LATVIJAS BANKA
EIROSISTĒMA

ISBN 978-9934-578-41-0

KONSTANTĪNS BEŅKOVSKIS, LUDMILA FADEJEVA

CHASING THE SHADOW: THE EVALUATION OF UNREPORTED WAGE PAYMENTS IN LATVIA

WORKING PAPER
1 / 2022

Latvijas Banka
K. Valdemāra iela 2A, Riga, LV-1050
Tel.: +371 67022300 info@bank.lv
http://www.bank.lv https://www.macroeconomics.lv

Chasing the Shadow: the Evaluation of Unreported Wage Payments in Latvia

Konstantīns Beņkovskis[*], Ludmila Fadejeva[†]

February 2, 2022

Abstract

We develop a novel way to evaluate the size of unreported wage payments at employee level. It is only the reported employer-employee income data combined with firm-level financial statements and survey information on various person-level indicators that are required for this purpose. We estimate the Mincer earning regression by the Stochastic Frontier Analysis approach, proxying the unreported wage payments by the non-negative inefficiency term. Our methodology is tested on the Latvian data: we find that small and young firms engage in illegal wage payments more than other firms. Unofficial payments to employees with small reported wages are more frequent and sizeable, revealing lower wage income inequality in Latvia when the unreported wage is taken into account.

Keywords: unreported wage, tax evasion, Mincer earning regression, income distribution

JEL Codes: E26, H26, J08, J31

[*]Monetary Policy Department, Latvijas Banka, K. Valdemāra iela 2A, Rīga, LV-1050, Latvija; Economic Department, Stockholm School of Economics in Riga, Strēlnieku iela 4A, Rīga, LV-1010, Latvija; e-mail: Konstantins.Benkovskis@bank.lv

[†]Monetary Policy Department, Latvijas Banka, K. Valdemāra iela 2A, Rīga, LV-1050, Latvija; e-mail: Ludmila.Fadejeva@bank.lv

1 Introduction

The shadow economy, comprising activities that have market value and would add to tax revenue and GDP were they recorded, is a widespread global phenomenon (International Monetary Fund, 2021). As with any fraudulent behaviour, the size of unreported cash payments is difficult to observe and measure. Such activities have significant economic and social implications across several dimensions. At macro level, weakened tax revenue materialises in lower provision of public goods. At firm level, the shadow economy results in distortion of resource allocation due to unfair competitive advantage of tax evading entities. At employee level, receiving unreported income worsens social protection and leads to low pension savings and unemployment benefits. Detailed employee and firm level data allows measuring unreported wages at disaggregated level, which is essential for better policy design.

Several approaches to quantify the size of unreported wage payments exist in the empirical literature. They can be classified into three broad groups. First, audit data from fiscal authorities or surveys provide direct information on the frequency and size of illegal payments. Audit information, as in Kleven et al. (2011), may represent the most effective way to evaluate the size of unreported payments, but such data is still rare. Also, this source of information can provide a biased picture in case of a non-random audit. Direct surveys of employees (see e.g. Eurobarometer 2020) or employers (see Putniņš and Sauka 2015) are an effective alternative in the absence of audit data. Possible untruthful answers combined with relatively small samples limit the usability of such results. The second strand of the literature – consumption-based analysis, taking its origins from Pissarides and Weber (1989) – evaluates the size of envelope payments based on discrepancies between (under)reported income and reported consumption. Despite a solid theoretical background and clear intuition, such evaluations rarely go beyond aggregate numbers: the availability and the size of consumption data at household level remains a serious constraint. The third approach relies on discrepancies between administrative data and self-reported income (see Kumler et al., 2020). Still, untruthful answers may inject a systematic bias, while respondent inattention induces additional noise in the evaluation of "envelope" cash payments.

We propose a novel approach to evaluate the size of unreported wage payments at employee level based on employer-employee administrative data on income as collected by the respective fiscal authority. Data on reported income is combined with a firm's financial statements and survey data containing various person-level indicators like education, experience and contract type. The increasing availability of such data allows implementing our approach internationally. The methodology consists of three steps, where the first two follow the very recent research by Gavoille and Zasova (2021b), who identify firms engaged in labour tax evasion using machine learning techniques and investigate the employment reaction to a minimum wage shock for compliant and tax-evading firms. The first stage of the methodology requires defining the treated group of firms ("definitely evading") and the control group ("definitely compliant"). Ideally, this should be done on the basis of fiscal audit information. In the absence of such data, we identify the latter group with state-owned firms, as well as enterprises whose owners are located in low-corruption countries. Here we rely on evidence that the owner may transfer business ethics and tax morale from her country of origin (see e.g. Braguinsky et al., 2014; DeBacker et al., 2015). The former group of "definite evaders" is identified with the firms that pay "suspiciously low" wages for a given level of occupation, region, age and gender. The second stage introduces a model trained/estimated on the sample of definitely compliant and evading firms. This model classifies the rest of the firms as compliant or tax-evading based on their financial reports. While Gavoille and Zasova (2021b) use the random forest algorithm, we employ a simple probit model instead to obtain the probability that each firm is involved in tax evasion.

Our major contribution to the literature on tax evasion is the third step of the methodology – evaluating the unreported wage. We estimate the Mincer earning regression (Mincer, 1958; Mincer, 1974) by the Stochastic Frontier Analysis (SFA) method, which allows us to include two stochastic components in the model. The first one is the traditional idiosyncratic error term accounting for the unobserved wage determinants like ability. The second – the non-negative inefficiency term – serves as a proxy for the unobserved and unreported wage payment. We introduce heterogeneity in the inefficiency term by linking its variance to the predicted probability that the respective firm will evade labour taxes (obtained in the second step). We thus restrict the unreported wage payments to zero for compliant firms, while allowing positive illegal cash payments for tax-evading firms. Since the Mincer earning regression is estimated at the employee level, we evaluate unreported wage payments for each worker included in our sample. Intuitively, the envelope payments are proxied by the gap between the potential gross wage (given the occupation, age, gender, education and many other employee- and firm-level indicators) and reported gross wage net of the idiosyncratic term.

We apply our novel methodology to Latvian data. Under-reported wage is a widespread phenomenon in Latvia investigated by several researchers. In particular, Putniņš and Sauka (2015) report that 21.1% of GDP is in the shadow economy in Latvia; 40% of the shadow economy consists of envelope wages. Eurobarometer (2020) suggests a larger frequency and size of unreported wages in Latvia compared to the EU. Gavoille and Zasova (2021a) conclude that workers in domestically-owned firms conceal 26% more income than employees in foreign-owned firms. As to the data, the good quality employer-employee earning information collected by the State Revenue Service (SRS) of Latvia combined with the firm-level financial statements is available for recent years. Additional person-level indicators can be obtained from the Labour Force Survey (LFS) and merged with the employer-employee database. The availability of this multifaceted micro level data and numerous previous evaluations based on different approaches makes Latvia an ideal testing ground for our study.

While the firm sample used in estimation does not include public institutions, commercial banks, micro-enterprises and the self-employed, it still covers the majority of economic activity in Latvia. According to our estimates, more that 40% of employees included in our sample were involved in labour tax fraud. For the employees with non-zero unreported wage, the size of such payments averaged nearly 30% of the reported gross wage, making the share of envelope wage payments close to 10% of total reported gross wages. These aggregate numbers hide substantial heterogeneity in labour tax evasion. Small and young firms engage in illegal envelope wage payments to a larger extent than other firms. Unofficial payments are more frequent and sizeable for employees with low reported wages, revealing lower wage income inequality in Latvia when the unreported wage is taken into account.

We review the approaches to evaluate the size of envelope payments in the next Section in more detail. Section 3 describes the Latvian data, while Section 4 goes through all three steps of our novel methodology. Section 5 briefly describes our results regarding the detection of tax-evading firms. The main results relate to the size and distribution of unreported wage payments, including validation and robustness checks reported in Section 6. The last Section concludes.

2 Literature review

The literature on under-reporting of income is vast and continuously growing. The under-reporting of income is usually documented in developing countries, for example, Gorodnichenko et al. (2009)

discover large changes in tax evasion following the flat tax reform in Russia, Tonin (2011) investigates the interaction between the minimum wage and labour tax evasion in Hungary, Kukk and Staehr (2014) estimate the relative under-reporting of income for households with business income in Estonia, and Kumler et al. (2020) find evidence of income under-reporting by firms in Mexico. This phenomenon is not specific to developing countries only, as Hurst et al. (2014) document income under-reporting by the self-employed in the US.

The issues of under-reported income and envelope wages in Latvia received a lot of attention in recent years from both international institutions (see World Bank, 2017 and Eurobarometer, 2020) and academic researchers. According to a recent (September 2019) survey by Eurobarometer (2020), 36% of Latvian respondents indicated that they know people who work without declaring all or part of their income, while 6% admitted that they carried out undeclared paid activities in the last 12 months (the share of positive answers in the EU was 33% and 3% respectively). Putniņš and Sauka (2015) estimate the size of the shadow economy for three Baltic States in 2009-2012, reporting 21.1% of GDP for Latvia in 2012: envelope wages contribute almost 40% of this number, while unreported employees contribute another 20%. Putniņš and Sauka (2021) construct the SSE Riga Shadow Economy Index based on the same methodology and report the gradual increasing trend in the size of the shadow economy in Latvia between 2016 and 2020. Gavoille and Zasova (2021a) use expenditure-based measures of income and conclude that households with the head working in domestically-owned firms conceal 26% more income on average than those whose head works in foreign-owned firms. Gavoille and Zasova (2021b) propose a novel approach to detect labour tax avoiding firms: according to their results, 37% of firms in 2011-2013 were classified as labour-tax-evading.

Despite obvious complications in evaluating the magnitude of the non-compliance, income under-reporting and the size of the envelope payments, several evaluation strategies exist in the literature. One strand of the literature relies on the direct surveying of the income under-reporting phenomena. Eurobarometer (2020) directly asks respondents about their involvement in the undeclared economic activities, both from the demand side (e.g. buying products that included undeclared work) and the supply side (e.g. receiving undeclared income). Putniņš and Sauka (2015) survey company managers, using their unique knowledge about under-reported business income and wages. Given the sensitivity of the topic, the survey-based measurements may underestimate the size of undeclared payments because of non-response or untruthful response. This risk can be reduced by several survey and data collection techniques. In particular, Putniņš and Sauka (2015) ask managers about the prevailing practice in their industry instead of their firm. Similarly, Eurobarometer (2020) asks whether respondents "personally know any people who work without declaring all or part of their income to tax or social security authorities". Running a survey is time consuming and expensive, thus the sample size typically remains modest (1006 interviews in Eurobarometer, 2020; 500-600 phone interviews in Latvia each year by Putniņš and Sauka, 2015) that does not allow studying the distribution of unreported income in detail.

Few papers exploit data on audits performed by fiscal authorities. For example, Kleven et al. (2011) exploit data from a tax enforcement field experiment in Denmark to find that under-reporting is widespread for self-reported income but not third-party reporting. Such data is still rare, and some papers make use of exogenous shifts in the threat of audit. For example, Almunia and Lopez-Rodriguez (2018) exploit the revenue threshold for more thorough monitoring in Spain, while Pomeranz (2015) investigate announcements of additional monitoring on transactions of Chilean firms.

Consumption-based analysis is another big strand of the empirical literature on income under-reporting. Established by Pissarides and Weber (1989), this approach is based on the assumption that all income groups report food expenditures correctly, and compare food consumption for a household group with respect to some compliant benchmark to reveal the level of income under-reporting. The consumption-based approach is typically applied to evaluate the under-reported income of the self-employed (see e.g. Hurst et al., 2014 for the US; and Kukk et al., 2020 for the EU countries; see also Kukk and Staehr, 2014 for households with business income in Estonia). This is not the only possible comparison, and Gavoille and Zasova (2021a) compare Latvian households whose heads work in domestically-owned firms with households whose heads work in foreign-owned firms (assumed to be compliant), suggesting that labour tax evasion by the former can provide an alternative explanation for the wage premium for employees in foreign-owned firms.

One can also rely on discrepancies between different income data sources, when they usually contain self-reporting or assessment data. For instance, Kumler et al. (2020) compare income data from the social security agency with self-reported income from the household survey and discovers substantial unreported income for the employees of Mexican firms, which is especially widespread among small firms. Artavanis et al. (2016) replicate the banks' estimate of household income to compare it with officially reported information.

3 Data

We use the availability of rich administrative and survey micro-level data in Latvia to perform our task. There are three main datasets that were linked using anonymized firm and employee identificators.[1] First, we use the employer-employee dataset provided by the State Revenue Service (SRS) and Central Statistical Bureau (CSB) of Latvia. This administrative dataset contains monthly gross wage for all employees reported by firms to Latvia's tax authorities (excluding the self-employed persons). In addition to income data, the dataset includes hours worked, age and gender of each employee. The monthly income dataset is accompanied by another employer-employee database that records any changes in an employee's status like hiring, firing, transfer to a new position, parental leave, etc. Starting from July 2015, each such entry should contain a six-digit employee occupation code corresponding to the Latvian profession classification.[2] Although information on changes in an employee's status arrives infrequently, we are able to impute occupation information for many observations in the monthly employer-employee gross income dataset. We assume unchanged occupation following the moment of hiring or transfer to a new position until a new change in status; similarly we trace back the occupation for months preceding the moment when the employee was fired. Table 1 indicates that we were able to trace the occupation status for almost 15% of all employer-employees pairs between 2016 and 2018.[3,4]

Despite the relatively low coverage for information on occupations, we still have enough observations to detect "suspiciously low" wages at the individual profession level, which is a crucial step in our methodology (see Section 4.1). The subsample of employees for whom we can identify occupations is representative of the population in terms of reported gross wage statistics, as shown in the last two columns of Table 1 comparing the average gross full time equivalent (FTE) wage for employer-employee pairs with identified occupations and the aggregate statistics provided by the statistical office (see also Figure A1 in Appendix for similar comparison by broad occupation groups).

[1]Firm registration numbers and personal codes are replaced by unique hash codes. While making the identification of firms and employees impossible, these unique hash codes are consistent across various datasets, allowing to link datasets together.

[2]The first four digits of the occupation code coincide with ISCO-08 classification.

[3]Information on occupation is missing if the employee's status remained unchanged after June 2015. In addition, a substantial fraction of firms does not report occupation data even after July 2015. Finally, we assigned occupation data only to employees working for at least four months within a respective calendar year to exclude short-term employees.

[4]Latvia's employer-employee level dataset is available for 2007-2020. However, firms only report occupations since July 2015. The sample ends in 2018 due to the lagged availability of firm-level financial statements.

Table 1: Employer-employee pairs with detected occupation

Year	Total number of unique employer-employee pairs in the dataset	Pairs with occupation	Share of observations with occupations	Average gross FTE wage for pairs with occupation, EUR	Average gross wage according to CSB of Latvia, EUR
2016	1'278'735	165'786	13.0%	875.53	859
2017	1'305'309	191'169	14.6%	961.34	926
2018	1'402'464	201'923	14.4%	1087.78	1004

Sources: SRS of Latvia, CSB of Latvia, own calculations.

The firm-level financial statements provided by the SRS and CSB of Latvia serve as the second major source of information. It contains firm balance sheets and profit and loss statements, including turnover, total compensation of employees and value added. It also indicates the four-digit sector of activity according to NACE 2 classification, the year of establishment, NUTS3 region and ownership code that allows detecting state-owned enterprises.[5] The firm-level financial data is available until 2018, which limits the sample period of our research to 2016-2018. The number of firms included in the dataset varies between 117'292 in 2016 and 107'880 in 2018; it does not include public institutions, commercial banks and the self-employed. Moreover, data availability is scarce for small enterprises due to non-reporting. To track the country of origin for foreign-owned firms in Latvia, balance sheets and profit and loss statements are also linked with the information about firms' foreign assets and liabilities provided by Latvijas Banka. The merchandise trade dataset provided by the CSB adds information on export and import operations.

The Labour Force Survey (LFS) collected and provided by the CSB is our third source of information. The survey uses a rotating panel, collecting the data at household level for people aged from 15 to 74. Each household is surveyed four times with intervals of 13, 39 and 13 weeks; the survey covers more than 10 thousand households per year. This information can be linked with administrative employer-employee data using an anonymised person identifier. Although the sample size of the survey is substantially smaller than for administrative data, the LFS provides critical information on personal characteristics like education, experience, contract type, family status, etc. We need this information to evaluate the potential gross wage in Section 4.3. The survey also provides another especially valuable information for our research – directly asking employed respondents to report their net income in the previous month. As in Kumler et al. (2020), this allows us to compare official wage data by tax authorities to a self-reported income data.

[5]The ownership code is available until 2016. We assume that enterprises that were labelled as state-owned in 2016 remained in state ownership also in 2017 and 2018.

4 Methodology

This section describes our novel methodology to evaluate the size of unreported wage payments. It consists of three stages. The first two stages are devoted to identifying tax-evading firms following the approach suggested by Gavoille and Zasova (2021b). The third stage contains the main methodological innovation by evaluating unofficial payments at the employer-employee level. We describe all three steps below.

4.1 Forming the treated and control group

First, we define the group of firms that are assumed to be "definitely compliant", and the group of firms assumed to be "definitely evading" the labour tax. The former serves as control group, while the latter serves as treatment group for the second stage model predicting the probability that a firm evades labour taxes.

Foreign and state-owned firms as a control group

The control group contains firms that is assumed to be "definitely compliant" and do not evade labour taxes. Ideally, one can use data from audits performed by fiscal authorities (e.g. Kleven et al., 2011) to define both control and treated groups. Such data is rarely available and we do not have it at hand for Latvia. Instead we follow the approach of Gavoille and Zasova (2021b) who assume full compliance of Scandinavian-owned firms in Latvia. They base this assumption on the existing empirical evidence of higher transparency among firms whose owners originate from countries with a low corruption environment. For instance, DeBacker et al. (2015) employ the US audit data to show that such firms are less likely to evade taxes, while Braguinsky et al. (2014) and Braguinsky and Mityakov (2015) use Russian administrative data to find higher wage transparency in multinational firms.

We use a similar approach, although we expand the set of firms that are assumed to be "definitely compliant" in order to increase the size of the control group and to improve the stability of estimates in the second step. We expand the set of compliant firms in two ways. First, instead of owners from Scandinavian countries only, we use the top countries from the Corruption Perception Index.[6]

[6] We used the top 15 countries from the Corruption Perception Index of 2016 (see https://www.transparency.org/en/cpi/2016/index/nzl) and removed the countries that have negligible direct investments in Latvia. The final list consists of 11 countries: Denmark, Finland, Sweden, Switzerland, Norway, the Netherlands, Canada, Luxembourg, Germany, the UK and Iceland.

While this includes all Scandinavian countries, it adds several other OECD countries, increasing the size of the control group considerably, as Germany and the UK are among the top sources of the FDI in Latvia. Second, we add state-owned firms to the "definitely compliant" group. These are subject to more rigorous control, in particular from the State Audit Office of Latvia, minimizing the opportunities for tax evasion. Also Braguinsky et al. (2014) suggest substantially higher compliance among state-owned enterprises.

Low-wage firms as a treated group

Defining the group of "definitely evading" firms is more complicated. Although various empirical studies point at certain groups of tax evaders, like the self-employed (e.g. Hurst et al., 2014; Kukk et al., 2020) or small enterprises (e.g. Kumler et al., 2020; Braguinsky and Mityakov, 2015), this does not mean that all firms within such a group evade labour taxes. In the absence of tax audit data, one feasible solution suggested by Gavoille and Zasova (2021b) is to select firms with "suspiciously low" wages. Due to the flexibility of the Latvian labour market (see Fadejeva and Opmane, 2016), it is reasonable to assume that employees only agree to receive an official wage substantially below the level considered to be normal if this difference is compensated by some unofficial payment in cash.

Gavoille and Zasova (2021b) use various firm-level information, as well as employee-level characteristics from the LFS to spot firms with "suspiciously low" wages, since the survey contains information on education, experience, occupation and other characteristics of the employee that are important determinants of the wage level. However, the LFS dataset imposes two serious restrictions. First, the sample size is small compared to the number of employees and enterprises. As a result, the decision to label a firm as "definitely evading" is typically based on a single employee receiving a very low wage, which can be an outlying observation within a firm. Second, the information on occupations in the LFS is not sufficiently detailed, containing only the two-digit occupation code.

We take the advantage of the very detailed occupation information at hand and modify the algorithm. Our occupation data comes from the administrative employer-employee dataset and contains a six-digit code. Despite the low coverage (see Table 1 above), we still have 150-200 thousand employer-employee pairs with very detailed occupation each year. The same administrative data source includes the age and gender of each employee. Thus, we can detect employees with "suspiciously low" reported wages controlling for detailed occupation, age and gender. The information on education and experience is absent in the dataset, but detailed occupation partially compensates, since many professions require a certain level of education.

We proceed as follows. We regress the log of reported gross wage on the set of occupation dummies controlling for gender and age. We also include the year-region fixed effects to account for region-specific changes in the labour market conditions. Then, we label employees as receiving a "suspiciously low" wage if legal gross wage falls below the 10th percentile for a respective occupation in a given year and region controlling for age and gender.[7] Several comments are necessary here. First, we use the four-digit occupation code instead of the six-digit code due to the small size of Latvia's economy and the self-imposed restriction of at least 100 observations for each occupation-year-region group. Nevertheless, the four-digit code provides 412 unique occupations, which is a much finer classification than in the LFS. Second, we exclude all enterprises subject to the micro-enterprise tax.[8] Although such firms form a non-negligible part of total employment, we cannot compare them to other firms due to a specific labour tax regime. Third, observations with reported gross wage below the 1st or above the 99th percentile were censored, as were employees with less than 500 hours worked in a given year. Finally, we did not control for the industry effect when identifying the "suspiciously low" wages. Implicitly we assume that employees with a certain occupation can freely move between different industries within one region, e.g. a bookkeeper (ISCO-08 code 2411) can freely move from manufacturing to construction or administrative services.

Our procedure detects 10-15 thousand employees with a "suspiciously low" wage each year. However, this does not mean that all corresponding firms evade labour taxes, as some wages may seem suspiciously low because of unobserved worker characteristics (like insufficient qualification). We want to avoid labelling firms as evading based only on one case, so we require two conditions to be satisfied: a) the share of employees with "suspiciously low" wage in a given year equals 50% or more of all classified employees in the firm (i.e. employees with imputed occupation data); b) at least one third of employees, or at least 10 employees in the respective firm have occupation

[7] To account for the possibility of a low wage variance within a given profession, year and region, we labelled employees as receiving a "suspiciously low" wage only when the reported wage was at least 10% below the average (controlling for age and gender). This additional condition, however, was never binding.

[8] The micro-enterprise tax is a single tax payment, which includes mandatory state social insurance contributions, personal income tax, corporate income tax, and personal income tax of the micro-enterprise owner. See the State Revenue Service for more details at https://www.vid.gov.lv/default.aspx?tabid=8&id=5831&hl=2. Although we do not have a variable pointing to the micro-enterprise tax payers in our dataset, we exclude firms that comply with all micro-enterprise tax requirements.

data.[9] These conditions ensure that we observe enough employees with occupation information in the respective firm, and a substantial part of them appears to receive under-reported wages.

4.2 Predicting the probability of evading labour taxes at firm level

The second step of our methodology assumes that labour tax evasion, as any illegal operation, should leave some traces in the firm's financial reports. Establishing a systematic relationship between the probability of fraud and reported financial information allows one to detect fraudulent firms (see Beneish, 1999; Cecchini et al., 2010; Hajek and Henriques, 2017). Gavoille and Zasova (2021b) apply a random forest algorithm based on pre-defined treated and control groups, using several financial indicators to detect evading firms. Validating the classification through LFS and Household Budget survey data proves the relevance of their approach.

Instead of a random forest algorithm, we stick to a simple probit model. Despite potential losses in predictive power, we have two reasons for this. First, probit models are more transparent, as we can report the sign and the significance for coefficients on the firm-level indicators. Second, and more importantly, the evaluation of unofficial payments in the next step requires an estimate of the probability that the firm is involved in tax evasion, not just a binary classification.

We run the following probit model:

$$Pr\left(E_{it}=1|X_{it}\right)=\Phi\left(X_{it}\beta+u_{it}\right), \tag{1}$$

where E_{it} denotes a dummy variable that equals 1 for the treated group ("definitely evading" firms) and 0 for the control group ("definitely compliant" firms); t denotes the year, but i stands for the firm, and $i \in C$, where C represents a set of firms labelled as "definitely evading" and "definitely compliant" during the previous step; X_{it} denotes the set of (primarily financial) firm-level characteristics that may correlate with labour tax evasion.

After estimating (1) we use the probit model to predict the out-of-sample probability of evading:

$$\hat{p}_{it}=\Phi\left(X_{it}\hat{\beta}\right), \tag{2}$$

where $\hat{p}_{it}$ is the predicted probability of labour tax evasion for the firm i in year t. Note that now

[9] For example, a firm with 15 employees would be labelled as evading if we have occupation information for 7 employees and 4 of them have "suspiciously low" wage (i.e. below the 10th percentile for the respective occupation-year-region after controlling for age and gender).

$i \in I$, where I is the set of all firms that provide the necessary financial information X_{it}. Finally, we assume that the probability of tax evasion equals zero for firms that were labelled as compliant during the first stage of our algorithm, namely state-owned firms and firms where the owner comes from countries with a low corruption environment ($\hat{p}_{it} = 0$ if $E_{it} = 0$).

4.3 Evaluating the size of the unreported wage payments

The third step of our methodology evaluates the size of the unreported wage payments at employee level. Intuitively it is clear that employees with very low wages have a higher probability of receiving unreported payments, as the size of such payments is linked to the gap between the observed and the "normal" gross wage. Mincer earning regressions (Mincer 1958; Mincer 1974) have a long tradition in empirical labour economics (see Heckmann et al. 2003; and Lemieux 2006) and serve as a natural choice to determine "normal" or potential wage levels using data on education, experience, occupation, industry, as well as other firm and employee-level characteristics.

Two problems arise. First, Mincer earning regressions use the reported wage rate as a dependent variable. However, we know that some wages are under-reported. One could use only compliant firms in Mincer earning regression estimates, but this reduces the sample size and creates a sample selection bias, since compliant firms differ systematically from evading firms (e.g. in size or sectoral composition). Unobserved determinants of earnings pose the second problem. Even with a very rich micro-level data at hand we are unable to explain differences in wage levels completely due to unobserved personal and firm-level characteristics (like the ability of the employee or technology of the firm). Consequently, we cannot solely attribute the gap between the observed and predicted wage to illegal cash payments solely. Also, it is not clear how to interpret the case when the reported wage rate exceeds the wage predicted by the Mincer earning regression.

Earning regression estimated by the Stochastic Frontier Analysis

The main methodological innovation we propose in this research is to use the Stochastic Frontier Analysis (SFA) to overcome the above problems. Although SFA models, introduced by Aigner et al. (1977) and Meeusen and van Den Broeck (1977) to estimate the production frontier, have a long history, to the best of our knowledge, they were not used to evaluate unreported wage payments. The SFA fits our needs perfectly, as it assumes two stochastic terms in the earning regression instead of one. The random or so called idiosyncratic error term would account for all unobserved firm- or employee-level factors influencing the potential wage level. The inefficiency term would account for unobserved illegal cash payments. The inefficiency term in the SFA is non-negative by definition, which corresponds to the fact that unreported wage payments cannot be negative.

We define the Mincer earning regression determining the potential level of gross wage as follows:

$$ln\,(w_{ijt}) = x_{ijt}\beta + v_{ijt} - u_{ijt}, \tag{3}$$

where w_{ijt} stands for the reported gross wage for the employee j working in the firm i in year t, and x_{ijt} denotes the observed determinants of potential gross wage. Note that the set of determinants includes employee-level characteristics, firm-level characteristics, as well as various fixed effects (time, sector, occupation). v_{ijt} corresponds to the idiosyncratic error term assumed to be independently normally $N(0, \sigma_v)$ distributed. It captures unobserved firm- and worker-level potential wage determinants not accounted by x_{ijt}. The inefficiency term u_{ijt} is the most important for us – it is assumed to be independently half-normally $N^+(0, \sigma_u)$ distributed. Since $u_{ijt} \geq 0$, it can be interpreted as the gap between the logarithm of the potential gross wage and the logarithm of the reported gross wage net of the idiosyncratic term. For the moment, we can approximate the size of the unreported payments (relative to the potential wage) by u_{ijt}.

Heterogeneity in the inefficiency and idiosyncratic term

Straightforward estimates of (3) would not produce meaningful results, since the default model with a homoskedastic inefficiency term ($E(\sigma_u^2|x_{ijt}) = const$) does not distinguish between compliant and evading firms. Heterogeneity in the inefficiency term should be allowed. On the one hand, unreported payments are expected to be zero if the firm is not involved in under-reporting wages (corresponding to $\sigma_u^2 = 0$ for compliant firms). On the other hand, we can expect some positive unreported payments in tax-evading firms (although the size of cash payments may differ for individual employees) and $\sigma_u^2 > 0$.

We relate the variance of the inefficiency term σ_u^2 to the predicted probability that firm i is involved in unreported payments ($\hat{p}_{it}$), obtained in the second step. In particular, we model the heterogeneity in the inefficiency term as:

$$\sigma_u^2 = e^{\psi(\hat{p}_{it})}, \tag{4}$$

where $\hat{p}_{it}$ stands for the predicted probability that firm i is non-compliant in the year t from (2). We approximate the function $\psi()$ by a second-degree polynomial, and we expect that $\sigma_u^2 \to 0$ when $\hat{p}_{it} = 0$. Note that the use of the estimated probability $\hat{p}_{it}$ instead of a binary variable allows the variance of the inefficiency term to differ for firms which are almost certain evaders ($\hat{p}_{it} \to 1$) and firms that look suspicious, but are not definite evaders (e.g. $\hat{p}_{it} = 0.7$).

As pointed by Kumbhakar and Lovell (2000, p. 117), failing to account for the heteroskedasticity in the idiosyncratic term may lead to biased estimates of the inefficiency term. We follow the theoretical and empirical observations in Mincer (1958) that income dispersion increases with age, education and occupational skills. Thus, we assume that the variance of the idiosyncratic term v_{ijt} depends on the employee-level characteristics:

$$\sigma_v^2 = e^{z_{jt}\gamma}, \tag{5}$$

where z_{jt} stands for employee-level determinants of the inefficiency variance: age, education and occupation.

Expected inefficiency term

After estimating the system of equations (3)-(5), obtaining the residual from (3) is straightforward; namely, $\hat{\epsilon}_{ijt} = ln\left(w_{ijt}\right) - x_{ijt}\hat{\beta}$. There is no unique way, however, to split this residual into inefficiency and idiosyncratic terms. The only possible evaluation for unreported wage payments comes from the mean of the conditional distribution $f(u_{ijt}|\epsilon_{ijt})$. The expected value of the inefficiency term is the following:

$$E\left(u_{ijt}|\epsilon_{ijt}\right) = \mu_{*i} + \sigma_* \left(\frac{\phi\left(-\mu_{*i}/\sigma_*\right)}{\Phi\left(-\mu_{*i}/\sigma_*\right)}\right), \tag{6}$$

where μ_{*i} and σ_* are defined for the SFA model with half-normally distributed inefficiency term as

$$\mu_{*i} = -\epsilon_{ij}\frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}; \quad \sigma_* = \frac{\sigma_u \sigma_v}{\sqrt{\sigma_u^2 + \sigma_v^2}},$$

but estimates of σ_u and σ_v come from (4) and (5) respectively (for more details, see Kumbhakar and Lovell, 2000). The use of the expected inefficiency term as a proxy for the unreported payment does not allow for strong conclusions about each individual employee. However, we can still evaluate the average size of unreported wage payments, as well as their distribution by different types of firms and workers.

5 Detecting tax-evading firms

Now, we briefly overview the results of the two first steps of our methodology yielding the firm-level estimate of the probability of labour tax evasion. Although this is not the main focus of our research, these intermediate results provide some useful insights into tax evasion in Latvia.

5.1 Compliant and evading firms

According to the procedure described in section 4.1, we assume that firms owned by foreign investors from the low-corruption countries (mostly Scandinavia and a few other OECD members) are "definitely compliant" and form the control group along with state-owned firms. Ownership of 10% or more is required for the foreign-owned compliant firms (similarly, with regards to the definition of foreign direct investments) to proxy for a lasting interest and a significant degree of influence. State-owned firms include state and municipal enterprises, as well as firms with non-zero state-owned share in equity, assuming that any state involvement imposes certain business ethics and compliance.[10] The treated group consists of firms that were classified as "definitely evading" due to a large share of "suspiciously low" wages: at least half of employees with occupation data have reported wages below the 10th percentile for the respective occupation-year-region group after controlling for gender and age.[11] We report the number of firms in each category in Table A1 in Appendix.

We also check the validity of our assumption about the compliance of foreign- and state-owned firms. Table 2 reports the share of workers with "suspiciously low" wages among employees with detailed occupation information. The share is close to 80% for firms labelled as "definitely evading" in all years. On the other hand, the share of employees that probably receive illegal cash payments appears to be very low in firms labelled as compliant. Although we did not use any wage-related information in defining the control group, the firms assumed to be compliant indeed have substantially higher reported wages compared to firms classified as definitely evading. The number of

[10] We lack detailed information on the share of state ownership, as we can only distinguish between firms with at least 50% of state ownership and below 50% state ownership. We check the robustness of our results by applying a stricter definition, requiring at least 50% foreign or state ownership for "definitely compliant" firms. Despite the smaller set of the control group, the main results of our analysis remained unchained to this alternative definition. Even more, the main findings remained almost unchanged when we restricted the set of control firms to foreign-owned firms with at least 50% ownership only. Results are available upon request.

[11] We also control for the square of age to account for possible non-linearities.

foreign- and state-owned firms classified as "definitely evading" by our algorithm is very small, and such firms were excluded from the control group.

Table 2: The share of "suspiciously low" wages in compliant and evading firms

Year	Share of "suspiciously low" wages in "definitely evading" firms	Share of "suspiciously low" wages in "definitely compliant" firms	Number of "definitely compliant" firms also classified as "definitely evading"
2016	78.2%	3.2%	3
2017	78.9%	8.1%	13
2018	79.1%	6.7%	13

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Figure A2 in Appendix shows the distribution of reported employee-level wages for the various categories of firms: the differences are striking. The wage distributions for firms owned by the state or by foreign investors from low-corruption countries are shifted to the right, and the share of very low reported wages remains negligible. The opposite can be observed for firms that are labelled as "definitely evading" – most wages are concentrated just above the minimum wage.[12] Table 3 reports the final size of the treated and control groups: we are left with more than seven thousand observations for the probit model (substantially larger than the sample used by Gavoille and Zasova, 2021b).

Table 3: The size of the treated and control groups

Year	Number of "definitely compliant" firms (control group)	Number of "definitely evading" firms (treated group)	Total number of firms
2016	563	1'953	28'791
2017	330	2'429	30'216
2018	315	2'873	31'838

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Tables A2 and A3 in Appendix report the distribution of compliant and evading firms by firm size and industry. Firms that were labelled as "definitely compliant", namely foreign- and state-owned firms, are well represented in all size categories. Most firms classified as evading have less than 10 employees, while none of the large firms (250 employees or more) paid low enough wages to be labelled as "definitely evading". The more active involvement of smaller firms in under-reporting wages is consistent with the results found by Putniņš and Sauka (2021). Regarding the distribution by industry, we have enough observations from both control and treated groups for most industries. The only exceptions are mining, energy, and water supply and sewerage industries that were excluded from further analysis.

[12] Almost half of employees from "definitely evading firms" received less than 110% of minimum wage in 2018.

5.2 Probability of paying unreported wages

Now we estimate the probit model on the subset of firms from the treated and control groups defined above, determining the probability that firms evade labour taxes. We keep the set of explanatory variables short, focusing on financial indicators like turnover or debt ratio. In addition, we control for the age of the firm and some other variables related to ownership and business activities. We avoid variables directly related to wages like the share of labour costs in total expenses[13] or the share of employees at the minimum wage, since our definition of the treated group was directly based on wage information. We do not control for the number of employees for the same reason – taken together with turnover it implicitly provides information on labour productivity that should be directly linked to wages. Finally, we also control for region, year and sector fixed effects.[14] The results of the probit model are reported in Table 4.

The signs on most coefficients coincide with economic theory. In particular, the probability of tax evasion is higher for small and young firms, as confirmed through the survey results by Putniņš and Sauka (2021) for Latvia. Also Beneish (1999) reports that evading firms tend to be smaller and more leveraged, which is consistent with our findings that a higher debt-to-assets ratio (in particular the short-term debt ratio) is associated with more probable tax evasion. In line with DeBacker et al. (2015), foreign ownership (in addition to owners from low-corrupted countries) serves as a good indicator of tax compliance, especially when owners are from OECD countries.[15] Participation in merchandise import operations is another compliance indicator that can be related to a more intense control of such companies by tax authorities. We observe a higher probability of tax evasion for firms with a higher turnover to assets ratio and a lower share of intermediate inputs to turnover (although the latter indicator is only of marginal statistical significance), which might be related to an abnormally high turnover of evading firms. The coefficient on the ratio of cash to assets suggests that firms with relatively large cash holdings tend to be more compliant.[16] The only

[13] We include the ratio of intermediate inputs to turnover, however. While this ratio is inversely related to the share of labour costs, this relationship is not strong due to variation in profits.

[14] In addition, we include sector-year dummies to control for any sector-specific developments over time. These fixed effects also address the problem of varying ratio of "definitely compliant" to "definitely evading" firms in Table 3. Although we use administrative data, the information on occupation is available only for a subset of employees, and the coverage tends be lower at the beginning of the sample. As a result, the number of firms labelled as "definitely evading" increases over time in absolute terms, as well as in comparison with the number of "definitely compliant" firms. This can be misinterpreted as a growing probability of evading labour taxes over time, so we control this by fixed effects.

[15] Note that both ownership dummies exclude foreign owners from low-corruption countries. Moreover, removing ownership dummies from the probit model has only a marginal effect on the performance of the model and the evaluation of the unreported wages. Results are available upon request.

[16] Firms that evade labour taxes and pay envelope wages obviously need cash, but this cash is not reported in the

counter-intuitive effect comes from the low profits dummy[17] – extremely low profits serve as a sign of compliance, which may be due to correlation with turnover and indebtedness variables.

Table 4: Probit model for the probability of tax evasion in 2016-2018

Variable	Coefficient	p-value
Log of turnover	-0.511***	0.000
Debt to assets	0.0550*	0.007
Short-term debt to assets	0.267***	0.001
Cash to assets	-0.339***	0.004
Turnover to assets	0.0104***	0.002
Intermediate inputs to turnover	-0.00475	0.118
Age of the firm	-0.0552***	0.000
Low profits dummy	-0.987***	0.000
Merchandise imports dummy	-0.298***	0.000
Foreign owner from other OECD (excl. low corruption countries) dummy	-1.311***	0.001
Foreign owner from non-OECD dummy	-0.451**	0.062
Region dummies	Yes	-
Macroeconomic sector	Yes	-
Year	Yes	-
Macroeconomic sector * Year dummies	Yes	-
Number of observations	7'494	-

Note: (*), (**), (***) indicate statistical significance levels of 10, 5 and 1 percent respectively.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

The results of the probit estimated on the subsample of 7'494 pre-classified compliant and evading enterprises are used to predict the probability of tax evasion for the entire set of firms with the required financial statements. Altogether we are able to predict the probability for about 60 thousand Latvian firms each year.[18] Although we are primarily interested in the probability of tax evasion in our next step, we can also report the share of firms classified as tax evaders by year, industry and firm size. A subjective decision about the probability threshold splitting compliant and evading firms should be made – we set the threshold at 0.84, namely firms with predicted probability above 84% are classified as evading and the others as compliant.[19] All numbers reported below should be interpreted only in relative terms, however, as the absolute share of evading firms depends crucially on the level of this threshold. Table 5 reports aggregate statistics by year.

official balance sheets.

[17]We define this dummy as an indication that the profit to turnover ratio is below the 20th percentile for a respective industry and year.

[18]As mentioned above, we exclude micro enterprises from this exercise due to the specific nature of labour taxation. Also, several industries were excluded due to the small number of observations in the treated and control groups.

[19]This choice was based on the predicting performance (AUC) of the probit model for different levels of thresholds. We run the 10-fold cross-validation, randomly splitting our sample into 10 folds of equal size, re-estimating the model using the data for 9 folds and classifying firms in the 10th fold (10 repetitions were made to classify all firms in the sample). We were able to classify correctly 86.5% of firms, 86.9% of evading firms were classified correctly by the model and 96.8% of firms classified as evading by the probit model were actually labelled as evading in the first step of our algorithm. The area under the receiver operating characteristics curve (AUC) equals 0.853, which, although below the performance of the model by Gavoille and Zasova (2021b), still indicates a good performance.

Table 5: Evading and compliant firms by years

Year	Number of classified firms	Number of compliant firms	Number of evading firms	Share of evading firms	Share of employees working in evading firms
2016	60’828	15’569	45’259	74.4%	25.2%
2017	56’779	13’372	43’407	76.4%	25.5%
2018	57’389	11’639	45’750	79.7%	30.3%

Note: we classify a firm as evading if the predicted probability of labour tax evasion exceeds 80%.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

For the given threshold of 0.84, the model classifies between 75 to 80% of Latvia’s firms as evading labour tax, depending on the year. The share of tax evaders increases in 2017-2018 compared to 2016, which is in line with evaluations of Putniņš and Sauka (2021). These numbers are considerably higher than those reported by Gavoille and Zasova (2021b) (37%), which can be attributed to several differences in the datasets.[20]

Table 6: Evading firms by firm size class in 2018

Size class	Number of evading firms	Share of evading firms	Share of employees working in evading firms
1 to 9 employees	42’246	88.6%	81.3%
10 to 19 employees	2’332	49.7%	48.5%
20 to 49 employees	764	25.9%	24.0%
50 to 249 employees	88	6.0%	5.1%
250 or more employees	0	0.0%	0.0%

Note: the size class of 1 to 9 employees does not include micro enterprises. We classify a firm as evading if the predicted probability of labour tax evasion exceeds 80%.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

The share of employees working in tax evading firms is much smaller, however, due to substantially better compliance among large firms – for the given level of threshold we estimate that around 25-30% of Latvia’s employees work in tax-evading enterprises. One should remember that these results do not include micro enterprises, the self-employed or several important industries, and these figures by no means represent the size of the shadow economy, or the overall size of unreported wage payments in Latvia. Still, these estimates are rather close to the evaluation by Gavoille and Zasova (2021b) for four major industries in 2011-2013 (23.7%) and do not contradict the 36% share of the respondents who reported they know someone working without declaring part of their income, see Eurobarometer (2020).

The tight relationship between labour tax evasion and firm size is confirmed by Table 6. While

[20] Gavoille and Zasova (2021b) focus on firms with at least 6 employees in four major industries (manufacturing, construction, trade and transportation), while we also include smaller firms that are not subject to the micro-enterprise tax and broaden the industry scope. The higher share of tax evading firms in our results appears natural given the prevalence of tax evasion behaviour among very small firms and several industries like professional or administrative services (see Tables 6 and 7).

most of small firms are classified as evading by the probit model, the share of tax-evading enterprises drops substantially already for medium-sized firms, and becomes negligible for large firms. Although the share of tax evading firms does not vary a lot by industries, the more substantial differences can be observed for the share of employees working in tax-evading firms (see Table 7): evasion is most common in construction and in professional and other services, while it is less common in manufacturing and ICT. Despite the subjective choice of threshold to classify evading firms, our conclusions about the distribution of tax-evading firms by size and sector are in line with previous findings by Putniņš and Sauka (2015), Putniņš and Sauka (2021), and Gavoille and Zasova (2021b).

Table 7: Evading firms by sectors in 2018

Industry	Number of evading firms	Share of evading firms	Share of employees working in evading firms
A: Agriculture, forestry and fishing	2'234	79.7%	30.1%
C: Manufacturing	4'444	75.2%	20.1%
F: Construction	4'585	85.6%	45.5%
G: Trade	13'019	78.9%	30.0%
H: Transportation	3'618	80.7%	26.5%
I: Hotels and restaurants	2'057	81.5%	40.6%
J: Information and communication	2'261	79.0%	20.5%
L: Real estate	3'525	69.8%	30.4%
M: Professional services	5'364	84.3%	45.0%
N: Administrative services	2'347	89.8%	42.9%
RST: Other services	1'824	90.2%	43.6%

Note: we classify a firm as evading if the predicted probability of labour tax evasion exceeds 80%.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

To validate the classification of firms into evading and compliant, we replicate the exercise performed by Gavoille and Zasova (2021b), although in a much simpler way. Table A4 in Appendix reports the effect of increasing the minimum wage on changes in employment. Gavoille and Zasova (2021b) found that firms with larger shares of workers at the minimum wage (or close to it) tend to reduce the number of employees after an increase in the minimum wage. However, this effect is much smaller or even absent for firms classified as evading, since such firms compensate by reducing unreported wage payments and only face the increase in tax-related costs. We use the fact that the minimum wage was increased twice within our sample period: from 370 to 380 EUR in 2017, and further to 430 EUR in 2018.[21] Our results lead to similar conclusions. First, the negative and significant coefficient on the interaction term between changes in the minimum wage and the share of employees close to the minimum wage confirms the negative effect of minimum wage increases on employment due to growing labour costs. Second, the positive and statistically

[21] We have no years in our sample without changes in the minimum wage, so our identification is based on the size of the adjustments.

significant coefficient on the triple interaction term (changes in the minimum wage, share of workers close to the minimum wage, and dummy indicating labour tax evasion), shows that evading firms tend to reduce employment less, adjusting unreported wages instead.

6 Evaluation of the unofficial wage payments

6.1 Evaluation of the potential gross wage

We estimate the Mincer earning equation (3)-(5) by the SFA to evaluate the potential gross wage for each employee.[22] The earning regression includes an extensive list of potential wage determinants, both employee- and firm-level. As the LFS remains the only available source of detailed employee-level information in Latvia, we can only estimate the potential wage for employees who were included in the LFS in 2016-2018, and who were employed by firms for which we have predicted the probability of labour tax evasion in Section 5.2.[23] This narrows our sample to approximately 11 thousand employees between 2016 and 2018, which is still enough to run the SFA. The set of employee-level variables includes the traditional Mincer earning equation determinants – education dummies, age and experience in the current working place (as well as squared terms; age is included in logarithmic form to reduce correlation with experience). In addition, our regression includes the usual demographic variables (gender, ethnicity, citizenship) and variables describing the type of working contract. Firm-level variables control for the size and the age of the enterprise, ownership type, international trade in goods and location. Finally, we also control for numerous fixed effects, including industry, region, occupation, year and month. Summary statistics for the employee- and employer-specific variables included in the potential gross wage regression are reported in the Table A6 in Appendix.

[22]Each observation is weighted according to the individual LFS annual weights.

[23]We use the data on gross reported wage and hours worked from the employer-employee data at the respective date to link it with LFS data. In order to avoid non-regular income or hours worked due to vacations or bonuses, we take the mode gross wage and hours worked during the centered three-month window.

Table 8: Potential wage regression with heteroskedastic inefficiency and idiosyncratic terms

Variable	SFA regression (1)	SFA regression with inverse Mills ratio (2)
Employee-level variables		
Education: secondary general	0.0483***	0.0229
Education: secondary professional	0.0854***	0.0612***
Education: professional	0.0789**	0.0492
Education: higher	0.183***	0.179***
Graduation year	0.00406***	0.00550***
Experience in the current working place	0.00821***	0.00800***
Experience in the current working place squared	-0.00017***	-0.00017***
Logarithm of age	1.449***	1.416***
Logarithm of age squared	-0.197***	-0.186***
Female	-0.142***	-0.145***
Non-Latvian	-0.0803***	-0.0770***
EU (but not Latvian) citizen	0.0853	0.159*
Non-EU citizens, aliens	0.0478***	0.0620***
Temporary contract	-0.118***	-0.120***
Partial time contract	-0.0863***	-0.0863***
Pensioner	-0.0459	-0.0465
Disabled person	-0.118**	-0.116*
Other employees	-0.0811*	-0.0794*
Firm-level variables		
Logarithm of number of employees	0.0124***	0.0124***
Logarithm of firm's age	0.0111	0.0121*
State owned	0.0406**	0.0418**
Foreign-owned (top non-corrupted countries)	0.0763*	0.0753*
Foreign-owned (Baltic countries)	-0.193***	-0.195***
Foreign-owned (other OECD countries)	0.164**	0.162**
Foreign-owned (other non-OECD countries)	0.0786	0.0798
Exporter of goods	-0.0532***	-0.0541***
Importer of goods	0.0266	0.0274
Located in averagely inhabited territory	-0.0541***	-0.0501***
Located in rarely inhabited territory	-0.0994***	-0.0938***
Inverse Mills ratio	-	-1.966***
Industry fixed effects	Yes	Yes
Region fixed effects	Yes	Yes
Occupation fixed effects (2-digit level)	Yes	Yes
Year fixed effects	Yes	Yes
Quarter fixed effects	Yes	Yes
Variance of inefficiency term: $ln(\sigma_u^2)$		
Constant	-9.258***	-9.267***
Probability to pay envelope wages	14.454***	14.443***
Probability to pay envelope wages square	-7.234***	-7.219***
Variance of idiosyncratic term: $ln(\sigma_v^2)$		
Constant	-8.302***	-8.242***
Logarithm of age	3.939***	3.907**
Logarithm of age squared	-0.558***	-0.554**
Professionals	-0.398***	-0.407***
Technicians and associate professionals	-0.423***	-0.421***
Clerical support workers	-0.638***	-0.634***
Services and sales workers	-1.159***	-1.156***
Skilled agricultural, forestry and fishery workers	-0.445**	-0.451**
Craft and related trades workers	-0.668***	-0.667***
Plant and machine operators and assemblers	-0.534***	-0.528***
Elementary occupations	-0.723***	-0.721***
Education: secondary general	0.0892	0.0914
Education: secondary professional	0.0607	0.0659
Education: professional	0.272***	0.282***
Education: higher	0.208***	0.216***
Number of observations	10'712	10'712

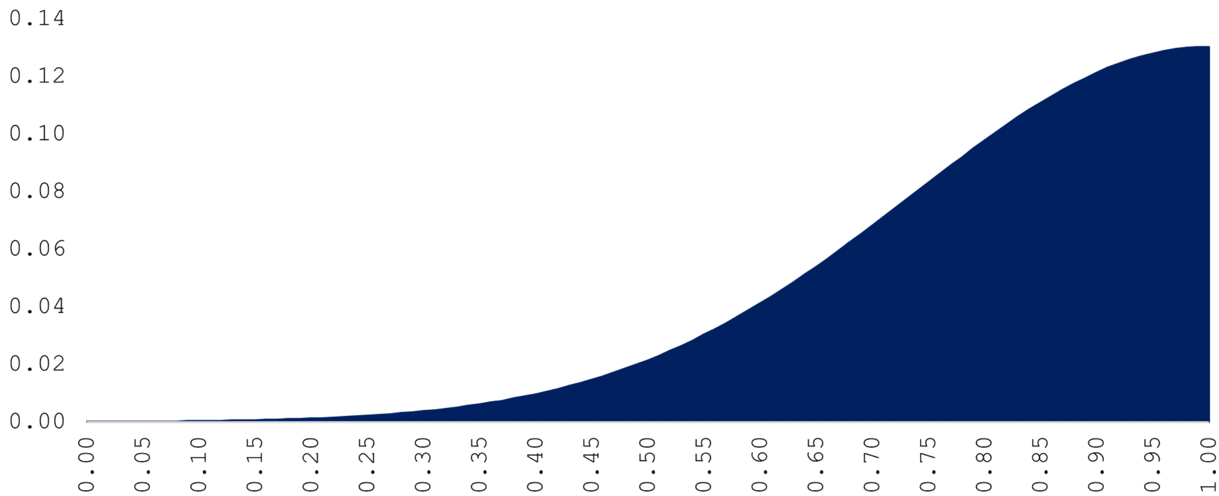
Note: (*), (**), (***) indicate statistical significance levels of 10, 5 and 1 percent respectively.
Sources: SRS of Latvia, CSB of Latvia, own calculations.

Table 8 reports the results of two SFA regressions. The first column contains the outcome of the SFA regression with heteroskedastic inefficiency and idiosyncratic terms. Although it includes an extensive list of determinants, we may face the classical problem of sample selection bias since only employed persons are included in the sample. To deal with this problem, we use the fact that the LFS also contains data on unemployed persons and apply Heckman's two-step approach (see Heckman 1979) to estimate equations (3)-(5). In the first step we estimate the probit model that explains the probability of employment for economically active individuals depending on their available personal characteristics. We use the number of children and family status (as well as interactions with gender) as instruments for the first stage – the results can be found in Table A5 in Appendix. Then, in the second stage, the inverse Mills ratio is added to the SFA model with heteroskedastic inefficiency and idiosyncratic terms to account for the potential sample selection bias. Overall, despite the significance of the inverse Mills ratio, the coefficients of the earnings regression do not change substantially (except for a few education-related variables). Below, we refer to the results of the simple SFA regression with heteroskedastic inefficiency and idiosyncratic terms as our baseline results.

According to Table 8, the potential gross wage depends significantly on various employee- and firm-level variables. Most of the signs are in line with prior expectations and economic theory. For instance, we observe higher potential gross wages for males, Latvians, employees with regular contracts, higher education, recently graduated and longer experiences in the current working place. As to firm-level variables, the potential gross wage tends to be higher in larger and older firms located in cities. Finally, foreign-owned firms (by investors from non-Baltic OECD members) and state-owned firms tend to have higher potential wages compared with domestic firms.

The last two blocks of Table 8 are devoted to the variance of idiosyncratic and inefficiency terms. Figure 1 reports the predicted level of the variance of the inefficiency term (σ_u^2). Although we imposed no specific restrictions on the function $\psi()$, except its polynomial form, the estimated inefficiency variance appears to be almost zero for firms with predicted probability of tax evasion $\hat{p}_i$ below 0.2, meaning that employees from firms predicted to be compliant are expected to receive negligible unreported payments (proxied by the inefficiency terms). The variance of inefficiency term increases rapidly when $\hat{p}_i$ exceeds 0.5-0.6 and flattens when probability of a firm to be evading exceeds 90%. Thus, employees working for tax-evading firms are expected to receive on average high unreported payments, while there are almost no unofficial wage payments for compliant firms.

Figure 1: Variance of the inefficiency term (σ_u^2) depending on the predicted probability to evade labour taxes ($\hat{p}_i$)



Note: the results come from the outcome of the SFA regression, see Table 8, column (1).
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Note, however, that our approach allows the size of unreported wage payments to differ for employees within a single firm, since it is driven by the employee-level characteristics and the reported gross wage.

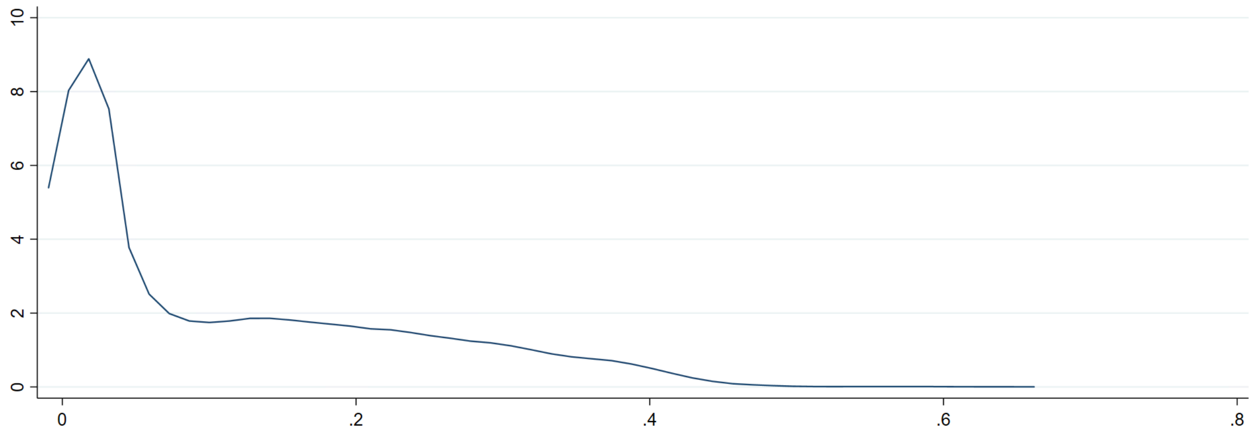
As to the variance of the idiosyncratic term, our results are in line with the classical observation by Mincer (1958). In particular, the variance of unexplained potential wage component (e.g. ability) increases with age. Managers have the highest variance of the idiosyncratic term, while service and sales workers – the lowest. Finally, the largest variance of idiosyncratic term is predicted for the employees with higher and professional education. In other words, the potential wage rate of elder managers with professional education has the largest variability conditional on observed firm and personal characteristics, while the salary for young and uneducated services and sales workers is more standardised.

6.2 The gap between the potential and reported wages

After estimating the equation for the potential gross wage, one can use equation (6) to evaluate the expected inefficiency, i.e. the gap between the predicted potential gross wage and reported gross wage net of the expected idiosyncratic term.

Figure 2 shows kernel density of the above gap in logarithmic terms. The density graph reveals

Figure 2: Distribution of the gap between the logarithm of potential and reported gross wages



Notes: the graph reports the kernel density function of the expected inefficiency term – the gap between the logarithm of the potential and reported gross wages net of the expected idiosyncratic term.
Individual weights from LFS are used to calculate the summary statistics.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

that approximately half of the employees have a very small positive gap (below 0.05), that is, the difference between the potential and reported gross wages associated with tax evasion does not exceed 5% of the reported wage. We should flag that the SFA model produces positive expected value of the inefficiency term for all observations, thus a very small gap most probably points to the absence of tax evasion for the respective employees. However, the expected inefficiency term is not negligible for all employees. The distribution is skewed right with the largest observations going up to 0.65: 25% of employees have expected inefficiency above 0.194 (corresponding to the 21.4% gap in terms of the reported gross wage), but 5% – above 0.352 (42.2% gap). This points to sizeable unofficial wage payments to some employees.

Figure A3 in Appendix shows that the distribution of the gap differ by firm size and industry. For instance, Putniņš and Sauka (2021) report that small firms engage in shadow activities more often than large firms. The left part of Figure A3 confirms this finding, since the proportion of employees with substantial gaps between the potential and reported wages is obviously higher for firms with less than 20 employees. As to industries, the distribution for construction and professional services (the ones with the highest share of tax evasion according to Table A4) in comparison with other industries demonstrates a lower share of employees with small gaps (presumably no unreported payments) and a higher share of employees with large gaps. Construction is mentioned as the most evading industry in Latvia by Putniņš and Sauka 2015, Putniņš and Sauka 2021 and Gavoille and Zasova 2021b).

One should remember, however, that in addition to unreported wage payments in cash, the gap between the potential and reported gross wages may include unobserved individual and firm characteristics. This can be related to the misspecification of the Mincer earning equation or the idiosyncratic term's variance. Also, we only obtain the expected rather than the actual value of the inefficiency and idiosyncratic terms. Keeping these caveats in mind, we now proceed to the evaluation of the unofficial wage payments.

6.3 From the expected inefficiency term to the size of the unofficial wage payments

While the expected inefficiency term corresponding to the gap between the potential and reported gross wages net of the expected idiosyncratic term may serve as a proxy for unofficial wage payments, it still provides biased estimates. First, the expected inefficiency term always exceeds zero. Second, the potential gross wage predicted by (3) corresponds to the potential legal gross wage that includes all necessary labour tax payments. The amount of labour tax payments reduces by paying some part of the income unofficially, providing an additional gain for employer and/or employee. The actual size of unreported payment depends on the way the gains from the unpaid labour taxes are split between the employee and the employer.

To solve the latter problem, we assume that the unpaid labour taxes are split fifty-fifty between the employer (in the form of reduced labour costs) and employee (in the form of unreported payments). A simple numerical example is necessary here. Assume that the reported gross wage equals 518 EUR, but the expected inefficiency term equals 0.24 that corresponds to the gap of 27.5% relative to the reported wage, so the potential gross wage for the expected level of the idiosyncratic term equals 660.57 EUR. Accounting for the labour tax in Latvia, the reported net wage equals 381.3 EUR, but the employee would get 506.32 EUR net wage if her potential wage of 660.57 EUR paid legally. Reporting 518 EUR gross wage instead of 660.57 EUR reduces the labour tax payments by 80.53 EUR. This sum is equally split between the employer and the employee (40.27 EUR each). Thus, the size of the unreported payment is $506.32 - 381.30 + 40.27 = 165.29$ EUR (the difference between the potential net wage in case everything is paid legally and the actual net wage plus 50% of gains from lower tax payments). Note that the size of unreported payment differs from the gap between the potential and legal gross wages (142.57 EUR). The fifty-fifty split is a strong but natural assumption: the extreme split of unpaid labour taxes would be unreasonable. Adding all gains from the unpaid labour taxes to the unreported payment does not reduce the costs of labour for the firm, while a risk of being audited and caught appears. On the other hand, leaving all gains from the unpaid labour taxes to the firm by reducing the labour costs would not satisfy the worker, as she gets the same net payments as in the case of 100% legal payments, but loses part of social protection that directly depends on the social security contributions. We provide the robustness check of our results for the alternative splits of unpaid tax gains in the robustness section.

To reduce the effect of strictly positive inefficiency term, we assume that any monthly unofficial cash payment below 50 EUR or below 10% of the reported gross wage is negligible (compliant worker with no illegal payments). The reasoning for such conditions is the following: the firm (and also the worker) has no incentive to receive small unreported payments, as gains from tax reductions become small compared to the potential costs if being audited and caught. The difference between the potential and reported gross wages should be large enough both in absolute and relative terms for gains to outweigh costs. There is no way to determine the thresholds based on some statistical criteria, so our decision is purely subjective and based on intuition. Later we check the robustness of our findings to these thresholds.

6.4 The distribution of unreported wage payments and the effect on inequality

This subsection reports our estimates of unreported payments given the previous predictions about the probability to evade labour taxes, the potential gross wage and the assumption on the split of unpaid labour tax gains between the employer and employee. Table 9 shows the share of employees receiving unreported wage payments and the average size of this payment by predicted probability to evade ($\hat{p}_{ijt}$), given the above assumption regarding the minimum absolute and relative size of unofficial payment. There is a clear threshold for the firms' probability to evade labour taxes at around 40-50% level. The share of employees with unreported payments is zero or just negligible for firms with $\hat{p}_{ijt} < 0.4$, while the vast majority of employees working in firms with $\hat{p}_{ijt} > 0.5$ receive some unreported wage payments. The average size of such payments for the employees who actually receive unreported payments varies depending on $\hat{p}_{ijt}$: the size is small (slightly above 10%) in the relative terms for averagely compliant firms, but approaches 40% of the reported gross wage or about 200 EUR when firms are predicted to be almost certain labour tax evaders. This average size of unreported payments may seem small, but one should remember that it masks substantial heterogeneity of illegal cash payments at individual level. Moreover, the relative size of unreported payments to net wage (reflecting an employee's perception) appears substantially higher, and for workers receiving minimum wage it exceeds 60%. Finally, we may still underestimate the importance of unreported payments, as some firms may also under-report the number of hours worked. The Mincer earning regression in Table 8 evaluates the potential full-time equivalent wage and does not account for such a kind of fraud.

Table 9: Estimated size of unreported payments by probability to evade in 2018

Predicted probability to evade for the firm ($\hat{p}_i$)	Number of employees in the respective firms	Average expected inefficiency term ($E\left(u_{ijt}\|\epsilon_{ijt}\right)$)	Share of employees receiving unreported wages	Average size of the unreported payment relative to the official gross wage for employees receiving unreported payments	Average size of the unreported payment in EUR for employees receiving unreported payments
$[0.0 - 0.1)$	825	0.0096	0.0%	0.0%	0.00
$[0.1 - 0.2)$	310	0.0212	0.0%	0.0%	0.00
$[0.2 - 0.3)$	264	0.0393	0.0%	0.0%	0.00
$[0.3 - 0.4)$	241	0.0640	0.7%	11.5%	50.16
$[0.4 - 0.5)$	189	0.0987	66.7%	12.8%	100.01
$[0.5 - 0.6)$	196	0.142	96.8%	17.3%	136.20
$[0.6 - 0.7)$	204	0.177	94.1%	22.1%	175.37
$[0.7 - 0.8)$	237	0.231	96.3%	29.5%	177.19
$[0.8 - 0.9)$	248	0.259	94.5%	33.5%	184.54
$[0.9 - 1.0)$	482	0.299	91.5%	39.9%	198.50

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

The aggregate statistics for unreported wage payments are shown in Table 10. Around 45% of Latvian employees receive unreported wages, and the average size of such payments (conditional on receiving them) is slightly below 30% of the reported gross wage in 2018. This means that the overall size of unofficial wage payments is close to 10% of the official gross wage fund. These results should be treated with caution. First, the number representing the share of unreported wage payments in the total official gross wage fund can by no means serve as estimates of the shadow economy in Latvia. Our research deals only with unreported payments but does not consider non-labour tax evasion or illegal employment. Both phenomena are sizeable and form a substantial part of the shadow economy in Latvia according to Putniņš and Sauka (2015) and Putniņš and Sauka (2021). Second, our estimates can be biased due to the sample issue. On the one hand, our estimates exclude the self-employed and the employees of micro enterprises. This most probably leads to the underestimation of overall unreported payments, since the self-employed tend to participate in labour tax evasion (see e.g. Kukk et al. 2020), but micro enterprises may act in much the same way as other small enterprises participating in labour tax evasion more actively than on average. On the other hand, several industries were not included in the analysis due to data constraints. The industries like finance, education, healthcare and energy are expected to be mostly compliant, which means that the numbers reported in Table 10 are overestimated. The net outcome of both effects is unclear. Finally, although the number of employees in the LFS for whom we were able to estimate the size of unofficial payments is large (more than 3 thousand every year), we cover less than a quarter of all employees who participated in the LFS due to missing data and a narrowed set of enterprises. This does not allow us to make strong conclusions about the changes in unreported wage payments over time. Finally, the levels reported in Table 10 may depend on the assumptions about the minimum level of unreported wage payments, as well as on other assumptions.

Table 10: Aggregate statistics on unreported wage payments by year

Year	Share of employees in the LFS with evaluated size of the unreported wage payments	Share of employees receiving unreported wage payments	Average size of the unreported wage payments relative to the official gross wage for employees with non-zero unreported payments	Share of unreported wage payments in the total reported gross wage fund
2016	23.4%	36.9%	29.6%	7.3%
2017	19.4%	40.7%	29.6%	8.5%
2018	19.0%	46.5%	29.5%	9.5%

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

The comparison of our finding with other evaluations for Latvia is not straightforward due to differences in methodologies, data sources and the ways to report the output. According to the survey conducted by Eurobarometer (2020), only 7% of respondents acknowledged that they received part of their salary in cash without declaring it, which is substantially below our numbers and may be related to untruthful responses. On the other hand, the size of cash payments was evaluated as approximately 40% of the gross yearly income, which exceeds our estimates of the average unreported wage payments. However, these estimates in Eurobarometer (2020) are based on the answers of only 24 respondents. According to Gavoille and Zasova (2021a), the households whose head works in a domestic firm conceal 26% more income compared with the households whose head works in a foreign-owned firm, while Putniņš and Sauka (2015) report that 34% of total wages in Latvia are paid unofficially. Even if we consider the difference in time periods and samples, the numbers in Table 10 suggest a smaller role for the unreported wage payments. One possible explanation for that relates to the difficulties in splitting the idiosyncratic and inefficiency term. To some extent this can also be driven by unaccounted informal employment.

The major advantage of our empirical approach is the possibility to see the difference in frequency and size of labour tax evasions by industry, firm or employee group. In particular, we clearly observe that both the frequency and size of unreported wage payments in small firms substantially exceed the ones in large firms. According to Table 11, two thirds of workers in firms with less than 10 employees receive unofficial wage payments representing around 18% of the total official gross wage fund. While the share of employees receiving unofficial wages, as well as the size of such payments is still non-negligible for very large enterprises, the overall amount of unreported payments is evaluated to be around 1% of the total gross wage fund.[24] These findings are in line with both the recent empirical results for Latvia (see Putniņš and Sauka, 2021) and other countries (see e.g. Kumler et al., 2020).

Table 11: Aggregate statistics on unreported wage payments by firm size in 2018

Firm size class	Share of employees receiving unreported wage payments	Average size of unreported wage payments relative to the official gross wage for employees with non-zero unreported payments	Share of unreported wage payments in the total reported gross wage fund
1 to 9 employees	66.2%	34.7%	17.5%
10 to 19 employees	59.0%	28.5%	12.7%
20 to 49 employees	60.0%	25.7%	11.1%
50 to 249 employees	25.1%	22.6%	4.2%
250 or more employees	7.2%	23.2%	1.1%

Note: the size class of 1 to 9 employees does not include micro enterprises.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Tables A7 and A8 in Appendix provide similar statistics by industry and firm age group. Although the average size of unreported wage (for the employees with unofficial payments only) is comparable across sectors and lies in the range of 25-33% of reported gross wage, the largest share of employees with illegal payments (and also the ratio of overall illegal payments to the reported gross wage bill) remains the highest in agriculture, construction, real estate, professional and other services. As to the age, young firms tend to be more involved in unreported wage payments, which is in line with findings by Putniņš and Sauka (2015).

The distribution of unreported payments can also be analysed by employee type, in particular by the level of reported gross wage (see Table 12). As expected, the largest share of evaders, as well as the largest unreported payments (both in absolute and relative terms) are observed for the employees receiving the wage equalling or slightly exceeding the minimum wage (430 EUR in 2018). More than two thirds of such workers receive some additional unofficial payment of almost 200 EUR on average. The share of evaders tends to diminish with the increase of reported gross wage. As to

[24]While Table 6 reports no tax evasion for large firms in Latvia, we should remember that it shows only firms with probability to evade exceeding 84%.However, Table 11 reports that 7.2% of employees in large firms receive an unreported wage (84% threshold is not binding in the third step of our methodology). These employees mostly concentrate in a few large firms. The smaller average size of unreported wage relative to the official gross wage can be explained both by a higher official gross wage (as large firms tend to be more productive) and a higher risk related to fraud due to larger attention from tax authorities.

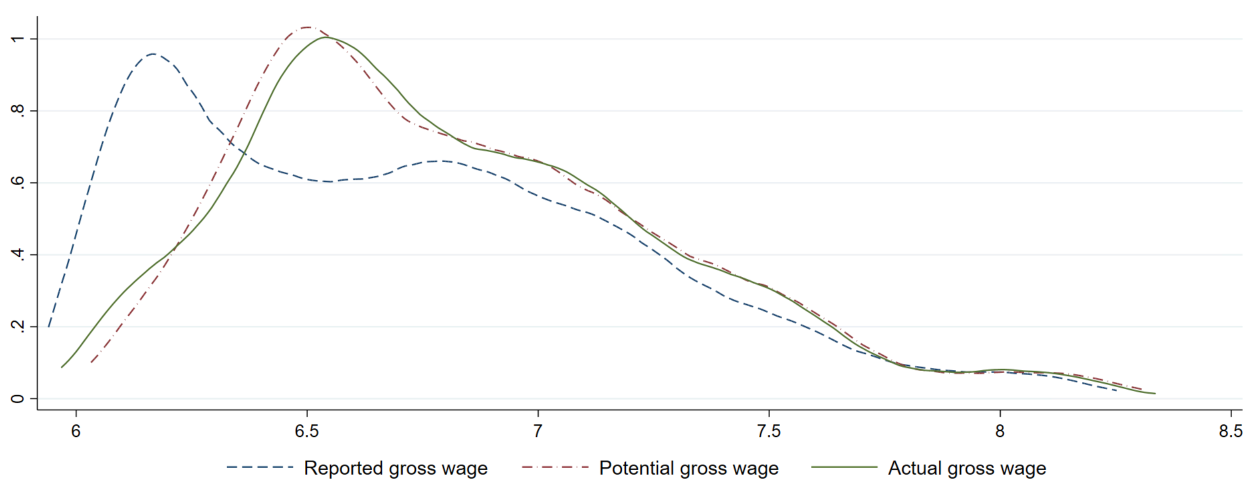
the size of unreported wage, it clearly diminishes in relative terms with higher reported gross wage, but stays roughly unchanged in absolute terms, ranging between 150 and 200 EUR on average for any reported wage interval (except for employees with the official wage exceeding 2000 EUR, where the estimated average unreported wage exceeds 300 EUR). This, however, does not account for the heterogeneity of unreported wage payments within each group of employees.

Table 12: Aggregate statistics on envelope payments by the official gross wage in 2018

FTE gross wage in EUR	Share of employees receiving unreported wage payments	Average size of the unreported wage payments relative to the official gross wage for employees with non-zero unreported payments	Average size of the unreported wage payments in EUR for employees with non-zero unreported payments
$[430, 435)$	69.0%	47.0%	185.13
$[435, 450)$	62.5%	38.5%	152.11
$[450, 500)$	64.3%	39.3%	161.14
$[500, 550)$	67.0%	34.7%	156.19
$[550, 600)$	50.8%	28.0%	139.94
$[600, 700)$	53.9%	27.6%	159.40
$[700, 800)$	44.8%	24.3%	158.81
$[800, 900)$	44.5%	21.9%	160.66
$[900, 1000)$	40.8%	21.5%	166.57
$[1000, 1250)$	30.1%	19.3%	188.47
$[1250, 1500)$	34.4%	19.6%	216.09
$[1500, 2000)$	23.8%	18.4%	207.73
$[2000, +\infty)$	24.5%	15.5%	337.46

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Figure 3: Distribution of reported, potential and actual wages (in natural logarithms) in 2018



Note: the distribution is based on the subsample of employees participating in the LFS, conditional on the estimation of the unreported wage.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Instead of focusing on the average size of unreported payments for particular groups of firms and employees, we can directly look at the distribution of the reported and actual gross wages (that also includes the estimated unofficial wage payments). Figure 3 also reports the distribution of the potential gross wage obtained from the model in Table 8 for comparison. The major difference between the distributions of the (logarithm of) reported gross wage and actual gross wage is observed in the left part of both distributions. A substantial share of reported gross wages is observed in the range of 430-500 EUR (corresponding to 6.06-6.21 in logarithmic terms) in 2018. Not all such workers get unofficial payments, but, as shown above, the majority of employees with low reported income receive additional 40-50% in cash, moving the mode of the actual gross wage to 665 EUR approximately (6.5 in logarithmic terms). There are few employees with the income below 600 EUR when the payments in cash are taken into account. Both distributions become very similar above the gross income of 1400 EUR, which corresponds to the smaller size of unreported payments in high gross wages. Figure 3 makes a very clear point of the effect of unreported payments on income inequality, i.e. it appears to be lower when unofficial wage payments are taken into account. Table 13 quantifies the difference in conventional Gini coefficient and S80/S20 income ratio. Note that these are the measures of the wage income that do not account for social transfers and undeclared employment, therefore, Figure 3 only provides a partial information on the income distribution in Latvia.

Table 13: Inequality measures for the official and actual gross wages

Year	Gini coefficient for the official gross wage	Gini coefficient for the actual gross wage	80/20 income ratio for the official gross wage	80/20 income ratio for the actual gross wage
2016	0.454	0.432	4.70	4.23
2017	0.532	0.506	4.84	4.49
2018	0.485	0.447	4.78	4.36

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Both measures point to a lower level of inequality when the unreported wage payments are taken into account. The difference is substantial and remains stable over time.[25] Two conclusions can be derived from these results. First, the real situation concerning wage income inequality appears to be better than it looks from the point of view of the official income statistics, since many seemingly low-paid workers receive part of their income in the form of unofficial payments. Second, the level of social protection (unemployment benefits, future pensions) for a significant share of employees with low-to-medium income is low due to unreported income.

[25] We expect smaller effect on inequality when social transfers are taken into account, since unreported wage payments cannot affect the disposable income of pensioners, the unemployed and inactive persons.

6.5 Validation of the estimates of unreported wage payments

The estimates of unreported wage payments are based on various strong assumptions at each step of our methodology. Although the actual size of unofficial payments is unobservable due to the absence of audit data, we can still check the validity of our results. The LFS contains a question to respondents on their net monthly income, which allows us to follow the approach of Kumler et al. (2020) and check the discrepancy between the self-reported net wage and the official information on legal net wage available from the administrative database. Although it is hard to expect that respondents will report their actual net income precisely and honestly, we still can expect that employees participating in labour tax evasion may report at least some part of these unofficial payments in answering the survey.

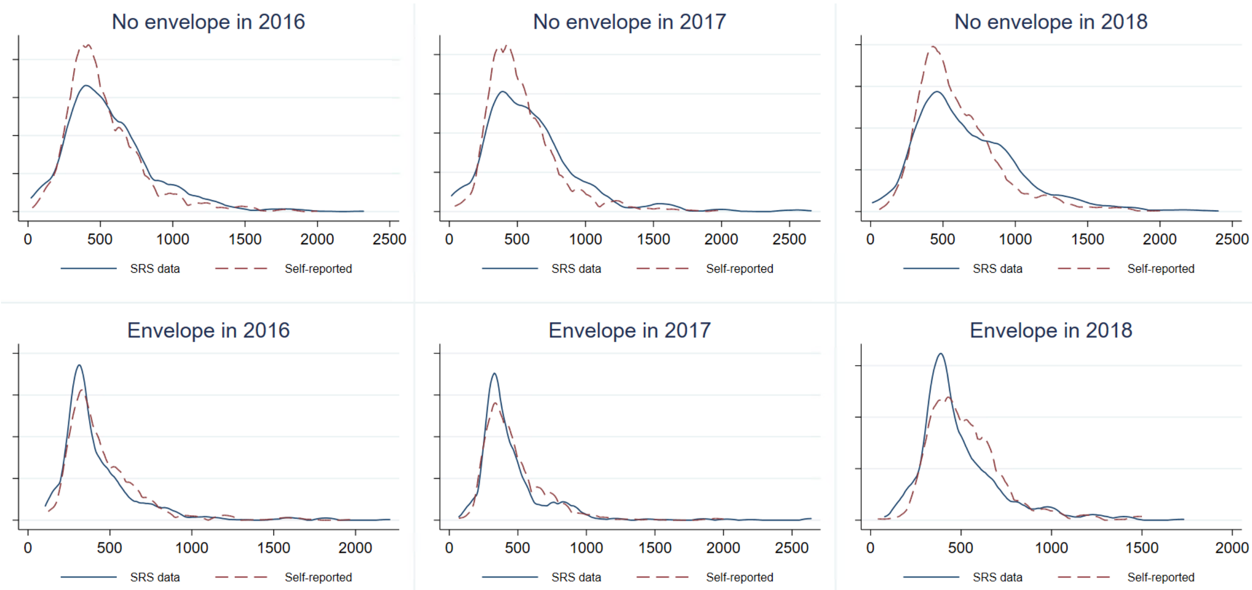
One hurdle we need to overcome before the validation of our results relates to the fact that respondents are asked to report their net income, while the administrative data contains gross wage. We calculate the legal net wage from the legal gross wage using the information on corresponding tax rates, the status of the employee (e.g. working pensioner) and the potential number of dependants that affect the size of deductible income. Our estimates can differ from the actual net legal wage[26], but we can validate our calculations by the information from the LFS. A large share of the LFS respondents (around one third) does not report their net income, and the CSB of Latvia imputes this information from the administrative data. Since CSB experts have access to the information on the reported net wage, we can compare our calculations with the net wage data imputed by the the CSB of Latvia. Figure A4 and Table A9 in Appendix show that despite some differences at individual level, we capture the distribution of net wages correctly.

Next we compare the distribution of self-reported net income by LFS respondents (only for the respondents who actually answered this question, excluding imputed answers) with the net wage data from the administrative dataset. We do this comparison separately for the employees who were estimated to receive an unreported payment and for the compliant employees. Figure 4 reports the outcome of validation.

Let's focus on the subset of compliant employees who, according to our analysis, do not receive any unreported payments. Presumably compliant LFS respondents tend to under-report their net income, which can be observed by comparing both distributions. Table A10 in Appendix confirms

[26] For example, we cannot precisely define the number of dependants for each employee in the family with two adult employees and two kids (potential dependants).

Figure 4: Distribution of official and self-reported data on net wage



Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

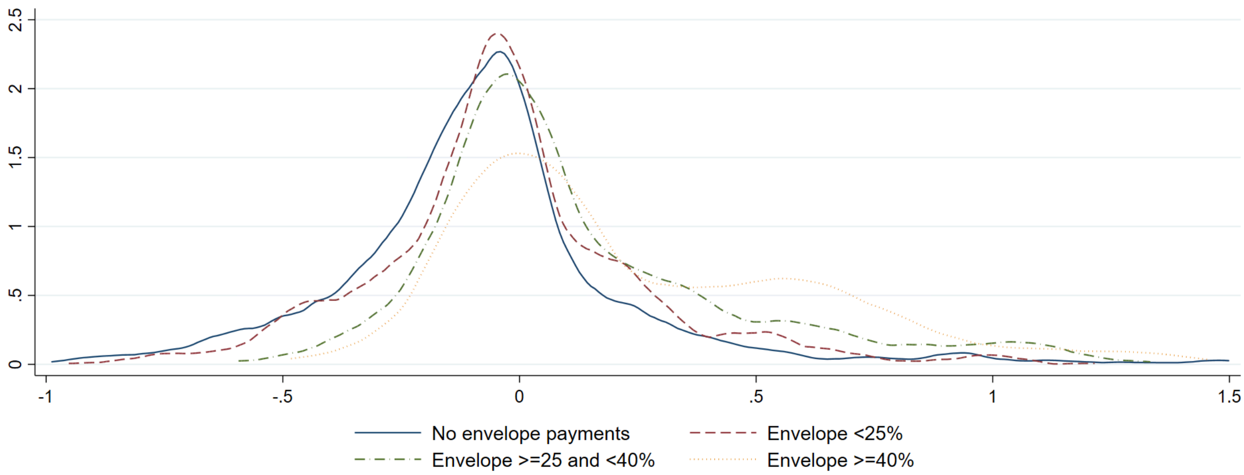
that the average compliant employee self-reports lower net income (by 7-10%) compared to the official information. The situation reverses when we focus on the employees that were predicted to receive unofficial payments. The distribution of self-reported net wage is now shifted to the right compared with official data, which is also confirmed by the results of the Kolmogorov-Smirnov test. Evading employees over-report their net income (by 3-7% on average). This difference between self-reported and official net income confirms validity of our classification into evading and compliant employees.

To check the validity of unreported payment estimates, we compare the gap between the self-reported and official net wage for the employees with different size of unreported wage payments. Figure 5 shows the corresponding distributions of the gap. Again, the results confirm our estimates of unreported payments. The higher relative size of predicted unofficial payment corresponds to a larger right shift in the distribution of the gap between the self-reported and official net wage. Table A11 indicates that the difference between the distributions is statistically significant.

6.6 Robustness check

Finally, we test the robustness of our findings to the changes in two assumptions made during the last step of our methodology. The left part of Figure 6 shows the sensitivity of wage distribution

Figure 5: Distribution of the gap between the self-reported and official net wage



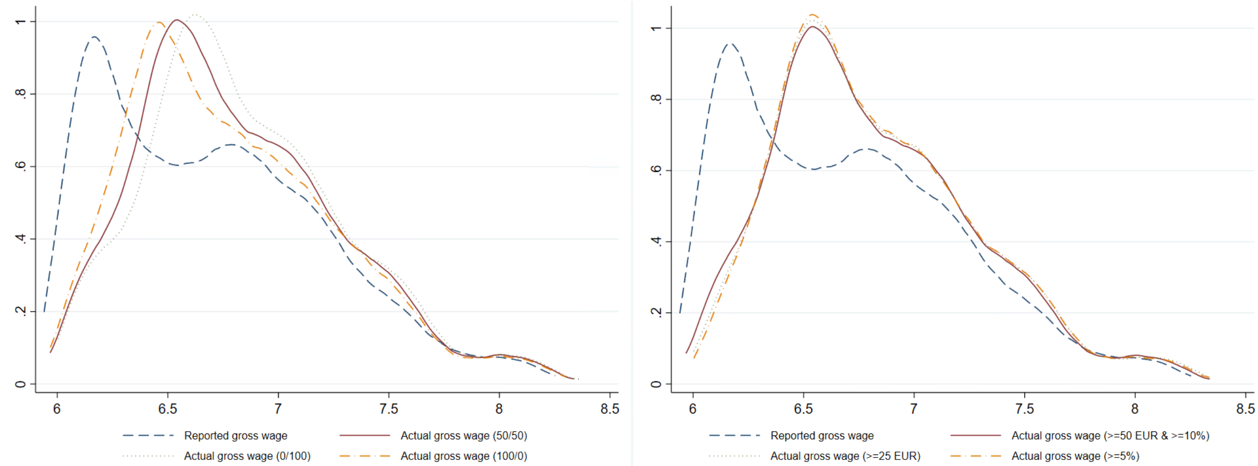
Note: the gap is calculated as the difference between the logarithm of self-reported net wage and the logarithm of offical net wage.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

to the assumption about the split of the gains from unpaid labour taxes between the employer and employee. We re-evaluate the actual wage for two extreme splits: (a) when all gains are acquired by employees (0/100), and (b) all gains are acquired by employers (100/0). Both extremes are unrealistic from the economic point of view: in the former case, a firm obtains no labour cost reduction, while in the latter case workers remain less socially protected for the same actual nominal income (although the reported income becomes lower than the actual one). However, these scenarios can serve as the upper and lower bounds for the estimates of unreported wage payments. Indeed, Figure 6 shows that the distribution of the actual wage remains shifted to the right compared with the distribution of the reported gross wage, the magnitude of this shift is sensitive to the split of the gains.

Table 14 reports a higher share of employees receiving unreported payments (as more unreported payments exceed the threshold), a higher average size of unreported payments and a larger share of unreported payments in the total gross wage fund when unpaid labour tax gains are solely obtained by employees. The outcome naturally reverses when all gains from unpaid labour taxes are acquired by firms. As mentioned above, both extreme splits can serve as the lower and upper bounds for the evaluation of unreported payments. In particular, the share of unreported wage payments in the total reported gross wage fund is evaluated to be in the range of 6.6-12.4%.

Another assumption was related to the minimum level of unreported wage payments when the

Figure 6: Robustness of the distribution of actual wage in 2018



Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

employee is classified as compliant. We relax the baseline threshold (at least 50 EUR and at least 10% of the reported gross wage) in two ways. One scenario assumes that the size of the unreported wage payment should be at least 25 EUR, the other – at least 5% of the reported wage. The distribution of the actual wage is almost insensitive to the change in the threshold: while the number of unreported payment receivers goes up, the sum of such payments is small and does not change the distribution of actual income (see the right part of Figure 6). The share of employees receiving unreported payments naturally increases, and the average size of unreported wage – declines (see Table 14), but alternative assumptions regarding the threshold do not affect our evaluation of the total sum of unreported wage payments.

Table 14: Robustness of aggregate envelope payments statistics in 2018

Estimate	Share of employees receiving unreported wage payments	Average size of the unreported wage payments relative to the official gross wage for employees with non-zero unreported payments	Share of unreported wage payments in the total reported gross wage fund	Gini coefficient for the actual gross wage
Official gross wage	-	-	-	0.485
Actual gross wage (baseline)	46.5%	29.5%	9.5%	0.447
Actual gross wage (all gains to employees)	50.3%	35.1%	12.4%	0.438
Actual gross wage (all gains to employers)	40.9%	23.6%	6.6%	0.456
Actual gross wage (envelope≥25 EUR)	67.3%	22.7%	10.9%	0.456
Actual gross wage (envelope≥5%)	60.2%	25.5%	10.5%	0.450

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Despite some substantial changes in the evaluation of the average size of unreported payments or the share of employees involved in unofficial wage payments, the last column of Table 14 shows little impact on the Gini coefficient. Any assumptions regarding the split of the gains or the threshold for unreported wage preserves the conclusion about lower wage inequality after accounting for the unofficial payments.

7 Conclusions

We propose a novel approach to evaluating the size of unreported wage payments at employee level based on the employer-employee official income data combined with the survey information on various person-level characteristics, including education, experience, occupation and contract type. After the detection of tax-evading firms, we estimate the Mincer earning equation by the SFA. This allows for two stochastic components: (a) the idiosyncratic error term accounting for unobserved wage determinants like ability, and (b) the inefficiency term representing the unobserved illegal wage payment. The key point of the methodology is to allow for the heterogeneity in the inefficiency term, relating the variance of the inefficiency with predicted probability of the respective firm to evade labour taxes. Thus, we restrict unreported payments to be negligible in compliant firms, while allowing positive unreported payments for the tax-evading enterprises.

Our approach represents an alternative to the existing methodologies evaluating the unreported wage payments: audit data and surveys, a consumption-based approach and detection of discrepancies between different income data sources. Compared with direct surveying approaches, our estimates use administrative data on income, thus reducing the impact of untruthful responses. A larger sample size is another potential advantage compared to the alternative evaluation techniques: the estimation of Mincer earning regression by the SFA does not require data on self-reported income or consumption. The reliance on administrative data combined with a larger sample leads to the main advantage of our approach – the opportunity to observe the distribution of unreported wage payments and the actual wage income.

We find substantial differences in the degree of labour tax evasion for different types of firms and employees in Latvia. In particular, we confirm the previous empirical findings that small and younger firms are involved in the labour tax fraud more often. Agriculture, construction, professional and other services are the least compliant industries. Regarding the difference between employees – the unofficial wage payments are more frequent and relatively more sizeable for the employees with a smaller reported wage. More than two thirds of workers receiving the minimum wage (according to the official information) get on average another 50% in cash on top of their reported gross wage. As a result, the wage income inequality in Latvia appears to be less pronounced if the unreported wage payments are taken into account. This, however, reflects only the current income situation, as workers with low-to-medium official wages are less socially protected and will receive lower pensions in future.

Measuring unobserved tax-evading behaviour is based on numerous strong assumptions. We need to define a set of truly compliant and evading firms to decide on financial indicators containing traces of tax evasion, to assume how the gains from unpaid labour taxes are split between the employer and the employee, to set the lower threshold for the unreported wage payments. The estimated level of labour tax evasion is sensitive to those assumptions and should be taken with caution, while the conclusion on the lower wage income inequality after accounting for the unreported payments remains robust.

Some of the above assumptions can be improved given the better data availability, i.e. the audit data can help forming the control and treated firm groups. Another limitation of our approach is related to the fact that there is no unique way of splitting the residual of the SFA earnings regression into idiosyncratic and inefficiency terms: we use the expected level of the inefficiency instead. This reduces the reliability of the unreported payment estimates for each individual employee and masks the heterogeneity of unreported payments at individual level, without affecting our conclusions about the distribution. Finally, one should remember that our approach works only for the under-reported wage, but does not capture unofficially employed workers and does not account for the possibility of under-reported working hours. All in all, our novel methodology gives an imperfect, but useful alternative way to reveal the fraudulent behaviour at disaggregated level, providing an additional tool to researchers and policy makers.

References

Aigner, D., C. A. K. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics 6*(1), 21–37.

Almunia, M. and D. Lopez-Rodriguez (2018). Under the radar: The effects of monitoring firms on tax compliance. *American Economic Journal: Economic Policy 10*(1), 1–38.

Artavanis, N., A. Morse, and M. Tsoutsoura (2016). Measuring income tax evasion using bank credit: Evidence from Greece. *Quarterly Journal of Economics 131*(2), 739–798.

Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal 55*(5), 24–36.

Braguinsky, S. and S. Mityakov (2015). Foreign corporations and the culture of transparency: Evidence from Russian administrative data. *Journal of Financial Economics 117*(1), 139–164.

Braguinsky, S., S. Mityakov, and A. Liscovich (2014). Foreign corporations and the culture of transparency: Evidence from Russian administrative data. *The Journal of Law and Economics 57*(2), 281–319.

Cecchini, M., H. Aytug, G. J. Koehler, and P. Pathak (2010). Detecting management fraud in public companies. *Management Science 56*(7), 1146–1160.

DeBacker, J., B. T. Heim, and A. Tran (2015). Importing corruption culture from overseas: Evidence from corporate tax evasion in the United States. *Journal of Financial Economics 117*(1), 122–138.

Eurobarometer (2020). Undeclared work in the European Union. Technical Report 498, European Commission.

Fadejeva, L. and I. Opmane (2016). Internal labour market mobility in 2005-2014 in Latvia: the micro data approach. *Baltic Journal of Economics 16*(2), 152–174.

Gavoille, N. and A. Zasova (2021a). Foreign ownership and labor tax evasion: Evidence from Latvia. *Economics Letters 207*(C), 1–4.

Gavoille, N. and A. Zasova (2021b). What we pay in the shadow: labor tax evasion, minimum wage hike and employment. Research Paper 6, SSE Riga/BICEPS.

Gorodnichenko, Y., J. Martinez-Vazquez, and K. Sabirianova Peter (2009). Myth and reality of flat tax reform: Micro estimates of tax evasion response and welfare effects in Russia. *Journal of Political Economy 117*(3), 504–554.

Hajek, P. and R. Henriques (2017). Mining corporate annual reports for intelligent detection of financial statement fraud – a comparative study of machine learning methods. *Knowledge-Based Systems 128*, 139–152.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Heckmann, J. J., J. Lochner, Lance, and P. E. Todd (2003). Fifty years of Mincer earnings regressions. Working Paper 9732, NBER.

Hurst, E., G. Li, and B. Pugsley (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *The Review of Economics and Statistics 96*(1), 19–33.

International Monetary Fund (2021). *The Global Informal Workforce: Priorities for Inclusive Growth.* Washington, DC: International Monetary Fund.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? Evidence from a tax audit experimentin Denmark. *Econometrica 79*(3), 651–692.

Kukk, M., A. Paulus, and K. Staehr (2020). Cheating in Europe: underreporting of self-employment income in comparative perspective. *International Tax and Public Finance 27*, 363–390.

Kukk, M. and K. Staehr (2014). Income underreporting by households with business income: Evidence from Estonia. *Post-Communist Economies 26*(2), 257–276.

Kumbhakar, S. C. and C. A. K. Lovell (2000). *Stochastic Frontier Analysis.* Cambridge: Cambridge University Press.

Kumler, T., E. Verhoogen, and J. Frias (2020). Enlisting employees in improving payroll-tax compliance: Evidence from Mexico. *The Review of Economics and Statistics 102*(5), 881–896.

Lemieux, T. (2006). The "Mincer equation" thirty years after schooling, experience, and earnings. In S. Grossbard (Ed.), *Jacob Mincer A Pioneer of Modern Labor Economics*, pp. 127–145. Boston, MA: Springer US.

Meeusen, W. and J. van Den Broeck (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review 18*(2), 435–444.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy 66*(4), 281–302.

Mincer, J. (1974). Education, experiene and the distribution of earnings and employment: An overview. In T. F. Juster (Ed.), *Education, Income, and Human Behavior*, Chapter 3, pp. 71–93. Cambridge, MA: NBER.

Pissarides, C. A. and G. Weber (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics 39*(1), 17–32.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review 105*(8), 2539–2569.

Putniņš, T. J. and A. Sauka (2015). Measuring the shadow economy using company managers. *Journal of Comparative Economics 43*(2), 471–490.

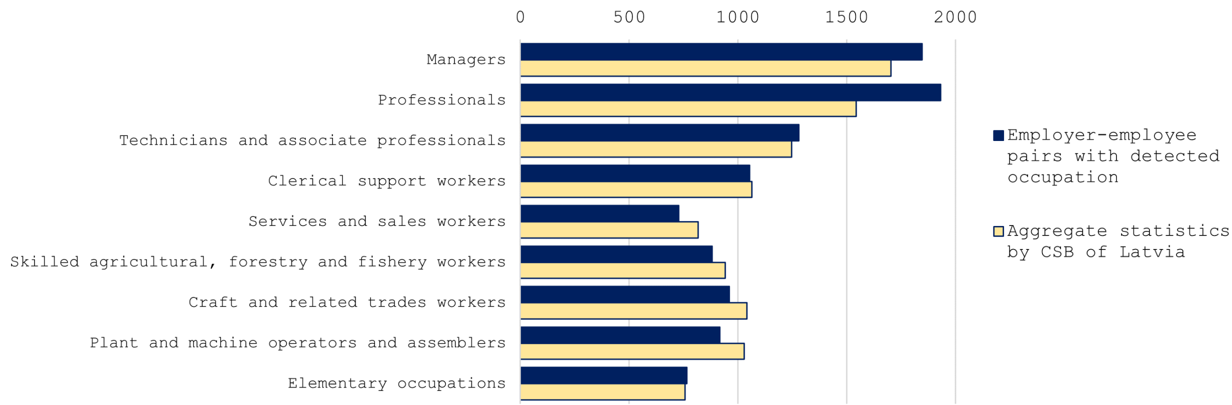
Putniņš, T. J. and A. Sauka (2021). SSE Riga shadow economy index 2009-2020. Technical report, SSE Riga.

Tonin, M. (2011). Minimum wage and tax evasion: Theory and evidence. *Journal of Public Economics 95*(11), 1635–1651.

World Bank (2017). Latvia tax review. Technical report, World Bank Group, Washington, DC.

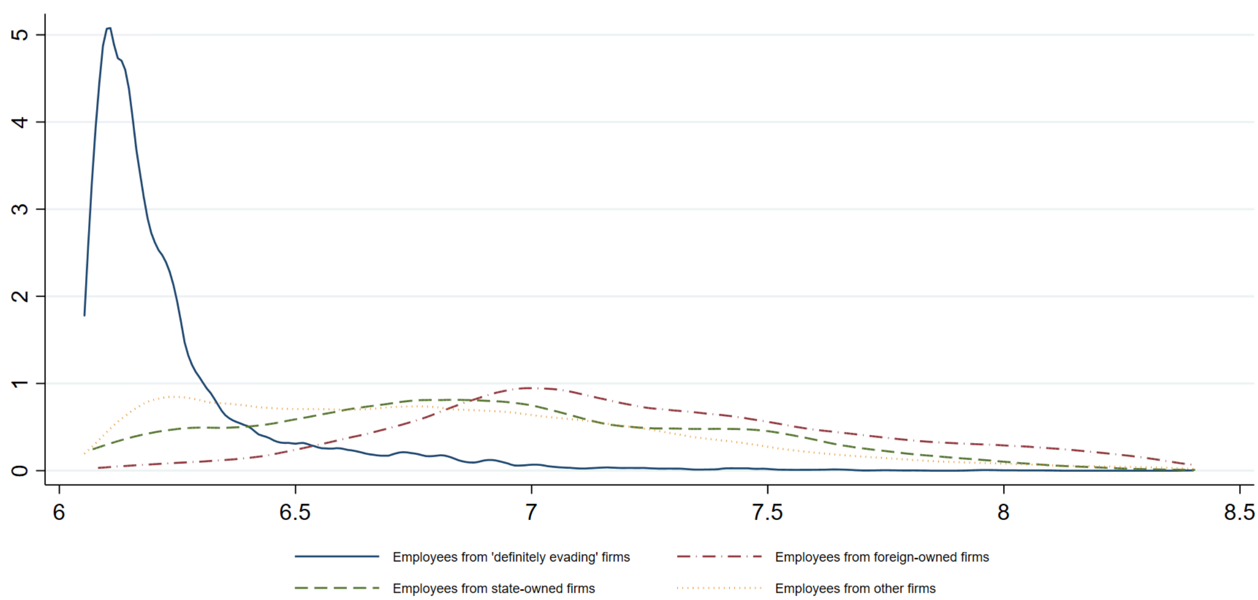
Appendix

Figure A1: Average gross FTE wage by broad occupation groups in 2018, %



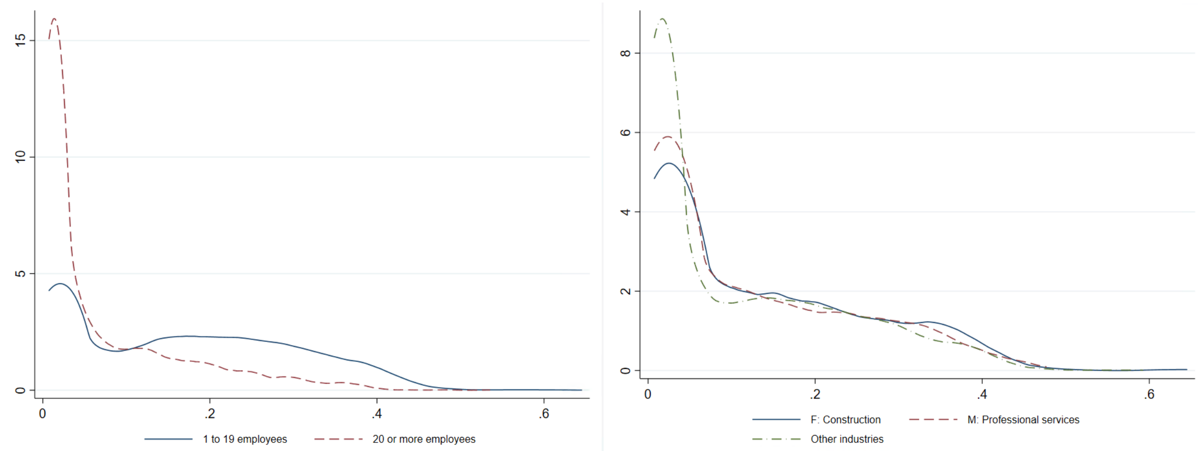
Sources: SRS of Latvia, CSB of Latvia, own calculations.

Figure A2: Distribution of the natural logarithm of wages by different employees in 2018



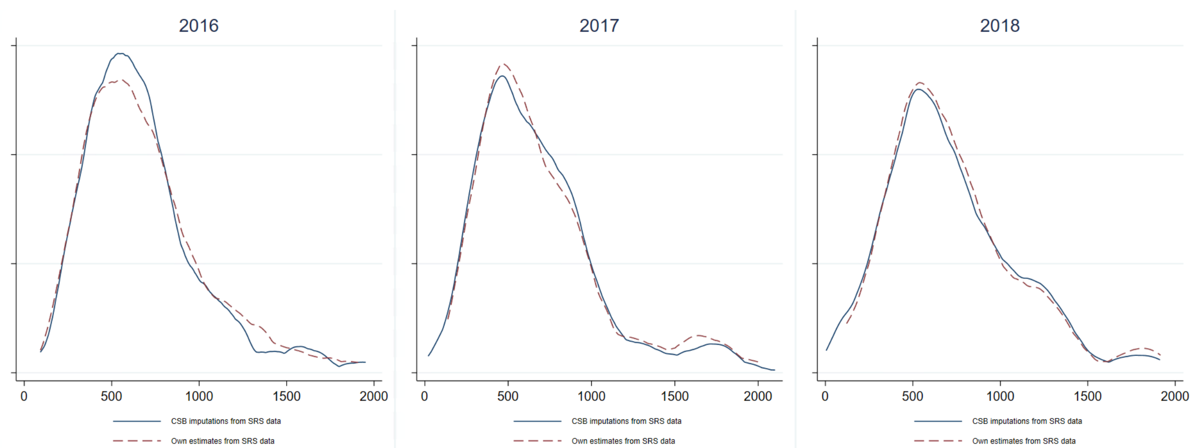
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Figure A3: Distribution of the gap between the reported and potential logarithms of gross wage by firm size and industries



Notes: the graph reports the kernel density function of the expected gap between the reported and potential logarithms of gross wage net of expected idiosyncratic term. Individual weights from the LFS are used to calculate the summary statistics.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Figure A4: Distribution of net wage: CSB imputations and own estimations from the SRS data



Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Table A1: Compliant and evading firms by year

Year	"Definitely compliant" firms: firms owned by investors from low-corruption countries	"Definitely compliant" firms: state-owned firms	"Definitely evading" firms: large share of "suspiciously low" wages
2016	219	348	1956
2017	190	153	2442
2018	174	156	2886

Note: low-corruption countries are Sweden, Finland, Norway, Denmark, Iceland, Switzerland, the Netherlands, Germany, Luxembourg, Canada and the UK.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Table A2: Compliant and evading firms by size class

Firm size class	Number of "definitely compliant" firms (control group)	Number of "definitely evading" firms (treated group)	Total number of firms
1 to 9 employees	124	2'578	22'235
10 to 19 employees	60	208	4'686
20 to 49 employees	65	78	3'149
50 to 249 employees	58	9	1'560
250 or more employees	8	0	208

Note: the size class of 1 to 9 employees does not include micro enterprises.
Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Table A3: Compliant and evading firms by industry

Industries	Number of "definitely compliant" firms (control group)	Number of "definitely evading" firms (treated group)	Total number of firms
A: Agriculture, forestry and fishing	17	120	1'479
B: Mining and quarrying	3	6	105
C: Manufacturing	50	327	3'851
D: Energy	2	5	209
E: Water supply, sewerage	1	8	178
F: Construction	19	245	3'068
G: Trade	114	1'038	9'115
H: Transportation	29	291	2'797
I: Hotels and restaurants	11	153	1'808
J: Information and communication	15	129	1'320
L: Real estate	12	97	1'569
M: Professional services	24	202	2'555
N: Administrative services	6	110	1'313
RST: Other services	5	105	1'201

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Table A4: Regressions for the effect of raising minimum wage on changes in employment in 2017 and 2018

Variable	Model 1	Model 2	Model 3
Share of minimum wage (t-1)	-	-0.119***	-0.0928***
Log-changes in minimum wage * Share of minimum wage (t-1)	-0.851***	0.234**	-0.201
Log-changes in minimum wage * Evading dummy (t-1)	-	-	-0.599***
Share of minimum wage (t-1) * Evading dummy (t-1)	-	-	-0.0375*
Log-changes in minimum wage * Share of minimum wage (t-1) * Evading dummy (t-1)	-	-	0.637**
Log of turnover (t-1)	0.0565***	0.0556***	0.0521***
Log of number of employees (t-1)	-0.179***	-0.179***	-0.183***
Merchandise exports dummy (t-1)	0.0466***	0.0453***	0.0405***
Debt to liabilities (t-1)	0.0000016	0.0000016	0.0000018
Profits to turnover (t-1)	0.0000006	0.0000005	0.0000004
Sector * Year fixed effects	Yes	Yes	Yes
Region * Year fixed effects	Yes	Yes	Yes
Number of observations	96'524	96'524	96'524
R^2	0.0872	0.0887	0.0897

Note: (*), (**), (***) indicate statistical significance levels of 10, 5 and 1 percent respectively.
Sources: SRS of Latvia, CSB of Latvia, own calculations.

Table A5: Probit regression for the probability to be employed

Variable	Probit model
Logarithm of age	-0.853
Logarithm of age squared	-0.00220
Not married	-0.563***
Number of kids	0.0689
Female	-0.242
Female * Number of kids	-0.0599
Female * Not married	0.511**
Non-Latvian	-0.0570
EU (but not Latvian) citizen	-0.743
Non-EU citizens, aliens	-0.221
Education: secondary general	0.454**
Education: secondary professional	0.379**
Education: professional	0.535*
Education: higher	0.0418
Graduation year	-0.0209*
Located in averagely inhabited territory	-0.0722
Located in rarely inhabited territory	-0.105
Region fixed effects	Yes
Years	Yes
Quarters	Yes
Number of observations	11'769

Note: (*), (**), (***) indicate statistical significance levels of 10, 5 and 1 percent respectively.
Sources: SRS of Latvia, CSB of Latvia, own calculations.

Table A6: Descriptive statistics of potential wage determinants

Variable	Mean	Standard deviation	Minimum	Maximum	Number of observations
Employee level variables					
Logarithm of official gross wage	6.63	0.509	5.86	8.28	10'712
Logarithm of age	3.72	0.308	2.71	4.32	10'712
Female	0.428	0.495	0	1	10'712
Non-Latvian	0.403	0.490	0	1	10'712
EU (but not Latvian) citizen	0.00209	0.0457	0	1	10'712
Non-EU citizens, aliens	0.148	0.355	0	1	10'712
Temporary contract	0.0162	0.126	0	1	10'712
Partial time contract	0.0557	0.229	0	1	10'712
Pensioner	0.0391	0.194	0	1	10'712
Disabled person	0.0112	0.105	0	1	10'712
Other employees	0.933	0.251	0	1	10'712
Education: secondary general	0.171	0.377	0	1	10'712
Education: secondary professional	0.236	0.425	0	1	10'712
Education: professional	0.0495	0.217	0	1	10'712
Education: higher	0.0998	0.300	0	1	10'712
Graduation year	1995	13.1	1959	2018	10'712
Experience in the current working place	7.25	7.30	0	54	10'712
Firm level variables					
Logarithm of number of employees	3.606	1.813	0.00	8.94	10'712
Logarithm of firm's age	2.36	0.700	0	3.33	10'712
State owned	0.0527	0.223	0	1	10'712
Foreign-owned (top non-corrupted countries)	0.0122	0.110	0	1	10'712
Foreign-owned (Baltic countries)	0.00436	0.0659	0	1	10'712
Foreign-owned (other OECD countries)	0.00429	0.0654	0	1	10'712
Foreign-owned (other non-OECD countries)	0.00648	0.0802	0	1	10'712
Exporter of goods	0.0905	0.287	0	1	10'712
Importer of goods	0.149	0.356	0	1	10'712
Located in averagely inhabited territory	0.222	0.416	0	1	10'712
Located in rarely inhabited territory	0.314	0.464	0	1	10'712

Sources: SRS of Latvia, CSB of Latvia, own calculations.

Table A7: Aggregate statistics on unreported wage payments by industry in 2016-2018

Industry	Share of employees receiving unreported wage payments	Average size of the unreported wage payments relative to the official gross wage for employees with non-zero unreported payments	Share of unreported wage payments in the total reported gross wage fund
A: Agriculture, forestry and fishing	46.2%	27.8%	9.1%
C: Manufacturing	39.7%	28.7%	7.9%
F: Construction	45.4%	31.4%	10.0%
G: Trade	40.6%	29.8%	8.2%
H: Transportation	34.3%	29.4%	6.9%
I: Hotels and restaurants	38.8%	30.0%	8.2%
J: Information and communication	40.2%	28.4%	7.8%
L: Real estate	46.4%	29.5%	10.4%
M: Professional services	43.5%	29.6%	8.9%
N: Administrative services	36.5%	29.0%	6.5%
RST: Other services	39.0%	30.4%	8.4%

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Table A8: Aggregate statistics on unreported wage payments by firm age in 2016-2018

Age of firms	Share of employees receiving unreported wage payments	Average size of the unreported wage payments relative to the official gross wage for employees with non-zero unreported payments	Share of unreported wage payments in the total reported gross wage fund
less than 3 years	70.4%	32.6%	19.4%
3 to 5 years	65.8%	31.5%	17.0%
more than 5 years	38.0%	29.0%	7.4%

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Table A9: CSB and own net wage data

Year	Average net wage, imputed by CSB	Average net wage, own imputations	Correlation	Kolmogorov-Smirnov statistics (p-value in parentheses)
2016	602.32	610.15	0.968	-0.0424 (0.396)
2017	596.48	610.09	0.939	-0.0491 (0.189)
2018	643.17	651.36	0.953	-0.0370 (0.390)

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Table A10: Self-reported and imputed net wage

Year	Average self-reported net wage	Average estimates for official net wage	Correlation	Kolmogorov-Smmirnov statistics (p-value in parentheses)
Employees without unreported payments				
2016	525.13	567.98	0.752	-0.137 (0.000)
2017	535.44	597.59	0.779	-0.138 (0.000)
2018	606.85	673.52	0.802	-0.140 (0.000)
Employees with unreported payments				
2016	456.49	428.86	0.787	0.169 (0.000)
2017	471.29	459.31	0.698	0.131 (0.000)
2018	528.37	492.86	0.610	0.197 (0.000)

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.

Table A11: Pairwise Kolmogorov-Smirnov test results for the distributions of the gap between the self-reported and official net wage

	Envelope 10-25%	Envelope 25-40%	Envelope ¿40%
No envelope	0.118 (0.000)	0.336 (0.000)	0.416 (0.000)
Envelope 10-25%		0.243 (0.000)	0.312 (0.000)
Envelope 25-40%			0.111 (0.064)

Sources: SRS of Latvia, CSB of Latvia, Latvijas Banka, own calculations.