloney, 2004; Perry *et al.*, 2007). The “exclusion” hypothesis explains that undeclared work is a form of inferior employment, which is characterised by low wages, insecurity, and poor working conditions (Harris and Todaro, 1970; Loyaza, 1994; Perry *et al.*, 2007).

As both of these causes of informal employment may be in play at the same time, on theoretical grounds it is not clear whether one should expect informal workers to earn more or less than those in the formal sector. The empirical evidence is also not conclusive. For Poland, there have been three studies conducted thus far, and they provide all three possible answers. They show that informal workers earn higher wages (Tyrowicz and Cichocki, 2011), lower wages (Cichocki and Tyrowicz, 2010a), or the same wages (Cichocki and Tyrowicz, 2010b), when compared to formally employed individuals.

The aim of this paper is to shed more light on this issue by using more recent and richer data for Poland. The Polish LFS dataset for the period 2009–2017 allows me to use two definitions of informal employment: a) employment without a written agreement and b) employment and official registration as unemployed at the same time. I use both the OLS and PSM methods to estimate the effect of being employed informally on monthly earnings, hourly wage, and working time. I find that informal workers – regardless of the definition used – earn on average lower monthly and hourly wages in the period 2009–2017. But the wage penalty to informal employment is substantially higher for individuals at the bottom of the wage distribution, which supports the hypothesis of the two-tier structure of the informal labour market in Poland.

This study contributes to the literature in at least two ways. First, when compared to the existing studies by Cichocki and Tyrowicz, this study provides a more in-depth analysis (using two definitions of informal employment and a richer set of control variables) based on more recent data (2009–2017) and reaches qualitatively different results on the wage gap between informal and formal workers in Poland. Second, this is one of only a few studies in the literature – and the first for Poland – that tries to test for the hypothesis of the two-tier structure of the informal labour market.

This paper is structured into five sections. In the first section, I present the theoretical and empirical literature. The second section presents the data, the third one – the methodology, and the fourth one – the results of the empirical analysis of the effects of working informally on monthly earnings and hourly wage in Poland. The paper ends with a summary that contains the most important conclusions.

1. Review of literature

The theoretical literature suggests that individuals may have two basic reasons for working informally. First, they may believe that they will earn more when employed informally because they will not have to pay income tax. Second, they may agree to work informally if they cannot find a job in the formal sector. These two arguments were formalised in the theoretical literature as the “exit” and “exclusion” hypotheses (Maloney, 2004; Perry *et al.*, 2007; Arias and Khamis, 2008). The exit hypothesis is based on the assumption that individuals choose informal employment to avoid paying taxes and earn more than the net-of-tax wage in the formal sector. The exclusion hypothesis explains that individuals choose to work informally even if their wage in the informal sector is lower than the net-of-tax wage in the formal sector because they cannot find a job in the formal sector. Exclusion from formal employment may result from an individual’s low productivity when compared to the minimum wage or the efficiency wage, from the activity of trade unions, or from discrimination based on the individual’s characteristics. The exclusion hypothesis is thus closely related to the segmentation hypothesis, which assumes that the labour market is divided into two sectors – the primary sector, which is characterised by high wages and stable employment, and the secondary sector, where wages are low and employment is unstable (Doeringer and Piore, 1971; Leontaridi, 1998). The two segments co-exist because – for some reasons, like an individual’s characteristics – it is difficult (costly) or impossible for individuals to move from one sector to the other. From this point of view, the informal sector may be regarded as the secondary or inferior sector of the labour market, with low wages, lack of stability, and poor working conditions (Loyaza, 1994; Perry *et al.*, 2007). Informal workers would like to move to the formal sector, but they cannot do it because of their education, age, gender, or place of residence.

In addition to the above-mentioned two opposite views, there is a third one that combines them. This third view emphasises that the informal sector is highly heterogeneous in its nature and consists of two tiers – an upper tier including those who are voluntarily informal and a lower tier including those who cannot afford to be unemployed but who also cannot find a formal job (Fields, 1990). Accordingly, wages in the upper tier should be higher than

in the formal sector (in line with the exit hypothesis) and wages in the lower tier should be lower than in the formal sector (in line with the exclusion hypothesis).

As for the empirical evidence, most of the early studies support the exclusion hypothesis by showing that informal workers earn less at the mean than those working in the formal sector (Heckman and Hotz, 1986; Pradhan and van Soest, 1995; Tansel, 1997; Gong and van Soest, 2002; Arias and Khamis, 2008; Badaoui *et al.*, 2008; Blunch, 2015). There are also a number of studies that provide evidence supporting the two-tier structure of the informal sector. These are based on two methodologies. Most of these studies use the quantile regression (QR) to test for the heterogeneity of the informal sector wage penalty along the wage distribution. Evidence showing a larger wage penalty in the lower part of the wage distribution is viewed as proof of the two-tier informal sector (Tannuri-Pianto and Pianto, 2002; Botelho and Ponczek, 2011; Lehmann and Zaiceva, 2013). But there are also a few studies using this method that do not find evidence supporting this hypothesis (Nguyen *et al.*, 2013; Bargain and Kwenda, 2014; Staneva and Arabsheibani, 2014; Tansel *et al.*, 2020). Another approach is to use information on the voluntary or involuntary nature of a respondent's informal employment, as the two-tier hypothesis assumes that the former will earn higher wages than formal workers and the latter will earn lower wages (Lehmann and Pignatti, 2018).

In order to properly identify the wage gap between informal and formal workers, one needs to address a possible bias resulting from the non-random selection of individuals into informal employment. Cross-sectional studies using OLS ignore unobservable characteristics and thus may suffer from omitted variable bias if unobservable worker characteristics simultaneously determine the sector choice and wages. This possibility precludes the causal interpretation of OLS estimates. Therefore, authors use other estimation techniques to address this problem.

One widely used strategy was to employ cross-sectional data and the Heckman two-stage correction procedure (Magnac, 1991; Tansel, 2000, 2002; Carneiro and Henley, 2001; Gong and van Soest, 2002; Arias and Khamis, 2008). The problem with this method is that it requires suitable instruments – that is, variables that are determinants of employment in the informal sector in the first stage equation but that are not correlated with wages and thus may be excluded from the wage equation in the second stage. If the instruments used are not strong predictors of informal sector choice, their suitability may be questioned.

Another solution, used recently in a few studies, is to employ panel data and estimate a fixed effect (FE) model. This method deals with the possible bias resulting from the selection on unobservables, provided that the unobserved individuals' characteristics are time-invariant. These studies find that the informal sector wage penalty gets smaller when compared to the OLS estimation (Botelho and Ponczek, 2011; Nguyen *et al.*, 2013; Bargain and Kwenda,

2014; Tansel and Kan, 2016; Tansel *et al.*, 2020), or that the penalty completely disappears (Pratap and Quintin, 2006; Badaoui *et al.*, 2008; Nordman *et al.*, 2016).

Another recent approach is to use the propensity score matching (PSM) technique with cross-sectional or panel data (Calderón-Madrid, 1999; Pratap and Quintin, 2006; Badaoui *et al.*, 2008; Cichocki and Tyrowicz, 2010a, 2010b; Tyrowicz and Cichocki, 2011; Bargain and Kwenda, 2014). This method deals with possible misspecifications that may occur due to the linearity assumption on the covariates.

Most of the empirical literature on the wage gap between informal and formal workers comes from developing countries, like Argentina (Pratap and Quintin, 2006; Arias and Khamis, 2008), Brazil (Carneiro and Henley, 2001; Tannuri-Pianto and Pianto, 2002; Henley *et al.*, 2009; Botelho and Ponczek, 2011), Colombia (Magnac, 1991), Côte d'Ivoire (Günther and Launov, 2012), Egypt (Tansel *et al.*, 2020), Madagascar (Nordman *et al.*, 2016), Mexico (Gong and van Soest, 2002), South Africa (Badaoui *et al.*, 2008), or Turkey (Tansel, 2000, 2002; Tansel and Kan, 2016). There are also a few studies on the post-Soviet states – Russia (Lehmann and Zaiceva, 2013), Ukraine (Lehmann and Pignatti, 2018), Tajikistan (Staneva and Arabsheibani, 2014), and the Baltic states (Meriküll and Staehr, 2010). As for Central and Eastern European countries, to the best of my knowledge, there are only studies for Poland (Cichocki and Tyrowicz, 2010a, 2010b; Tyrowicz and Cichocki, 2011). These last three studies are most closely related to my research.

Cichocki and Tyrowicz (2010a) use the PSM technique and cross-sectional data from a survey conducted in 2007 by the Ministry of Labour and Social Affairs of Poland on a sample of approximately 19,000 individuals. Their definition of informal workers includes individuals employed without a written employment agreement and those whose earnings were not declared to the social security and tax authorities. They estimate propensity scores based on a relatively small set of explanatory variables representing workers' characteristics, including gender, age, education, and marital status (and their interactions). They find that the raw wage gap between formal and informal workers is 29% and it narrows down to 23% after using the PSM. This outcome is consistent with the exclusion hypothesis.

Cichocki and Tyrowicz (2010b) use the same data and identification strategy as Cichocki and Tyrowicz (2010a). The only difference between the two studies is that the former uses a broader definition of informal employment, including not only fully informal employees, but also partially informal employees – those where a portion of their earnings are not declared to the social security and tax authorities or whose formal earnings are lower than their actual ones. Importantly, the total amount of earnings of the partially informal workers (formal wage + envelope wage) was analysed. Both the raw and PSM estimates show that there is no wage gap at the mean between informal and formal workers.

Tyrowicz and Cichocki (2011) employ the PSM technique and data from the Polish Labour Force Survey (LFS) for the period 1995–2007. Due to data limitations, they define informal workers as employed individuals who are officially registered as unemployed at labour offices. The study finds that the raw wages of informal workers were 30–50% lower than those of formal workers, but the PSM estimates surprisingly show that these are informal workers who earn 40–50% more than their formal sector counterparts. Thus, the results of this study support the exit hypothesis.

In summary, the evidence for Poland is thus far inconclusive. The three studies by Cichocki and Tyrowicz that use different datasets and different informal employment definitions provide qualitatively different results. Thus, based on the existing evidence, it is not possible to conclude which one of the three hypotheses explaining the determinants of informal employment holds for Poland.

2. Data

The analysis is based on cross-sectional data from the Polish LFS for the years 2009–2017. The LFS is a representative sample survey of Polish residents that is conducted quarterly by Statistics Poland (GUS). Approximately 50,000 individuals aged 15 years or more are surveyed every quarter. They provide detailed information on their economic activity, as well as on a large set of background characteristics[1]. Importantly for this study, in 2009, a question aimed at identifying informal workers was added to the LFS questionnaire. Namely, the employed respondents are asked whether they have a written agreement with their employer. Additionally, the questionnaire includes information on whether respondents are registered as unemployed at a labour office, which was used by Tyrowicz and Cichocki (2011) to identify informal workers. Thus, I define informal workers in two alternative ways in my study – as:

1. employed individuals without a written agreement with their employer,
2. employed individuals who are officially registered as unemployed at a labour office.

For the purpose of identifying the wage gap between informal and formal workers, I restricted the sample to employed individuals who reported the amount of their net earnings and their working hours on their main job on the month prior to the survey. In other words, I dropped all individuals who were not employed or did not report their earnings or working hours. Importantly, I also dropped the self-employed, as they are not asked to report their income in the LFS. I also restricted the sample to individuals at the so-called productive age, which is 18–59 for women and 18–64 for men, because only this group was asked to provide information on their unemployment registration status. The sample size, subject to all of the above restrictions, is 332,183 observations, including 10,054 individuals without a written employment agreement and 2,690 employed individuals who are registered as unemployed.

3. Method

While trying to assess the wage gap between informal and formal workers, one needs to address the problem of possible selection into informal employment. Previous studies for Poland show that informal workers are indeed different from their formal sector counterparts – undeclared work is more common among men (Beręsewicz and Nikulin, 2018), low-skilled individuals, those working in small or micro firms, as well as in the construction, agriculture, or trade sector (Cichocki and Tyrowicz, 2011). Thus, the raw wage gap may result – at least to some extent – from the fact that formal and informal workers are different in terms of various observable and unobservable characteristics. I address this problem by employing the PSM technique. This method was used in a few recent studies, including the three above-mentioned studies for Poland (Cichocki and Tyrowicz, 2010a, 2010b; Cichocki and Tyrowicz, 2011). However, there are four novelties in my approach when compared to those studies. First, my analysis is based on a more recent data set (2009–2017). Second, I estimate the wage gap based on two different definitions of informal employment, which was not possible with the LFS data prior to 2009. Third, I improve the quality of matching by employing a much larger number of individual characteristics to estimate the propensity scores. Fourth, I estimate the wage gap not only at the means but also for each quartile along the wage distribution.

First, as a baseline for my analysis I estimate a wage regression using OLS with robust standard errors. The wage equation is the following:

$$\ln w_i = INF_i \beta_1 + X_i \beta_2 + \varepsilon_i \quad (1)$$

where $\ln w_i$ is the natural logarithm of the net monthly earnings or the hourly wage rate^[2], INF_i is a dummy variable equal to 1 for informal workers and 0 otherwise, and X_i is a vector of the other explanatory variables, including individuals' demographic, educational, and employment-related characteristics, as well as controls for place of residence, region,

and survey year. A complete list of variables included in the wage equation and descriptive statistics of the sample are presented in Tables A1 and A2, respectively, in the Appendix.

Second, I employ the PSM technique. This method involves matching informal workers (treatment group) with individuals having a similar propensity to work in the informal sector, although they actually work in the formal sector (control group). The propensity scores are estimated using a probit model on the basis of the same set of individuals' and labour market characteristics (X_i) that was used in the OLS estimation. Thus, I estimate the following equation:

$$INF_i = X_i \beta + \vartheta_i \quad (2)$$

Separate estimations of the model are made for each definition of informal employment. The full set of estimates based on the first definition of informal employment, that is employment without a written agreement, is presented in Table A3 in the Appendix.

Then, I use the nearest neighbour method to match formal and informal workers on the basis of propensity scores. In order to reduce the standard error of estimation I match five individuals from the control pool (the group of formal workers) to each individual in the treatment group (informal workers), which may be referred to as the NN5 matching. As the size of the control pool is far larger than the treatment group, I use the matching procedure with replacement – an individual from the control pool may be matched to only one individual in the treatment group[3].

The key issue in the matching procedure is to obtain a balanced distribution of observable characteristics of individuals included in the treatment and control groups. As the analysis covers a relatively long period of time (2009-2017), when substantial changes in the Polish labour market took place, the matching of observations from different survey years might result in a bias. Hence, I use a mixed matching pattern – exact matching (1:1) based on the survey year and NN5 matching based on the propensity to work informally. As a result, the matched sample is well balanced. In the pooled sample (2009-2017), before matching the mean values of almost all (68 out of 77) variables in the treatment group and the control pool were significantly different (at the 5% level), while after matching most of these differences disappeared (only 24 remained significant). Thus, the formal workers included in the control group are very similar in terms of their observable characteristics to informal workers in the treatment group. The quality of matching is also illustrated in Figure A1 in the Appendix, where the standardised percentage bias (Rosenbaum and Rubin, 1985) and the ratio of the variance of the residuals in the treatment group to the control group (Rubin, 2001) are presented. This figure clearly shows that both these indicators were

substantially reduced as a result of matching, and the ratio of the variance of the residuals with respect to almost all variables is within the interval [0.8; 1.25], as suggested by Rubin (2001).

In the final step of the PSM procedure, I estimate the average treatment effect on the treated (ATT) as the difference between the mean log of monthly earnings or hourly wage rate in the treatment and control groups. Standard errors are bootstrapped by performing 500 replications.

4. Results

The baseline for my analysis are the OLS estimates presented in Table 1, which are based on the pooled data (2009–2017) and informal employment defined as working without a written agreement[4]. They clearly show that informal workers in Poland earn less than their formal counterparts, both in terms of their raw wages and after controlling for their observable characteristics. The raw wage gap, presented in column 1, is substantial and it amounts to 24.7% of monthly earnings and 22% of the hourly wage rate. Then, in columns 2–8, I gradually add control variables to see which of the individuals' characteristics explain the wage gap. I find that the wage gap reduces when variables representing workers' human capital and firm characteristics are added to the wage equation. This result is not surprising as the descriptive statistics of the sample show that informal workers are less educated, less tenured, and they work in smaller firms, which typically pay lower wages. When the full set of control variables is included in the model, the wage penalty to informal employment is reduced to 9.1% of monthly earnings and 5.4% of the hourly wage (see column 8). The fact that the wage penalty in terms of monthly earnings is higher than in terms of hourly wage may be a sign of the shorter working time of informal workers. To check this presumption formally, I regressed working time on the full set of explanatory variables included in the wage regression (see Table A4 in the Appendix for the full specification). The estimate in column 8 shows that informal workers indeed have shorter working hours than their formal sector counterparts, but the difference is not substantial – it amounts to less than 1 hour per week.

Table 1. The effects of employment without a written agreement (OLS estimates, pooled sample 2009–2017)

DEPENDENT VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Earnings (ln)	-0.247*** (0.007)	-0.243*** (0.006)	-0.251*** (0.006)	-0.164*** (0.006)	-0.123*** (0.006)	-0.109*** (0.006)	-0.088*** (0.006)	-0.091*** (0.006)
Hourly wage (ln)	-0.220*** (0.005)	-0.214*** (0.005)	-0.210*** (0.005)	-0.119*** (0.005)	-0.076*** (0.005)	-0.066*** (0.005)	-0.051*** (0.005)	-0.054*** (0.005)
Working time (hours)	0.158 (0.129)	0.058 (0.128)	-0.351*** (0.125)	-0.535*** (0.125)	-0.882*** (0.123)	-0.738*** (0.123)	-0.619*** (0.123)	-0.628*** (0.123)
Control variables								
Gender, age, age square, civil status characteristics		yes	yes	yes	yes	yes	yes	yes
Place of residence, region			yes	yes	yes	yes	yes	yes
Education level				yes	yes	yes	yes	yes
Firm size, firm ownership, economic sector					yes	yes	yes	yes
Occupation						yes	yes	yes
Job tenure, job tenure square							yes	yes
Internship, trial period								yes

Notes: Each coefficient comes from a separate estimation of the model. Each specification additionally includes survey year. The full set of estimates coming from models presented in column 8 are reported in Table A4 in the Appendix.

***/**/* stand for 1%, 5% and 10% significance, respectively. Standard errors in brackets.

Source: Author's own analyses based on unit data from the Polish LFS, 2009–2017.

Another interesting issue is whether the wage penalty to undeclared employment is stable over the analysed time period (2009–2017). Thus, I estimated the full specification of the model (column 8 in Table 1) for each survey year separately. The estimates presented in Table

2 show that informal workers earn less than formal workers both in terms of monthly earnings and hourly wage in every single year in the analysed period. Both measures of the wage penalty are quite stable over time, although there seems to be a slight upward trend from 2009 to 2013 and a downward trend afterwards. These trends coincide with similar trends in the unemployment rate in Poland, which increased from 8.2% in 2009 to 10.3% in 2013 and then declined to 4.9% in 2017 (Statistics Poland, 2018). A possible explanation for these trends in the wage penalty could be that during economic downturns, it is easier for employers to cut the wages of informal workers because they are not bound by any written employment agreement that states the wage rate or the period of employment. However, I will not attempt to verify this hypothesis in this paper.

Table 2. Effects of employment without a written agreement (OLS estimates)

DEPENDENT VARIABLES	CONTROL VARIABLES	POOLED, 2009–2017	2009	2010	2011	2012	2013	2014	2015	2016	2017
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Earnings (ln)	Only survey year	-0.247*** (0.007)	-0.179*** (0.017)	-0.251*** (0.018)	-0.241*** (0.017)	-0.254*** (0.020)	-0.281*** (0.021)	-0.244*** (0.016)	-0.255*** (0.018)	-0.263*** (0.021)	-0.236*** (0.027)
	Full set of controls	-0.091*** (0.006)	-0.048*** (0.014)	-0.080*** (0.015)	-0.091*** (0.014)	-0.096*** (0.016)	-0.123*** (0.019)	-0.086*** (0.014)	-0.105*** (0.016)	-0.081*** (0.018)	-0.081*** (0.022)
Hourly wage (ln)	Only survey year	-0.220*** (0.005)	-0.174*** (0.015)	-0.218*** (0.014)	-0.223*** (0.014)	-0.224*** (0.015)	-0.241*** (0.016)	-0.227*** (0.014)	-0.226*** (0.015)	-0.241*** (0.018)	-0.186*** (0.022)
	Full set of controls	-0.054*** (0.005)	-0.033*** (0.012)	-0.034*** (0.013)	-0.061*** (0.012)	-0.054*** (0.014)	-0.073*** (0.014)	-0.054*** (0.012)	-0.062*** (0.013)	-0.051*** (0.016)	-0.034* (0.019)
Working time (hours)	Only survey year	0.158 (0.129)	0.599* (0.315)	-0.199 (0.316)	0.529 (0.361)	0.231 (0.419)	-0.094 (0.434)	0.299 (0.340)	0.038 (0.363)	0.142 (0.368)	-0.180 (0.506)
	Full set of controls	-0.628*** (0.123)	-0.035 (0.296)	-1.047*** (0.299)	-0.300 (0.334)	-0.641 (0.390)	-0.931** (0.420)	-0.679** (0.336)	-0.858** (0.355)	-0.530 (0.358)	-0.488 (0.487)

Note: Each coefficient comes from a separate estimation of the model. The full set of controls includes all independent variables listed in Table A1 in the Appendix.

***/**/* indicate a 1%, 5%, and 10% significance level, respectively. Standard errors in brackets.

Source: Author's own analyses based on unit data from the Polish LFS, 2009–2017.

Table 3 presents the results of the PSM estimation, including both the unmatched and the *average treatment effect on the treated* (ATT) estimates of the wage gap. As expected, the unmatched wage differentials, which represent the raw wage penalty to informal employment, are basically equal to the corresponding OLS estimates[5]. But the main interest should be focused on the matched wage differentials – that is, on the ATT estimates of the wage penalty – and these should be compared with the OLS estimates coming from the models with the full set of controls (see Table 2). First, the PSM estimates based on the pooled data show that informal workers earn 11.7% lower monthly earnings and a 7.9% lower hourly wage than workers in the formal sector. These wage penalties are 2–3 percentage points higher than those coming from the OLS estimation. Second, when I estimate the model for single survey years, I find that the wage penalties to undeclared employment – both in terms of monthly earnings and hourly wage – are persistent over the entire analysed period (2009–2017). Although there are small variations of these estimates over time, the estimates are negative and strongly significant in every single survey year. Some of them are a bit higher, while others are a bit lower than the corresponding OLS estimates, but altogether I do not find any substantial or systematic differences between the PSM and OLS estimates. Importantly, the PSM estimates do not confirm the observation based on the OLS estimates that the wage penalty is correlated with the unemployment rate.

Table 3. Effects of employment without a written agreement (PSM estimates)

DEPENDENT VARIABLES	CONTROL VARIABLES	POOLED, 2009–2017	2009	2010	2011	2012	2013	2014	2015	2016	2017
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Earnings (ln)	Unmatched	-0.247*** (0.005)	-0.175*** (0.012)	-0.242*** (0.012)	-0.224*** (0.012)	-0.231*** (0.014)	-0.259*** (0.015)	-0.222*** (0.012)	-0.245*** (0.013)	-0.256*** (0.015)	-0.225*** (0.017)
	ATT	-0.117*** (0.007)	-0.076*** (0.017)	-0.102*** (0.016)	-0.120*** (0.017)	-0.108*** (0.024)	-0.132*** (0.022)	-0.088*** (0.018)	-0.141*** (0.019)	-0.128** (0.022)	-0.131*** (0.024)
Hourly wage (ln)	Unmatched	-0.231*** (0.004)	-0.164*** (0.045)	-0.217*** (0.011)	-0.222*** (0.012)	-0.224*** (0.015)	-0.231*** (0.014)	-0.217*** (0.012)	-0.231*** (0.013)	-0.233*** (0.014)	-0.183*** (0.016)
	ATT	-0.079*** (0.006)	-0.049*** (0.015)	-0.047*** (0.014)	-0.094*** (0.016)	-0.082*** (0.020)	-0.072*** (0.019)	-0.063*** (0.015)	-0.104*** (0.015)	-0.088** (0.018)	-0.075*** (0.020)
Working time (hours)	Unmatched	0.446*** (0.070)	0.410* (0.208)	0.047 (0.198)	0.886*** (0.197)	0.894*** (0.221)	0.353 (0.241)	0.678*** (0.200)	0.458** (0.201)	-0.291 (0.219)	-0.143 (0.243)
	ATT	-0.700*** (0.134)	-0.430 (0.316)	-1.217*** (0.338)	-0.187 (0.360)	-0.228 (0.469)	-1.267** (0.540)	-0.459 (0.380)	-0.722** (0.362)	-0.669 (0.457)	-0.894* (0.491)

Note: ***/**/* indicate a 1%, 5%, and 10% significance level, respectively. Standard errors in brackets.

Source: Author's own analyses based on unit data from the Polish LFS, 2009–2017.

In order to check for the robustness of the above presented results, I re-estimate the model using a different definition of unregistered employment. Namely, I follow Tyrowicz and Cichocki (2011) and define informal workers as the employed individuals who are – at the same time – officially registered as unemployed at a labour office. The rationale for defining informal employment in this way is that formal workers are not allowed to register at labour offices and they have no reason to do so, while informal workers may be willing to register as unemployed in order to be covered by health insurance. But, importantly, some informal workers may have no incentive to register at labour offices because they may be entitled to health insurance based on reasons other than registered unemployment. For example, they may apply for health insurance if their family member (husband, wife, or child) is formally employed. Hence, we should be aware that this definition of undeclared employment most likely does not cover all informal workers. This may lead to overestimation of the wage gap. Nevertheless, this definition was used by Tyrowicz and Cichocki (2011) as there was not any other variable representing informal employment in the Polish LFS at that time. I use this definition only as a robustness check, as I am aware that my base definition – which covers individuals who are employed without a written agreement – is a much better measure of informal employment.

Table 4 presents the results of the robustness check that come from the PSM estimation. As expected, both the unmatched and the ATT estimates of the wage penalty are higher than those based on the base definition of informal employment (in Table 3). The PSM estimation on the pooled data (2009–2017) shows that the raw wage penalty to informal employment amounts to 54% of monthly earnings and 42% of the hourly wage, while the corresponding ATT estimates are much lower – they amount to 19.0% and 9.7%, respectively. The lower wage penalty in terms of hourly wage (when compared to the one for monthly wage) may result from the shorter working time of informal workers (by approximately 2 hours per week). When we look at the estimates of the wage gap for single years in the period 2009–2017, we can notice substantial time variation of all the measures of the wage gap. But these estimates are stable in qualitative terms – for every single survey year, the ATT estimates of the wage gap between informal and formal workers are negative and strongly statistically significant. This outcome is at odds with the findings of Tyrowicz and Cichocki (2011), who show that the raw wage penalty amounts to 40–50% of monthly wages in the period 1995–2007, but after matching this raw wage gap surprisingly translates into the ATT wage premium of 40–60%. The results of my study presented in Table 4, which are based on the same definition of informal employment, are very similar in terms of the raw wage gap, but entirely different in terms of the ATT wage gap. Importantly, the results shown in Table 4 are consistent with those in Table 3, which are based on a more reliable definition of informal employment. Therefore, I am convinced that undeclared workers on average do not enjoy a wage premium, but they rather incur a wage penalty.

Table 4. Effects of employment while registered as unemployed (PSM estimates)

DEPENDENT VARIABLES	CONTROL VARIABLES	POOLED, 2009–2017	2009	2010	2011	2012	2013	2014	2015	2016	2017
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Earnings (ln)	Unmatched	-0.542*** (0.009)	-0.449*** (0.029)	-0.593*** (0.019)	-0.527*** (0.021)	-0.491*** (0.024)	-0.446*** (0.023)	-0.463*** (0.021)	-0.337*** (0.031)	-0.330*** (0.035)	-0.526*** (0.046)
	ATT	-0.190*** (0.013)	-0.113*** (0.042)	-0.190*** (0.027)	-0.242*** (0.034)	-0.182*** (0.036)	-0.176*** (0.038)	-0.202*** (0.030)	-0.118** (0.049)	-0.185*** (0.057)	-0.377*** (0.083)
Hourly wage (ln)	Unmatched	-0.425*** (0.008)	-0.390*** (0.028)	-0.447*** (0.018)	-0.358*** (0.021)	-0.367*** (0.023)	-0.360*** (0.023)	-0.371*** (0.020)	-0.321*** (0.030)	-0.222*** (0.030)	-0.301*** (0.040)
	ATT	-0.097*** (0.009)	-0.084** (0.034)	-0.080*** (0.019)	-0.096*** (0.024)	-0.096*** (0.024)	-0.114*** (0.028)	-0.131*** (0.022)	-0.110*** (0.037)	-0.095** (0.038)	-0.191*** (0.053)
Working time (hours)	Unmatched	-2.059*** (0.135)	-0.632 (0.491)	-3.518*** (0.312)	-3.356*** (0.353)	-2.331*** (0.394)	-1.222*** (0.402)	-1.203** (0.353)	1.040** (0.511)	-1.358** (0.573)	-3.483*** (0.737)
	ATT	-1.835*** (0.278)	-0.097 (0.869)	-2.535*** (0.559)	-3.039*** (0.683)	-1.578** (0.792)	-0.928 (0.823)	-1.175 (0.767)	0.412 (1.154)	-1.487 (1.297)	-3.096* (1.872)

Note: ***/**/* indicate a 1%, 5%, and 10% significance level, respectively. Standard errors in brackets.

Source: Author's own analyses based on unit data from the Polish LFS, 2009–2017.

The negative wage gap between informal and formal workers is consistent with the exclusion hypothesis, which states that individuals agree to work in the informal sector and earn lower wages because they cannot find a job in the formal sector where wages are higher. But the coefficients presented in Tables 1–4 are estimated at the means of the wage distributions, which does not allow to test for the two-tier structure of the informal sector. Hence, I check for the heterogeneity of the wage penalty along the wage distribution. The results of this analysis based on the pooled sample are presented in Table 5. I find that only the individuals in the first quartile of the wage distribution suffer from the informal sector wage penalty in terms of monthly earnings, which is 13.7% for workers without a written agreement and 19.1% for those registered as unemployed. Undeclared workers in the second, third, or fourth quartile of the wage distribution earn on average the same monthly wage as formal workers. The lower monthly earnings of informal workers in the first quartile may be partly explained by their shorter working time (by 2–3 hours per week), but still their hourly wage is 5–6% lower when compared to their counterparts in the formal sector. Undeclared workers in the second to fourth quartiles also incur a penalty in terms of hourly wages (1–3%), but they work more than formal workers (by 1–2 hours per week) and this is why their monthly earnings are on average the same as those of formal workers.

Table 5. Effects of informal employment along the wage distribution (PSM estimates of ATT, pooled sample)

DEPENDENT VARIABLES	DEFINITION OF INFORMAL EMPLOYMENT	TOTAL	QUARTILE OF THE WAGE DISTRIBUTION			
			Q1	Q2	Q3	Q4
Earnings (ln)	No written agreement	-0.117*** (0.007)	-0.137*** (0.008)	0.003 (0.002)	-0.001 (0.002)	0.003 (0.009)
	Registered as unemployed	-0.190*** (0.013)	-0.191*** (0.013)	-0.0002 (0.004)	-0.013** (0.005)	0.002 (0.025)
Hourly wage (ln)	No written agreement	-0.079*** (0.006)	-0.054*** (0.007)	-0.013*** (0.005)	-0.021*** (0.005)	-0.030*** (0.009)
	Registered as unemployed	-0.097*** (0.009)	-0.062*** (0.009)	-0.026** (0.010)	-0.030** (0.014)	-0.049* (0.029)
Working time (hours)	No written agreement	-0.700*** (0.134)	-2.081*** (0.227)	0.947*** (0.187)	1.060*** (0.227)	1.710*** (0.327)
	Registered as unemployed	-1.835*** (0.278)	-2.714*** (0.363)	1.223*** (0.438)	1.040* (0.587)	2.472*** (0.924)

Note: Each coefficient comes from a separate estimation of the model. The full set of controls includes all independent variables listed in Table A1.

***/**/* indicate a 1%, 5%, and 10% significance level, respectively. Standard errors in brackets.

Source: Author's own analyses based on unit data from the Polish LFS, 2009–2017.

Overall, it seems that informal workers in the first quartile of the wage distribution and those above the first quartile are in two partially different segments of the informal labour market, which supports the two-tier hypothesis. The former suffer from substantially lower monthly earnings than formal workers, which is a result of both a lower hourly wage and a shorter working time, while the latter have the same monthly earnings as formal workers, as they make up for a slightly lower hourly wage rate with a longer working time. Importantly, if I define undeclared workers as those without a written employment agreement, those in the first quartile incur a higher hourly wage penalty (5.4%) than those above the first quartile (1.3–3.0%). On the other hand, support for the two-tier hypothesis does not seem very strong, as individuals above the first quartile incur an hourly wage penalty, while following Fields (1990), one could expect that informal employment in the upper tier would be voluntary and hence it would yield a wage premium rather than a wage penalty. In fact, only Botelho and Ponczek (2011) provide evidence for Brazil that supports this view, as they find that informal workers in the first two quartiles of the wage distribution incur a

wage penalty, while those in the fourth quartile gain a wage premium. My results are more similar to those of Tannuri-Pianto and Pianto (2002), who find that in Brazil the wage penalty to informal employment exists over the entire wage distribution, but is higher at the bottom of the distribution, or to those of Lehmann and Zaiceva (2013), who find that in Russia informal workers suffer from a wage penalty only as long as they are in the first two quartiles of the wage distribution. To the best of my knowledge, these are the only studies that confirm the two-tier hypothesis. There are a few other studies that do not find evidence supporting this hypothesis (Nguyen *et al.*, 2013; Bargain and Kwenda, 2014; Staneva and Arabsheibani, 2014; Tansel *et al.*, 2020).

5. Conclusions

The goal of this study was to determine whether individuals working in the informal sector in Poland incur a wage penalty or rather enjoy a wage premium when compared to their counterparts in the formal sector. Based on this, I wanted to determine whether it is the exit, exclusion, or two-tier hypothesis that best explains the determinants of informal employment in Poland. To answer these questions, I employed the PSM technique and examined data from the Polish LFS for the period 2009–2017. I used two definitions of informal employment: employment without a written agreement and employment while officially registered as unemployed at a labour office.

After controlling for a rich set of individual characteristics, I find that informal workers earn on average less than formal workers, both in terms of monthly earnings and hourly wage. When informal workers are defined as employed individuals without a written agreement, the monthly and hourly wage penalties amount to 11.7% and 7.9%, respectively. When I use the second definition of undeclared employment based on official registration as unemployed, these wage penalties are even higher (19% and 9.7%, respectively). Importantly, the results presented above, which are based on a pooled sample (2009–2017), are also stable over time – that is, I find that informal workers, irrespective of how they are defined, suffer from monthly and hourly wage penalties in every single year over the period 2009–2017. Interestingly, this outcome is at odds with the findings of Tyrowicz and Cichocki (2011), who show that employed individuals who were registered as unemployed enjoyed a wage premium of 40–60% in the period 1995–2007.

In order to test for the two-tier hypothesis, which is that some individuals work in the informal sector because they are excluded from formal employment whereas others are voluntarily informal, I also checked for the heterogeneity of the wage penalty along the wage distribution. I find that informal workers in the first quartile of the wage distribution and those above the first quartile are in two partially different segments of the informal labour market, which supports the two-tier hypothesis. The former suffer from substantially lower monthly earnings than formal workers, which is a result of both a lower hourly wage and a shorter working time, while the latter have the same monthly earnings as formal workers because they make up for a slightly lower hourly wage rate with a longer working time.

Overall, my results are consistent with a few recent studies that find evidence of the two-tier structure of the informal sector in Brazil (Tannuri-Pianto and Pianto, 2002; Botelho and Ponczek, 2011) and Russia (Lehmann and Zaiceva, 2013).

One should also keep in mind some limitations of my findings. First, my analysis is based on data coming from a population survey – that is, on information declared by respondents – which is obviously subject to a number of measurement problems such as non-reporting or misreporting. This issue, which is common for all studies using this type of data, may be even more important in cases of studies on this topic, as some respondents may be unwilling to reveal the informal nature of their employment or the amount of their informal earnings. This is why I used two different methods of measuring informal employment status. Importantly, the wording of both questions in the LFS questionnaire, on which these two definitions are based, does not suggest that they are aimed to identify informal employment. Second, although I used the PSM technique and a rich set of variables to control for observable heterogeneity, the wage gap between informal and formal workers may still be biased due to unobservable heterogeneity.

Notes

- [1] For more information on the Polish LFS, see Statistics Poland (2018).
- [2] The nominal amounts of monthly earnings and hourly wage were deflated with the CPI (base year=2009).
- [3] The computations were performed in Stata/SE 13.0, using the psmatch2 command, version 4.0.11 22 oct2014.
- [4] The full set of estimates is reported in Table A4 in the Appendix.
- [5] Small differences between the PSM and OLS estimates of the raw wage gap result from the fact that I used weights in the OLS estimation, while the PSM unmatched estimates are – by definition – not weighted. The weighting in the PSM estimation is applied only in the matching process when the propensity scores are estimated. Thus, only the ATT estimates are weighted.

References

Arias, O. and Khamis, M. (2008), "Comparative advantage, segmentation and informal earnings: a marginal treatment effects approach", IZA Discussion Paper no. 3916, Institute of Labor Economics (IZA), Bonn.

Badaoui, E., Strobl, E., and Walsh, F. (2008), "Is there an informal employment wage penalty? Evidence from South Africa", *Economic Development Cultural Change*, Vol. 56 No. 3, pp. 683–710.

Bargain, O. and Kwenda, P. (2014), "The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on Panel Data", *Economic Development and Cultural Change*, University of Chicago Press, Vol. 63 No. 1, pp. 117–153.

Beręsewicz, M., and Nikulin, D. (2018), "Informal employment in Poland: An empirical spatial analysis", *Spatial Economic Analysis*, Vol. 13 No. 3, pp. 338–355.

Blunch, N.H. (2015), "Bound to lose, bound to win? The financial crisis and the informal-formal sector earnings gap in Serbia", *IZA Journal of Labor Development*, Vol. 4 No. 1, pp. 13–34.

Botelho, F. and Ponczek, V. (2011), "Segmentation in the Brazilian labor market", *Economic Development Cultural Change*, Vol. 59 No. 2, pp. 437–463.

Calderón-Madrid, A. (1999), "Job stability and labor mobility in Mexico during the 1990's", Centro de Estudios Económicos working paper, Mexico.

Carneiro, F. and Henley, A. (2001), "Modelling formal vs. informal employment and earnings: micro-econometric evidence for Brazil", Management and business working paper No. 2001–2015, University of Wales at Aberystwyth.

Cichocki, S. and Tyrowicz, J. (2010a), "Shadow employment in post-transition - Is informal employment a matter of choice or no choice in Poland?", *Journal of Socio-Economics*, Vol. 39, pp. 527–535.

Cichocki, S. and Tyrowicz, J. (2010b), "Źródła zatrudnienia nierejestrowanego w Polsce [Sources of informal employment in Poland]", *Bank i Kredyt*, Vol. 41 No. 1.

Cichocki, S. and Tyrowicz, J. (2011), "Determinanty zatrudnienia nierejestrowanego w Polsce w okresach wysokiej i niskiej koniunktury gospodarczej [Determinants of un-

declared employment in Poland in periods of high and low economic prosperity]”, *Gospodarka Narodowa*, Vol. 3/2011, pp. 1–27.

Doeringer, P. and Piore, M. (1971), *Internal Labor Markets and Manpower Analysis*, Heath Lexington Books, Lexington, MA.

European Commission (2014), “Undeclared Work in the European Union, Report”, Special Eurobarometer 402 – September 2013, European Union, DOI: 10.2767/37041.

European Commission (2020), “Undeclared Work in the European Union, Report”, Special Eurobarometer 498 – September 2019, European Union, DOI: 10.2767/018492.

Fields, G.S. (1990), “Labour market modelling and the urban informal sector: Theory and evidence”, Turnham, D., Salomé, B., Schwarz, A. (Eds.), *The informal sector revisited*, OECD, Paris, pp. 49–69.

Gong, X. and van Soest, A. (2002), “Wage differentials and mobility in the urban labor market: a panel data analysis for Mexico”, *Labour Economics*, Vol. 9 No. 4, pp. 513–529.

Günther, I. and Launov, A. (2012), “Informal employment in developing countries: Opportunity or last resort?”, *Journal of Development Economics*, Vol. 97 No. 1, pp. 82–98.

Harris, J. and Todaro, M. (1970), “Migration, Unemployment and Development: A Two-Sector Analysis”, *American Economic Review*, Vol. 60 No. 1, pp. 126–142.

Heckman, J. and Hotz, V. (1986), “An investigation of the labor market earnings of Panamanian males”, *Journal of Human Resources*, Vol. 21 No. 4, pp. 507–542.

Henley, A., Arabsheibani, G.R., and Carneiro, F.G. (2009), “On defining and measuring the informal sector: evidence from Brazil”, *World Development*, Vol. 37 No. 5, pp. 992–1003.

Lehmann, H. and Pignatti, N. (2018), “Informal Employment Relationships and the Labor Market: Is There Segmentation in Ukraine?”, IZA Discussion Papers, No. 11256, Institute of Labor Economics (IZA), Bonn.

Lehmann, H. and Zaiceva, A. (2013), “Informal employment in Russia: Incidence, determinants and labor market segmentation”, Working paper No. 930, University of Bologna, Bologna, Italy.

Leontaridi, M. (1998), "Segmented Labour Markets: Theory and Evidence", *Journal of Economic Surveys*, Vol. 12 No. 1, pp. 103–109.

Loyaza, N. (1994), "Labor Regulations and the Informal Economy", Policy Research Working Paper 1335, World Bank.

Magnac, T. (1991), "Segmented or competitive labor markets", *Econometrica*, Vol. 59 No. 1, pp. 165–187.

Maloney, W. (2004), "Informality revisited", *World Development*, Vol. 32 No. 7, pp. 1159–1178.

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*, Vol. 52, pp. 637–670.

Nguyen, H.C., Nordman, C.J., and Roubaud F. (2013), "Who suffers the penalty? A panel data analysis of earnings gaps in Vietnam", *Journal of Development Studies*, Vol. 49 No. 12, pp. 1694–1710.

Nordman, C.J., Rakotomanana, F., and Roubaud, F. (2016), "Informal versus formal: a panel data analysis of earnings gaps in Madagascar", *World Development*, Vol. 86, pp. 1–7.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A., and Saavedra-Chanduvi, J. (2007), "Informality - Exit and Exclusion", The World Bank, Washington DC.

Pradhan, M. and van Soest, A. (1995), "Formal and informal sector employment in urban areas of Bolivia", *Labour Economics*, Vol. 2 No. 3, pp. 275–297.

Pratap, S. and Quintin, E. (2006), "Are labor markets segmented in developing countries? A semiparametric approach", *European Economic Review*, Vol. 50 No. 7, pp. 1817–1841.

Rosenbaum, P. R. and Rubin, D. B. (1985), "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score", *The American Statistician*, Vol. 39, pp. 33–38.

Rubin, D.B. (2001), "Using propensity scores to help design observational studies: Application to the tobacco litigation", *Health Services and Outcomes Research Methodology*, Vol. 2, pp. 169–188.

Staneva, A.V. and Arabsheibani, G.R. (2014), "Is there an informal employment wage premium? Evidence from Tajikistan", *IZA Journal of Labor Development*, Vol. 3 No. 1, pp. 1-24.

Statistics Poland (GUS) (2018), *Labour force survey in Poland IV quarter 2017*, Statistical information, Warsaw.

Statistics Poland (GUS) (2019), *Unregistered Employment in Poland in 2017*, Statistical information, Warsaw.

Tannuri-Pianto, M. and Pianto, D. (2002), "Informal employment in Brazil-a choice at the top and segmentation at the bottom: a quantile regression approach", *Serie de Textos para Discussao No. 236* Universidade de Brasilia, Brasilia, Brazil.

Tansel, A. (1997), "Informal sector earnings determination in Turkey", (in:) *Economic research forum (ERF), regional trade, finance and labor markets in transition, conference proceedings*, September, 7-9, 1997, Beirut, Lebanon, pp. 153-161.

Tansel, A. (2000), "Formal and informal sector choice of wage earners and their wages in Turkey", (in:) Bulutay, T. (ed.) *Informal Sector I*, State Institute of Statistics, Ankara, Turkey, pp. 125-150.

Tansel, A. (2002), "Wage earners, self-employment and gender in the informal sector in Turkey", (in:) *Economic research forum (ERF) working paper No. 0102*, Cairo, Egypt. Background paper for Engendering Development, World Bank, Washington, DC.

Tansel, A. and Kan, E.O. (2016), "The formal/informal employment earnings gap: evidence from Turkey", (in:) Bishop, J.A. and Rodríguez J.G. (Eds.), *Inequality after the 20th century: papers from the Sixth ECINEQ meeting (research on economic inequality)*, Vol. 24. Emerald Group Publishing Limited, Bingley, pp. 121-154.

Tansel, A., Keskin, H.I., and Ozdemir, Z.A. (2020), "Is there an informal employment wage penalty in Egypt? Evidence from quantile regression on panel data", *Empirical Economics*, Vol. 58 No. 6, pp. 2949-2979.

Tyrowicz, J. and Cichocki, S. (2011), "Employed unemployed? On shadow employment in transition", *Empirica*, Vol. 38 No. 2, pp. 259–281.

Williams, C.C., Bejakovic P., Mikulic D., Franic J., Kedir A., and Horodnic I.A. (2017), "An evaluation of the scale of undeclared work in the European Union and its structural determinants: estimates using the Labour Input Method", European Commission, Brussels, DOI: 10.2767/98169.

Appendix

Table A1. Variables in the wage equations

VARIABLES	DEFINITION / VALUE CLASSES
Dependent variables	
Earnings (ln)	The natural logarithm of net amount earned last month on the main job (in PLN) deflated with CPI (base year=2009)
Working time	The number of hours typically worked in a week on the main job
Hourly wage (ln)	The natural logarithm of net amount earned last month on the main job per one hour of working time (in PLN) deflated with CPI (base year=2009)
Independent variables	
<i>Informal employment</i>	
Employment without a written agreement	1 – yes; 0* – no
Employment while registered as unemployed	1 – yes; 0* – no
Controls	
Gender	1 – woman; 0* – man
Age, age square	continuous variables
Civil status	1 – single; 0* – married
Education level	1 – tertiary; 2 – post-secondary; 3 – secondary vocational; 4 – secondary general; 5 – basic vocational; 6* – primary or less
Firm size	1* – up to 10 employees; 2 – 11-49 employees; 3 – 50-250 employees; 4 – 251 employees or more; 5 – unknown
Firm ownership	1 – private; 0* – public
Firm economic sector	Binary variables for 21 sections (Level 1) of economic activity according to NACE-08.
Occupation	Binary variables for 10 major (one digit) groups of occupations according to ISCO-08.
Job tenure	continuous variable (years)
Internship	1 – employed as an intern; 0* – otherwise
Trial period	1 – employed on a trial period; 0* – otherwise
Place of residence	1* – town >100,000 inhabitants; 2 – town 50,000-100,000 inhabitants; 3 – town 20,000-50,000 inhabitants; 4 – town 10,000-20,000 inhabitants; 5 – town 5,000-10,000 inhabitants; 6 – town 2,000-5,000 inhabitants; 7 – town < 2,000 inhabitants; 8 – rural
Region (voivodship)	1* – Dolnośląskie; 2 – Kujawsko-pomorskie; 3 – Lubelskie; 4 – Lubuskie; 5 – Łódzkie; 6 – Małopolskie; 7 – Mazowieckie; 8 – Opolskie; 9 – Podkarpackie; 10 – Podlaskie; 11 – Pomorskie; 12 – Śląskie; 13 – Świętokrzyskie; 14 – Warmińsko-mazurskie; 15 – Wielkopolskie; 16 – Zachodniopomorskie
Survey year	Binary variables for the years 2009-2017

Note: asterisks indicate the base category. Source: Author's own elaboration.

Table A2. Descriptive statistics of the sample

VARIABLES	EMPLOYED INDIVIDUALS WITHOUT A WRITTEN AGREEMENT				EMPLOYED INDIVIDUALS WITH A WRITTEN AGREEMENT			
	MEAN	STD. DEV.	MIN	MAX	MEAN	STD. DEV.	MIN	MAX
Earnings (PLN)	1428.53	766.51	126.70	7288.89	1784.41	903.34	120.86	10666.67
Working time (hours per week)	40.39	10.60	0	140	39.94	6.80	0	140
Hourly wage (PLN)	8.44	4.51	0.70	72.59	10.61	5.58	0.51	273.26
Woman	0.319	0.466	0	1	0.489	0.500	0	1
Age	38.9	11.8	18	64	40.9	11.3	18	64
Single	0.444	0.497	0	1	0.305	0.460	0	1
Education level	4.325	1.521	1	6	3.278	1.699	1	6
Firm size	1.925	1.213	1	5	2.661	1.159	1	5
Firm ownership	0.886	0.318	0	1	0.668	0.471	0	1
Occupation	6.610	2.177	0	9	5.158	2.558	0	9
Job tenure	4.654	7.155	0	45.3	9.681	9.735	0	49.8
Internship	0.032	0.176	0	1	0.028	0.166	0	1
Trial period	0.042	0.200	0	1	0.024	0.152	0	1
Place of residence	5.146	3.932	0	9	4.277	3.880	0	9
Number of observations	10,054				322,129			

Source: Author's own analyses based on unit data from the Polish LFS, 2009–2017.

Table A2. Descriptive statistics of the sample (cont.)

VARIABLES	EMPLOYED INDIVIDUALS WHO ARE REGISTERED AS UNEMPLOYED				EMPLOYED INDIVIDUALS WHO ARE NOT REGISTERED AS UNEMPLOYED			
	MEAN	STD. DEV.	MIN	MAX	MEAN	STD. DEV.	MIN	MAX
Earnings (PLN)	1086.24	614.01	140.19	6250.00	1779.25	901.39	120.86	10666.67
Working time (hours)	37.92	12.44	2	140	39.97	6.88	0	140
Hourly wage (PLN)	6.81	3.29	1.39	37.09	10.57	5.57	0.51	273.26
Woman	0.407	0.491	0	1	0.485	0.500	0	1
Age	36.7	12.4	18	64	40.9	11.3	18	64
Single	0.544	0.498	0	1	0.307	0.461	0	1
Education level	4.310	1.505	1	6	3.301	1.702	1	6
Firm size	1.873	1.231	1	5	2.645	1.165	1	5
Firm ownership sector	0.781	0.414	0	1	0.674	0.469	0	1
Occupation	6.770	2.090	0	9	5.189	2.559	0	9
Job tenure	1.839	4.212	0	40.5	9.592	9.713	0.0	49.8
Internship	0.203	0.402	0	1	0.027	0.162	0	1
Trial period	0.044	0.205	0	1	0.024	0.153	0	1
Place of residence	5.320	3.786	0	9	4.295	3.884	0	9
Number of observations	2,690				329,493			

Source: Author's own analyses based on unit data from the Polish LFS, 2009–2017.

Table A3. Determinants of employment without a written agreement (probit estimates)

Model specification	COEFFICIENT	STANDARD ERROR	STATISTIC Z	PR > Z
Woman	-0.077	0.015	-5.10	0.000
Age	0.018	0.004	4.53	0.000
Age square	0.000	0.000	-3.83	0.000
Single	0.152	0.013	11.58	0.000
Education level: tertiary	-0.449	0.030	-15.03	0.000
Post-secondary	-0.455	0.042	-10.87	0.000
Secondary vocational	-0.341	0.021	-16.26	0.000
Secondary general	-0.286	0.026	-11.00	0.000
Basic vocational	-0.212	0.018	-11.98	0.000
Firm size: 11-49 employees	0.467	0.024	19.85	0.000
50-250 employees	-0.001	0.024	-0.04	0.971
251 employees or more	-0.144	0.026	-5.59	0.000
Firm size: unknown	-0.185	0.029	-6.32	0.000
Firm ownership: private	0.127	0.025	5.17	0.000
Firm economic sector: non-individual agriculture, forestry or fishing	-0.608	0.045	-13.61	0.000
Mining and Quarrying	-1.212	0.074	-16.31	0.000
Manufacturing	-1.000	0.033	-29.92	0.000
Electricity, Gas, Steam and Air Conditioning Supply	-0.874	0.083	-10.48	0.000
Water Supply; Sewerage, Waste Management and Remediation Activities	-1.045	0.058	-18.03	0.000
Construction	-0.510	0.033	-15.37	0.000
Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles	-1.106	0.036	-30.37	0.000
Transportation and Storage	-1.056	0.041	-25.51	0.000
Accommodation and Food Service Activities	-0.961	0.044	-21.62	0.000
Information and Communication	-1.165	0.066	-17.61	0.000
Financial and Insurance Activities	-1.179	0.066	-17.92	0.000
Real Estate Activities	-1.141	0.078	-14.58	0.000
Professional, Scientific and Technical Activities	-1.001	0.058	-17.12	0.000
Administrative and Support Service Activities	-1.209	0.044	-27.31	0.000
Public Administration and Defence	-0.889	0.049	-17.99	0.000
Education	-1.022	0.046	-22.22	0.000
Human Health and Social Work Activities	-0.930	0.044	-20.94	0.000
Arts, Entertainment and Recreation	-1.030	0.072	-14.37	0.000
Other service activities	-0.473	0.045	-10.57	0.000
Unknown	-0.733	0.190	-3.85	0.000

Model specification	COEFFICIENT	STANDARD ERROR	STATISTIC Z	PR > Z
Occupation: managers	-0.427	0.094	-4.56	0.000
Professionals	-0.197	0.086	-2.29	0.022
Technicians and associate professionals	-0.346	0.085	-4.08	0.000
Clerical support workers	-0.312	0.086	-3.63	0.000
Service and sales workers	-0.336	0.086	-3.92	0.000
Skilled agricultural, forestry and fishery workers	-0.078	0.096	-0.81	0.416
Craft and related trades workers	-0.147	0.084	-1.75	0.080
Plant and machine operators and assemblers	-0.363	0.085	-4.26	0.000
Elementary occupations	-0.019	0.084	-0.22	0.823
Job tenure	-0.058	0.002	-25.78	0.000
Job tenure square	0.001	0.000	20.65	0.000
Internship	-0.156	0.035	-4.48	0.000
Trial period	0.052	0.030	1.76	0.079
Place of residence:				
town 50,000-100,000 inhabitants	-0.083	0.024	-3.42	0.001
Town 20,000-50,000 inhabitants	-0.091	0.022	-4.18	0.000
Town 10,000-20,000 inhabitants	0.004	0.024	0.15	0.880
Town 5,000-10,000 inhabitants	-0.039	0.034	-1.14	0.256
Town 2,000-5,000 inhabitants	-0.078	0.038	-2.05	0.041
Town < 2,000 inhabitants	-0.149	0.101	-1.47	0.142
Rural	-0.069	0.017	-4.13	0.000
Region: Kujawsko-pomorskie	-0.019	0.031	-0.63	0.532
Lubelskie	0.194	0.027	7.08	0.000
Lubuskie	0.049	0.032	1.51	0.130
Łódzkie	0.088	0.029	3.00	0.003
Małopolskie	0.017	0.034	0.49	0.626
Mazowieckie	0.203	0.026	7.88	0.000
Opolskie	-0.206	0.034	-6.13	0.000
Podkarpackie	-0.234	0.037	-6.35	0.000
Podlaskie	-0.061	0.028	-2.13	0.033
Pomorskie	0.131	0.029	4.58	0.000
Śląskie	0.160	0.030	5.38	0.000
Świętokrzyskie	-0.228	0.033	-6.90	0.000
Warmińsko-mazurskie	0.002	0.029	0.08	0.939
Wielkopolskie	-0.398	0.034	-11.85	0.000
Zachodniopomorskie	-0.007	0.033	-0.20	0.842
Survey year: 2010	-0.214	0.022	-9.88	0.000
2011	-0.180	0.022	-8.10	0.000

Model specification	COEFFICIENT	STANDARD ERROR	STATISTIC Z	PR > Z
2012	-0.232	0.024	-9.80	0.000
2013	-0.272	0.024	-11.12	0.000
2014	-0.293	0.022	-13.07	0.000
2015	-0.271	0.024	-11.44	0.000
2016	-0.309	0.024	-12.61	0.000
2017	-0.408	0.026	-15.46	0.000
Constant	-0.572	0.126	-4.55	0.000
Pseudo R2		0.1998		
No of observations		330,889		

Notes: ***/**/* stand for 1%, 5% and 10% significance, respectively; standard errors in brackets.

Source: Author's own analyses based on unit data from the Polish LFS, 2009–2017.

Table A4. Effects of employment without a written agreement (full set of OLS estimates)

MODEL SPECIFICATION	EARNINGS	HOURLY WAGE	WORKING TIME
	(1)	(2)	(3)
Employment without a written contract	-0.091*** (0.006)	-0.054*** (0.005)	-0.628*** (0.123)
Woman	-0.201*** (0.002)	-0.161*** (0.001)	-1.421*** (0.030)
Age	0.029*** (0.001)	0.020*** (0.001)	0.274*** (0.012)
Age square	-0.000*** (0.000)	-0.000*** (0.000)	-0.004*** (0.000)
Single	-0.041*** (0.002)	-0.036*** (0.001)	-0.138*** (0.033)
Education level: tertiary	0.292*** (0.004)	0.288*** (0.004)	-0.064 (0.079)
Post-secondary	0.154*** (0.005)	0.141*** (0.004)	0.229** (0.094)
Secondary vocational	0.147*** (0.004)	0.129*** (0.003)	0.402*** (0.069)
Secondary general	0.145*** (0.004)	0.137*** (0.003)	0.158** (0.080)
Basic vocational	0.073*** (0.003)	0.057*** (0.003)	0.426*** (0.067)
Firm size: 11-49 employees	0.087*** (0.002)	0.063*** (0.002)	0.546*** (0.047)
50-250 employees	0.140*** (0.002)	0.098*** (0.002)	1.028*** (0.048)
251 employees or more	0.217*** (0.003)	0.170*** (0.002)	1.184*** (0.050)
Firm size: unknown	0.083*** (0.004)	0.071*** (0.003)	0.305*** (0.074)
Firm ownership: private	0.022*** (0.003)	0.004* (0.002)	0.846*** (0.049)
Firm economic sector: non-individual agriculture, forestry or fishing	0.006 (0.010)	0.033*** (0.009)	-1.057*** (0.209)
Mining and Quarrying	0.191*** (0.010)	0.256*** (0.010)	-2.281*** (0.202)

MODEL SPECIFICATION	EARNINGS	HOURLY WAGE	WORKING TIME
	(1)	(2)	(3)
Manufacturing	-0.047*** (0.009)	0.010 (0.008)	-2.168*** (0.183)
Electricity, Gas, Steam and Air Conditioning Supply	0.077*** (0.010)	0.146*** (0.009)	-2.400*** (0.199)
Water Supply; Sewerage, Waste Management and Remediation Activities	-0.031*** (0.010)	0.018** (0.009)	-1.864*** (0.202)
Construction	0.101*** (0.009)	0.089*** (0.008)	0.577*** (0.189)
Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles	-0.052*** (0.009)	0.001 (0.008)	-1.984*** (0.187)
Transportation and Storage	0.044*** (0.009)	0.070*** (0.008)	-0.569*** (0.191)
Accommodation and Food Service Activities	-0.054*** (0.010)	0.013 (0.008)	-2.146*** (0.210)
Information and Communication	-0.020* (0.011)	0.060*** (0.010)	-2.824*** (0.219)
Financial and Insurance Activities	0.037*** (0.010)	0.083*** (0.009)	-1.677*** (0.204)
Real Estate Activities	-0.055*** (0.011)	0.033*** (0.010)	-3.031*** (0.221)
Professional, Scientific and Technical Activities	-0.082*** (0.011)	-0.020** (0.009)	-2.190*** (0.211)
Administrative and Support Service Activities	-0.182*** (0.009)	-0.109*** (0.008)	-2.432*** (0.202)
Public Administration and Defence; Compulsory Social Security	-0.030*** (0.009)	0.027*** (0.008)	-1.808*** (0.193)
Education	-0.115*** (0.009)	0.048*** (0.008)	-4.962*** (0.194)
Human Health and Social Work Activities	-0.150*** (0.009)	-0.097*** (0.008)	-1.614*** (0.194)
Arts, Entertainment and Recreation	-0.151*** (0.011)	-0.019* (0.010)	-3.818*** (0.239)
Other service activities	-0.130*** (0.013)	-0.017* (0.010)	-3.318*** (0.263)
Unknown	-0.057* (0.031)	0.023 (0.027)	-2.768*** (0.513)
Occupation: managers	0.077*** (0.007)	0.037*** (0.007)	1.556*** (0.085)

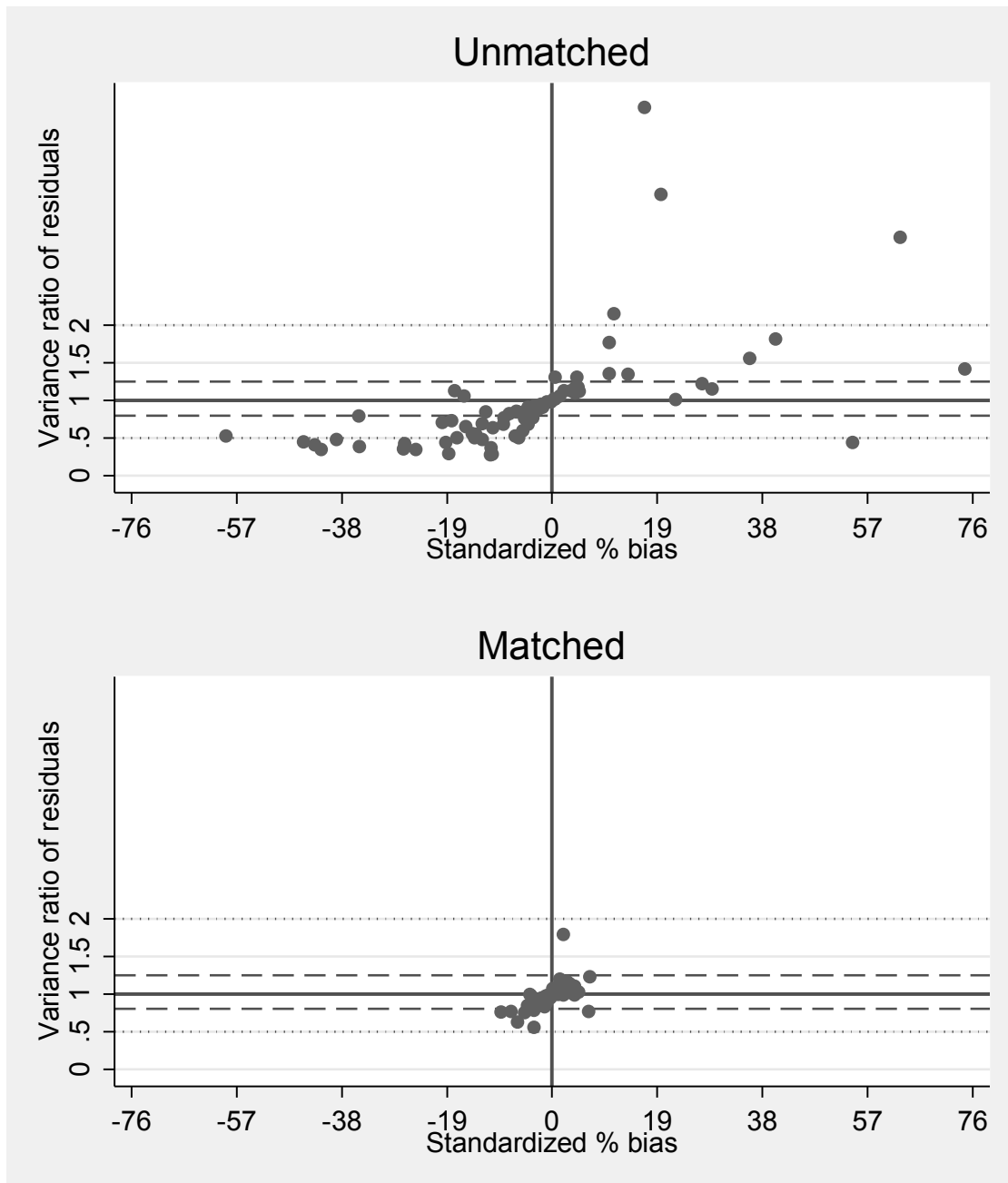
MODEL SPECIFICATION	EARNINGS	HOURLY WAGE	WORKING TIME
	(1)	(2)	(3)
Professionals	-0.085*** (0.006)	-0.059*** (0.006)	-0.413*** (0.077)
Technicians and associate professionals	-0.202*** (0.006)	-0.208*** (0.006)	0.399*** (0.070)
Clerical support workers	-0.358*** (0.006)	-0.364*** (0.006)	0.228*** (0.076)
Service and sales workers	-0.376*** (0.006)	-0.402*** (0.006)	1.076*** (0.083)
Skilled agricultural, forestry and fishery workers	-0.383*** (0.013)	-0.400*** (0.011)	0.774*** (0.224)
Craft and related trades workers	-0.335*** (0.006)	-0.355*** (0.006)	0.913*** (0.077)
Plant and machine operators and assemblers	-0.314*** (0.006)	-0.344*** (0.006)	1.384*** (0.078)
Elementary occupations	-0.463*** (0.007)	-0.447*** (0.006)	-0.155* (0.089)
Job tenure	0.013*** (0.000)	0.009*** (0.000)	0.086*** (0.005)
Job tenure square	-0.000*** (0.000)	-0.000*** (0.000)	-0.002*** (0.000)
Internship	-0.212*** (0.006)	-0.203*** (0.005)	-0.495*** (0.095)
Trial period	-0.037*** (0.005)	-0.045*** (0.004)	0.145 (0.095)
Place of residence: town 50,000-100,000 inhabitants	-0.034*** (0.003)	-0.035*** (0.002)	-0.007 (0.050)
Town 20,000-50,000 inhabitants	-0.041*** (0.002)	-0.046*** (0.002)	0.130*** (0.046)
Town 10,000-20,000 inhabitants	-0.042*** (0.003)	-0.048*** (0.002)	0.170*** (0.050)
Town 5,000-10,000 inhabitants	-0.049*** (0.004)	-0.054*** (0.004)	0.137** (0.070)
Town 2,000-5,000 inhabitants	-0.039*** (0.004)	-0.045*** (0.004)	0.229*** (0.083)
Town < 2,000 inhabitants	-0.039*** (0.011)	-0.043*** (0.011)	0.244 (0.266)

MODEL SPECIFICATION	EARNINGS	HOURLY WAGE	WORKING TIME
	(1)	(2)	(3)
Rural	-0.050*** (0.002)	-0.056*** (0.002)	0.191*** (0.038)
Region: Kujawsko-pomorskie	-0.070*** (0.004)	-0.063*** (0.003)	-0.189*** (0.072)
Lubelskie	-0.101*** (0.003)	-0.086*** (0.003)	-0.490*** (0.068)
Lubuskie	-0.036*** (0.004)	-0.034*** (0.004)	-0.117 (0.075)
Łódzkie	-0.054*** (0.003)	-0.053*** (0.003)	0.005 (0.070)
Małopolskie	-0.032*** (0.004)	-0.016*** (0.004)	-0.585*** (0.074)
Mazowieckie	0.054*** (0.003)	0.038*** (0.003)	0.676*** (0.065)
Opolskie	-0.030*** (0.004)	-0.011*** (0.003)	-0.700*** (0.071)
Podkarpackie	-0.129*** (0.003)	-0.116*** (0.003)	-0.703*** (0.065)
Podlaskie	-0.066*** (0.003)	-0.088*** (0.003)	0.704*** (0.065)
Pomorskie	0.017*** (0.004)	0.015*** (0.003)	0.326*** (0.074)
Śląskie	-0.054*** (0.004)	-0.035*** (0.003)	-0.441*** (0.072)
Świętokrzyskie	-0.090*** (0.004)	-0.094*** (0.003)	0.141** (0.072)
Warmińsko-mazurskie	-0.054*** (0.003)	-0.052*** (0.003)	-0.099 (0.070)
Wielkopolskie	-0.048*** (0.004)	-0.038*** (0.003)	-0.190*** (0.071)
Zachodniopomorskie	0.004 (0.004)	-0.003 (0.004)	0.366*** (0.080)
Survey year: 2010	0.046*** (0.003)	0.049*** (0.003)	-0.108* (0.060)
2011	0.088*** (0.003)	0.089*** (0.003)	-0.081 (0.061)

MODEL SPECIFICATION	EARNINGS	HOURLY WAGE	WORKING TIME
	(1)	(2)	(3)
2012	0.133***	0.129***	0.053
	(0.003)	(0.003)	(0.061)
2013	0.159***	0.152***	0.124**
	(0.003)	(0.003)	(0.062)
2014	0.194***	0.182***	0.186***
	(0.003)	(0.003)	(0.058)
2015	0.225***	0.211***	0.207***
	(0.003)	(0.003)	(0.058)
2016	0.259***	0.246***	0.198***
	(0.003)	(0.003)	(0.062)
2017	0.318***	0.307***	0.127**
	(0.003)	(0.003)	(0.062)
Constant	6.885***	1.909***	35.538***
	(0.017)	(0.015)	(0.322)
Number of observations	330,889	330,303	330,305
R2	0.479	0.493	0.115

Notes: ***/**/* stand for 1%, 5% and 10% significance, respectively; standard errors in brackets.
Source: Author's own analyses based on unit data from the Polish LFS, 2009–2017.

Figure A1. Independent variables balance (PSM, NN5, pooled sample)



Source: Author's own analyses based on unit data from the Polish LFS, 2009-2017.