Journal of Public Economics 228 (2023) 105027

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

What we pay in the shadows: Labor tax evasion, minimum wage hike and employment

Nicolas Gavoille [a,*], Anna Zasova [b]

[a] *Stockholm School of Economics in Riga and Baltic International Centre for Economic Policy Studies, Strelnieku iela 4a, Riga LV 1010, Latvia*
[b] *Baltic International Centre for Economic Policy Studies, Strelnieku iela 4a, Riga LV 1010, Latvia*

ARTICLE INFO

JEL classification:
J08
J38
H26

Keywords:
Minimum wage
Employment
Tax evasion

ABSTRACT

The interactions between minimum wage policy and tax evasion remain largely unknown. We study the firm-level employment effects of a large and biting minimum wage increase in the context of widespread wage underreporting. We apply machine learning to classify firms as either tax-compliant or tax-evading. We then show that firms engaged in labor tax evasion are insensitive to the minimum wage shock. Our results indicate that these firms use wage underreporting as an adjustment margin, converting part of their formerly undeclared cash payments into official wages. Increasing the minimum wage improves tax enforcement, but comes at the cost of negative employment consequences for compliant firms.

1. Introduction

How do firms respond to minimum wage shocks? The vast literature studying the employment effect of minimum wage hikes remains largely inconclusive. The few papers examining firm-level employment responses document relatively small effects (Machin et al., 2003; Mayneris et al., 2018; Harasztosi and Lindner, 2019). A possible explanation is that firms may use margins other than employment to absorb the shock, such as a price pass-through (Harasztosi and Lindner, 2019; Renkin et al., 2020; Allegretto and Reich, 2018), profits (Draca et al., 2011; Harasztosi and Lindner, 2019; Bell and Machin, 2018; Drucker et al., 2021), substitution towards higher-skilled labor (Clemens et al., 2021), fringe benefits (Clemens et al., 2018) but also compliance (Clemens and Strain, 2022). The interaction between minimum wage policy and labor tax evasion, however, remains largely unexplored.

This paper studies the firm-level employment effects of a minimum wage hike, conditional on labor tax evasion. We focus on a prevalent form of labor tax evasion known as "envelop wages", which consists of unreported cash-in-hand complements to the official wage. Income underreporting is a widespread phenomenon documented in various countries, including Russia (Gorodnichenko et al., 2009), the Baltics (Putniņš and Sauka, 2015), Hungary (Tonin, 2011), Turkey (Pelek and Uysal, 2018), Argentina (Perry et al., 2007), and the US self-employment sector (Hurst et al., 2014). In this context, minimum wage policy can serve as a fiscal tool: a minimum wage hike pushes firms to convert part of the envelope wage into official wages to comply with the new level, allowing them to remain under the tax authorities' radar. From an employment perspective, unreported wages may act as a buffer to absorb minimum wage shocks (Tonin, 2011). At the same time, not all firms are tax-evaders, and some firms genuinely pay their employees the minimum wage.

We focus on a sequence of two minimum wage hikes in Latvia in 2014 and 2015, representing a combined 26% increase in the nominal minimum wage and impacting more than 20% of the workforce. Envelope wages are a significant issue in the Latvian labor market, with more than one in ten employees in Latvia admitting to receiving envelope wages (Special Eurobarometer, 2014). Putniņš and Sauka (2015) estimate that 34% of total wages are paid in envelopes in Latvia. Exploiting a unique combination of administrative and survey data structured around a matched employer-employee dataset with monthly frequency for the 2011–2017 period, we construct a firm-level binary tax compliance classification. We then use this classification to investigate whether compliant and tax-evading firms have a different employment response following a minimum wage hike. We show that in firms heavily impacted by the reform, the employment response is three times greater for compliant firms than for evading ones in the year after the hike.

Our empirical analysis is composed of three main steps. First, we introduce a methodology for classifying firms as either compliant or tax-evading using supervised machine learning. This approach relies

* Corresponding author.
E-mail addresses: nicolas.gavoille@sseriga.edu (N. Gavoille), anna@biceps.org (A. Zasova).

[1] See Athey and Imbens (2019) for a brief introduction to the method for economists.

https://doi.org/10.1016/j.jpubeco.2023.105027
Received 31 October 2022; Received in revised form 3 November 2023; Accepted 6 November 2023
Available online 28 November 2023

on three key components: (i) a classification algorithm; (ii) firm-level information to be used as predictors and (iii) a sample of firms for which we know the "true" type to train the algorithm. For the algorithm, we apply gradient boosting decision trees (Friedman, 2001; Chen and Guestrin, 2016).[1] To select predictors, we draw upon a substantial body of literature in accounting and computer science that focuses on fraud detection, using information from firms' balance sheets and income statements (Beneish, 1999; Cecchini et al., 2010; West and Bhattacharya, 2016). The primary challenge lies in obtaining a sample of firms for which we know their compliance status. We construct a sample based on (strong) assumptions. First, we consider firms owned by Nordic companies to be compliant. DeBacker et al. (2015) provide evidence that tax morale is imported into foreign-owned firms. Denmark, Finland, Norway and Sweden are regarded as benchmarks for legal compliance and consistently stand at the top of rankings such as the Corruption Perception Index. Second, for the subsample of labor tax-evading firms, we use firms that are paying a "suspiciously low wage" to their employees. In practice, we link the matched employer-employee dataset to multiple waves of the Labor Force Survey (LFS). This enables us to estimate a wage equation for a subsample of employees, regressing the administrative wage with a wide range of individual characteristics. We can then identify firms employing workers paid far below predictions and consider these firms as tax-evaders.

Applying the model to the entire population of firms, we estimate that 34% of the companies in our sample, covering 18% of employees, engaged in labor tax evasion during the 2011–2013 period. These results are aligned with both aggregate and survey estimates. In particular, smaller firms exhibit a higher likelihood of underreporting, with the construction sector being one of the most affected, as documented by Putniņš and Sauka (2015).

In a second stage, we conduct three main checks to ensure the validity of our classification, which is based on admittedly strong assumptions. First, comparing employee-level survey data with *administrative* data, we observe that employees of firms classified as tax-evading report higher wages on average in the survey. The wage discrepancy is much smaller for employees of compliant firms. Second, we study the *administrative* wage of individuals who switch between firm types. We observe that workers transitioning from tax-evading to tax-compliant firms benefit, on average, from the largest wage increase. Third, we implement a consumption-based underreporting analysis *à la* Pissarides and Weber (1989) and do not find any sign of underreporting for households led by employees of compliant firms. However, we estimate that households led by individuals working in tax-evading firms underreport about 35% of their total income.

In a third stage, we examine the impact of the minimum wage hike on firm-level employment. Following Machin et al. (2003) and Harasztosi and Lindner (2019), we estimate the relationship between the share of workers affected by the hike and the percentage change in employment between a post-reform period t and a pre-reform reference period. The interaction between the bite of the minimum wage increase and the tax-evasion indicator allows us to investigate whether tax-evading firms have a different reaction from compliant firms. Our findings indicate that tax-evading firms remain largely unaffected by the hike, the level of exposure to the minimum wage hike not being significantly related to changes in employment. At the same time, we observe that one year after the reform, compliant firms employing only minimum wage workers experienced 11% lower employment growth compared to compliant firms with no workers affected by the policy. This negative employment effect is driven by both extensive (firm exit) and intensive margins (hiring/firing decisions). These results hold for alternative measures of firm-level treatment intensity.

The difference in firm-reaction persists over time: three years after the initial minimum wage hike, employment in tax-evading firms remains insensitive to their initial level of exposure to the minimum wage shock, whereas employment growth in exposed compliant firms experience a large decrease. At the same time, the average income increases for exposed firms, regardless of their type, as do the average income tax and social security contributions collected per employee. The minimum wage hike had a direct effect on the wage of affected workers, supporting the envelope wage conversion mechanism. A trade-off emerges for policy-makers in the context of widespread labor tax evasion: increasing the minimum wage has a positive effect on the enforcement of tax rules, which offers employees additional social protections. This, however, comes at the cost of negatively affecting tax-compliant firms exposed to this hike.

This paper sits at the intersection of the literature on minimum wage and the literature on labor tax evasion. We build upon research aimed at estimating the employment effect of a minimum wage hike (Machin et al., 2003; Dube et al., 2010; Allegretto et al., 2011; Draca et al., 2011; Meer and West, 2016; Harasztosi and Lindner, 2019; Manning, 2021; Gopalan et al., 2021; Jardim et al., 2022 among many others). We contribute to this literature in several dimensions. First, we study an episode of a minimum wage hike that is particularly large and biting. In comparison, the US minimum wage hikes studied in Cengiz et al. (2019) affect, on average, 8.6% of the workforce with an average increase of 10.1% (as opposed to 20% and 26% in our case).[2] This allows us to study firm reactions without the need to narrow our focus to a specific sector, which could lead to sample selection issues (Manning, 2021). Second, the monthly frequency of our data enables us to precisely observe firm reaction timing in the short and medium run.[3] In particular, the employment response begins to exhibit a negative trend between the announcement and the implementation of the reform.[4] Third, while the literature on minimum wages has mostly focused on the US and other advanced countries, such as the UK (Machin et al., 2003) and Germany (Dustmann et al., 2022), we contribute to a growing body of research examining the effects of minimum wage policy in less developed countries.[5] Fourth, several papers investigate how minimum wage hikes impact different *types* of firms. For instance, Chava et al. (2023) show that small firms experience higher financial stress in the aftermath of a minimum wage increase. Ahn et al. (2022) document that multinational firms are more sensitive to minimum wage hikes than domestic firms. Luca and Luca (2019) provide evidence that high-quality restaurants are more likely to survive a minimum wage increase than low-quality ones. We contribute to this literature by examining another key characteristics of firms. Finally, our results also contribute to explaining the large heterogeneity in the employment effect of minimum wage hikes across countries (Neumark and Corella, 2021). For instance, the very low employment elasticity documented by Harasztosi and Lindner (2019) in Hungary could be (partially) explained by firms absorbing the shock through the "envelope margin", as labor tax evasion is well documented in Hungary (Tonin, 2011; Bíró et al., 2022). The effect estimated using all firms is indeed a weighted average of the reaction of tax-compliant and tax-evading firms.

Second, our paper contributes to the literature that examines the interaction between labor tax evasion and minimum wage policy. A series of papers investigate potential non-employment margins that firms can use to absorb minimum wage shocks (see Clemens, 2021

[2] The magnitude and the bite of the increase in minimum wage have both been shown to be related to the size of the employment response (Clemens et al., 2021; Clemens and Strain, 2022; Gregory and Zierahn, 2022).

[3] To our knowledge, the only other paper on minimum wage using monthly frequency data is Georgiadis and Manning (2020).

[4] The minimum wage hike taking place on January 1, 2014 was announced in May 2013.

[5] Several papers study the impact of minimum wages on the informal sector (see for instance Lemos, 2009, Bosch and Manacorda, 2010 and Meghir et al., 2015). In this literature, individuals have to choose whether to work in the formal or the informal sector. The type of informality that we study in this paper is different, as informality here is located at the intensive margin of formality.

for an overview). Among these margins, several papers document that a higher minimum wage is associated with a higher level of non-compliance, measured by the prevalence of subminimum wage payments (e.g., Ashenfelter and Smith, 1979; Goraus-Tańska and Lewandowski, 2019; Clemens and Strain, 2022; Basu et al., 2010; Garnero and Lucifora, 2022). In this paper, we explore the flip side of the same coin: how a minimum wage can actually *improve* tax enforcement. Tonin (2011) provides a theoretical framework showing that when labor tax evasion is prevalent at the intensive margin (as is the case in our context), minimum wage policy becomes a fiscal policy tool. This model yields two key results: (1) labour tax evaders bunch at the minimum wage, generating a spike in the wage distribution; (2) unreported income acts as a buffer to absorb the minimum wage shock. Several papers provide empirical support for the first point (Tonin, 2013; Bíró et al., 2022; Gavoille and Zasova, 2022).[6] Whether firms can use wage underreporting as an adjustment margin, however, remains unknown. Our aim is to fill this gap.

The paper is structured as follows: Section 2 describes the institutional framework of the Latvian labor market and the data. In Section 3, we introduce our methodology to classify firms between compliant and tax-evading ones, and provide several validation checks. Section 4 provides estimates of the impact of the minimum wage hike. Section 5 concludes.

2. Institutional context and data

2.1. Institutional context

The minimum wage in Latvia applies to all employees, across all regions and industries.[7] It is determined on a monthly basis for full-time jobs through special government decrees following consultations with social partners. The government is not required to regularly revise the minimum wage. Between 2010 and 2020, the ratio of the minimum wage to the median wage of full-time workers in Latvia fluctuated between 0.47 and 0.52 (the average ratio in OECD countries with a statutory minimum wage ranged from 0.50 to 0.54 during the same period). The gross minimum wage from January 2011 to December 2013 was 200 Lats, equivalent to 285 euros[8]. In June 2013, the government approved a proposal to raise the minimum wage to 320 euros on January 1, 2014, with support from social partners and the finance ministry. In 2014, the Finance Ministry suggested another minimum wage increase to 330 euros, but it was not approved. After the general election in November, the new governing coalition approved a minimum wage raise to 360 euros in December 2014, to take effect the following month. The minimum wage thus increased to 360 euros on January 1, 2015. The minimum wage remained stable (with a yearly update to adjust for inflation, which remained very low over the last decade) until January 1, 2018, when it increased to 430 euros. Figure Fig. 1 displays the wage distribution in January 2013. The dashed line at EUR 285 represents the (gross) minimum wage in place in 2013. This figure illustrates how hard the minimum wage hike is biting. With the minimum wage increasing from 285 to 320 euros in 2014 and then to 360 in 2015, the two consecutive hikes represent a cumulative nominal increase of 26%, affecting more than 20% of jobs.

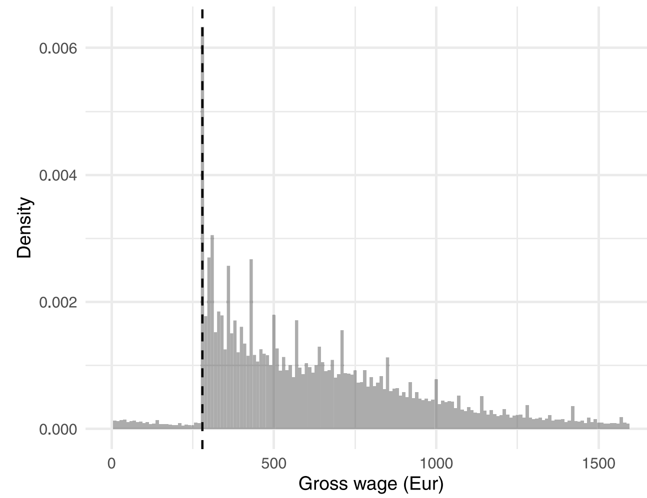


Fig. 1. Wage distribution in Latvia. Note: This figure displays the monthly wage distribution for full-time employment in Latvia in January 2013, expressed in Euro. The vertical dashed line indicates the (gross) minimum wage in 2013.

One of the explicit motivations for the Finance Ministry to support this series of wage hikes was to reduce the size of the shadow economy by curbing underreporting behavior. As formalized by Tonin (2011), setting a minimum wage imposes a constraint on the decision to underreport, as full-time contracts officially paid below the minimum wage are more easily detected by fiscal authorities. Consequently, a spike in the wage distribution emerges at the minimum wage level. Tonin (2013) therefore argues that the magnitude of this spike is correlated to the prevalence of income underreporting. Thus, the very large spike visible in Fig. 1 suggests the prevalence of labor tax evasion in Latvia. Envelope wages are indeed a significant issue in the Latvian labor market. Non-declared cash payments in addition to the official wage are considered the most significant tax fraud issue (World Bank, 2017; OECD, 2019), whereas the share of employed workers without any contract is smaller than in the majority of other European countries (Hazans, 2011). In a European-wide survey focusing on income underreporting, Latvia ranks first for the share of respondents who say they have received envelope wages (Special Eurobarometer, 2014). This is likely a lower-bound estimate since respondents earning part of their wage in undeclared cash may either untruthfully answer or refuse to answer. Using indirect survey methods, Putniņš and Sauka (2015) estimate that 34% of total wages in Latvia are paid in envelopes. Jascisens and Zasova (2021) provide evidence of a sharp *increase* in pregnant women's wages during the time period when parental benefits are calculated, which they interpret as a shift from envelope to official wages.

Regarding taxation, personal income tax was imposed at a flat-rate (24% in 2013 and 2014, then reduced to 23%) before the introduction of different tax brackets in 2018. To introduce a degree of tax progressivity, income below a certain threshold is exempt from personal income tax. This non-taxable threshold increased from 50 euros per month in 2009 to 100 euro in 2016, but it consistently remained far below the minimum wage. The total labor cost also includes a flat-rate social security contribution shared between employers and employees (employers and employees contribute 23.59% and 10.5% of gross earnings respectively). Social security contributions are applied from the first euro of wages, resulting in a high tax wedge even for low-wage earners. Income tax and social security contributions are remitted by employers, and wages are reported to tax authorities by employers. The marginal tax wedge for minimum wage earners was approximately 40% during the period under study, implying significant potential benefits from wage underreporting.

[6] Tonin (2013) documents a correlation between the magnitude of the spike at the minimum wage and the prevalence of income underreporting in EU countries. Examining this spike in Latvia, Gavoille and Zasova (2022) present a body of suggestive evidence highlighting the prevalence of wage underreporting among minimum wage earners. Bíró et al. (2022) study the introduction of a "double minimum wage rule" in Hungary, which required firms to pay social security contributions for their employees based on twice the amount of the minimum wage or to face an increased probability of audit. They detect patterns consistent with the preexistence of income underreporting.

[7] Since 2019, some occupations in the construction sector benefit of a specific minimum wage, set above the general minimum wage.

[8] Latvia formally became part of the Eurozone on January 1, 2014. The Latvian currency was pegged to the Euro since 2005.

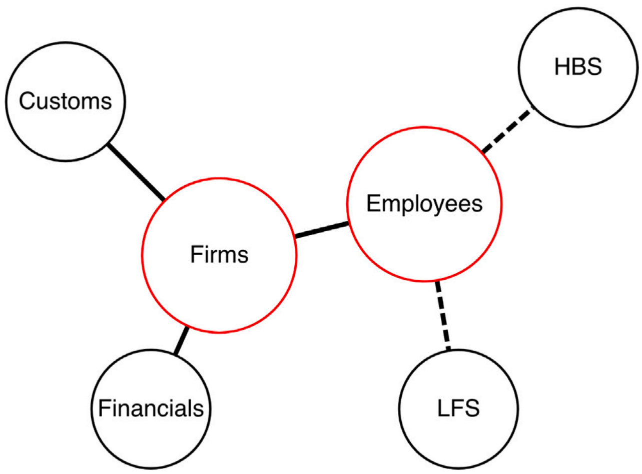


Fig. 2. Data map. Note: This figure shows the links between the different datasets that we combine. Dashed lines indicate links for a subsample of the population. Employer–employee data is available at the monthly level. All other sources are at the yearly level.

Three main changes in the policy environment took place during the period under consideration (2011–2017) that could potentially interact with our empirical analysis. First, Latvia joined the Eurozone on January 1, 2014. As mentioned earlier, the Lat had been pegged to the Euro since 2005. This formal change is unlikely to have had a significant impact on firms' business operations. Second, the 2014 Russian crisis had a significant impact on Baltic exporting firms, as evidenced by Lastauskas et al. (2021). Russia, Latvia's third trading partner at the time, imposed a trade ban on imports of various food and agricultural products from the EU in August 2014. This trade shock may have affected employment concurrently with the 2015 minimum wage hike. We address this possible issue in two ways: (i) agriculture and farming companies are not included in the analysis, as detailed below; (ii) our analysis relies on monthly employment data. We do not detect any visible drop or kink in employment dynamics in the aftermath of the ban. Third, Latvia's accession to the OECD in 2016 led to the incorporation of OECD anti-money laundering and anti-tax evasion packages into the Latvian legal framework. These packages, however, primarily focus on the transparency of international bank operations and anti-bribery policies, rather than directly targeting labor tax evasion.

2.2. Data

Our analysis relies on a combination of administrative and survey data. Fig. 2 maps the connections between the six main data sources. The core component is an (anonymized) matched employer-employee dataset with a monthly frequency. This dataset provides information on gross wages, personal income tax and social security payments for all employees. We have access to the 2010–2019 period but focus on January 2011 to December 2017, as the minimum wage changed in both January 2011 and January 2018. This dataset is collected by the Latvian State Revenue Service. It covers the entire population of firms (with some exceptions, such as micro-enterprises and certain sectors like banking and finance). It includes approximately 800,000 unique employees per month on average. This dataset enables us to measure the number of employees employed by a firm in each given month, as well as the firm-level average wage, personal income tax and social security contributions. We calculate the intensity of the minimum wage bite at the firm level using this information, as described in Section 4.

Thanks to the firm ID, we can link this data to various firm-level attributes. First, we link it to a set of general firm characteristics, including sector, date of creation, and juridical status. This dataset also contains a binary indicator of foreign ownership, indicating whether a firm is controlled by foreign capital. The country of origin of the capital is also reported. Second, we add yearly firm financial data, as reported to the tax authorities. We use it to detect firms involved in labor tax evasion, as will be explained in the next section. Third, we also link our dataset to customs data. This dataset includes information on monthly firm-level export activities, providing details on export value and destination country. We use export share (export/turnover) as a control in the employment analysis presented in Section 4.

Employees' anonymized IDs allow us to combine this administrative dataset with two national surveys. First, we link our dataset to several waves of the Labor Force Survey (LFS). This allows us to obtain highly detailed individual characteristics for a subset of employees, such as education, experience and occupation. Second, we can also link our main dataset to the Household Budget Survey (HBS), which provides information on household composition, consumption and living conditions. However, the Latvian Central Statistical Bureau (CSB) started to gather household members' individual IDs from the 2020 round of this survey (covering 2019). We will nevertheless use this information as a validation check for firm classification, as explained in the next section.

Our analysis focuses on four sectors: manufacturing, wholesale and retail, construction, and transport. We exclude state-owned firms. Similar to Harasztosi and Lindner (2019), we retain all firms that were active in January 2013 (our reference period) and already existed in January 2011, three years before the 2014 minimum wage hike. We track these firms until December 2017, the month preceding the next significant minimum wage increase. We keep all firms that exited between the reference period and the final period of the sample, assigning them 0 employees. This results in a final sample of 5,524 firms, representing 247,000 employees, which account for approcimately 30% of the total workforce in Latvia.

3. Detecting tax-evading firms

This section begins with a detailed description of the methodology we employ to classify firms as either tax-compliant or tax-evading. Given that this classification relies on a set of admittedly strong assumptions, we then proceed with a series of validation checks.

3.1. Description of the approach

The central question this paper aims to answer is whether labor tax-evading firms absorb minimum wage shocks differently from tax-compliant firms. Addressing this question requires disentangling compliant from non-compliant firms as a preliminary step. Since tax evasion is by nature largely unobservable, we would like to *predict* the firm type for all the observations in our sample. Machine learning tools can be highly relevant when the goal is predictive accuracy (Varian, 2014).[9] Implementing a (supervised) classification task requires three key ingredients: (1) a set of variables to be used as predictors; (2) an algorithm learning to classify firms, and (3) a subsample of firms for which we know the "true" type, containing instances of both compliant and labor tax evading firms.

Predictors - A substantial body of accounting and computer science literature on fraud detection has demonstrated that good prediction performance can be achieved using variables derived from firms' annual financial reports (Cecchini et al., 2010; Hajek and Henriques, 2017; Huang et al., 2014). This approach builds on accounting research that establishes systematic relationships between the probability of manipulation and financial statement items (Beneish, 1999). Labor tax evasion constitutes a form of financial manipulation and is likely to

[9] See Mullainathan and Spiess, 2017 and Athey and Imbens, 2019 and for a general comparison of the goals and methods between machine learning literature and "traditional" econometrics.

result in specific balance sheet patterns, such as understatement of revenue, assets, costs, or liabilities. Economic theory does not provide a formal model for these patterns. We therefore adopt a purely data-driven approach and use a set of balance sheets and income statement items that have previously been used in the literature. A comprehensive list of features is provided in Table A.3 in Appendix A.1.

Algorithm - We implement gradient boosting decision trees (Friedman, 2001).[10] The general idea of gradient boosting is as follows: let y_i denote the realized outcome for observation i, and $f(x_i)$ denote the prediction based on the vector of predictors x. The objective is to minimize a chosen loss function $\mathcal{L}(y, f(x_i))$ with respect to $f(x_i)$ using gradient descent. The first step involves formulating an initial naïve prediction, typically the average outcome for the sample. In the second step, we compute the negative gradient $-g(x_i) = -\frac{\partial \mathcal{L}(y, F(x_i))}{\partial F(x_i)}$.[11] The third step is to fit a regression tree h to the negative gradient. Finally, in the fourth step, we partially update the initial prediction based on the learning rate ρ, so that $f \equiv f + \rho h$. Steps 2 to 4 are repeated until a predetermined number of iterations is reached. This framework allows for high flexibility, especially regarding the loss function, which only needs to be differentiable. For binary classification tasks, a Log Loss function is generally used. To prevent overfitting, a regularization term is introduced in the objective function to penalize model complexity. As each additional model partially corrects the error of its predecessor, this approach enables excellent predictive performance (Hastie et al., 2009). One drawback of this method is its black box nature, which makes it uninformative about the relationship between outcomes and predictors. For detailed implementation information, see Appendix A.2.

Training Sample - Obtaining a sample of firms for which we know the true type is not trivial. In the absence of clear-cut measures, we rely on assumptions to build this subsample. For the subset of tax-compliant firms, we assume that firms owned by a Nordic company (Denmark, Finland, Norway and Sweden) are tax-compliant. This assumption is motivated by several observations. First, Braguinsky et al. (2014), Braguinsky and Mityakov (2015) document a greater transparency of wage reporting in foreign-owned firms operating in Russia. Gavoille and Zasova (2021) obtain similar results in Latvia: employees of foreign-owned firms are less likely to receive envelope wages. Second, DeBacker et al. (2015) document a strong correlation between a foreign-controlled owner's cultural norms and illicit corporate activities, in particular regarding tax compliance. Liu (2016) further shows that a CEO's cultural background is linked to corporate misconduct. At the same time, Fisman and Miguel (2007) show that misconduct by United Nations officials in Manhattan is correlated with the corruption and legal enforcement norms in the country of origin. The Nordic countries regularly top international rankings of tax compliance and control of corruption, such as the Corruption Perceptions Index.[12] Nordic-controlled firms account for almost 30% of foreign-owned firms in Latvia, and operate in various sectors. Nordic-controlled firms account for almost 30% of foreign-owned firms in Latvia and operate in various sectors. We focus on the four most represented sectors: manufacturing, wholesale and retail, construction, and transport.[13]

For the sample of tax-evading firms, we use a subset of firms that pay a "suspiciously low wage" to their employees. To spot these wage anomalies, we estimate a wage equation regressing (the log of) administrative wages on individual employee characteristics. Once identified, we track the firms for which these individuals work and consider them as tax-evading.

This process involves linking the matched employer–employee dataset to the Labor Force Survey for the 2011–2013 period, which precedes the minimum wage hike. LFS provides a wealth of individual characteristics not included in the employer–employee data, such as occupation, experience, and education. Using this additional information, we can estimate a wage equation for a subsample of employees, regressing the *administrative* wage on a large amount of individual characteristics. Pooling these three years, excluding self-employed, employees of public firms and keeping only full-time employees in the four sectors of interest, we estimate this wage equation using approximately 5300 employees. The R^2 is equal to 0.25, which aligns with findings in other papers that estimate this type of wage regression (see Table A.1 in Appendix A.1 for more details). We consider that employees in the bottom 10% of the residual distribution are suspiciously low-paid, and treat the firm employing them as tax-evading.[14] This is of course a strong assumption. In support of this assumption, we know that a large share of the employees bunching at the minimum wage (see Fig. 1) do receive envelope wages, as evidenced by Tonin (2011). About two-thirds of the anomalies we identify involve individuals being paid at the minimum wage, whereas our wage model predicts a much higher income. Second, the Latvian State Revenue Service uses a simpler version of this approach when deciding whether to audit a firm (among other indicators). From aggregate data, however, we know that among firms audited following suspicion of underreported earnings, irregularities are detected in 90% of cases. Hence this approach may be crude, but rather effective.

To sum up, we train a gradient boosting algorithm to distinguish between compliant and tax-evading firms using firms' financial variables as inputs and a sample of firms composed of (i) Nordic-owned firms, and (ii) firms employing workers paid much below their predicted wage as examples of the two classes. Descriptive statistics for the two groups are displayed in Table A.2 in Appendix A.1. This appendix also provides a detailed description of the procedure used to construct the training sample.

Equipped with these three components, the general procedure is as follows: First, the subsample of firms for which we know the true type is randomly split into two parts: the training sample (80% of observations) and the test sample (the remaining 20%). Second, we set the number of trees contained in the ensemble to 100 and fine-tune the model using four hyperparameters: the maximum tree depth, the minimum number of data points in a node that is required for the node to be split further, the learning rate ρ, and the reduction in the loss function needed to trigger further splitting. We use 10-fold cross-validation to find the optimal set of hyperparameters, adopting the Precision-Recall AUC as the performance metric. We choose this metric for two main reasons. First, Precision-Recall AUC focuses on predicting the positive class, prompting the model to exercise caution when making positive predictions and hence reducing the risk of false positives. Second, this metric is particularly suitable in the context of class imbalance (Saito and Rehmsmeier, 2015).[15]

The tuned model is then applied to the firms in the test sample, which have never been "seen" by the algorithm at this stage, to

[10] We use *Extreme Gradient Boosting* (*Chen and Guestrin, 2016*) via its R implementation XGBoost. Alternatively, we also implemented random forest, support vector machine and a logit model. Gradient boosting surpassed these alternatives in all out-of-sample performance metrics.

[11] As an illustration, note that in the case of a squared loss function like in OLS, the negative gradient is simply the vector of residuals.

[12] We do not assume that Nordic-owned firms are *more* likely to comply with labor tax in Latvia than other foreign-owned firms. Rather, we simply assume that Nordic-owned firms are likely to be compliant, without making any reference nor comparison with other foreign-owned firms. Results obtained using a broader set of Western countries are available in Appendix C.

[13] Several Nordic-owned banks operating in the Baltic States have been involved in money laundering scandals in the recent period. Banking and financial sectors are not included in our sample.

[14] A potential concern could be that here we capture low-productivity firms rather than tax-evading firms. If that were true, we would expect employment in these firms to be more sensitive to minimum wage hikes. However, we show in Section 4 that this is not the case.

[15] Alternatively, fine-tuning the model with respect to ROC AUC metrics provides similar performance on the test set.

Table 1
Out-of-sample performance.

		Actual	
		0	1
Prediction	0	112	20
	1	2	33
PR-AUC	0.949		
ROC-AUC	0.914		
F_1	0.911		
Accuracy	0.868		
Kappa	0.666		

Note: This table displays classification performance metrics evaluated on the test sample. Prediction $= 0$ and Prediction $= 1$ respectively indicate a firm classified as tax-compliant and a firm classified as tax-evading. See Appendix A for technical details.

Table 2
Classification results.

	Overall	Construction	Trade	Manufacturing	Transportation
% firms	0.341	0.39	0.298	0.229	0.584
% employees	0.182	0.226	0.227	0.069	0.258

Note: This table displays the share of firms classified as labor tax evading for the four sectors under consideration, and the respective within-sector share of workers employed in these firms.

assess out-of-sample performance. One straightforward approach is to compare the predicted class with the actual class for observations in the test set. Gradient boosting provides an output score in the range $[0, 1]$, which can loosely be interpreted as the probability of being classified as positive (i.e., an evader in our case). We then need to apply a cutoff to this score to obtain a binary classification. In the baseline results, we choose a cutoff that maximizes the F_1 measure, which is the harmonic mean of precision and recall (bounded between 0 and 1, with larger values indicating better performance). This choice directly aligns with our focus on the Precision-Recall curve. Table 1 displays the classification table using this threshold. Note that the model is quite conservative and classifies very few compliant firms as tax-evading, as expected. Several standard performance metrics are included in the bottom part of this table. Both Precision-Recall AUC and ROC AUC are remarkably high, a coefficient of 0.9 denoting very good performance. Accuracy is also high, even though the base rate (the accuracy we would obtain by naively classifying all the observations as compliant) is already high due to class imbalance. Appendix A.2 contains additional information for performance assessment: the Precision-Recall curve (Fig. A.2), the ROC curve (Fig. A.3), and the density of the estimated scores in the test set for the two classes (Fig. A.1). Finally, we evaluate the relative importance of inputs for the classification using the permutation procedure (see Greenwell et al., 2020). Results are provided in Fig. A.4 in Appendix A.2.

The final step is to classify all the firms in our analysis. To be on the safe side and remain as conservative as possible, in our definition of tax-evading firms, we label a firm as an evader if it is classified as evading for all three years from 2011 to 2013. The results of this classification are provided in Table 2. Overall, 34% of the 5,524 firms in our dataset are considered to be tax-evading. The fact that these firms cover 18.2% of the employees indicates that evasion mostly occurs in small firms, consistent with existing survey evidence (Putniņš and Sauka, 2015). The proportion largely varies across sectors. The relative prevalence of tax evasion across sectors is also in line with the literature, the construction sector often being reported as particularly affected by envelope wages, whereas manufacturing is known to be one of the least impacted (Putniņš and Sauka, 2015). This suggests that the classification procedure provides reasonable results. Note that we do not mean that 18.2% of workers are receiving envelope payments. Such a statement would require assuming that (i) a non-compliant firm underreports wages for *all* its employees, (ii) absolutely no employee working in compliant firms receives envelope wages.

Table 3
Difference administrative/reported wage.

Statistic	N	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
Compliant firms	1,759	2.085	186.965	−80.496	−3.178	75.587
Tax evading firms	421	76.270	193.056	−24.832	42.045	147.865

Note: This table displays the distribution of differences between the wage stated in LFS and administrative wage (both in Euro, at the monthly level), for employees of compliant firms and tax evading firms separately. The period covered is 2011–2013 (pre-minimum wage increase).

3.2. Validation of classification

The firm classification between compliant and labor tax-evading relies on strong assumptions. In this subsection, we implement three checks to demonstrate the relevance of this approach.

First, we exploit the fact that LFS provides a self-reported measure of income. We know which firm LFS respondents work for, and we have access to their administrative wages. We can thus compare the difference between survey income and administrative income for employees of compliant firms to the difference for employees of tax-evading firms. For the classification to make sense, we should observe on average a larger positive discrepancy between the survey income and the official income for employees of evading firms than for employees of compliant firms. In the Mexican context, Kumler et al. (2020) implement a similar discrepancy approach, comparing survey income to administrative income to estimate wage underreporting. Their objective is to investigate the dependence of the discrepancy on firm size. In our case, we simply want to know whether the discrepancy is larger for employees of firms classified as tax-evading than for employees of firms classified as tax-compliant. Closer to us geographically, Paulus (2015) examines underreporting in Estonia using a similar comparison, but focuses on underreporting heterogeneity across the income distribution.

In LFS, income is right-censored at 1500 Lats (approximately 2000 Euro). We thus exclude individuals with administrative incomes exceeding 1500 Lats. We also exclude individuals with zero earnings and focus on individuals who report working full-time. Table 3 provides descriptive statistics (in euros) for the discrepancy, conditional on the type of employer (based on pre-minimum wage hike observations). The median and the mean difference for employees of evading firms amount to 14% and 26% of the monthly minimum wage, respectively. In contrast, the mean and median differences are of much smaller magnitude, with the latter even being negative for employees of compliant firms.[16] This is supportive evidence that employees of tax-evading firms are more likely to engage in income underreporting compared to employees of compliant firms.[17]

Second, we can track individuals across firms thanks to the matched employer–employee structure of the dataset. Intuitively, when workers consider changing job, they should compare their true wage to the true alternative wage. If the firm classification is meaningful, we would expect to observe, on average, a significant increase in wage when an individual switches from a tax-evading firm to a compliant one. Conversely, we should not observe much difference in wages when an employee switches from a compliant firm to a tax-evading one, as the wage increase is likely to be paid in an envelope. To examine whether

[16] In Estonia, Paulus (2015) also reports a negative difference between survey income and administrative income for employees working in sectors where income underreporting is constrained.

[17] Initially, we considered using firms employing workers reporting a higher wage in LFS than in administrative data as instances of tax evading to train the algorithm. LFS wage data is very noisy (e.g., respondents mixing gross and net wage, mixing months, mixing wage from a given employer and overall income). The difference between LFS and administrative wage hence cannot reasonably inform about the type of the employer at the *individual* level.

Table 4
Change in wage.

Statistic	N	Mean	Pctl(25)	Median	Pctl(75)
Δ w from E to C	9,281	86.023	−28.237	49.377	180.737
Δ w from C to E	9,302	−16.825	−99.737	0.000	92.935
Δ w from E to E	3,956	26.072	−40.449	10.363	88.107
Δ w from C to C	40,275	38.853	−70.595	26.387	157.830

Note: This table displays average changes in monthly wage for employees switching employers (in Euro). E and C respectively stand for $evader$ and $compliant$. The period covered is 2011–2013 (pre-minimum wage increase).

this holds true, we study all workers who changed jobs over the 2011–2013 period for which we have classifications for both the original and destination employers. We calculate the average wage over the last three months in the firm of origin (excluding the very last month, which may be truncated or subject to a departing bonus), and the average wage over the first three months in the new firm (excluding the very first month for similar reasons). We then compute the difference $\bar{Y}_{post} - \bar{Y}_{pre}$ for the four different types of transitions (evading to compliant, compliant to evading, evading to evading and compliant to compliant). The results are provided in Table 4. Employees switching from a compliant to an evading firm experience on average a *decrease* in their reported wage. On the other hand, individuals switching from an evading to a compliant firm benefit on average from a large wage increase. For employment changes within the same type of employer, the average wage change is modest and falls between these two cases. The ordering of the four transitions is the same when comparing median changes.

Third, to validate the classification we implement an expenditure method *à la* Pissarides and Weber (1989). This approach allows us to estimate the extent of underreporting in a group of households vis-à-vis a reference group of non-evading households using household survey data. The Engel curve describes the relationship between food expenditure and income. Assuming that there is no incentive to misreport food expenditure, any systematic difference in the propensity for food consumption between the two groups can be interpreted as wage underreporting.[18] This methodology has been applied in various contexts, typically comparing self-employed individuals to employees (see for instance Hurst et al., 2014 in the US, Kukk and Staehr, 2014 in Estonia, Engström and Hagen, 2017 in Sweden, Nygård et al., 2019 in Norway).

We implement a similar approach by comparing the propensity for food consumption between households where the "main breadwinner" is an employee of a compliant firm and households where the main breadwinner is an employee of a tax-evading firm. If there are no systematic differences in (true) food consumption among people working for different types of employers, but employees of evading firms tend to underreport income, individuals working in non-compliant firms will report, on average, a higher marginal propensity to consume.

We link a wave of Household Budget Survey (HBS) data to our main dataset. This Europe-wide survey was conducted annually (now every three years) in Latvia by the Central Statistical Bureau. However, the individual identifier of a household member has been collected only since the 2020 wave, which covers the year 2019. Hence, we cannot merge HBS with the rest of the data for previous years. To overcome this issue, we use the 2018–2019 firms' balance sheet to obtain a firm classification for the 2019 period. If the algorithm performs well in 2013, it should still be relevant for classifying firms in 2019. With this 2019 firm classification, we can classify households in three groups depending on the head: public sector, compliant firms, and tax-evading firms. We also include self-employed in order to compare the results to the rest of the literature. Appendix B provides a comprehensive

[18] See Appendix B for a detailed presentation of the methodology.

Table 5
Misreporting regression.

	Dependent variable:	
	ln(food expenditure)	
	(1)	(2)
β (consumption propensity)	0.401***	0.382***
	(0.072)	(0.072)
γ (compliant firms)	−0.068	−0.034
	(0.053)	(0.057)
γ (tax-evading firms)	0.168**	0.151*
	(0.077)	(0.086)
γ (self-employed)	0.255***	0.266***
	(0.107)	(0.114)
Controls	No	Yes
Observations	444	444

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Robust standard errors in parentheses. This table displays the 2SLS regression estimates of Eq. (1), instrumenting reported income by education and region. Controls are included in the specification displayed in column 2). See the list of controls in Appendix B).

description of the model and of sample construction. We estimate the following equation:

$$ln\ c_i = \alpha + \beta ln\ y_i^{reported} + \gamma D_i + X_i'\phi + \xi_i, \tag{1}$$

where c_i is the food consumption of household i and $y^{reported}$ is the total household income reported in HBS. D is a set of three dummies, each denoting the household head's employment status: employed by a compliant firm, employed by a tax-evading firm or self-employed. We use households where the household head is a public sector employee as a benchmark (as in Paulus, 2015). X_i is a set of controls (age, number of children, etc.). The reported income is endogenous by construction in the model. This equation is thus estimated via 2SLS, instrumenting reported income by education and region (as for instance in Hurst et al., 2014 and Kukk and Staehr, 2014). As is usually done in this literature, we restrict the sample to households consisting of two adults, with children or not.

We report the main results in Table 5. We do not observe any difference in propensity for food consumption between public sector employees and employees of compliant firms. However, the underreporting coefficient for employees of tax-evading firms is both positive and significant, indicating underreporting. The share of underreported income stands at 34%.[19] Note that the underreporting coefficient for employees of tax-evading firms is lower than the coefficient for the self-employed, suggesting greater underreporting in the latter group, which is often assumed to be the most prone to underreporting.[20]

4. The impact of the minimum wage hike

4.1. Empirical strategy

Equipped with this firm-level binary classification of tax evasion, we now turn to estimating the employment response to the minimum wage hike. The identification strategy closely follows Machin et al. (2003) and Harasztosi and Lindner (2019).[21] This strategy exploits heterogeneity in the impact of the minimum wage: firms employing a large share of minimum wage workers are more affected by the reform than firms employing none. This allows us to implement a difference-in-differences analysis, using the intensity of the impact as a continuous

[19] See the details of the calculus in Appendix B.

[20] Underreporting estimates for self-employed are similar to those obtained by Kukk et al. (2020) for Latvia. We obtain a larger estimate because our benchmark group is restricted to public employees whereas they use all employees.

[21] See also Draca et al. (2011) and Jardim and Van Inwegen (2019) for other papers with a similar approach.

treatment. We estimate the following regression models, which allows time effect and the impact of firm characteristics to vary over time:

$$\frac{y_{it} - y_{i,Jan13}}{y_{i,Jan13}} = \alpha_t + \beta_t FA_i + \delta_t D_i + \lambda_t FA_i \times D_i + \gamma_t X_{it} + \epsilon_{it} \quad (2)$$

The left-hand side of the equation is the percentage change in employment in firm i between period t and January 2013.[22] D_i is a binary variable indicating whether firm i is a tax-evader, FA is a measure of the exposure to the minimum wage hike for this firm and X is a set of controls. The main coefficient of interest is λ_t, which indicates to what extent the employment response differs between impacted compliant and impacted evading firms. Omitting D and its interaction with FA results in the same specification as in Harasztosi and Lindner (2019). In this case, at period t, β_t represents the difference in employment growth since January 2013 between a 100% affected firm and a non-affected firm.

As explained in Section 2, we focus on firms employing a minimum of 6 workers throughout a period of three years before the hike. Firms that shut down, experiencing a 100% decline in employment compared to the reference period, are preserved in the sample. The estimated employment effect hence captures both the intensive and extensive margins.

We measure firm exposure to the minimum wage shock in two alternative complementary ways. First, FA, the fraction of affected workers, is computed as the share of (full-time) workers receiving in January 2013 a wage below 320 euros (which is the post-hike minimum wage).[23] Second, GAP measures the increase in total wage bill required for a firm to comply with the next minimum wage level:

$$GAP_i = \frac{\sum_j max(w_{ji}^{min} - w_{ji}, 0)}{\sum_j w_{ji}}.$$

Each of these two measures is computed in two ways. The minimum wage increased by 12% in January 2014, followed by a 12.5% increase in January 2015. To examine the short-run employment effect of the 2014 hike, we calculate FA_{2014} and GAP_{2014} based on the 2014 minimum wage level. This enables us to observe employment dynamics during 2013 and 2014, excluding the impact of the subsequent minimum wage increase. For medium-run analysis, we compute FA_{2015} and GAP_{2015} based on the 2015 minimum wage level.[24]

The set of controls X contains firm age, legal status, NACE sector, average export share, average share of labor, average profit as well as the square of the latter three variables, measured between 2011 and 2013. Descriptive statistics are provided in Table 6. Fig. C.1 in Appendix C displays the distribution of FA for compliant and tax-evading firms. All regressions are estimated using the logarithm of average turnover over the 2011–2013 period as weights, similar to Harasztosi and Lindner (2019).[25]

4.2. Results

4.2.1. Baseline results - short-run employment effect

We begin with simply regressing the percentage change in employment on FA and the set of controls using the subsample of compliant firms and the subsample of tax-evading firms separately. Fig. 3(a) displays the short-run estimates of β for the subsample of tax-compliant firms. Period 0 indicates January 2013. The minimum wage hike is announced in period 5 (May 2013) and is implemented in period 12 (January 2014). A negative trend appears in the second half of 2013, between the announcement of the reform and its implementation, although the effect turns significant in the first few months following the implementation. One year after the implementation of the reform, employment growth in a compliant firm with all of its employees affected by the minimum wage hike is on average 11% lower than in a similar-but-not-affected firm. Considering that the average employment change for non-impacted compliant firms (i.e., with $FA = 0$) is 3.4%, this suggests that largely impacted compliant firms do not only grow more slowly, but also experience a *decrease* in size. Fig. 3(b) reports the results of the same exercise but with the subsample of tax-evading firms. The results sharply contrast with those in the previous figure: the coefficient associated with FA is consistently insignificant, and the point estimates are considerably smaller than those for compliant firms. In other words, the share of affected workers is irrelevant to explaining the change in employment in evading firms. These results are consistent with the hypothesis that envelope wages are used as a shock absorber.

An identifying assumption is that in the absence of a change in the minimum wage, employment in affected firms would have evolved in the same way as in non-affected firms. To verify whether this assumption is credible, Fig. 3 also reports the estimates for β when we compare employment in January 2013 to employment in past periods. For instance, the point estimate at period −12 indicates the percentage change in employment between January 2013 and January 2012 (a positive estimate indicating a decrease in employment, as per Eq. (2)). The share of affected workers is not a determinant of change in employment in the pre-minimum wage reform period, which supports the validity of the parallel trend assumption.

We now proceed with the estimation of Eq. (2) on the entire sample. We first study the short run effect, focusing solely on the 2014 minimum wage hike before considering both hikes altogether. Columns 1 and 2 of Table 7 report the regression results using the percentage change in employment between January 2013 and December 2014 with and without the set of controls, using FA as the treatment variable. These results confirm that, for a given share of affected workers, the employment response has been much more pronounced in tax-compliant firms than in tax-evading ones. Fig. 4 plots the predicted change in employment growth between January 2013 and December 2014 conditional on FA and the tax-compliance indicator.[26] This change is similar for tax-compliant and tax-evading firms not affected by the policy change. However, for firms with 100% of their employees affected, the point estimate is three times smaller for evading firms than for compliant ones (−10.8% vs −3.7%). To put the magnitude of this effect into perspective, an average compliant firm with a value of FA equivalent to the 75th percentile of the FA distribution has an expected employment growth 6 p.p. lower than an identical firm at the 25th percentile. Similar conclusions are drawn when using GAP instead of FA, as displayed in columns 3 and 4.[27] Eq. (2) assumes a linear

[22] We set January 2013 as the reference period as it is exactly a year before the implementation of the minimum wage hike and precedes discussions related to this increase.

[23] The data source does not cleanly indicate whether an employee is full-time or part-time. Appendix C describes the methodology used to disentangle the two.

[24] Several papers similarly focus on a multi-stage increase of minimum wage, studying the sequence of hikes as a whole event (e.g., Harasztosi and Lindner, 2019).

[25] Some papers use weights, others do not (e.g., Machin et al., 2003; Draca et al., 2011). All the results presented below are qualitatively insensitive to this choice.

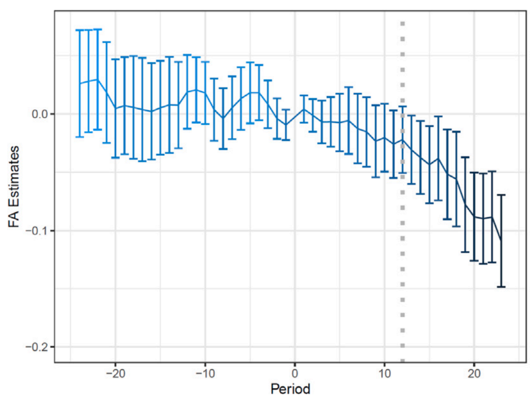
[26] A graphic representation of the interaction using GAP is provided in Fig. C.2 in Appendix C.

[27] As mentioned in the previous section, if the classification was merely capturing productivity rather than labor tax compliance, we would expect a larger employment response in the evading (low-productivity) group rather than in the compliant (high-productivity) group: low-productivity firms, with low-productivity workers, should experience a larger job destruction rate following a minimum wage hike than high-productivity firms (Dustmann et al., 2022) and should have higher probability to exit (Mayneris et al., 2018). We observe the opposite. If we consider that the set of evading firms contains *some* compliant but low-productivity firms, then the estimated employment response for evading firms is actually over-estimated, and the true response even closer to 0.

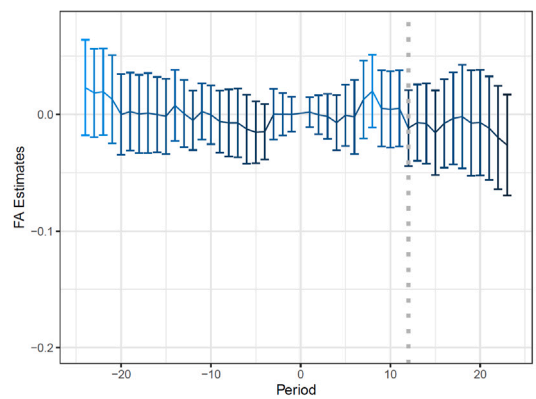
Table 6
Firm descriptive statistics.

	All N = 5,524		Compliant N = 3,683		Tax evaders N = 1,841	
	Mean	Median	Mean	Median	Mean	Median
# employees	44.634	17	54.768	21	24.360	14
Average wage	477.289	332.346	554.572	393.527	322.681	276.318
FA_{2014}	0.275	0.1	0.229	0.03	0.368	0.222
FA_{2015}	0.390	0.2	0.329	0.1	0.512	0.545
GAP_{2014}	0.023	0.036	0.018	0.001	0.031	0.010
GAP_{2015}	0.062	0.015	0.051	0.007	0.085	0.054
Firm age	14.421	15	14.225	15	14.813	16
Profitability (profit/revenue)	0.021	0.018	0.017	0.016	0.028	0.024
Export share (export/revenue)	0.118	0	0.156	0.000	0.045	0.000
Labor share (labor cost/revenue)	0.136	0.102	0.136	0.103	0.140	0.102

Note: Firm-level descriptive statistics in January 2013. FA_{2014} and GAP_{2014} indicate the bite of the 2014 minimum wage hike, whereas FA_{2015} and GAP_{2015} the bite of the 2014 and 2015 minimum wage increases altogether.



(a) Tax-compliant firms



(b) Tax-evading firms

Fig. 3. Employment effect - short-run. Note: These figures show the effect of the share of affected workers FA on firm-level employment growth, controlling for firm's age, legal status, NACE sector, average export share and its square, average share of labor and its square, average profit and its square. The samples are restricted respectively to the sample of firms classified as labor tax compliant (panel (a)) and labor tax evaders (panel (b)). Each period represents a month. Period 0 indicates January 2013, the reference period. The vertical bar indicates January 2014, when the new minimum wage was introduced.

relationship between FA and the employment response. Alternatively, we provide a binscatter plot using the approach of Cattaneo et al. (2019) in Fig. C.3 in Appendix C.

These short run estimates imply an employment elasticity with respect to the minimum wage for directly affected workers amounting to −1.01 for compliant firms and −0.35 for evading firms (with respective bootstrapped standard errors of 0.17 an 0.14). These elasticities are calculated using the entire population of employees, not only categories of employees particularly affected by the minimum wage hike, such as teenage workers in the US literature. To make our results comparable with this literature, we need to scale our elasticities by 0.25, which is the share of directly impacted teenage workers in the US. This adjustment results in elasticities of −0.25 and −0.09, which correspond to the two bounds of the −0.3/−0.1 elasticity range reported by Neumark and Wascher (2010) and Brown (1999). The variation in employment effect based on tax compliance is thus substantial. When focusing on middle income and developing countries, Neumark and Corella (2021) suggest an employment elasticity of -0.102. Part of the large cross-country heterogeneity they document may be explained by the underestimation of the effect in countries where envelope wages are prevalent (e.g., Poland and Turkey). Harasztosi and Lindner (2019) obtain employment estimates close to 0 following a large minimum wage hike in Hungary in 2001, at a time when envelope wages were widespread (Tonin, 2011). Our results suggest that this estimate could be a lower bound.

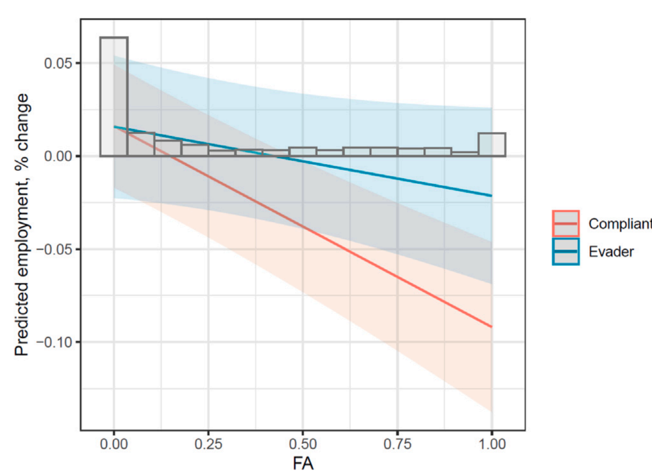


Fig. 4. Employment effect - Interaction. Note: This figure displays the estimated employment response to the 2014 minimum wage hike conditional on the share of affected workers FA and labor tax compliance based on Eq. (2) (calculated between January 2013 and December 2014). Confidence interval are represented at the 95% level. The histogram in the backdrop represents the distribution of FA.

4.2.2. Robustness checks

We perform a battery of robustness checks to evaluate the sensitivity of the baseline estimates. All the results are provided in Appendix C. The first series of tests is related to the measurement of FA and GAP. First, the number of employees and wages may, in some instances, substantially vary from one month to the next. Instead of using January 2013 as a reference period, we alternatively calculate the average of

Table 7
Short-run employment regressions.

	Dependent variable:					
	% change Jan 13 - Dec 14				% change Jan 13 - Jan 12	
	(1)	(2)	(3)	(4)	(5)	(6)
Evader	−0.002	−0.0004	−0.001	−0.001	−0.015*	−0.018**
	(0.013)	(0.014)	(0.012)	(0.013)	(0.009)	(0.008)
FA	−0.141***	−0.108***			0.018	
	(0.018)	(0.018)			(0.014)	
FA × Evader	0.086***	0.071***			−0.022	
	(0.028)	(0.027)			(0.018)	
GAP			−1.457***	−1.091***		0.098
			(0.189)	(0.190)		(0.127)
GAP × Evader			0.979***	0.838***		−0.117
			(0.273)	(0.268)		(0.175)
Controls	No	Yes	No	Yes	Yes	Yes
Observations	5,524	5,524	5,524	5,524	5,524	5,524
R^2	0.011	0.047	0.011	0.047	0.026	0.026

Note: ***p < 0.01, **p < 0.05, *p < 0.1 Robust standard errors in parentheses. The set of controls includes firm age, legal status, NACE sector, labor share and its square, profitability and its square, export share and its square.

these variables for the year 2012 (i.e., we calculate FA and GAP for each month of 2012 and then compute the average value). A year after the reform, point estimates for compliant and evading firms are respectively −9.7% and −2.6%, very close to those presented in Table 7. Second, in our main specification, the measurement of FA and GAP is based on the simple comparison of workers' wage in January 2013 and the 2014 minimum wage level. However, in the absence of the reform, some workers may have had a wage increase that would have pushed their wage above the new minimum wage level. In that case, FA and GAP would overstate the bite of the hike. To mitigate this concern, we construct alternative measures, assuming a wage increase between 2013 and 2014 equal to the national average wage increase over the 2010–2013 period (4.23%, Latvian Central Statistical Bureau). Using this alternative approach, the difference in employment response a year after the reform between compliant firms and evaders is actually *larger* than in the baseline approach: −11.6% for compliant firms, −2.3% for evading firms.

The second group of robustness checks is related to the compliant/evader classification. First, as explained in Section 3.1, the output of gradient boosting is a score in the $[0, 1]$ range, which is converted into a binary classification based on a chosen cutoff. We assess the sensitivity of the results by using alternative thresholds. Results are presented in Table C.3. In a second test, we use an alternative sample to train the algorithm. Instead of using Nordic-owned firms as examples of compliant firms, we use a broader set of Western-owned firms. The results are in line with those obtained using the baseline sample, as shown in Table C.4. Finally, besides a standard binary classification approach, we also implement another type of binary classification technique: One-Class Classification (OCC, see Fernández et al., 2018 for an introduction). OCC is an approach allowing for binary classification when only instances of one class are available for training. We train it using exclusively the set of evaders, then reproduce this exercise using only the set of compliant firms (see Appendix C for details). We then reproduce this exercise using only the set of (assumed) compliant firms. Training the model on evading firms, the FA estimate is −0.085 and the coefficient associated with the interaction FA × Evader is 0.062 (standard error: 0.035), which is quantitatively fairly close to the baseline estimates (although less precisely estimated). When using GAP instead of FA, the estimates are of similar magnitude as the baseline, but the interaction term is less precisely estimated (0.467, s.e.=0.309). When using the set of Nordic-owned firms to train the algorithm, the point estimate for FA is stable (−0.084). The interaction term is of smaller magnitude and imprecisely estimated (0.025, s.e. = 0.027). The same occurs when using GAP. Considering that one-class is a more challenging task than standard binary classification, the results using this alternative approach are overall weakly supporting our baseline estimates.

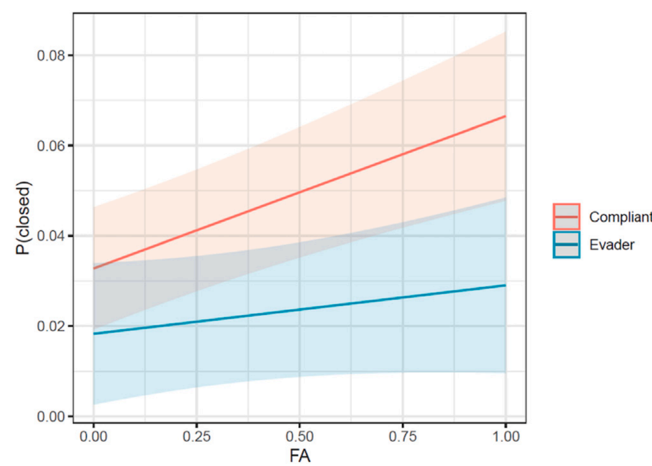


Fig. 5. Firm exit - Interaction. Note: This figure displays the estimated probability to shut down after the 2014 minimum wage hike conditional on the share of affected workers FA and labor tax compliance based on Eq. (2) (calculated between January 2013 and December 2014). Confidence interval are represented at the 95% level.

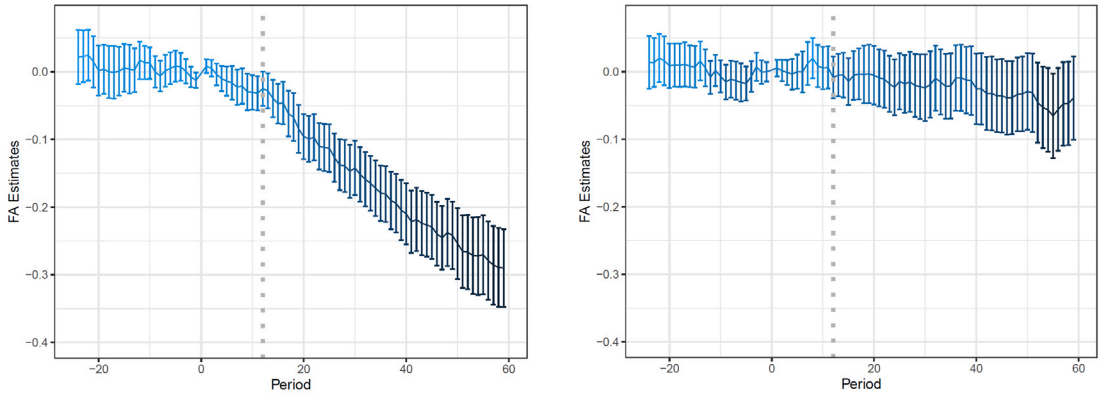
4.2.3. *Intensive vs extensive margin*

The point estimates capture both the extensive and intensive margins, as firms closing in the aftermath of the minimum wage hike are kept in the sample. To disentangle the relative importance of these two margins, we estimate the probability of a firm closing using a linear probability model and the same left-hand side variables as in Eq. (2). The results, provided in Table C.6 in Appendix C, indicate that compliant firms do adjust at the extensive margin: the probability of a severely affected compliant firm closing is much larger than for a compliant non-affected firm. This contrasts with tax-evading firms, for which the probability of closing remains about the same irrespective of the intensity of the shock.

Second, we estimate Eq. (2) keeping only firms surviving throughout the 2011–2017 period (hence excluding all the firms that shut down after the hike). The results, displayed in Table C.7 in Appendix C, show that compliant firms also adjust employment at the intensive margin.

4.2.4. *Medium-run employment effect*

We document a clear difference in employment response between tax-compliant and tax-evading firms in the short run. Does this difference between compliant and tax-evading firms persist over time? We now investigate the medium run effect, considering the two consecutive



(a) Tax-compliant firms **(b)** Tax-evading firms

Fig. 6. Employment effect - medium-run. Note: These figures show the effect of the share of affected workers FA on firm-level employment growth, considering altogether the 2014 and 2015 minimum wage hikes, controlling for firm's age, legal status, NACE sector, average export share and its square, average share of labor and its square, average profit and its square. The samples are restricted respectively to the sample of firms classified as labor tax compliant (panel (a)) and labor tax evaders (panel (b)). Each period represents a month. Period 0 indicates January 2013, the reference period. The vertical bar indicates January 2014, when the first minimum wage increase was implemented.

Table 8
Medium-run employment regressions.

	Dependent variable:							
% change Jan 13 -	Jan 12 (1)	Dec 15 (2)	Dec 16 (3)	Dec 17 (4)	Jan 12 (5)	Dec 15 (6)	Dec 16 (7)	Dec 17 (8)
Evader	−0.012	−0.038**	−0.045**	−0.092***	−0.016*	−0.022	−0.024	−0.066***
	(0.010)	(0.019)	(0.021)	(0.025)	(0.009)	(0.016)	(0.018)	(0.023)
FA_{2015}	0.016	−0.180***	−0.243***	−0.288***				
	(0.011)	(0.020)	(0.023)	(0.026)				
FA_{2015} × Evader	−0.022	0.150***	0.190***	0.237***				
	(0.016)	(0.031)	(0.034)	(0.039)				
GAP_{2015}					0.040	−0.851***	−1.154***	−1.374***
					(0.055)	(0.098)	(0.110)	(0.122)
GAP_{2015} × Evader					−0.069	0.662***	0.839***	1.053***
					(0.076)	(0.140)	(0.156)	(0.176)
Observations	5,524	5,524	5,524	5,524	5,524	5,524	5,524	5,524
R^2	0.026	0.062	0.085	0.077	0.026	0.060	0.083	0.075

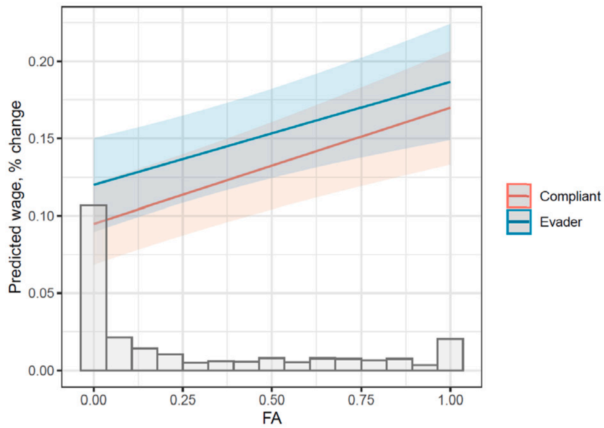
Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Robust standard errors in parentheses. FA_{2015} and GAP_{2015} measure the intensity of the 2014 and 2015 minimum wage hikes altogether with respect to firm-level employment in January 2013. All regressions include the following controls: firm age, legal status, NACE code, labor share and its square, profitability and its square, export share and its square.

minimum wage hikes altogether. First, we proceed as above, and study tax-evading and compliant firms separately, regressing the percentage change in employment on a measure of treatment intensity and the set of controls. The results are displayed in Fig. 6, using FA_{2015}, the share of full-time employees paid below the 2015 minimum wage in January 2013, as the measure of treatment intensity. The employment change for affected compliant firms plummeted, whereas even in the medium run, tax-evading firms remain resilient to the hit. In December 2017, three years after the January 2015 minimum wage hike, there is still no significant employment effect for tax-evading firms (albeit negative point estimates), indicating that they absorbed the shock differently from compliant firms. Table 8 provides regression results using the whole sample, when comparing the percentage change in employment between January 2013 and December 2015, 2016 and 2017 (the last month of our sample). Between January 2013 and December 2017, employment elasticity with respect to the minimum wage is equal to −1.1 for compliant firms and −0.5 for tax-evading ones (bootstrapped standard errors: 0.09 and 0.07, respectively), thus remaining largely different. The results using GAP_{2015} depict a similar picture.

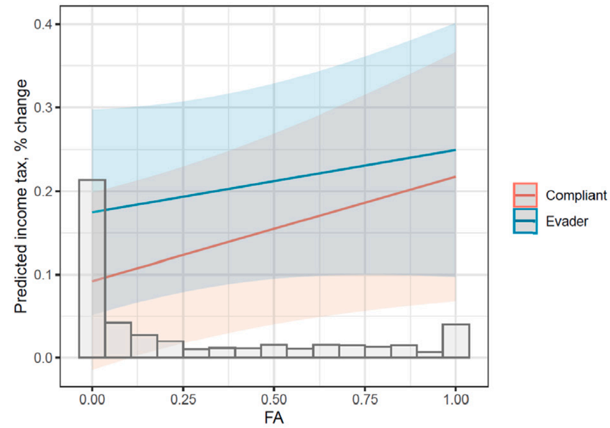
4.2.5. Other outcomes: wage, personal income tax and social security contributions

The absence of employment reaction for tax-evading firms could also be explained by non-compliance with the new minimum wage, as documented in Basu et al. (2010). For example, this could take the form of full-time employees artificially switching to part-time. To investigate this alternative channel, we estimate the effect of the minimum wage hike on average gross wages. Since wages are only observable for surviving firms, we restrict the sample accordingly. We examine changes in average gross wage between January 2013 and four periods: June 2012 (one year before announcement of the hike, placebo check), June 2013 (announcement of the hike), January 2014 (first month after the first hike) and January 2015 (first month after the second hike). In addition, we study tax-evading and compliant firms separately, regressing the percentage change in wage on FA and the set of controls, and provide the point estimates associated with FA in Fig. C.4(b) in Appendix C.

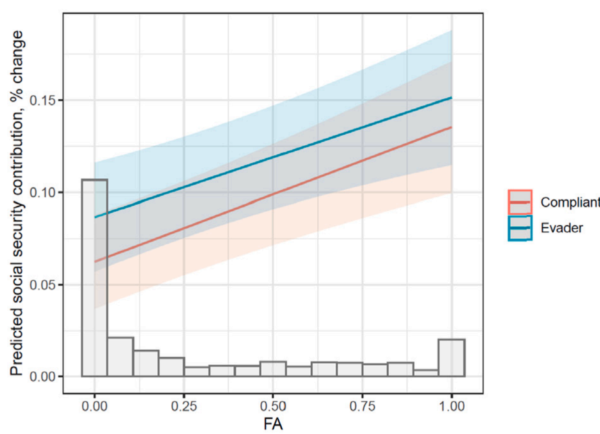
The first panel of Table 9 displays the results, alternatively using FA_{2015} and GAP_{2015}. We do not observe much difference in employment growth in the pre-hike period (at least in the two-year period preceding the minimum wage increase, see Figs. C.4(b) and C.4(a) in Appendix C). The average wage does not change in the direct aftermath of the minimum wage announcement. However, average wages sharply increase directly after implementation of the new minimum wage levels, both in 2014 and in 2015. The change in average between January 2013 and January 2014 conditional on minimum wage bite and tax evasion is plotted in Fig. C.4. The average wage increases in a parallel way in the two groups, suggesting that tax-evading firms did not avoid the increase in labor costs differently from compliant firms. The conclusion is similar when focusing on the 2015 hike.



(a) Effect on wage



(b) Effect on income tax



(c) Effect on social contribution

Fig. 7. Effect on other outcomes. Note: These figures display respectively the wage response (panel (a)), personal income tax response (panel (b)) and social security contribution response (panel (c)) conditional on the share of affected workers FA and labor tax compliance (calculated between January 2013 and January 2014, when the minimum wage increases was implemented. Confidence interval are represented at the 95% level. Histograms in the backdrop represent the distribution of FA.

To further study this point, we proceed to the same analysis using firm-level average income tax and average social contributions. The results are displayed respectively in panel 2 and panel 3 of Table 9. As above, we study tax-evading and compliant firms separately, regressing the percentage change in wage on FA and the set of controls, and provide the point estimates associated with FA in Figs. C.5(b) and C.5(a) in Appendix C. Both exhibit a sharp increase immediately after the minimum wage hike. Firms' reactions were similar irrespective of their type, as shown in Fig. 7. Therefore, increasing the minimum wage did have a negative employment effect, but also increased income tax revenue and social contributions. This rules out the explanation that tax-evading firms reduce their use of labor at the intensive margin differently from tax-compliant firms.

5. Conclusion

This paper provides a firm-level analysis of a minimum wage hike in the context of prevalent labor tax evasion. We provide two main contributions to the literature. First, we propose a novel methodology for classifying firms as either compliant or labor tax evaders using machine learning. This methodology relies on strong assumptions, but several validity checks confirm the relevance of the approach. This classification method can be used in other countries where envelope wages are a widespread phenomenon provided that two necessary (but not sufficient) conditions are met: (i) access to firms' balance sheets for the population of firms and employee-level information for a subsample of workers; (ii) a context where a subsample of firms can be considered as compliant. The trend towards more widespread access to administrative data will make condition (i) easier to satisfy in the future. An alternative path could be to use audit data to construct a sample of compliant and non-compliant firms. However, audits are rarely random, and this non-randomness would imply that some types of evading firms are not captured. Our approach is not immune to this criticism either, as what we capture likely emphasizes labor tax evasion at the bottom of the wage distribution. Note also that our classification does not aim to serve as an evasion detection device, but rather as a proxy for labor tax evasion for the purposes of our study.

Equipped with this classification tool, we study the conditional employment effect of a large minimum wage hike. Compared to the existing literature, the magnitude and the bite of this episode are both in the upper bound. We find that employment in firms paying envelope wages is much less sensitive than in compliant firms. This finding is consistent with the model proposed by Tonin (2011), in which firms that pay envelope wages can accommodate a minimum wage increase, at least in part, by reducing the degree of their underreporting rather than by increasing their workers' net pay. Raising the minimum wage thus contributes to enforcement of the tax policy and helps to increase

Table 9
Wage regressions.

	Dependent variable:							
% change Jan 13 -	Jun 12 (1)	Jun 13 (2)	Jan 14 (3)	Jan 15 (4)	Jun 12 (5)	Jun 13 (6)	Jan 14 (7)	Jan 15 (8)
Wage								
Evader	0.014	0.020*	0.025**	0.052**	0.012	0.019**	0.027**	0.057***
	(0.009)	(0.010)	(0.012)	(0.021)	(0.008)	(0.009)	(0.011)	(0.018)
FA	0.002	−0.019	0.075***	0.153***				
	(0.009)	(0.012)	(0.015)	(0.021)				
$FA\times$ Evader	−0.021	−0.024	−0.008	−0.036				
	(0.016)	(0.019)	(0.023)	(0.033)				
GAP					0.064	−0.162	0.761***	0.761***
					(0.091)	(0.140)	(0.166)	(0.114)
$GAP\times$ Evader					−0.201	−0.276	−0.147	−0.254
					(0.150)	(0.185)	(0.222)	(0.156)
Income tax								
Evader	0.068**	0.102***	0.082**	0.186***	0.057**	0.097***	0.082***	0.161***
	(0.027)	(0.037)	(0.033)	(0.060)	(0.024)	(0.032)	(0.029)	(0.047)
FA	0.046	0.040	0.125***	0.296***				
	(0.036)	(0.043)	(0.043)	(0.049)				
$FA\times$ Evasion	−0.074	−0.112	−0.050	−0.174*				
	(0.048)	(0.104)	(0.062)	(0.091)				
GAP					0.610*	0.640	1.431***	1.543***
					(0.321)	(0.418)	(0.410)	(0.253)
$GAP\times$ Evasion					−0.585	−1.256	−0.613	−0.747**
					(0.431)	(1.034)	(0.583)	(0.404)
Social contribution								
Evader	0.014	0.019*	0.024**	0.510**	0.012	0.018**	0.026**	0.056***
	(0.009)	(0.010)	(0.012)	(0.020)	(0.008)	(0.009)	(0.011)	(0.017)
FA	0.002	−0.020*	0.073***	0.150***				
	(0.009)	(0.012)	(0.015)	(0.021)				
$FA\times$ Evasion	−0.021	−0.024	−0.008	−0.036				
	(0.016)	(0.019)	(0.022)	(0.032)				
GAP					0.065	−0.168	0.739***	0.744***
					(0.092)	(0.140)	(0.162)	(0.110)
$GAP\times$ Evasion					−0.203	−0.274	−0.142	−0.250*
					(0.150)	(0.184)	(0.216)	(0.152)
Observations	5,391	5,391	5,391	5,391	5,391	5,391	5,391	5,391

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Robust standard errors in parentheses. In column 4) and 8). FA and GAP measure the intensity of the 2014 and 2015 minimum wage hikes altogether with respect to firm-level employment in January 2013. All regressions include the following controls: firm age, legal status, NACE code, labor share and its square, profitability and its square, export share and its square.

social protection coverage of workers, in addition to increasing the average wage.

However, policy makers do face a trade-off, as a minimum wage hike can have a negative effect on compliant firms employing minimum wage workers. To quantify this trade-off between job loss in compliant firms and wage and fiscal gain from the minimum wage hike (including from reduced evasion in evading firms) at the aggregate level, we perform a simple back-of-the-envelope analysis. Using Eq. (2), we estimate what the number of employees in each firm would have been in the absence of the reform (i.e., setting $FA = 0$ for all firms), aggregate these firm-level figures and finally compare the total with the *actual* total employment in December 2014 (see Appendix C for details). We repeat this exercise for wages, and derive from it an estimate of the fiscal consequences of the minimum wage hike. This raw approach suggests that a year after the hike, the total number of employees is 1.5% lower than what it would have been in the absence of the minimum wage increase. On the other hand, the actual aggregate wage bill is 0.5% larger than its counterfactual level, implying a modest positive fiscal effect (considering exclusively personal income tax and social security contributions). As a comparison, Meer and West (2016) estimate that a permanent 10% minimum wage increase leads to a circa 1.5%–2% decline in total employment in US states over the medium term. The employment response is more swift in the Latvian context, as a similar magnitude is reached within a year despite evading firms absorbing the shock through the envelope margin (thus mitigating the overall employment response). A potential explanation is the magnitude of the bite, which is (much) larger in Latvia.

We only document short- and medium- run partial equilibrium effects. One of the negative effects of labor tax evasion is competition distortion. If raising the minimum wage increases compliance, in the long run all compliant firms will benefit from a reduction in these distortions. Similarly, our results are obtained focusing on firms that existed before the reform, and we do not address the question of firm entry. Also, we do not study the reallocation effect of the minimum wage hike. Dustmann et al. (2022) provide evidence of such reallocation effects studying the introduction of a minimum wage in Germany. Finally, even if increasing the minimum wage leads to an increase in workers' social protection and average *reported* wage, Tonin (2011) provides evidence that affected workers can actually see a *decrease* in their total disposable income: the decrease in underreporting leads to a larger share of their wages being taxed, while the tax incidence falls on them. In future work, all these additional effects should be integrated to a unified framework to estimate the meaningful welfare consequences of a minimum wage hike.

We have no relevant or material financial interests that relate to the research described in the paper.

Data availability

The paper uses a combination of administrative and survey data from Latvia, owned either by the Latvian State Revenue Service (SRS) or the Latvian Central Statistical Bureau (CSB). We access this data via a secure server provided by CSB. We therefore cannot share the data ourselves. Nevertheless, the access to the various datasets we use can be obtained via a request to CSB. See the readme file accompanying the code for a precise list.

Declaration of competing interest

We have no relevant or material financial interests that relate to the research described in the paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

For valuable and constructive comments, we would like to thank the editor Jeffrey Clemens, three anonymous referees, Konstantins Benkovskis, Anikó Biró, Zainab Iftikhar, Vitalijs Jascisens, Marc Sangnier, Konstantin Sonin, Giancarlo Spagnolo, Ija Trapeznikova, and seminar participants at ULB Bruxelles, IESEG Paris, Stockholm Institute of Transition Economics, Bank of Latvia, Bank of Ukraine, Kyiv School of Economics, UC3 de Madrid, IZA workshop on labor market institutions, IZA workshop on transition economics, ZEW Public Finance conference, Trier Workshop on Labour Economics, at Shadow and LAGV conferences, and at the annual meetings of SOLE, EALE, EEA, FREE Network, and Baltic Economic Association. We also would like to thank Uldis Ainars (Latvian Central Statistical Bureau) for his help with the data. Financial support by the EEA and Norway grant project "Micro-level responses to socio-economic challenges in face of global uncertainties" (Grant No. S-BMT-21-8 (LT08-2-LMT-K-01-073)) is gratefully acknowledged.

Appendix A. Classification details

This appendix provides additional details on the classification procedure. We begin with a description of the procedure implemented to construct the sample used for training and testing the algorithm. In a second stage, we detail our machine learning procedure.

A.1. Creation of the training sample

This sample consists of two parts: a subset of firms considered as labor tax-compliant, and a subset of firms considered as tax-evaders.

Tax-compliant firms - First, for the subsample of tax-compliant firms, we use the set of firms owned by a Nordic company (Denmark, Finland, Norway, Sweden) during the 2011–2013 period. Information regarding foreign ownership is obtained from the Latvian Central Statistical Bureau (CSB). In 2013, foreign-owned firms represented 13.1% of the entire population of firms, and Nordic-owned companies 3.5%, equivalent to over 25% of all foreign-owned firms. We keep a firm/year observation only if the firm is Nordic-owned during that specific year (as this information is only available at the yearly level). Throughout the analysis, we restrict the sample to firms with a minimum of 6 employees at any point in time between 2011 and 2013, operating in the manufacturing, trade/retail, construction or transport sector.

Tax-evading firms - Second, to obtain a set of tax-evading firms, we proceed as follows. We combine three waves of Labor Force Survey data (2011–2013), which cover the pre-reform period. We extract the reference month of each interview from LFS data and match it with *administrative* employer–employee data for that specific month. We then filter the sample to include only full-time employees working in any of the four sectors of interest, keeping only one observation per individual (in case of multiple surveys). After excluding observations with missing values, we are left with 5,319 individuals. We then estimate a standard wage equation where the dependent variable is the natural logarithm of the *administrative* wages (i.e., wages reported to the tax authorities). The set of regressors is composed of:

- Individual characteristics: age, age^2, experience, experience^2 (all expressed in years) and a dummy taking the value 1 if the respondent is a woman.

Table A.1
Wage regression.

	Dependent variable: log(wage)
Age	0.014***
	(0.004)
Age^2	−0.0002***
	(0.00005)
Experience	0.019***
	(0.002)
Experience ^{2}	−0.0002***
	(0.0001)
Woman	−0.184***
	(0.019)
Construction	−0.040
	(0.027)
Trade	−0.128***
	(0.026)
Transportation	0.070***
	(0.024)
Pieriga	−0.087***
	(0.025)
Vidzeme	−0.140***
	(0.027)
Kurzeme	−0.060***
	(0.023)
Zemgale	−0.068***
	(0.023)
Latgale	−0.203***
	(0.021)
Moderate density	0.024
	(0.023)
Sparse density	0.004
	(0.018)
Year 2012	0.034*
	(0.018)
Year 2013	0.058***
	(0.018)
Observations	5,297
R^2	0.247

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$: Robust standard errors in parentheses. The dependent variable is the *administrative* wage of the respondent. The regression also includes occupation and education dummies.

- A set of dummy variables indicating the sector in which the respondent works (reference group: Manufacturing).
- A set of dummy variables indicating the region in which the respondent works (reference group: capital city Riga), and a set of dummies indicating the level of urbanisation at the respondent's residence (reference group: densely populated area).
- A set of dummies indicating the year of the reference period (reference group: 2011).
- A series of categorical variables indicating the level of education attained (12 categories) and the type of occupation (41 categories).

All the regressors come from LFS data, except age, which we gather from administrative data. Table A.1 displays the results. The signs and magnitudes of personal characteristics align with typical wage regression findings. The wage penalty for women appears substantial, but is in line with papers studying the gender wage gap in the Baltics (see for instance Christofides et al., 2013 and Meriküll and Tverdostup, 2023). Regional dummies also show the expected pattern, with the highest wages in Riga and the lowest in the Eastern region of Latgale. The R^2 of 0.25 resembles that of similar studies estimating wage equations (e.g., 0.27 in Harmon and Walker, 1995).

We keep observations in the bottom 10% of the residual distribution. From these 540 observations, almost two thirds (64.1%) earn no more than 110% of the minimum wage. We then retrieve firms employing these "suspiciously low paid" workers and classify them

Table A.2
Sample statistics.

	N	Pctl(25)	Median	Pctl(75)
Tax evading firms				
# employees	261	19	37	95
Turnover	261	490,122	1,182,634	3,308,750
Profit	261	41,918	141,281	459,988
Manufacturing	90			
Trade	65			
Transportation	68			
Construction	38			
Compliant firms				
# employees	567	15	34	76
Turnover	567	1,169,667	3,070,452	7,751,814.0
Profit	567	232,973	586,216	1,430,006.0
Manufacturing	260			
Trade	237			
Transportation	51			
Construction	19			

Note: This table displays descriptive statistics of the sample used for training, for tax evading and compliant firms separately.

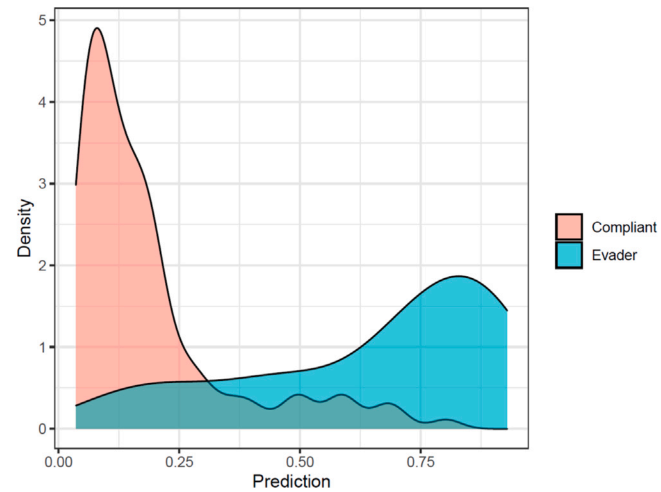


Fig. A.1. Classification score - Densities. Note: This Figure displays the density of the scores obtained for each observation in the test set, for tax-evading and compliant firms separately.

as tax-evaders. Firms with at least one employee in the bottom of the residual distribution in a given year is classified as evader that year. We keep all such firm/year positive cases. For instance, if an LFS respondent in 2012 is "suspiciously" low paid, we identify the firm employing this worker in 2012 and keep this 2012 firm/year observation. Anecdotally, we find (and drop) only two Nordic-owned firms with an employee in the bottom of the residual distribution.

Combining the two subsets of compliant firms and evaders, merging it with firms' financial data (yearly level) and after dropping firms with missing values, we are left with a sample containing 828 firm/year observations (567 compliant and 261 evaders). While not particularly large, this sample size aligns with standard practices in the financial fraud detection literature (see Hajek and Henriques, 2017 for a survey covering sample sizes). Some descriptive statistics are provided in Table A.2. The four sectors are fairly represented across the two subsamples. We initially considered additional sectors (e.g., food and accommodation services), but the number of foreign-owned firms operating in these sectors satisfying our size threshold was very small.

The sample contains all the information that we use as predictors, based on the fraud detection literature (Beneish, 1999; Kotsiantis et al., 2006; Kirkos et al., 2007; Cecchini et al., 2010; Ravisankar et al., 2011; West and Bhattacharya, 2016; Hajek and Henriques, 2017). The complete list is displayed in Table A.3.

A.2. Machine learning procedure - additional details

We randomly divide the sample into two parts: 80% of the observations are assigned to the training set, the remaining 20% to the test set. We apply gradient boosting using the R package XGBoost implementation, via the Tidymodels interface. All numeric inputs (see the list in the previous subsection) are centred and rescaled, and inputs that have large absolute correlations with other variables are automatically dropped (> 0.9). We set the number of trees contained in the ensemble to 100, and tune the model using four hyperparameters (using a parameter grid composed of 100 points): the maximum tree depth, the minimum number of data points in a node that is required for the node to be split further, the learning rate ρ, and the reduction in the loss function required to split further. We use 10-fold cross-validation to find the optimal set of hyperparameters, aiming at maximizing Precision-Recall AUC. Precision is the ratio of true positive to the total number of predicted positive, whereas recall is the ratio of true positives to the total number of actual positives (true positives + false negatives). There is usually a trade-off between precision and recall: increasing the model's threshold for classifying instances as positive increases precision, but is likely to decrease recall. The Precision-Recall curve represents this trade-off by plotting precision against recall for any classification threshold value. The overall performance of the classifier can be assessed in a single metric by computing Precision-Recall AUC, which is the area under this curve (AUC standing for "area under the curve"). We select this metric for two main reasons. First, Precision-Recall AUC focuses on predicting the positive class, encouraging the model to be cautious when making positive predictions and thus reducing the chances of a false positive. Second, this metric is particularly appropriate in the context of class imbalance (Saito and Rehmsmeier, 2015).[28]

After tuning the model on the training set, we apply it to the test set to assess out-of-sample performance. First, if the model performs correctly, evading firms should have on average a higher probability to be classified as such than compliant firms. Fig. A.1 displays the density of the scores obtained for firms in the test set. With a perfect classifier, the two groups should not overlap, whereas this overlap would be total with a no-skill classifier (i.e., a classifier completely unable to distinguish between the two classes). The vast majority of compliant firms obtain a very low score. Conversely, the mode for tax-evading firms is above 0.75. Second, Fig. A.2 displays the Precision-Recall curve discussed above. Good classifiers should come as close as possible to the top right corner: at this point, precision and recall are both equal to 1, so the model perfectly distinguishes the two classes. The dotted line represents the sample average of evading firms, which is equivalent to the Precision-Recall curve of a naïve classifier assigning the value 1 to all observations.

Finally, a disadvantage of gradient boosting is that it does not allow to study the functional form linking inputs to outputs. A way to evaluate the importance of each variable for the classification is the permutation procedure (Greenwell et al., 2020). The idea is the following: after training the model and computing the performance metrics of interest, the values of an explanatory variable are randomly permuted. The performance metrics are recomputed after this alteration and compared to the original. The larger the degradation of the performance metrics, the larger the importance of the variable for the classification. The procedure is then repeated for each explanatory variable. Fig. A.4 displays the 10 variables contributing the most to the classification outcome. Administrative cost is the variable exerting

[28] Alternatively, fine-tuning the model with respect to ROC AUC metrics provides similar performance on the test set.

Table A.3
List of variables.

Variable	Definition	Used in
Turnover	–	Hajek and Henriques (2017)
Profit	–	Ravisankar et al. (2011)
Long-term investment	–	Cecchini et al. (2010), Dechow et al. (2011)
Fixed asset	–	Beneish (1999), Cecchini et al. (2010), Dechow et al. (2011)
Current asset	–	Beneish (1999), Cecchini et al. (2010), Dechow et al. (2011)
Receivables	–	Beneish (1999), Cecchini et al. (2010), Dechow et al. (2011), Ravisankar et al. (2011)
Cash	–	Beneish (1999), Cecchini et al. (2010), Dechow et al. (2011), Ravisankar et al. (2011)
Liabilities	–	Beneish (1999), Cecchini et al. (2010), Dechow et al. (2011)
Administrative cost	–	Beneish (1999), Cecchini et al. (2010)
Selling cost	–	Beneish (1999), Cecchini et al. (2010)
Other cost	–	Beneish (1999), Cecchini et al. (2010)
Other revenue	–	Beneish (1999), Cecchini et al. (2010)
Equity	–	Cecchini et al. (2010), Dechow et al. (2011)
EBIT	–	Hajek and Henriques (2017)
Margin	profit/turnover	Hajek and Henriques (2017), Ravisankar et al. (2011)
Return on equity	profit/equity	Hajek and Henriques (2017)
Total asset to revenue	current asset/turnover	Hajek and Henriques (2017)
Cash to revenue	cash/turnover	Hajek and Henriques (2017)
Non-cash working capital to revenue	(inventories+receivables- short term liabilities)/turnover)	Hajek and Henriques (2017)
Fixed asset to total asset	fixed asset/current asset	Hajek and Henriques (2017)
Selling and administrative cost to revenue	(selling cost + administrative cost + other cost)/turnover)	Hajek and Henriques (2017)
Non-cash working capital	(selling cost + admin cost + other cost)	Hajek and Henriques (2017)
Liquidity	cash/current asset	Hajek and Henriques (2017)
Debt to total asset	(long term liabilities + short term liabilities)/current asset)	Hajek and Henriques (2017)
Cash growth	% change in cash between year t and $t-1$	Kotsiantis et al. (2006)
Revenue growth	% change in turnover between year t and $t-1$	Kotsiantis et al. (2006)
NACE sector	–	–

Note: This table lists the variables used as inputs for the classification algorithm, their definitions and previous work using them.

the largest influence on the classification. Administrative expenses are on average lower for evading firms than for compliant firms, both in absolute value and in expenses per employee. Administrative costs largely consist of salaries of administrative staff, in contrast to other cost categories, such as selling costs, which may also include items such as storage and transportation. Hence, this category is the most sensitive to wage underreporting.

Appendix B. The expenditure-based method

This appendix provides a description of the expenditure-based method originally proposed by Pissarides and Weber (1989). This approach allows us to estimate the extent of underreporting for a group of households using reported income and data on expenditures. It relies on two main assumptions: (i) food expenditures (or any other item) are accurately reported for all groups; (ii) income reporting is accurate for at least one group in the population. Pissarides and Weber (1989) consider two population groups: employees and the self-employed, respectively denoted $k = W$ and $k = S$. Employees are assumed to correctly report their wage. Food expenditures and true permanent income are related by the Engel curve:

$$ln\ c_i = \alpha + \beta ln\ y_i^P + X_i'\phi + \epsilon_i \tag{B.1}$$

where c_i denotes food expenditure in household i, y_i^P is the permanent income, X_i is a set of household characteristics (number of children, etc.) and ϵ_i is an error term. Consumption is based on permanent

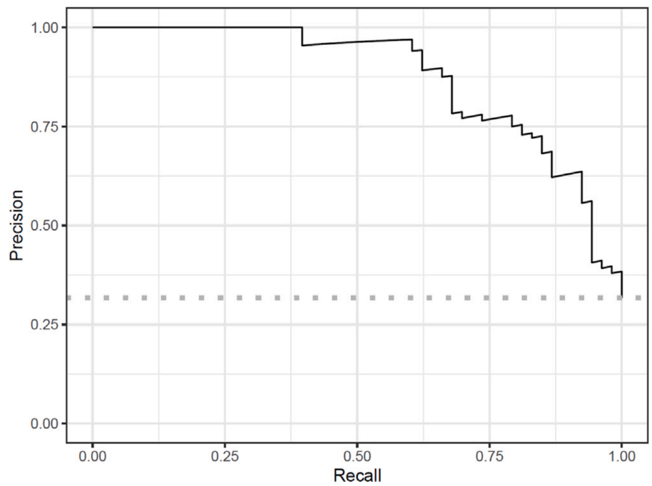


Fig. A.2. Precision-Recall curve. Note: This figure displays the Precision-Recall curve based on the test set.

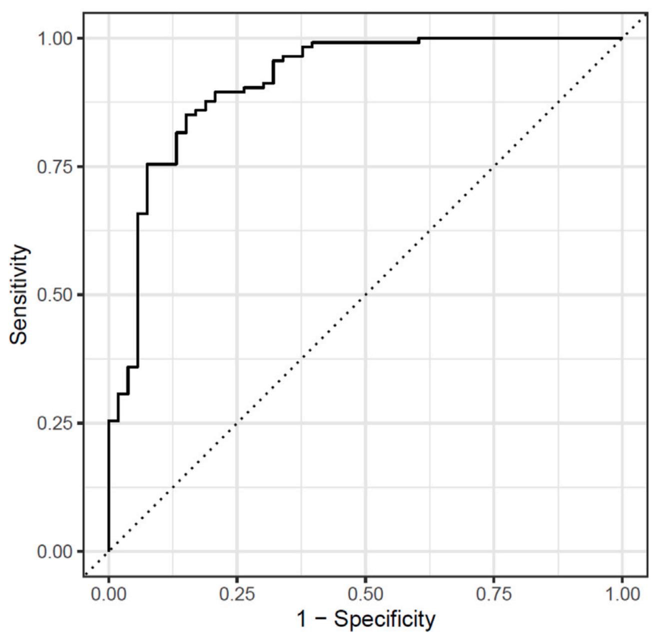


Fig. A.3. ROC curve. Note: This figure displays the ROC curve based on the test set.

income, but surveys usually enquire about current income. Denoting p_i the fraction of true current income to true permanent income, we can write:

$$ln\ y_i = ln\ p_i + ln\ y_i^P. \tag{B.2}$$

where p_i is assumed to follow a log-normal distribution so that $ln\ p_i = \mu_p + u_i$, where μ_p is the sample mean of $ln\ p_i$ and u_i is a disturbance with $E[u_i] = 0$ and $\sigma_u^2 = Var(u_i)$.

Self-employed respondents have incentives not to correctly report their income in the survey (e.g., worries that survey answers are shared with tax authorities). The relationship between reported current income y_i and true current income y_i^* is

$$ln\ y_i = ln\ \kappa_i + ln\ y_i^*. \tag{B.3}$$

κ_i is the underreporting factor, which is assumed to be log-normal so that $ln\ \kappa_i = \mu_\kappa + \upsilon_i$ with $E[\upsilon_i] = 0$ and $\sigma_\upsilon^2 = Var(\upsilon_i)$. All employee households are assumed to have $\kappa = 1$: they do not underreport their

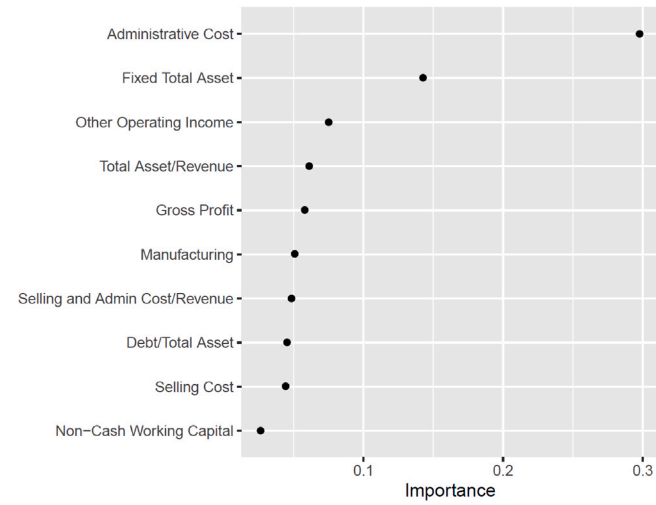


Fig. A.4. Variable Importance Plot. Note: This Figure displays a variable importance plot, using the permutation method. It represents the 10 variables contributing the most to the classification outcome.

income. This implies that $\sigma_{\upsilon|W}^2 = 0$ for employees whereas $\sigma_{\upsilon|S}^2 > 0$ for self-employed.

Combining Eqs. (B.2) and (B.3), the true permanent income can be written as:

$$ln\ y_i^P = ln\ y_i^* + ln\ \kappa_i - ln\ p_i = ln\ y_i^* + (\mu_\kappa + \upsilon_i) - (\mu_p + u + i) \tag{B.4}$$

Inserting Eq. (B.4) in Eq. (B.1), the Engel curve is expressed as:

$$ln\ c_i = \alpha + \beta ln\ y_i^* + \beta(\mu_\kappa - \mu_p) + X_i'\phi + \beta(\upsilon_i - u_i) + \epsilon_i \tag{B.5}$$

This equation corresponds to the empirical specification described in Eq. (1) in Section 3.2, where $\beta(\upsilon_i - u_i)$ is substituted with γD_i, so that $\gamma = \beta(\mu_\kappa + \frac{1}{2}(\sigma_{u|S}^2 - \sigma_{u|W}^2))$, and the three error terms are encompassed in ξ_i. The coefficient γ captures the systematic food expenditure shift of self-employed with respect to employees.

The mean underreporting factor $\bar{\kappa}$ can thus be computed as follows:

$$\bar{\kappa} = exp\left(\frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{\upsilon|S}^2 + \sigma_{u|W}^2 - \sigma_{u|S}^2)\right) \tag{B.6}$$

The variance terms usually being unknown, Pissarides and Weber (1989) show that we can obtain lower and upper bound estimates for $\bar{\kappa}$ such that:

$$\bar{\kappa} = \left[exp\left(\frac{\gamma}{\beta} - \frac{1}{2}(\sigma_{\xi|S}^2 - \sigma_{\xi|W}^2)\right), exp\left(\frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{\xi|S}^2 - \sigma_{\xi|W}^2)\right)\right] \tag{B.7}$$

Assuming that the variance term is equal in the two groups, as for instance in Hurst et al. (2014), then the underreporting estimate simplifies to $\bar{\kappa} = exp\left(\frac{\gamma}{\beta}\right)$. In the framework of this paper, we consider four groups of households: where the head works in the public sector, in a compliant firm, in a tax-evading firm or whether the head is self-employed. The public sector is the base group, where envelope wages presumably do not exist. This assumption is consistent with the Latvian context, where several waves of public sector reforms were implemented in the 2000's. Public sector employees are also used as the reference group in a study in Estonia (Paulus, 2015). Note also that the underreporting estimates are *relative*. In other words, this approach estimates the *additional* underreporting of a group of households compared to another one. It provides an absolute measure only if the reference group does indeed correctly report their income.

Turning to the estimation of Eq. (1), note that reported income is endogenous by construction. To overcome this issue, we instrument current income by education of the household head (three levels) and region, since the average wage differs across regions in Latvia. The

Table B.1
HBS - descriptive statistics.

Statistic	N	Mean	St. Dev.	Min	Max
Ln food consumption	444	7.939	0.456	6.470	9.287
Ln income	444	9.157	0.756	4.981	11.579
Public sector	224				
Compliant firm	143				
Tax-evading firm	57				
Self-Employed	20				

Note: This table displays descriptive statistics for the main variables used in the expenditure-based validation check. Public sector, compliant firms, tax-evading firms and self-employed indicate the type of employment of the household head.

underlying assumption is that these variables do not affect food expenditure other than through wages. As for the set of controls included in both stages, this contains the number of children, the number of working adults, the head's age and gender, and a dummy indicating whether the household rents its dwelling. As is standard in the literature, we keep only households composed of two adults and children (if any). Descriptive statistics are displayed in Table B.1. From the regression results displayed in Table 5, we can thus compute that for employees of tax-evading firms, $\bar{\kappa} = exp\left(\frac{0.168}{0.401}\right) \approx 1.52$. The share of underreported income is thus $\frac{\bar{\kappa}-1}{\bar{\kappa}} \approx 0.34$, as reported in section 3.2.

Appendix C. Minimum wage effect - additional results and details

This section provides additional details regarding estimation of the minimum wage effect. We first detail the construction of our two variables measuring the bite of the minimum wage increase, FA and GAP.

C.1. Construction of FA and GAP measures

First, as mentioned in Section 4, the two measures FA and GAP are computed using firm-level full-time employees. The motivation for this choice is that before 2014 reporting the number of hours was not mandatory. In addition, the reported number of hours takes only effective hours into account. Employees taking holidays will exhibit lower working hours than usual in that month. This prevents us from computing hourly wage, and computing the magnitude of the bite using all employees.

To disentangle part-time from full-time employees, we start by computing the mode of working hours distribution for each month. This provides us with the legal number of hours for a full-time equivalent job for each month. This number changes every month depending on the number of working days. We then consider an employee in a given firm as full-time for a given month if: (i) the reported number of hours is equal to or greater than 90% of the monthly mode or if (ii) the number of hours is missing but the wage is greater than 90% of the monthly minimum wage. These assumptions are likely to overestimate the number of full-time employees. To compute FA and GAP, in order to mitigate this effect we restrict full-time employees as workers satisfying one of these conditions for at least half of the months spent in a given firm within a year. For instance, an employee working 6 months in a given firm in 2012 earning 4 months more than the minimum wage but two months less than the minimum wage will be considered as a full-time employee. The support for FA conditional on firm type is displayed in Fig. C.1. For both types of firm, the support ranges from 0 to 1.

Fig. C.2 reproduces Fig. 4 using GAP instead of FA. It displays the interaction of GAP and the tax-evasion dummy from column 4) of Table 7.

Fig. C.3 displays the unconditional relationship between FA and the percentage change in employment between January 2013 and December 2014, for compliant and tax-evading firms separately. The number of bins is computed using the data-driven approach of Cattaneo et al. (2019). This figure shows the difference in employment change at different percentiles of the FA distribution for both types of firms, adjusting for controls. Linearity does not appear to be a too strong assumption: results assuming linearity provided in Fig. 4 are fairly similar, and point estimates for highly impacted firms are very close to the baseline.

C.2. Robustness checks - Alternative FA and GAP measurement

Table C.1 displays the results obtained when FA and GAP are not based on the information in January 2013, but rather on the average number of employees over 2012. The results are in line with those displayed in Table 7 in Section 4.

Table C.2 displays the results obtained when FA and GAP are computed assuming that 2013 wages would have increased at the rate of the 2011–2013 average national wage increase (4.23%, Latvian Central Statistical Bureau). Using this alternative approach, the difference in employment response between compliant firms and evaders is actually $larger$ than in the baseline approach, non-compliant firms showing a point estimate even closer to 0 ($-0.116 + 0.093 = -0.023$).

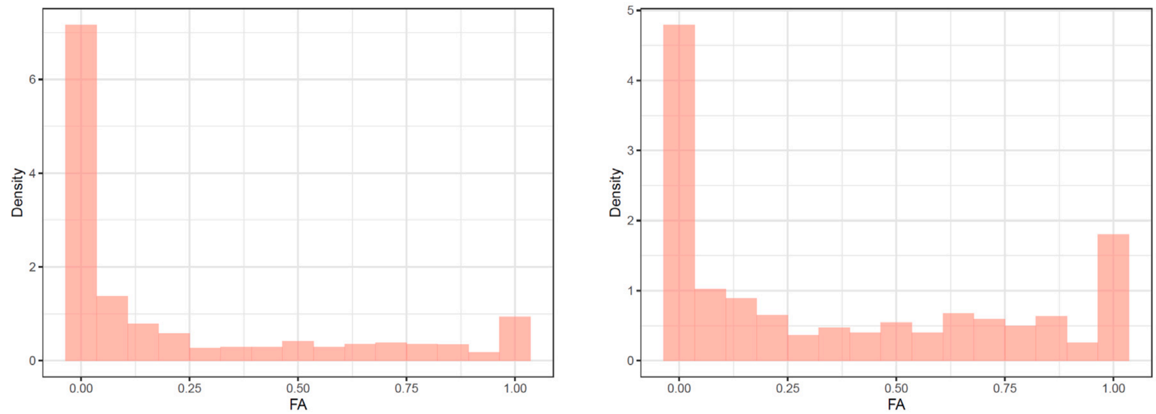
C.3. Robustness checks - Alternative classification

Table C.3 displays the results obtained using alternative thresholds for class classification. Gradient boosting assigns a score between 0 and 1 to each observation, which is then used to classify firms as tax-evaders or compliant. A higher threshold implies a lower share of firms classified as tax-evaders. The columns are arranged in increasing order of cutoff stringency from left to right. The baseline results, achieved using the threshold that maximizes the F_1 performance measure (cutoff = 0.684, see Appendix A.2 for details), are reported in Column (3). Column (1) displays the results when applying a naïve 0.5 cutoff. Column (2) and (4) show the results when slightly altering this optimal cutoff, by respectively adding -0.1 and $+0.1$. Finally, column (5) reports the results obtained when dropping the classification. The model represented in this last column is thus similar to the baseline models in Machin et al. (2003) and Harasztosi and Lindner (2019). The employment response estimated is a weighted average of the reaction of tax-compliant and tax-evading firms. The lower the threshold, the smaller the number of firms classified as evaders. The estimated coefficient for FA is thus smaller than in specifications disentangling compliant from tax-evading firms. The results do not qualitatively change when altering the cutoff value. With a 0.5 cutoff, employment elasticity with respect to minimum wage is -1.19 for compliant firms and -0.41 for evaders (with respective bootstrapped standard errors: 0.19 and 0.12).

Table C.4 displays the results obtained using an alternative sample to train the algorithm. Instead of considering Nordic-owned firms as examples of compliant firms, we use a set of firms "Western-owned" firms, i.e., firms owned by an investor from the US, Canada, UK, Germany, France, the Netherlands, Belgium or from Nordic countries. The results are in line with those obtained using the baseline sample.

C.4. One-class classification

Our approach to classifying firms relies on standard supervised machine learning binary classification techniques, where the algorithm is trained on a sample containing *both* types of firms (compliant and tax-evading). As discussed in Section 3, we must make two (admittedly strong) assumptions to create this sample: (i) assuming that Nordic-owned firms represent instances of compliant firms, (ii) assuming that firms with "suspiciously" low wage employees are instances of evading firms. Alternatively, we implement another type of binary classification technique: One-Class Classification (OCC, see Fernández et al., 2018 for an introduction). OCC allows binary classification when only instances of one class are available for training. The algorithm aims to capture



(a) Compliant firms (b) Tax evading firms

Fig. C.1. Distribution of FA. Note: These figures display the distribution of FA respectively for the set of tax compliant firms (panel (a)) and the set of non-compliant firms (panel (b)), calculated in January 2013.

Table C.1
Employment effect - average FA and GAP.

	Dependent variable:					
	% change Jan 13 - Dec 14				% change Jan 13 - Jan 12	
	(1)	(2)	(3)	(4)	(5)	(6)
Evader	−0.007	−0.005	−0.002	−0.001	−0.022**	−0.024***
	(0.014)	(0.014)	(0.013)	(0.013)	(0.009)	(0.009)
FA_{mean}	−0.122***	−0.097***			−0.001	
	(0.018)	(0.018)			(0.012)	
FA_{mean} ×Evader	0.086***	0.071***			0.003	
	(0.028)	(0.027)			(0.016)	
GAP_{mean}			−1.205***	−0.937***		−0.099
			(0.176)	(0.179)		(0.108)
GAP_{mean} ×Evader			0.808***	0.667**		0.129
			(0.266)	(0.263)		(0.159)
Controls	No	Yes	No	Yes	Yes	Yes
Observations	5,524	5,524	5,524	5,524	5,524	5,524
R^2	0.009	0.046	0.009	0.046	0.026	0.026

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$ Robust standard errors in parentheses. FA_{mean} and GAP_{mean} are alternative bite measures, based on the average number of impacted employees over 2012. The set of controls include the following controls: firm age, legal status, NACE code, labor share and its square, profitability and its square, export share and its square.

Table C.2
Employment effect - Alternative FA and GAP.

	Dependent variable:					
	% change Jan 13 - Dec 14				% change Jan 13 - Jan 12	
	(1)	(2)	(3)	(4)	(5)	(6)
Evader	−0.004	−0.004	−0.001	−0.001	−0.019**	−0.019**
	(0.013)	(0.013)	(0.012)	(0.012)	(0.008)	(0.008)
FA_{alt}	−0.155***	−0.116***			0.007	
	(0.020)	(0.020)			(0.013)	
FA_{alt} ×Evader	0.108***	0.093***			−0.007	
	(0.029)	(0.029)			(0.019)	
GAP_{alt}			−2.234***	−1.664***		0.109
			(0.300)	(0.299)		(0.209)
GAP_{alt} ×Evader			1.565***	1.348***		−0.108
			(0.429)	(0.420)		(0.280)
Controls	No	Yes	No	Yes	Yes	Yes
Observations	5,524	5,524	5,524	5,524	5,524	5,524
R^2	0.011	0.047	0.010	0.046	0.026	0.026

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$ Robust standard errors in parentheses. FA_{alt} and GAP_{alt} are alternative bite measures, considering that in the absence of the minimum wage hikes, wages would have increased at the rate of the 2011–2013 average national wage increase. The set of controls include the following controls: firm age, legal status, NACE code, labor share and its square, profitability and its square, export share and its square.

the characteristics of instances in the training set to distinguish them from anomalies (i.e., observations belonging to the alternative class). This type of classification task is more difficult, as the algorithm learns from only one single class. We apply the most common algorithm for this approach, the one-class support vector machine (Schölkopf et al., 1999; Tax and Duin, 1999), using its implementation from the kernlab

Table C.3
Employment effect - Alternative classification thresholds.

	Dependent variable:				
	% change Jan 13 - Dec 14				
	(1)	(2)	(3)	(4)	(5)
Threshold	0.5 cutoff	Baseline −0.1	Baseline	Baseline +0.1	No cutoff
Evader	−0.003	−0.008	−0.000	−0.003	
	(0.013)	(0.013)	(0.014)	(0.015)	
FA	−0.130***	−0.120***	−0.108***	−0.092***	−0.075***
	(0.022)	(0.020)	(0.018)	(0.017)	(0.015)
FA ×Evader	0.087***	0.083***	0.071***	0.053*	
	(0.028)	(0.027)	(0.027)	(0.030)	
Observations	5,524	5,524	5,524	5,524	5,524
R^2	0.048	0.047	0.047	0.046	0.046

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$ Robust standard errors in parentheses. This table displays the results obtained with alternative cutoffs for binary classification. The baseline cutoff is 0.684. All regressions include the following controls: firm age, legal status, NACE code, labor share and its square, profitability and its square, export share and its square.

Table C.4
Employment effect - Alternative training set.

	Dependent variable:					
	% change Jan 13 - Dec 14				% change Jan 13 - Jan 12	
	(1)	(2)	(3)	(4)	(5)	(6)
Evader	0.011	−0.001	0.009	−0.003	−0.011	−0.014
	(0.015)	(0.015)	(0.014)	(0.014)	(0.010)	(0.010)
FA	−0.132***	−0.105***			0.006	
	(0.017)	(0.017)			(0.013)	
FA ×Evader	0.077***	0.085***			0.003	
	(0.029)	(0.029)			(0.019)	
GAP			−1.378***	−1.071***		0.022
			(0.174)	(0.175)		(0.113)
GAP ×Evader			0.970***	1.014***		0.134
			(0.281)	(0.275)		(0.184)
Controls	No	Yes	No	Yes	Yes	Yes
Observations	5,524	5,524	5,524	5,524	5,524	5,524
R^2	0.011	0.047	0.010	0.046	0.025	0.025

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$ Robust standard errors in parentheses. This table displays the results obtained using a broader set of Western-owned countries for training. The set of controls include the following controls: firm age, legal status, NACE code, labor share and its square, profitability and its square, export share and its square.

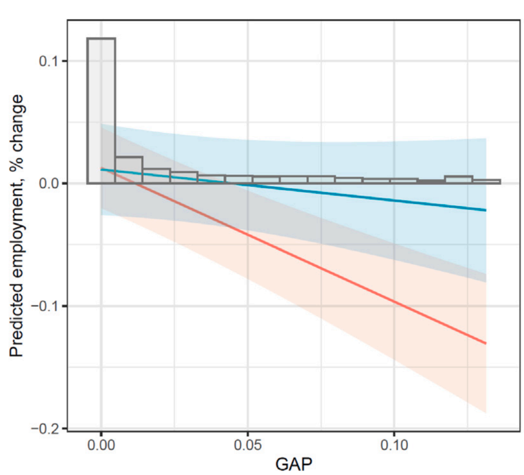


Fig. C.2. Employment effect - Interaction. Note: This figure displays the employment response to the 2014 minimum wage hike conditional on GAP and labor tax compliance (calculated between January 2013 and December 2014). Confidence intervals are represented at the 95% level. The histogram represents the distribution of GAP.

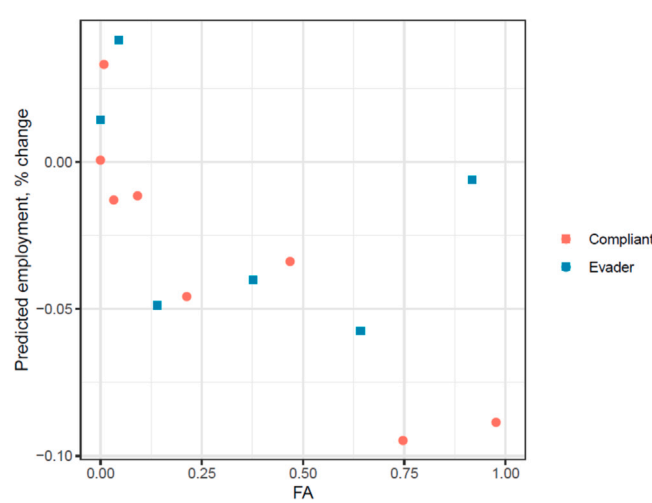


Fig. C.3. Binscatter plot. Note: This figure displays the relationship between FA and the percentage change in employment between January 2013 and December 2014, for compliant and tax-evading firms, adjusting for controls using the approach of Cattaneo et al. (2019).

R package. The idea is to determine the smallest possible hypersphere encompassing observations in the training set. Unlabeled observations within the boundaries of this hypersphere are considered normal, while those outside are considered anomalies. The main parameter to choose is the so-called ν-parameter, in the range $[0, 1]$, which governs how "soft" the margin of the hypersphere is, allowing for misclassified cases in the training set. We set this parameter to 0.5. The results are displayed in Table C.5. Columns (1) and (2) report the results when using evaders to train the algorithm, while columns (3) and (4) display the outcomes obtained using compliant firms. The conditional effect of non-compliance on employment response becomes significant

Table C.5
Employment effect - One-class classification.

	Dependent variable:			
	% change Jan 13 - Dec 14			
	One-class: Evading		One-class: Compliant	
	(1)	(2)	(3)	(4)
Evader	0.022	0.028*	−0.023	−0.022
	(0.015)	(0.014)	(0.015)	(0.014)
FA	−0.085***		−0.084***	
	(0.015)		(0.018)	
FA ×Evader	0.062*		0.025	
	(0.035)		(0.027)	
GAP		−0.757***		−0.774***
		(0.154)		(0.153)
GAP ×Evader		0.467		0.228
		(0.309)		(0.228)
Observations	5,524	5,524	5,524	5,524
R^2	0.047	0.046	0.046	0.045

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$ Robust standard errors in parentheses. This results displays the results obtained with a one-class support vector machine. Column 1) and 2) use the set of evading firms as a training sample, column 3) and 4) use the set of Nordic-owned firms. The set of controls includes the following controls: firm age, legal status, NACE code, labor share and its square, profitability and its square, export share and its square.

Table C.6
Firm exit.

	Dependent variable: P(close)							
	Dec 14	Dec 15	Dec 16	Dec 17	Dec 14	Dec 15	Dec 16	Dec 17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Evader	−0.014***	−0.010	−0.002	0.008	−0.014***	−0.014*	−0.011	−0.003
	(0.005)	(0.009)	(0.011)	(0.013)	(0.004)	(0.008)	(0.010)	(0.012)
FA	0.034***	0.044***	0.078***	0.105***				
	(0.011)	(0.012)	(0.015)	(0.016)				
FA ×Evader	−0.023*	−0.053***	−0.100***	−0.125***				
	(0.013)	(0.016)	(0.020)	(0.023)				
GAP					0.354***	0.220***	0.393***	0.542***
					(0.117)	(0.065)	(0.077)	(0.087)
GAP ×Evader					−0.278**	−0.259***	−0.482***	−0.618***
					(0.133)	(0.080)	(0.097)	(0.112)
Observations	5,524	5,524	5,524	5,524	5,524	5,524	5,524	5,524
R^2	0.022	0.053	0.075	0.081	0.022	0.053	0.075	0.082

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$ Robust standard errors in parentheses. The dependent variable is a binary variable indicting that the firm has exited by December of the year specified in each column. In column 2-4 and 6-8, FA and GAP measure the intensity of the 2014 and 2015 minimum wage hikes altogether with respect to firm-level employment in January 2013. All regressions include the following controls: firm age, legal status, NACE code, labor share and its square, profitability and its square, export share and its square.

(albeit only at the 10% level) when using a training sample composed of evaders and FA as a measure of the minimum wage bite. The sign of the coefficients remains, however, in line with the baseline. Considering that (i) one-class classification is a more difficult task than standard binary classification, and that (ii) the performance of a standard support vector machine, when evaluated on the test set, underperforms gradient boosting in our case, we nevertheless find these results consistent with the baseline.

C.5. Additional results

Table C.6 shows the effect of the interaction between a minimum wage hike and tax evasion on firm exit. The dependent variable takes the value 1 if the firm has exited by the given period. Fig. 5 in Section 4 represents the interaction in column (1).

Table C.7 is similar to Table 7 in Section 4 except that the models are estimated using only firms that survived throughout the whole period covered in the sample. The results thus provide information on the employment effect at the intensive margin. Together with Table C.6, the results indicate that firms use both the intensive and the extensive employment margins.

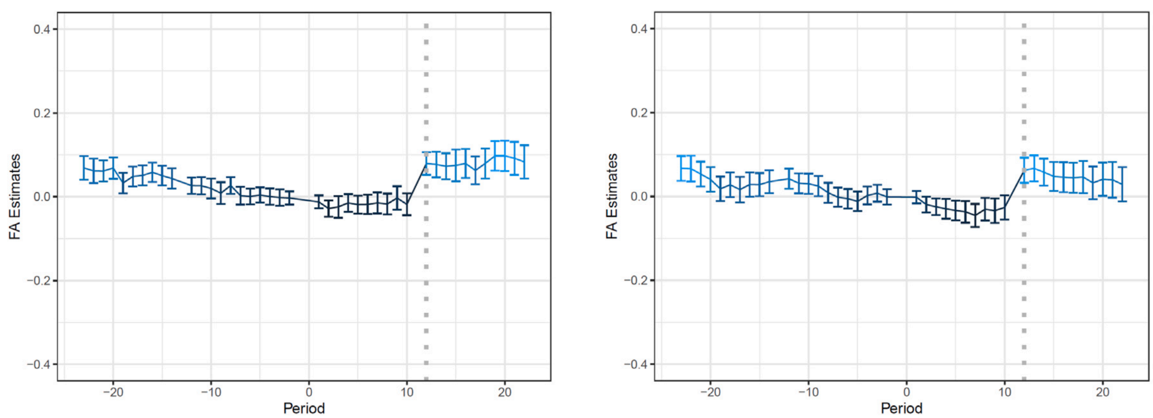
Figs. C.4, C.5 and C.6 are the counterparts of Fig. 3 respectively for wage, income tax and social security contributions. They display FA estimates (and the associated standard errors) when respectively regressing firm-level average wage, average personal income tax and average social security contributions on FA and the set of controls, on the subsample of compliant and evading firms separately.

Estimates for wage and social security contributions are more precise than for personal income. This is due to the fact that personal income tax may benefit from exemptions, delayed payments, or refunds from overdue payments (e.g., following a change in family status). In contrast, social security payments are calculated from gross wages at a perfectly flat rate, without any exemptions or non-taxable income. For all three variables, no clear adjustment is visible following the announcement of the minimum wage increase, except for a slight slowdown in wage and contributions in the third quarter of 2013 for the sample of evading firms. However, a sharp increase is visible following the implementation of the reform, both for compliant and evading firms. This provides evidence that the minimum wage hike did increase the wages of workers at the bottom of the income distribution. Average wages increased by approximately 8 p.p. more in largely impacted firms than in non-impacted firms. The 2014 minimum wage increase was equivalent to a 12% raise, but some workers already receiving pay greater than the previous minimum wage (but smaller than the new one), thus requiring a smaller wage increase to comply with the new level. Additionally, wages in non-impacted firms have been growing for reasons unrelated to the minimum wage. These two facts explains why the FA estimates are lower than 0.12. For wage and social security

Table C.7
Employment effect - intensive margin.

	Dependent variable:					
	% change Jan 13 - Dec 14				% change Jan 13 - Jan 12	
	(1)	(2)	(3)	(4)	(5)	(6)
Evader	−0.018	−0.015	−0.017	−0.016	−0.011	-0.013^{*}
	(0.013)	(0.013)	(0.012)	(0.012)	(0.009)	(0.008)
FA	-0.106^{***}	-0.077^{***}			0.007	
	(0.016)	(0.016)			(0.012)	
FA × Evader	0.061^{**}	0.050^{**}			−0.013	
	(0.026)	(0.025)			(0.017)	
GAP			-1.088^{***}	-0.778^{***}		0.056
			(0.164)	(0.163)		(0.128)
GAP × Evader			0.698^{***}	0.600^{**}		−0.083
			(0.250)	(0.245)		(0.175)
Control	No	Yes	No	Yes	Yes	Yes
Observations	5,391	5,391	5,391	5,391	5,391	5,391
R^2	0.007	0.041	0.007	0.041	0.020	0.020

Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$ Robust standard errors in parentheses. This table displays the results obtained when restricting the sample to surviving firms one year after the minimum wage increase. The set of controls includes the following controls: firm age, legal status, NACE code, labor share and its square, profitability and its square, export share and its square.



(a) Tax-compliant firms (b) Tax-evading firms

Fig. C.4. Wage effect. Note: These figures show the effect of the share of affected workers FA on firm-level average wage growth, controlling for firm's age, legal status, NACE sector, average export share and its square, average share of labor and its square, average profit and its square. The samples are restricted respectively to the sample of firms classified as labor tax compliant (panel (a)) and labor tax evaders (panel (b)). Each period represents a month. Period 0 indicates January 2013, the reference period. The vertical bar indicates January 2014, when the new minimum wage was introduced.

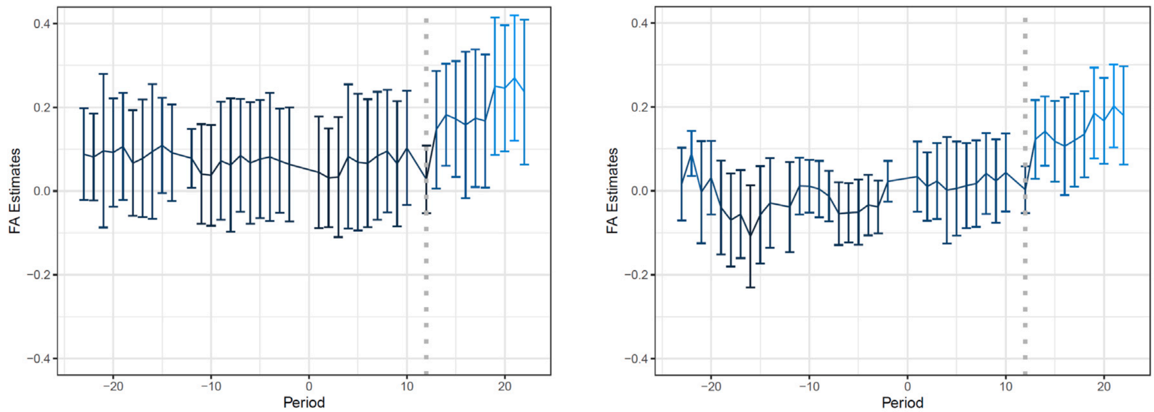
contributions, the magnitude of the estimated coefficients decreases over time for evading firms. This can be attributed to the relatively rapid increase in wages in non-impacted evading firms. Finally, in terms of pre-trends, there is no significant difference between impacted and non-impacted firms over 2012 and 2013. Going further back in 2011, a small difference appears for all firms, where wages and social security contributions were *lower* in impacted firms (the dependent variable is measured as $\frac{y_{i,t}-y_{Jan2013}}{y_{Jan2013}}$).

C.6. Back-of-the-envelope estimation

The main contribution of our analysis is to document a trade-off between job loss in compliant firms and fiscal gain from reduced evasion in evading firms. We provide a back-of-the-envelope analysis to quantify this trade-off, proceeding as follows:

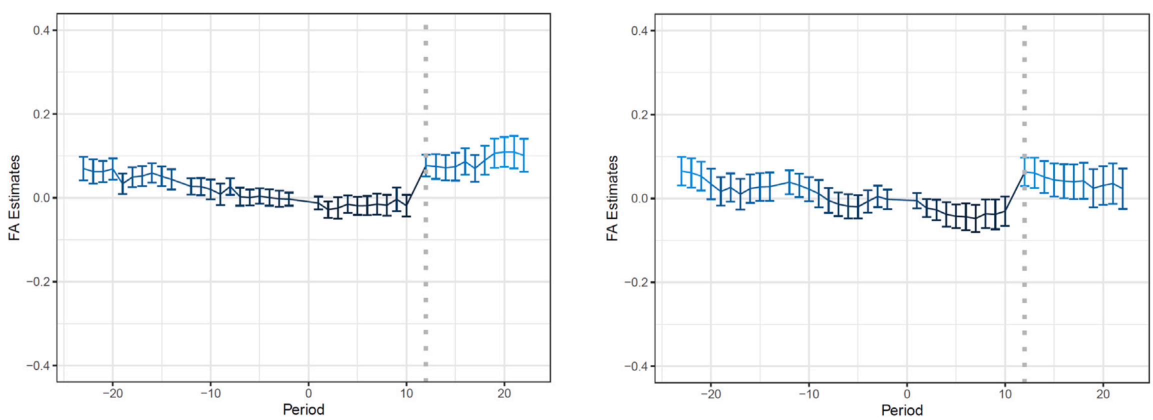
1. We estimate Eq. (2), comparing December 2014 to January 2013 using the whole sample.
2. We set the value of FA to 0 for all firms, as if the minimum wage hike did not occur.
3. We predict what employment would have been in each compliant firm using this alternative, counterfactual dataset, and assume a null employment response for tax-evading firms.
4. We use this estimated employment growth to compute the counterfactual number of employees in compliant firms in the absence of the reform, and obtain an aggregate number summing employees in all these firms.
5. Finally, we compare this counterfactual aggregate number to the actual aggregate number that we observe in the data.

Using this raw approach, we estimate that actual total employment is 1.5% lower than what it would have been in the absence of the minimum wage hike. Regarding fiscal implications, personal income tax (PIT) and social security contributions (SSC) are both linearly determined by wage (flat rate for PIT and SSC, even though several PIT wedges exist for low income, making the relationship with wage more fuzzy than for SSC). We thus estimate what would have been the total wage bill for all the firms in our sample in the absence of the reform, and then calculate the total PIT and SSC. When comparing these estimates to the actual aggregate figures, the actual aggregate wage bill is 0.3% larger than what it would have been in the absence of the reform. Since personal income tax and social contributions are (almost) perfectly linear functions of wages, the results are identical.



(a) Tax-compliant firms (b) Tax-evading firms

Fig. C.5. Personal income tax effect. Note: These figures show the effect of the share of affected workers FA on firm-level average personal income tax growth, controlling for firm's age, legal status, NACE sector, average export share and its square, average share of labor and its square, average profit and its square. The samples are restricted respectively to the sample of firms classified as labor tax compliant (panel (a)) and labor tax evaders (panel (b)). Each period represents a month. Period 0 indicates January 2013, the reference period. The vertical bar indicates January 2014, when the new minimum wage was introduced.



(a) Tax-compliant firms (b) Tax-evading firms

Fig. C.6. Social security contribution effect. Note: These figures show the effect of the share of affected workers FA on firm-level average social security growth, controlling for firm's age, legal status, NACE sector, average export share and its square, average share of labor and its square, average profit and its square. The samples are restricted respectively to the sample of firms classified as labor tax compliant (panel (a)) and labor tax evaders (panel (b)). Each period represents a month. Period 0 indicates January 2013, the reference period. The vertical bar indicates January 2014, when the new minimum wage was introduced.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jpubeco.2023.105027.

References

Ahn, J., Choi, J., Chung, S., 2022. Labor Market Rigidity at Home and Multinational Corporations' Flexible Task Reallocation Abroad. Technical Report, World Bank Policy Research Working Paper No. 10114.

Allegretto, S.A., Dube, A., Reich, M., 2011. Do minimum wages really reduce teen employment? Accounting for heterogeneity and selectivity in state panel data. Ind. Relat.: J. Econ. Soc. 50 (2), 205–240.

Allegretto, S., Reich, M., 2018. Are local minimum wages absorbed by price increases? Estimates from internet-based restaurant menus. ILR Rev. 71 (1), 35–63.

Ashenfelter, O., Smith, R.S., 1979. Compliance with the minimum wage law. J. Polit. Econ. 87 (2), 333–350.

Athey, S., Imbens, G.W., 2019. Machine learning methods that economists should know about. Annu. Rev. Econ. 11, 685–725.

Basu, A.K., Chau, N.H., Kanbur, R., 2010. Turning a blind eye: Costly enforcement, credible commitment and minimum wage laws. Econ. J. 120 (543), 244–269.

Bell, B., Machin, S., 2018. Minimum wages and firm value. J. Labor Econ. 36 (1), 159–195.

Beneish, M.D., 1999. The detection of earnings manipulation. Financ. Anal. J. 55 (5), 24–36.

Bíró, A., Prinz, D., Sándor, L., 2022. The minimum wage, informal pay and tax enforcement. J. Public Econ. 215, 104728.

Bosch, M., Manacorda, M., 2010. Minimum wages and earnings inequality in urban Mexico. Am. Econ. J.: Appl. Econ. 2 (4), 128–149.

Braguinsky, S., Mityakov, S., 2015. Foreign corporations and the culture of transparency: Evidence from Russian administrative data. J. Financ. Econ. 117 (1), 139–164.

Braguinsky, S., Mityakov, S., Liscovich, A., 2014. Direct estimation of hidden earnings: Evidence from Russian administrative data. J. Law Econ. 57 (2), 281–319.

Brown, C., 1999. Minimum wages, employment, and the distribution of income. In: Handbook of Labor Economics, vol. 3, Elsevier, pp. 2101–2163.

Cattaneo, M.D., Crump, R.K., Farrell, M.H., Feng, Y., 2019. On binscatter. arXiv preprint arXiv:1902.09608.

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

Cengiz, D., Dube, A., Lindner, A., Zipperer, B., 2019. The effect of minimum wages on low-wage jobs. Q. J. Econ. 134 (3), 1405–1454.

Chava, S., Oettl, A., Singh, M., 2023. Does a one-size-fits-all minimum wage cause financial stress for small businesses? Manage. Sci..

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Christofides, L.N., Polycarpou, A., Vrachimis, K., 2013. Gender wage gaps,'sticky floors' and 'glass ceilings' in Europe. Labour Econ. 21, 86–102.

Clemens, J., 2021. How do firms respond to minimum wage increases? Understanding the relevance of non-employment margins. J. Econ. Perspect. 35 (1), 51–72.
Clemens, J., Kahn, L.B., Meer, J., 2018. The Minimum Wage, Fringe Benefits, and Worker Welfare. Technical Report, National Bureau of Economic Research.
Clemens, J., Kahn, L.B., Meer, J., 2021. Dropouts need not apply? The minimum wage and skill upgrading. J. Labor Econ. 39 (S1), S107–S149.
Clemens, J., Strain, M.R., 2022. Understanding "wage theft": Evasion and avoidance responses to minimum wage increases. Labour Econ. 102285.
DeBacker, J., Heim, B.T., Tran, A., 2015. Importing corruption culture from overseas: Evidence from corporate tax evasion in the United States. J. Financ. Econ. 117 (1), 122–138.
Dechow, P.M., Ge, W., Larson, C.R., Sloan, R.G., 2011. Predicting material accounting misstatements. Contemp. Account. Res. 28 (1), 17–82.
Draca, M., Machin, S., Van Reenen, J., 2011. Minimum wages and firm profitability. Am. Econ. J.: Appl. Econ. 3 (1), 129–151.
Drucker, L., Mazirov, K., Neumark, D., 2021. Who pays for and who benefits from minimum wage increases? Evidence from Israeli tax data on business owners and workers. J. Public Econ. 199, 104423.
Dube, A., Lester, T.W., Reich, M., 2010. Minimum wage effects across state borders: Estimates using contiguous counties. Rev. Econ. Stat. 92 (4), 945–964.
Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., Vom Berge, P., 2022. Reallocation effects of the minimum wage. Q. J. Econ. 137 (1), 267–328.
Engström, P., Hagen, J., 2017. Income underreporting among the self-employed: A permanent income approach. Eur. Econ. Rev. 92, 92–109.
Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F., 2018. Learning from Imbalanced Data Sets, Vol. 10. Springer.
Fisman, R., Miguel, E., 2007. Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. J. Polit. Econ. 115 (6), 1020–1048.
Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Statist. 1, 1189–1232.
Garnero, A., Lucifora, C., 2022. Turning a blind eye? Compliance to minimum wages and employment. Economica 89, 84–907.
Gavoille, N., Zasova, A., 2021. Foreign ownership and labor tax evasion: Evidence from Latvia. Econom. Lett. 207.
Gavoille, N., Zasova, A., 2022. Minimum wage spike and income underreporting: A back-of-the-envelope-wage analysis. J. Comp. Econ. 51 (1), 372–402.
Georgiadis, A., Manning, A., 2020. The impact of the UK minimum wage: Evidence from high-frequency firm-level data. Available at SSRN 3560897.
Gopalan, R., Hamilton, B.H., Kalda, A., Sovich, D., 2021. State minimum wages, employment, and wage spillovers: Evidence from administrative payroll data. J. Labor Econ. 39 (3), 673–707.
Goraus-Tańska, K., Lewandowski, P., 2019. Minimum wage violation in Central and Eastern Europe. Int. Labour Rev. 158 (2), 297–336.
Gorodnichenko, Y., Martinez-Vazquez, J., Sabirianova Peter, K., 2009. Myth and reality of flat tax reform: Micro estimates of tax evasion response and welfare effects in Russia. J. Polit. Econ. 117 (3), 504–554.
Greenwell, B.M., Boehmke, B.C., Gray, B., 2020. Variable importance plots-an introduction to the vip package. R J. 12 (1), 343.
Gregory, T., Zierahn, U., 2022. When the minimum wage really bites hard: The negative spillover effect on high-skilled workers. J. Public Econ. 206, 104582.
Hajek, P., Henriques, R., 2017. Mining corporate annual reports for intelligent detection of financial statement fraud–A comparative study of machine learning methods. Knowl.-Based Syst. 128, 139–152.
Harasztosi, P., Lindner, A., 2019. Who pays for the minimum wage? Amer. Econ. Rev. 109 (8), 2693–2727.
Harmon, C., Walker, I., 1995. Estimates of the economic return to schooling for the United Kingdom. Amer. Econ. Rev. 85 (5), 1278–1286.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: DAta Mining, Inference, and Prediction. Springer Science & Business Media.
Hazans, M., 2011. Informal workers across Europe: Evidence from 30 European countries. World Bank Policy Research Working Paper 5912.
Huang, S.Y., Tsaih, R.H., Yu, F., 2014. Topological pattern discovery and feature extraction for fraudulent financial reporting. Expert Syst. Appl. 41 (9), 4360–4372.
Hurst, E., Li, G., Pugsley, B., 2014. Are household surveys like tax forms? Evidence from income underreporting of the self-employed. Rev. Econ. Stat. 96 (1), 19–33.
Jardim, E., Long, M.C., Plotnick, R., Van Inwegen, E., Vigdor, J., Wething, H., 2022. Minimum-wage increases and low-wage employment: Evidence from seattle. Am. Econ. J.: Econ. Policy 14 (2), 263–314.
Jardim, E., Van Inwegen, E., 2019. Payroll, Revenue, and Labor Demand Effects of the Minimum Wage. Technical Report, Upjohn Institute working paper No. 19-298.
Jascisens, V., Zasova, A., 2021. Million dollar baby: Should parental benefits depend on wages when the payroll tax evasion is present? BICEPS/SSE Riga Research papers, No.9.
Kirkos, E., Spathis, C., Manolopoulos, Y., 2007. Data mining techniques for the detection of fraudulent financial statements. Expert Syst. Appl. 32 (4), 995–1003.
Kotsiantis, S., Koumanakos, E., Tzelepis, D., Tampakas, V., 2006. Forecasting fraudulent financial statements using data mining. Int. J. Comput. Intell. 3 (2), 104–110.
Kukk, M., Paulus, A., Staehr, K., 2020. Cheating in Europe: Underreporting of self-employment income in comparative perspective. Int. Tax Public Finance 27 (2), 363–390.
Kukk, M., Staehr, K., 2014. Income underreporting by households with business income: Evidence from Estonia. Post-Communist Econ. 26 (2), 257–276.
Kumler, T., Verhoogen, E., Frías, J., 2020. Enlisting employees in improving payroll tax compliance: Evidence from Mexico. Rev. Econ. Stat. 102 (5), 881–896.
Lastauskas, P., Prosukute, A., Zaldokas, A., 2021. How do firms adjust when trade stops? Unpublished manuscript.
Lemos, S., 2009. Minimum wage effects in a developing country. Labour Econ. 16 (2), 224–237.
Liu, X., 2016. Corruption culture and corporate misconduct. J. Financ. Econ. 122 (2), 307–327.
Luca, D.L., Luca, M., 2019. Survival of the Fittest: The Impact of the Minimum Wage on Firm Exit. Technical Report, National Bureau of Economic Research.
Machin, S., Manning, A., Rahman, L., 2003. Where the minimum wage bites hard: Introduction of minimum wages to a low wage sector. J. Eur. Econom. Assoc. 1 (1), 154–180.
Manning, A., 2021. The elusive employment effect of the minimum wage. J. Econ. Perspect. 35 (1), 3–26.
Mayneris, F., Poncet, S., Zhang, T., 2018. Improving or disappearing: Firm-level adjustments to minimum wages in China. J. Dev. Econ. 135, 20–42.
Meer, J., West, J., 2016. Effects of the minimum wage on employment dynamics. J. Hum. Resour. 51 (2), 500–522.
Meghir, C., Narita, R., Robin, J.-M., 2015. Wages and informality in developing countries. Amer. Econ. Rev. 105 (4), 1509–1546.
Meriküll, J., Tverdostup, M., 2023. The gap that survived the transition: The gender wage gap in Estonia over three decades. Econ. Syst. 101127.
Mullainathan, S., Spiess, J., 2017. Machine learning: An applied econometric approach. J. Econ. Perspect. 31 (2), 87–106.
Neumark, D., Corella, L.F.M., 2021. Do minimum wages reduce employment in developing countries? A survey and exploration of conflicting evidence. World Dev. 137, 105165.
Neumark, D., Wascher, W., 2010. Minimum Wages. MIT Press, Cambridge, MA.
Nygård, O.E., Slemrod, J., Thoresen, T.O., 2019. Distributional implications of joint tax evasion. Econ. J. 129 (620), 1894–1923.
OECD, 2019. OECD economic surveys: Latvia.
Paulus, A., 2015. Tax Evasion and Measurement Error: an Econometric Analysis of Survey Data Linked with Tax Records. Technical Report, ISER Working Paper Series.
Pelek, S., Uysal, G., 2018. Envelope wages, underreporting and tax evasion: The case of Turkey. In: 17th Louis-Andre-Gerard-Varet International Conference in Public Economics.
Perry, G.E., Arias, O., Fajnzylber, P., Maloney, W.F., Mason, A., Saavedra-Chanduvi, J., 2007. Informality: Exit and Exclusion. The World Bank.
Pissarides, C.A., Weber, G., 1989. An expenditure-based estimate of Britain's black economy. J. Public Econ. 39 (1), 17–32.
Putniņš, T.J., Sauka, A., 2015. Measuring the shadow economy using company managers. J. Comp. Econ. 43 (2), 471–490.
Ravisankar, P., Ravi, V., Rao, G.R., Bose, I., 2011. Detection of financial statement fraud and feature selection using data mining techniques. Decis. Support Syst. 50 (2), 491–500.
Renkin, T., Montialoux, C., Siegenthaler, M., 2020. The pass-through of minimum wages into US retail prices: Evidence from supermarket scanner data. Rev. Econ. Stat. 1–99.
Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10 (3), e0118432.
Schölkopf, B., Williamson, R.C., Smola, A., Shawe-Taylor, J., Platt, J., 1999. Support vector method for novelty detection. Adv. Neural Inf. Process. Syst. 12.
Special Eurobarometer, 2014. Undeclared Work in the European Union. Technical Report, European Commission.
Tax, D.M., Duin, R.P., 1999. Support vector domain description. Pattern Recognit. Lett. 20 (11–13), 1191–1199.
Tonin, M., 2011. Minimum wage and tax evasion: Theory and evidence. J. Public Econ. 95 (11–12), 1635–1651.
Tonin, M., 2013. Underreporting of earnings and the minimum wage spike. IZA J. European Labor Stud. 2 (1), 1–18.
Varian, H.R., 2014. Big data: New tricks for econometrics. J. Econ. Perspect. 28 (2), 3–28.
West, J., Bhattacharya, M., 2016. Intelligent financial fraud detection: A comprehensive review. Comput. Secur. 57, 47–66.
World Bank, 2017. Latvia Tax Review. World Bank Group, Washington, D.C..