



Detecting envelope wages with e-billing information

Andrea Lopez-Luzuriaga¹ · Monica Calijuri² · Carola Pessino² ·
Simeon Schächtele² · Ubaldo Gonzalez² · Carla Chamorro³

Accepted: 22 October 2023 / Published online: 22 December 2023
© The Author(s) 2023

Abstract

We use information from the electronic billing system to estimate the underreporting of income of private sector employees. We follow an expenditure-based methodology using the consumption of public and private sector employees for similar levels of reported income. We find that private sector employees underreport between 7 and 9% of their income in Ecuador. The size of the underreporting gap is negatively correlated with the number of employees at the firm, consistent with different risks and administrative costs of ‘envelope wages’ in small versus large firms.

Keywords Income tax · Evasion · Electronic billing

JEL Classification H24 · H26 · D83

✉ Andrea Lopez-Luzuriaga
andrea.lopez@urosario.edu.co

Monica Calijuri
monicasc@iadb.org

Carola Pessino
cpessino@iadb.org

Simeon Schächtele
s.schaechtele@gmail.com

Ubaldo Gonzalez
ubaldog@iadb.org

Carla Chamorro
cnchamorro@sri.gob.ec

¹ Universidad del Rosario, Bogotá, Colombia

² Inter-American Development Bank, New York, USA

³ Servicio de Rentas Internas Ecuador, Quito, Ecuador

1 Introduction

Informality is a key feature of emerging markets and developing economies (Ohnsorge & Yu, 2021). It constitutes an extraordinary challenge for their development as the shadow economy eschews taxation and social security contributions while hindering the state's ability to deliver benefits and enforce regulations. An understudied form of quasi-informality is the payment of envelope wages. Envelope wages refers to the practice of firms paying formal workers part of their remuneration off the books. Taxpayers thus avoid payroll and income tax but also lose out on entitlements.

A great deal of hope for curbing evasion and increasing formality is put into technology. First, the withholding of income and payroll taxes by the employer is believed to make the underreporting of wages nearly impossible under the right circumstances (Kleven et al., 2011; Jensen, 2022; Slemrod, 2019), leaving self-employment as the main source of unreported labor income. Second, by increasing traceability and reducing transaction costs, digital payments could make tax evasion and informality a memory of the past. However, withholding may not work equally well in all countries; hence, it may not be correct that for employees, "tax authority income records can be regarded as the 'gold standard'" (Cabral et al., 2021). Moreover, the extent of envelope wage payments is not known in all countries, and to estimate their size, researchers have resorted to survey data, which have been shown to yield biased results (Cabral et al., 2021; Paulus, 2015).

This paper seeks to answer the question, "To what extent do employees under-report income?" using a novel approach that matches electronic billing data (a third-party reported measure of consumption) with income tax records. Our study sheds light on how well firms comply with the withholding of income and payroll taxes. In addition, we provide empirical results on the theoretical prediction that collusive tax evasion is easier and more likely in small firms.

We use an expenditure-based methodology to estimate the gap in reported income between public and private sector employees in Ecuador. Briefly, we estimate the consumption and income relationship, controlling for individuals' demographic characteristics for public and private sector employees. We assume that the relationship between consumption and real income is independent of the employment sector, and any differences observed are due to a difference between real and reported income. Ecuador has a comprehensive electronic billing system. We match this detailed data on consumption to employees' income tax records. Our study focuses on wage earners, so we exclude all individuals that have self-employed income. There are two reasons for making this empirical choice. First, the main focus of this study is to understand the extent to which envelope wages exist in a context where a third-party reporting system comprehensively covers the relationship between employers and employees. The second reason is practical; including these individuals would add noise to the estimation, given the characteristics of the tax system in Ecuador. In the Ecuadorian tax system, self-employed individuals only have to report their gross income and expenses. The expenses are deductible from their taxable income. As a result, self-employed individuals have an

extra margin to evade that, in the Ecuadorian case, is easily used to decrease taxable income when gross income is third-party reported as shown by Carrillo et al. (2017).

At the core of our estimation is the assumption that individuals with similar demographic characteristics and real income will have similar consumption patterns, particularly of food, independently of the source of their income. If we find differences in the relationship between consumption and reported income, those differences are consistent with income misreporting. Crucially, public employers have no incentive or opportunity to misreport their employees' wages, so we can use employees in the public sector as the benchmark. Using survey data, we find evidence that if there is a systematic difference in public and private sector employees, the public sector employees report higher consumption; therefore, any bias would be against our estimation, and the difference in reported income between otherwise similar employees in the public and private sectors is a lower bound estimate for underreporting of employee income in the private sector.

Overall, we find that for a given consumption level, the reported income from private sector employees is smaller than that reported by public sector employees. Specifically, estimates center at around 8% of underreporting. However, once we look at heterogeneity by firm size, the underreported income is between 25% for small firms and 12% for middle-size firms using food consumption and between 40% and 13% using total consumption, and the gap is statistically significant in all cases. We calculate the size of the underreported amount using our estimate for the whole sample to be conservative and find that the amount of wages not reported amounts to 3% of Ecuador's GDP. The income tax loss is relatively small due to the tax's progressive nature, around 1% of total tax revenue. However, the unpaid social security contributions are sizable and equivalent to 9% of the total contributions.

The heterogeneity by firm size is striking: the reporting gap is largest—up to 40% of income—for small firms with 3 employees. It decreases with firm size until it vanishes for larger firms with more than 50 employees. This finding confirms theoretical predictions that collusive tax evasion is less likely in large firms under more scrutiny from tax authorities and where many employees may be exposed to the practice.

Our empirical results are consistent with the theoretical prediction that collusive evasion is more likely in smaller firms (Kleven et al., 2016; Barth & Ognedal, 2018). From a practical viewpoint, our methodology may allow tax authorities to target their taxpayer education and enforcement measures.

The rest of the paper is organized as follows. Section 2 presents a literature review. Section 3 presents an overview of the background. Section 4 lays out our sources of information and empirical strategy. Section 4.1 discusses our main assumption and possible biases if it fails, and 5 presents the results. Section 6 concludes.

2 Literature review

Measuring evasion is a complex problem due to the nature of evasion itself, similar to measuring other illegal activities. There are two main ways of evading taxes: staying in the informal sector and not registering with the tax authority (extensive margin) or underreporting transactions to the tax authority (intensive

margin). We focus on the latter problem and make several contributions to the literature.

First, we focus on employees, whose tax compliance is most of the time taken for granted in the expenditure-based literature studying income underreporting and tax evasion. We do not assume that employees' income is reported perfectly due to third-party reporting but instead challenge this assumption as our starting point. We analyze the intensive margin of underreporting wages (i.e., envelope wages). We follow a 'traces-of-income' or consumption-based methodology in the spirit of the one pioneered by Pissarides and Weber. They assumed that the source of income does not have systematically different effects on the consumption of food by the self-employed and employees; therefore, any systematic difference between those two groups is due to the difference in underreporting opportunity (Pissarides & Weber, 1989). This literature has grown using different groups, types of consumption, and contexts from this seminal work. The mainstay of this literature has focused on the underreporting of self-employed people, benchmarking the consumption-income patterns of the self-employed against employees, who are assumed to report income reliably. Instead of comparing the consumption-income relation of self-employed versus employed taxpayers, we compare two groups of employees with different opportunities to evade—private and public sector employees.

Second, this paper is the first to use consumption data from electronic billing (matched to income tax records), avoiding biases inherent in survey data. In contrast, most of the existing literature relies on self-reported income or consumption data or survey data for both income and consumption, which has been shown to be unreliable for this purpose (Cabral et al., 2021; Paulus, 2015). Cabral et al. (2015), using survey data on food consumption and income, compare self-employed with employees in Great Britain and find that the self-employed report around 81% of their income. Engström and Holmlund (2009), using survey data on consumption, find that in Sweden, the self-employed report around 70% of their income. Dunbar and Fu (2015) use the Survey of Financial Security and the Survey of Household Spending in order to estimate that between 35 and 50% of Canadian Households underreport income. The estimated underreported income is equivalent to between 14 and 19 percent of the GDP. Artavanis et al. (2016) detect systematic differences in reported income between employees and self-employed individuals by comparing their access to credit in Greece, one exception that uses third-party information for the consumption proxy. In this context, self-employed individuals report around 55% of their actual income. Other authors have used consumption reported on the same tax form. Feldman and Slemrod (2007) estimate the underreporting of income in the USA using donations as a proxy for income. Their identifying assumption is that the donation-income relationship is the same for both employees and the self-employed. They find that self-employed individuals report 65% of their income. Torregrosa-Hetland (2020) measures the same gap for Spain, using donations, and finds that the self-employed report between 50% and 70% of their true income (the lower estimation corresponds to the top 10% of earners); and Domínguez-Barrero et al. (2017) show that compliance changes with the economic cycle.

Third, we contribute to the literature on underreporting employee income and envelope wages. There is some evidence that there is a possibility of underreported wages when part of the income received by the employee is kept outside the books and is not reported either to social security or the tax authority (i.e., envelope wages). The presence of these envelope wages varies widely with context. For instance, in Denmark, the evasion of third-party reported income is close to zero (Kleven et al., 2011). From survey data, there is evidence that low evasion on reported wages is not the norm in all European countries. Barth and Ognedal (2018) present survey evidence that, in the European Union countries, around 5% of employees received part of their wages on the side, and this extra income is not reported to the tax authority. Still, there is significant heterogeneity across countries, in Romania, the share is 15%, Bulgaria 10%, and Spain between 5% and 7% (Di Nola et al., 2019). Williams and Horodnic (2017) use information from the Eurobarometer and find that 3% of workers over the 28 European countries covered by the survey received underreported salaries, and the percentage is more significant for unskilled workers, although there is a considerable variation between Eastern and Western European countries. A few studies rely on variation in incentives created by a social security reform. Bergolo and Cruces (2014) show that when a social insurance reform that tied the benefits to the reported wages was introduced in Uruguay, employees of small firms increased their reported income by about 25%. Kummer et al. (2020) measure the underreporting of wages to evade income and payroll taxes in Mexico. They compare two sources of information: individual wages reported by their employers to social security and a household-labor survey. They cannot measure underreporting at the individual level but rather at cells defined by the metropolitan area, sector, firm size, and employees' age group. They take advantage of the change in the incentive structure to truthfully report one's wages generated by a 1997 social security reform. After the reform, there was an increase in reported wages, especially among smaller firms and younger workers. Using the Pissarides and Weber type method, Ekici and Besim (2016) estimate that private employees in North Cyprus report 86% of their true income. Gorodnichenko et al. (2009), also find that the workers of smaller firms in Russia are more likely to underreport income than those of larger firms using a similar methodology. They provide as possible mechanisms the different levels of monitoring according to firm size.

We contribute to this literature by estimating the underreporting of employees and documenting a gradient with firm size using third-party reported electronic billing data on consumption in combination with income tax records in Ecuador.

3 Institutional background and data

Ecuador is a middle-income country with a sizable part of the economy in the informal sector. The National Institute of Statistics and Censuses (INEC) defines the people involved in informal activities as economic units that are not legally incorporated

in a company.¹ Following this definition, employees in the formal sector work in registered firms, public or private, and by a constitutional mandate, there are no part-time workers; all formal employees have a 40-hour workweek. Hence, individuals cannot adjust their hours or take a (formal) second job in order to change their income.

In 2017, people employed in the urban area were approximately 50.4% in the formal sector.² The informal sector consists of all the economic activity made by agents who do not report to the government, pay taxes, or contribute to social security.

In this context, Ecuador has a progressive income tax with nine tax brackets and marginal tax rates from 0% to 35%. The taxable income for employees is their pay less the payroll tax (a flat rate of around 9% paid by the employee and around 12% paid by the employer³) and less deductions. For the fiscal year 2017, everyone with a taxable income less than \$11,290⁴ was in the first tax bracket and paid zero tax. All taxpayers are entitled to a deduction for personal expenses in education, clothing, health care, housing, and food. The deduction is capped at \$14,677. All taxpayers who made a deduction for personal expenses larger than \$ 5,645 had to fill out an extra tax form itemizing their consumption and had access to the information about their purchases from firms that were part of the electronic billing system. Seniors and disabled people are entitled to an extra deduction. Self-employed individuals fill out the same tax form and pay on the same tax schedule as wage earners.⁵ Self-employed individuals only have to report their gross income and expenses. The expenses are deductible from their taxable income.

Employers have to withhold taxes monthly. The tax year coincides with the calendar year. Employers have to fill out an income tax return on behalf of their employees in February of the following year; if adjustments need to be made to this report, they can report an income tax return until the end of March. We identify the employee's sector using her employer's tax registry information on the withholding forms and the social security contributions.

A taxpayer is a public sector employee if her employer is a public sector entity according to its tax ID. An employee is defined as a private sector employee if her employer is a private sector business according to its tax ID. An employee can have more than one job in a fiscal year; in those cases, her employers can be in different sectors.

Ecuador started implementing an electronic billing system in 2012. By 2017, the system included all incorporated and non-incorporated firms required to keep accounting books and taxpayers who can print sales receipts through computerized systems (instead of pre-printed bills). The electronic billing system stores each transaction, including the information of consumers and sellers, its tax ID, the location of the store where the transaction was registered, the total amount of the purchase, and

¹ International Labour Office 2013

² Instituto Ecuatoriano de Estadísticas y Censos 2017.

³ The percentage is slightly different depending on the type of contract.

⁴ Ecuador has used the United States dollar as currency since 2000.

⁵ Unless their gross income is larger than \$300,000.

the date. Due to the existence of deductions for personal expenses and the design of the electronic billing system, the default in Ecuador is to get a tax receipt that includes the consumer's tax ID.

We use two primary sources of information: income tax returns and information from the electronic billing system. A consumption proxy for each individual is generated using the information from the electronic billing system. We analyze the information from 2017 to take advantage of the fact that, as of this year, the electronic billing system covers a large portion of businesses in Ecuador; in practical terms, only the small unincorporated businesses that use pre-printed paper bills are not included. In fact, close to 75% of the sales reported in the value-added tax form—that covers all the formal transactions—in the country pass through the electronic billing system. Therefore, it is unclear whether consumers would be perfectly informed which stores would report the transaction to the tax authority using the electronic billing system.

Using the seller's economic activity code, we can identify a proxy for each individual's food consumption, and we calculate the proxy for total consumption using all the sellers. For instance, if Person A buys from Store B and B is registered as a grocery store, we categorize the consumption as food consumption. The transaction is part of the total consumption if registered as a clothes store. We do not have access to the line items of the transactions; therefore, there is some measurement error in the categorization of consumption. For instance, if a person buys a mattress in a department store, that consumption is included even though traditionally, a mattress is a durable good and would be excluded from this kind of calculation. Similarly, if a person buys cleaning products at the grocery store, that purchase would be added to food consumption. However, there is no reason to believe that this measurement error differs for public and private sector employees, so it should not bias our estimation. However, our point estimate of the portion of income consumed will be large. This should not be problematic for calculating the evasion gap because it is a relative measure of the reporting of income of the private sector employees against the public sector employees.

From the income tax returns, we have access to the reported wage, paid tax, and employer-employee relationships that we use to identify public and private sector employees. We use the income tax returns for all the employees of the country that were reported as such by their employers or report their income tax return themselves. We exclude from our sample all individuals who have self-employed income for two reasons. First, the main focus of this study is to understand the extent to which envelope wages exist in a context where a third-party reporting system comprehensively covers the relationship between employer and employee. Second, including individuals with self-employed income would add noise to the estimation because self-employed can use all their declared expenses as a deduction with poor monitoring (Carrillo et al., 2017) and minimize their taxable income.

In addition, we use the tax registry to recover demographic characteristics, tenure in the job in months, and the employer's characteristics (e.g., public or private and firm size). Unfortunately, tax returns in Ecuador (as in many developing countries) do not include an address for the taxpayer. As a result, we do not have information about the canton where each individual lives, which may affect the employment

opportunities available to that person. However, since we know the location of each seller she purchases from, we assume each person lives in the canton where she purchased the most by dollar amount during the year.

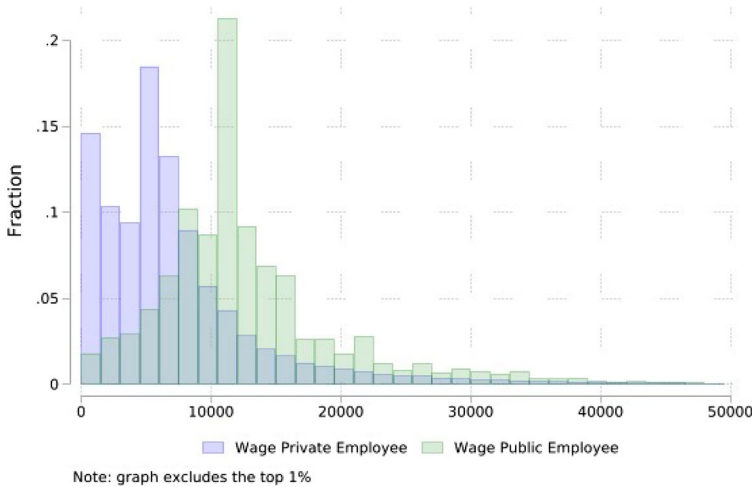


Fig. 1 Histogram of public and private sector employees wages

Table 1 Demographic characteristics of the estimation samples of employees

	(1) All mean/S.D.	(2) With food consumption information mean/S.D.	(3) With total consumption information mean/S.D.
Women	0.37 (0.48)	0.42 (0.49)	0.37 (0.48)
Married	0.40 (0.49)	0.43 (0.49)	0.40 (0.49)
Age in 2017	36.76 (11.54)	37.14 (10.99)	36.83 (11.48)
Finished high school	0.40 (0.49)	0.40 (0.49)	0.40 (0.49)
Finished college	0.36 (0.48)	0.46 (0.50)	0.37 (0.48)
Observations	2,761,600	1,798,204	2,706,427

Means are reported along with standard deviations, which are shown in parentheses. The categories *Finished high school* and *Finished college* do not overlap. *Finished high school* means that the employee’s highest level of education is a high school diploma, whereas *Finished college* indicates that the employee holds at least a bachelor’s degree. It should be noted that individuals who have graduated from college are not included in the ‘Finished high school’ category. We were unable to obtain the demographic characteristics of 1,260 individuals. In 2017, there were 2,762,860 public and private sector employees who reported wages and non-self-employed income and filed an income tax form. We were able to construct the variable of total consumption for 2,707,161 of these individuals, and the variable of food consumption for 1,798,517 of them

Table 2 Demographic characteristics of public and private sector employees

	(1) Public sector employees mean/S.D.	(2) Private sector employees mean/S.D.	(3) Both sector employees mean/S.D.
Women	0.46 (0.50)	0.33 (0.47)	0.42 (0.49)
Married	0.46 (0.50)	0.36 (0.48)	0.56 (0.50)
Age in 2017	39.21 (11.26)	35.36 (11.36)	41.95 (11.12)
Finished high school	0.31 (0.46)	0.45 (0.50)	0.21 (0.41)
Finished college	0.62 (0.48)	0.24 (0.42)	0.74 (0.44)
Observations	596,816	1,924,024	240,760

Means are reported along with standard deviations in parentheses. The *Finished high school* and *Finished college* categories do not overlap. *Finished high school* means that the highest degree the employee has is a high school diploma. *Finished college* means the employee holds at least a bachelor's degree. The college-graduated individuals are not included in the *Finished high school* category. We were not able to get the demographic characteristics of 1260 individuals

Our sample consists of all public and private sector employees with an income tax form for 2017 who were reported in the electronic billing system as buyers. Public sector employees are around 25% of the sample, and their annual wage is on average \$13,195. Private sector employees have a lower annual average wage of \$8,185 but a higher variance, as can be observed in the histogram of wages for each group of employees (Fig. 1). In 2017, there were 2,762,860 such employees. We can construct the variable of total consumption for 2,707,161 of them and the variable of food consumption for 1,798,517 of them (See Table 1 for descriptive statistics for each sample). The proportion of public sector employees that are women, married, and have finished college is larger than the proportion of private sector employees. The average age of public sector employees is also higher (See Table 2). We control for all of those demographic characteristics in our estimations.

4 Empirical strategy

To guide our empirical estimation, we consider a standard tax evasion model, where the taxpayer is an employee that decides to report their income. The probability of detection is affected by the presence of a third-party reporting mechanism like in Kleven et al. (2011). Still, the third-party reporting mechanism does not work perfectly and depends on the firm's size. We assume that the larger a firm, the more likely it is to report its withholdings to the tax authority correctly, and the less likely the rise of informal contracts where the employee is paid "envelope wages." Several rationales can support this assumption. Assume an individual is willing to

underreport income; there is no reason for him to choose to be an employee over being self-employed unless the former can provide a higher pay-off. Small firms might not pay as much as larger firms, but the employee might be able to underreport her income to keep a larger after-tax income (Barth and Ognedal, 2018). Alternatively, suppose part of the contract is an “envelope wage” that goes unreported. In that case, there is always the chance that some employee will be a whistle-blower, and that probability increases as the number of employees increases (Kleven et al., 2016; Barth and Ognedal, 2018). Under these assumptions, it is straightforward to realize that the marginal cost of evasion increases with firm size, but the marginal benefit is constant. Hence, employees of larger firms will evade to a lesser degree. A full model is in Section in the “Appendix.”

We estimate the income reporting gap between public and private sector employees using a methodology in the spirit of Pissarides and Weber. In a nutshell, the estimation consists of estimating a consumption expenditure equation based on the reported income and the individuals’ demographic characteristics. At the core of our estimation is the assumption that individuals with similar demographic characteristics and similar levels of real income have similar consumption patterns, particularly of food, independently of the source of their income (in this case, the wages reported by the public and private sectors). If we find differences in the relationship between consumption and income, those differences would be due to a misreporting of income. Public employers have no incentive to misreport their employees’ wages. All the taxpayers with self-employed income are excluded from our estimation because we want to explore the possibility of envelope wages in a context where a withholding system creates a third-party report for all wage earners. In addition, self-employed individuals have two extra margins to decrease their taxable income—underreport their self-employed gross income or over-report their self-employed expenses. In Ecuador, self-employed individuals do not pay income tax in a different schedule or have to fill out an extra tax form,⁶ they only have to report their gross income and expenses. The expenses are deductible from their taxable income in full. As shown by Carrillo et al. (2017), in the case of Ecuador, expenses are very costly to monitor, so self-employed individuals have a less risky evasion margin to use. In addition, there is no straightforward way to disentangle the business expenses of a self-employed person from their personal consumption. As a result, including self-employed individuals does not contribute to answering the main question of this study and can bias our estimation, so we exclude individuals with self-employed income.

We choose to use the electronic billing information to calculate the consumption proxies because there is no reason to think that public and private sector employees had incentives to select the store they consume from based on the availability of the electronic billing system. Also, 2017 was the first year all incorporated firms and a large portion of the non-incorporated firms were included in the electronic system. Therefore, it is unclear whether consumers would be perfectly informed which stores would report the transaction to the tax authority using the electronic billing system.

⁶ Unless their gross income is larger than \$300,000.

Although individuals could be maximizing their personal expenses deduction, the deduction is the same for all taxpayers; if anything, this behavior might introduce measurement error and the corresponding attenuation bias to our estimation.

The measurement error in consumption does not bias our estimation if it is uncorrelated with the categories of employees we are comparing—public and private sector employees. To address concerns that our consumption proxy could be correlated with employment type, we calculate the probability of being reported in the electronic billing system when buying food, based on the employment sector, demographic characteristics, and personal expenses deduction status. First, we note that individuals who reached the maximum deduction for food in the previous month are more likely to have been reported in the electronic billing system in the current month. In other words, those who consume larger amounts are more likely to appear in the electronic system. Being a private employee does not significantly affect the probability of being reported in the electronic billing system. This suggests that individuals do not systematically attempt to hide their consumption from tax authorities and may not even be aware that the system creates a third-party reported channel for consumption (see Table 3).

Another potential concern is if the type of store a particular group of workers frequents differs and is correlated with electronic billing system reporting. Specifically, if one group of workers is more likely to spend at chain stores, which report in the electronic billing system, this could bias our estimation. To address this concern, we calculate the probability of buying from a chain store and the amount purchased from chain stores for each individual using electronic billing data.

We define a chain store as a large taxpayer unit in the retail sector with more than the 90th percentile of the number of stores in that category (22 stores). We find that private sector employees are 2% less likely to buy from a chain store, but the total purchase amount is not statistically different (see Table 4). As a result, we do not believe that using purchases reported in the electronic billing system to calculate the consumption proxy introduces systematic bias into our estimation. We explore the validity of our assumption and possible biases in the following sections.

We follow the estimation proposed by Feldman and Slemrod (2007), in which they compared a metric of true income across individuals with different opportunities for evasion. In their case the metric of true income was charitable contributions, whereas in this work the metric is consumption. In their work, the opportunity for evasion had to do with whether the individual was required to file certain tax forms related to self-employment, whereas in our case we look at whether an individual is employed in the private or public sector. For our metric of true income, we constructed two variables of consumption—food consumption and total consumption—with information from the electronic billing system. For considering the opportunity of evasion, the estimation consists of a log-log estimation using a nonlinear procedure that allows us to include in the sample individuals who, during the same year, work for both private and public sector employers. We estimate the relationship between the log of consumption and the individual's real income. Real income has two components: visible and non-visible. The visible income is the one that cannot be underreported. We assume the wage of public sector employees is always visible income, and we allow for the wage of private sector employees to be non-visible. If

Table 3 Correlation between being reported on the e-billing system and reaching the maximum food deduction for private and public employees—LPM

	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Max. food ded	0.32*** (0.03)	0.33*** (0.04)	0.31*** (0.02)	0.33*** (0.02)	0.33*** (0.02)	0.31*** (0.02)	0.31*** (0.02)	0.29*** (0.02)	0.30*** (0.02)	0.30*** (0.02)	0.23*** (0.01)
Private employee	0.01 (0.01)	0.01 (0.01)	0.01** (0.01)	0.00 (0.00)	0.01** (0.00)	0.00 (0.00)	0.00 (0.00)	0.01 (0.00)	0.00 (0.00)	0.00 (0.00)	-0.01** (0.00)
Max. ded. × Private emp	-0.01 (0.06)	-0.01 (0.03)	-0.02 (0.02)	-0.03** (0.01)	-0.00 (0.01)	-0.01 (0.01)	-0.00 (0.01)	0.01** (0.01)	0.01*** (0.00)	0.00 (0.00)	0.02*** (0.00)
Canton of residency F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster Unit	Canton	Canton	Canton	Canton	Canton	Canton	Canton	Canton	Canton	Canton	Canton
Observations	2,102,371	2,102,371	2,102,371	2,102,371	2,102,371	2,102,371	2,102,371	2,102,371	2,102,371	2,102,371	2,102,371

The dependent variable is a binary variable equal to one if the taxpayer had food purchases reported on the e-billing system. The maximum food consumption is a binary variable equal to one if the taxpayer reached the maximum food deduction from income tax, equivalent to \$ 3670 for 2017, by the previous month. Taxpayers have been divided into two disjoint groups: public employees and private employees. Taxpayers who can fit in more than one category have been excluded because they have more evasion opportunities since they have different reporting margins. Demographic controls such as education, age, marital status and gender are included. Each column represents a model for a different month; the month is indicated on the column label. Canton of residency is described as the canton where the taxpayer had the largest portion of her total reported consumption. Standard errors are in parentheses and clustered at the canton level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 Purchases made at chain stores as reported in the E-billing system

	Extensive margin		Intensive margin	
	(1)	(2)	(3)	(4)
1 if private sector employee	-0.07*** (0.01)	-0.02** (0.01)	24.98 (33.35)	62.84 (40.74)
Wage	0.00*** (0.00)	0.00*** (0.00)	0.03*** (0.01)	0.02*** (0.01)
Woman		0.06*** (0.00)		68.51*** (17.38)
Married		0.02*** (0.00)		67.72*** (19.67)
Age		0.00** (0.00)		1.78 (1.36)
Finished high school		0.19*** (0.01)		110.70*** (31.48)
Finished college		0.26*** (0.01)		219.36*** (68.26)
Constant	0.67*** (0.02)	0.04 (0.02)	131.54*** (32.22)	-337.70** (152.30)
Canton of residency F.E	No	Yes	No	Yes
Controls	No	Yes	No	No
Observations	2,707,525	2,706,783	2,707,525	2,706,783

The dependent variable in the first two columns is a dichotomous variable, equal to one if the individual purchased from a chain store. The dependent variable in the next two columns is the dollar amount purchased from a chain store. A chain store is defined as a large taxpayer unit in the retail sector with more than the 90th percentile of the number of stores in that category (22 stores). The consumption variable is calculated using third-party reported purchases made by the taxpayer to a company that reports to the e-billing system. Wages in the public and private sectors are third-party reported by the employer, whether a government or private entity. All monetary values are in dollars. Canton of residence is defined as the canton where the taxpayer reported the largest portion of their total consumption. Standard errors are provided in parentheses and are clustered at the canton level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

income is not underreported, the real income will coincide with the reported income for both groups. In particular, we estimate the following using a nonlinear weighted least squares estimation:

$$\ln(C_i) = \beta_0 + \beta_1 \ln(V_i + kW_i + \rho_i S_i) + \gamma X_i + \mu_i \quad (1)$$

where C_i is total consumption or food consumption depending on the specification; V_i is the visible income, in this case, the wage of the public sector employees; W_i is the private sector wage; S_i is a dichotomous variable equal to one if the individual has a private sector wage; X_i is a vector of demographic characteristics such as age, level of education, marital status, gender, and canton of residency; and μ_i is the error term.

The null hypothesis is that k is equal to one. If k is similar to one, there would be no evidence of consumption differences between public and private sector employees, and there would not be underreporting of income (the wages of both public and private sector employees would be visible). If k were larger than one, the private sector employees would be underreporting their income compared with the public sector employees. There are some individuals that within the same fiscal year switch from a job in the public to the private sector (or vice versa), in which case the comparison is not across individuals but across sources of income like in the case of Feldman and Slemrod (2007)

If there were differences in the reporting of income between both groups of employees, then for each dollar that a public sector employee reported, the private sector would have reported $\frac{1}{k}$ dollars. A positive coefficient for S_i indicates that being a private sector employee has a positive income of ρ_i .

We run all estimations twice, once with food consumption and once with total consumption, where food consumption is our preferred estimation. We analyze subgroups to detect differences by firm size; our conceptual framework (See “Conceptual framework” Section in the “Appendix”) guides this later specification.

4.1 Limitation of the identifying assumption

The primary assumption of our estimation is that individuals with similar demographic characteristics and similar levels of real income exhibit comparable consumption patterns, particularly in relation to food, regardless of their income source. In this context, wages are reported by both public and private sectors. If this assumption holds, observed differences in the relationship between consumption and income would be attributable to a misreporting of income. In this section, we discuss the validity of this assumption using survey data and consider potential biases that might arise if the assumption does not hold.

We utilize the *Encuesta de Condiciones de Vida (ECV)* conducted by the *Instituto Ecuatoriano de Estadísticas y Censos*.⁷ The ECV of 2014 is the most recent survey providing information about employment and consumption. The ECV is representative at both national and provincial levels. The employment question is “What was

⁷ The equivalent of the Census Bureau in the USA.

your work last week?” with answer choices including public and private worker, self-employment, and non-paid laborer. Two additional questions ascertain formality: “Do you have a formal contract?” and “Are you affiliated with social security?”. The survey solicits information about main and secondary jobs, but the formality questions pertain only to the main occupation. Most of the consumption questions are household-based. The questionnaires gather information about the consumption of food of 111 items, how frequently they buy each items, the quantities they purchase each time adjusted for units of measure, and their total expenditure for each item. Using this information, we create a variable of monthly consumption of food. There is extra information about the non-food consumption on non-durable good that includes goods such as: newspapers, magazines, books, products for home and personal care, home services, entertainment, clothing and footwear. This variable is calculated by the *Instituto Ecuatoriano de Estadísticas y Censos*

Comparing the survey data with the administrative data is not straightforward. First, the administrative data lacks information about households; consequently, income and consumption information is available only at the individual level. Second, the tax returns data exclusively include information about formal occupations; thus, it is conceivable that some employees in the administrative data possess an additional informal source of income that cannot be observed but is partially discernible in the survey data (the formality questions pertain only to main occupation). We construct consumption variables at the household level and occupation variables for the household head. We conduct two exercises: we compare households with heads employed in the public and private sectors, and where possible we make comparisons at the individual level. We also attempt to reproduce our main result using the survey data.

We utilize the survey information to compare households with similar demographic characteristics as featured in our primary estimations. We conduct the following regression using a sample of individuals who hold a single formal job in either the public or private sector:

$$Y = \beta_0 + \beta_1 \text{Public}_i + \beta_2 \ln(\text{wage})_i + \gamma X_i + \epsilon_i \quad (2)$$

where Y can represent household consumption, composition or individual consumption depending on the sample, Public is a dichotomous variable equal to one if the individual is employed in the public sector, $\ln(\text{wage})$ is the natural logarithm of the nominal wage, and X encompasses demographic characteristics.

Looking at household information for public and private sector employees, we first find that household composition is similar for both; there is no significant difference in number of members or number of breadwinners. Second, non-food consumption is not statistically different between the two groups. Food consumption is 6% higher for public sector employees, which could indicate that public sector employees consume more than those in the private sector, suggesting that our estimation is a lower bound of the evasion gap. The results are presented in Table 5. In Fig. 2 is presented the average monthly food and non-food consumption by monthly wage percentile, visually there are no large differences in the consumption for this two groups at any level of income.

The same survey provides some information about personal expenses and time usage at the individual level. We calculate the expected value for each individual outcome using the same characteristics as before and repeat the calculations. The results are presented in Table 6.

Public sector employees spend more on food outside the household and on transportation. This also suggests a potential bias against our estimation. One might conjecture that the underreported income originates from a different source that is not reported. This additional income source would have to correlate with the sector of the main job to bias our estimation. The available survey data do not enable us to test for differences in additional sources of income. However, information on time usage can be used to examine potential differences in additional income sources. If an individual has a second job, this should detract from time available for other

Table 5 Differences in survey responses between households with heads working in the public and private sectors

	(1) Household size	(2) Number of breadwinners	(3) ln(Non-food consumption)	(4) ln(Food consumption)
1 if public sector employee	0.01 (0.06)	-0.03 (0.04)	0.06 (0.04)	0.06** (0.03)
ln(wage)	0.04 (0.05)	-0.09*** (0.03)	0.77*** (0.03)	0.15*** (0.03)
Woman	0.65*** (0.09)	0.26*** (0.07)	0.16** (0.08)	0.52*** (0.06)
Married	1.90*** (0.08)	0.66*** (0.06)	0.33*** (0.05)	0.87*** (0.05)
Age	0.01*** (0.00)	0.02*** (0.00)	0.00* (0.00)	0.01*** (0.00)
Finished high school	-0.16** (0.06)	-0.17*** (0.04)	0.20*** (0.04)	0.09*** (0.03)
Finished college	-0.31*** (0.07)	-0.10** (0.05)	0.48*** (0.05)	0.06 (0.04)
Constant	1.70*** (0.31)	1.27*** (0.21)	-0.42** (0.21)	2.99*** (0.17)
Province of residence FE	Yes	Yes	Yes	Yes
Population observations	611,073	611,073	1,081,604	1,072,273
Observations	3566	3566	6069	6021

The information is sourced from the ‘Encuesta de Condiciones de Vida’ (Consumption Survey) of 2014, conducted by the Instituto Ecuatoriano de Estadísticas y Censos. All monetary values are reported monthly and in USD. The dependent variable is listed in the column. ‘ln(Non-food consumption)’ and ‘ln(Food consumption)’, presented in the first two columns, are the natural logarithm of the consumption in USD. Meanwhile, the third and fourth columns, representing ‘Household size’ and ‘Number of breadwinners’, are reported in terms of individual counts. Standard errors are provided in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

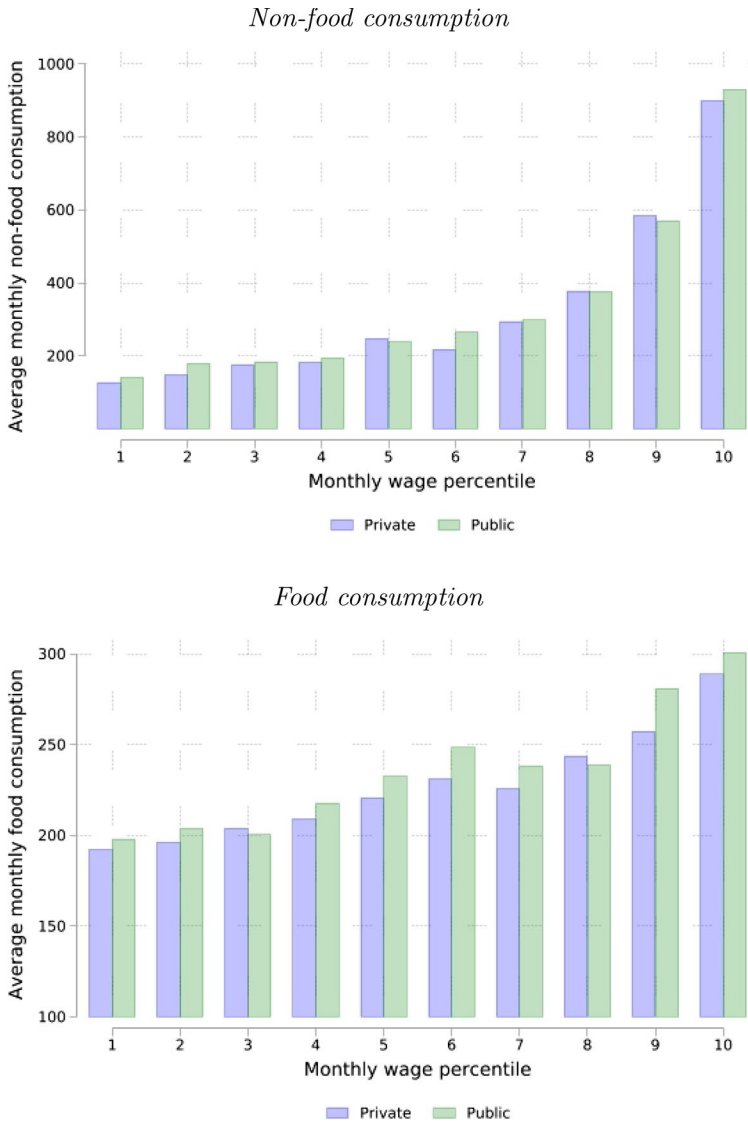


Fig. 2 Consumption by wage percentile between households with heads working in the public and private sectors

activities, which does not seem to be the case. Time usage information suggests that public sector employees spend more time on household chores and less time sleeping. These differences are minimal and do not suggest a systematic distinction between public and private employees' engagement in an additional occupation.

To reproduce our main results with the survey data, we restrict our sample to individuals who live alone (those with a household size of one). We do this to construct a dataset with income and consumption information at the individual level, as in the administrative data. We estimate Eq. (1), the only difference being

Table 6 Differences in survey responses between individuals working in the public and private sectors

	(1)	(2)	(3)	(4)	(5)
	ln(Food)	ln(Transportation)	ln(Entertainment)	Chores	Sleeping
I if public sector employee	0.16*** (0.04)	0.06* (0.03)	0.10 (0.11)	1.42*** (0.21)	-0.80*** (0.26)
ln(wage)	0.41*** (0.04)	0.30*** (0.03)	0.33*** (0.08)	-1.31*** (0.15)	-1.04*** (0.19)
Woman	-0.21*** (0.03)	-0.01 (0.03)	0.03 (0.10)	9.61*** (0.25)	-0.79*** (0.20)
Married	-0.19*** (0.03)	0.02 (0.03)	0.03 (0.10)	1.56*** (0.18)	-1.03*** (0.22)
Age	-0.00 (0.00)	-0.00** (0.00)	-0.01 (0.01)	0.07*** (0.01)	-0.02** (0.01)
Finished high school	0.16*** (0.04)	0.03 (0.03)	0.39** (0.15)	0.82*** (0.20)	-0.61** (0.25)
Finished college	0.34*** (0.05)	0.16*** (0.04)	0.50*** (0.16)	0.52** (0.25)	-0.77** (0.33)
Constant	1.15*** (0.27)	1.13*** (0.19)	1.58*** (0.44)	8.77*** (0.92)	63.10*** (1.28)
Province of residence FE	Yes	Yes	Yes	Yes	Yes
Population observations	1,720,630	1,358,049	170,854	2,127,275	2,127,275
Observations	9156	7149	692	11,540	11,540

The information is sourced from the ‘Encuesta de Condiciones de Vida’ (Consumption Survey) of 2014, conducted by the Instituto Ecuatoriano de Estadísticas y Censos. All monetary values are reported monthly and in USD. The dependent variable is listed in the column. The ‘ln(Food)’, ‘ln(Transportation)’, and ‘ln(Entertainment)’ displayed in the first three columns represent the natural logarithm of food consumption outside the household, transportation, and entertainment expenses, respectively, expressed in USD. Meanwhile, the final two columns indicate the number of hours spent on household chores and sleeping per week. Standard errors are provided in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the inclusion of the province of residence instead of the canton due to the information provided in the survey. The results are presented in Table 7. The first two columns pertain to food consumption, and the third and fourth to total consumption of non-durable goods as reported in the survey. The k is not statistically different from one in any case.

By using survey data, we do not observe substantial and systematic differences in consumption between public and private sector employees. The differences we note suggest that, at the same income level, public sector employees consume more than private sector employees. This implies that if the identifying assumption does not hold, the bias is against our estimation, indicating that we are measuring a lower bound of the evasion gap.

Table 7 Reported income compliance of private sector employees based on expenditures using survey information

	(1)	(2)	(3)	(4)
	ln(Food C.)	ln(Food C.)	ln(Total C.)	ln(Total C.)
β_1 : wage in the public sector (visible income)	1.06 (1.46)	1.89 (1.16)	-3.21* (1.31)	-2.19* (1.30)
k_1 : wage in the private sector	0.40 (0.32)	0.78*** (0.22)	0.47 (0.51)	0.98*** (0.35)
ρ_1 : private sector employee	270.60 (323.44)	165.70 (145.42)	2.64 (88.51)	220.65 (226.83)
Age			0.01** (0.01)	-0.01 (0.01)
Woman			0.50*** (0.15)	0.31** (0.15)
Finished high school			0.28 (0.24)	0.27 (0.21)
Finished college			0.16 (0.27)	0.73*** (0.22)
Sample	Everybody	Everybody	Everybody	Everybody
Wald test $H_0 : k = 1$				
p -value	0.16	0.39	0.38	0.96
Province of residency F.E	No	No	Yes	Yes
Controls	No	No	Yes	Yes
Sample obs.	290	290	290	290
Population obs.	46,368	46,368	46,368	46,368

The dependent variable in the first two columns is the natural logarithm of food expenditure, and in the last two columns, it is the total consumption. The data are sourced from the 2014 Encuesta de Condiciones de Vida (Living Conditions Survey) conducted by the Instituto Ecuatoriano de Estadísticas y Censos. The ECV collects employment information at the individual level and consumption data at the household level. This sample includes data from households with a size of one to ensure comparability with the administrative data. Standard errors are provided in parentheses. Both survey frequency weights and sample units were used in the analysis

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5 Results

This section presents and discusses our results for the whole sample. Guided by our conceptual framework, we present the analysis by firm size, comparing all public sector employees with private sector employees that work in firms of different sizes. Finally, we offer some robustness checks that verify whether marital status changes the estimation and if job stability plays a role in the different consumption patterns.

5.1 Main results

We estimate Eq. (1) using food consumption and total consumption (Tables 8 and 9 show the results, respectively). In general, our estimation shows that there is little

Table 8 Reported income compliance of private sector employees based on expenditures on food
 Dependent variable: $\ln(\text{food consumption})$

	(1)	(2)	(3)	(4)
β_1 : wage in the public sector (visible income)	0.96*** (0.04)	0.97*** (0.04)	0.87*** (0.02)	0.89*** (0.02)
k_1 : wage in the private sector	1.02*** (0.05)	1.04*** (0.06)	1.08*** (0.05)	1.10*** (0.05)
ρ_1 : private sector employee	1,875.29*** (142.14)	1,967.86*** (158.40)	2,169.14*** (166.14)	2,276.30*** (183.56)
Age			0.01*** (0.00)	0.01*** (0.00)
Woman			0.19*** (0.02)	0.20*** (0.02)
Married			0.17*** (0.01)	0.17*** (0.01)
Finished high school			0.34*** (0.05)	0.34*** (0.05)
Finished college			0.62*** (0.08)	0.61*** (0.08)
Ratio of private sector compli- ance			0.93	0.91
Wald test $H_0 : k = 1$				
<i>p</i> -value	0.69	0.51	0.11	0.05
Sample	Everybody	Single sector	Everybody	Single sector
Canton of residency F.E	No	No	Yes	Yes
Controls	No	No	Yes	Yes
Observations	1,798,517	1,742,676	1,798,204	1,742,537

The dependent variable is the natural logarithm of the expenditure on food reported to the e-billing system. The consumption variable is calculated using the third-party reported purchases made by the taxpayer to a company that reports to the e-billing system. The sample includes only individuals who purchased from a company with an e-billing system and worked in the formal sector as an employee. Public sector and private sector wages are third-party reported by the employer (government, or private entity). Columns (1) and (3) include all individuals. Columns (2) and (4) include only individuals who worked solely in the private or public sector (i.e., excluding individuals who worked in both). All monetary values are in dollars. Canton of residency is defined as the canton where the taxpayer had the largest portion of their total reported consumption. Standard errors are in parentheses and are clustered at the canton level. The number of employees that worked in the formal sector and did not report self-employment income was 2,762,860. We had information of food consumption for 1,798,517 of those individuals

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

to no evasion on reported wages on average when we consider all the sample of employees independently of the firm’s size. Using food consumption, we cannot reject the null hypothesis that k is equal to 1 (See Fig. 3). Using total consumption, we find a small gap: on average, for each dollar a public employee reports, a private employee reports 91 cents.

Table 9 Reported income compliance of private sector employees based on expenditures in all categories dependent variable: $\ln(\text{total consumption})$

	(1)	(2)	(3)	(4)
β_1 : wage in the public sector (Visible income)	1.28*** (0.03)	1.30*** (0.03)	0.99*** (0.02)	1.01*** (0.02)
k_1 : wage in the private sector	0.85*** (0.01)	0.86*** (0.01)	1.10*** (0.02)	1.11*** (0.02)
ρ_1 : private sector employee	1,138.76*** (101.67)	1,216.72*** (112.88)	1,192.51*** (154.43)	1,270.84*** (167.63)
Age			0.01*** (0.00)	0.01*** (0.00)
Woman			0.11** (0.04)	0.11*** (0.04)
Married			0.19*** (0.02)	0.19*** (0.02)
Finished high school			0.71*** (0.04)	0.71*** (0.04)
Finished college			1.20*** (0.05)	1.18*** (0.05)
Ratio of private sector compliance			0.91	0.90
Wald test $H_0 : k = 1$				
<i>p</i> -value	0.00	0.00	0.00	0.00
Sample	Everybody	Single sector	Everybody	Single sector
Canton of residency F.E	No	No	Yes	Yes
Controls	No	No	Yes	Yes
Observations	2,707,161	2,640,457	2,706,427	2,639,933

The dependent variable is the natural logarithm of the expenditure reported to the e-billing system. The consumption variable is calculated using the third-party reported purchases made by the taxpayer to a company that reports to the e-billing system. The sample includes only individuals who purchased from a company with an e-billing system and worked in the formal sector as an employee. Public sector and private sector wages are third-party reported by the employer (government, or private entity). Columns (1) and (3) include all individuals. Columns (2) and (4) include only individuals who worked solely in the private or public sector (i.e., excluding individuals who worked in both). All monetary values are in dollars. Canton of residency is defined as the canton where the taxpayer had the largest portion of their total reported consumption. Standard errors are in parentheses and are clustered at the canton level. The number of employees that worked in the formal sector and did not report self-employment income was 2,762,860. We had information of total consumption for 2,707,161 of those individuals

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To understand the importance of this gap, we follow Ekici and Besim (2016) to calculate the size of the shadow economy due to this intensive margin under-reporting. We calculate how much larger the country's gross domestic product (GDP) would be if all the wages were reported truthfully. In general, the shadow economy has three components: the economic activity of individuals who do not report any information to the government and are fully in the informal sector, the economic activity of self-employed individuals who are registered but hide some income from the government, and the wages that are paid to formal employees

Table 10 Implications for national accounts, tax and social security contribution gap

Row	Source/calc	Private sector employees		
		Food consumption	Total consumption	
1	Reported income (\$)	Section 5.1	8,185.53	8,185.53
2	Degree of under-reporting (%)	1-1/k	0.07	0.09
3	True income (\$)	R1/(1-R2%)	8,840.37	9,004.08
4	Unreported income (\$)	R3-R1	654.84	818.55
5	Registered contributors		2,235,000	2,235,000
6	Unreported income (\$)	R4*R5	1,463,572,764	1,829,465,955
7	Shadow economy—intensive margin employees (\$)	R6*(1/α)	2,975,005,723	3,718,757,153
8	Shadow economy—intensive margin employees % of GDP	R7/GDP	2.85%	3.57%
9	Tax losses		92,313,343	122,744,788
10	Tax losses % of total tax revenue		0.71%	0.95%
11	Unpaid social security contributions		301,495,989	376,869,987
12	Unpaid social security contributions % Total social security contributions		7.37%	9.22%

The calculation follows Ekici and Besim (2016). Comments to specific rows follows:

R1: Average annual salary reported to the tax authority for each group

R2: Estimates of *k* on Table 7 and 8, row 2 column 3

R5: There were 2,762,860 employees in the formal sector in 2017. We use 2,235,000 as a proxy of the registered contributors for this calculation because survey data suggests that this is approximately the number employed at any given time

R7: According to the national accounts by sector published by the Ecuadorian Central Bank <https://www.bce.fin.ec/index.php/informacioneconomica/sector-real> household gross disposable income is 69% of GDP. It is not possible to tell which part of that income corresponds to households with self-employed income, private employees' wages, or public employees' wages. However, if we assume the proportion of households corresponds to the proportion of registered taxpayers, private sector employees' gross disposable income is 49% of GDP. This is because employees are roughly 92% of the registered taxpayers of income tax, the other 8% have only self-employed income. Private employees are around 71% of registered payers of income tax

R8: The GDP of Ecuador in 2017 was 104,295,862 thousands of dollars

R9: We calculated the tax for each individual in our sample with their reported and calculated income. The tax losses are the aggregate of those differences

R10: Total tax collection Excluding import and export duties. in 2017 was 12,925,955 thousand dollars <https://www.sri.gob.ec/estadisticas-generales-de-recaudacion-sri>

R11: Social security contribution (similar to a payroll tax in the USA) is a flat rate of 20.6% over the salary. Employees pay 9.45% end employers 11.15% <https://www.iess.gob.ec/documents/10162/33703/C.D.+501>

R12: The total social security contribution in 2017 was 4,088,719 thousand dollars <https://www.iess.gob.ec/documents/10162/33703/C.D.+545>

that are not fully reported to the government (envelope wages). The last two components constitute quasi-formality. We calculate that the portion of the shadow economy generated by envelope wages is between 2% and 4% of GDP. To make

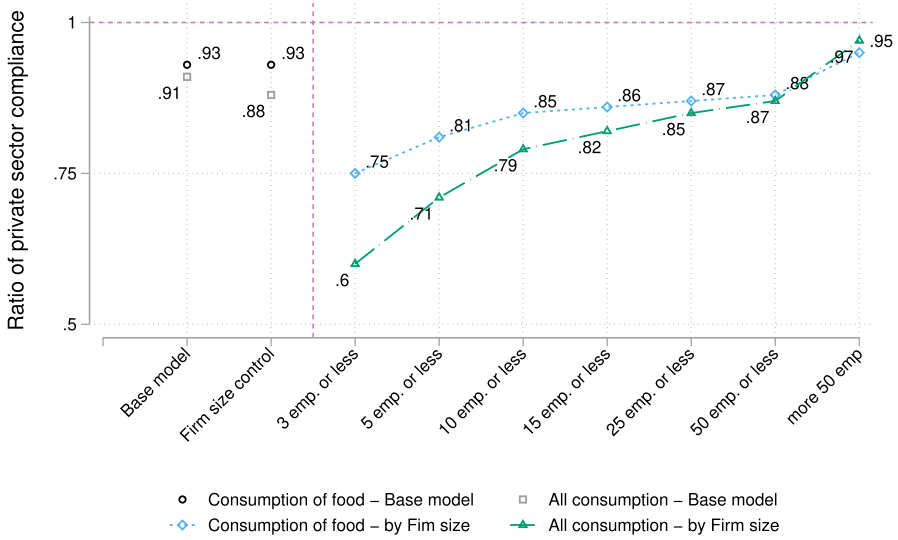


Fig. 3 Ratio of private sector compliance. The ratio of private sector compliance is $\frac{1}{k}$ from Eq. (1). A ratio of one means that there is no gap between the income reported by public and private sector employees. A ratio of 0.85 means that for every dollar a public sector employee reports, a private sector employee reports 85 cents

this calculation, we assume that every private employee would increase their income based on their underreported income gap (between a 7% - 9% increase). Using the national accounts, we estimate that the private sector employees' gross disposable income is 49% of GDP; keeping that proportion constant, we estimate how much larger the reported GDP would be if all the wages were reported.

Income tax in Ecuador is progressive, so instead of multiplying the underreported income by the average marginal tax rate as did Ekici and Besim (2016), we calculate the tax loss for each individual. In particular, we calculate the income tax with the reported income and with the calculated income considering the evasion gap. The tax loss is the difference between those calculations aggregated across all individuals. The income tax loss is between 0.7 and 1% of the total tax revenue. The social security contribution is a payroll tax with a flat tax rate. We calculate the unpaid contributions by multiplying the underreported income by the payroll tax rate. The unpaid contributions are fairly sizable and equivalent to over 7–9% of the total contributions. (All the calculations are presented and explained in Table 10)

In general terms, considering all the employees, the withholding system creates incentives to report income truthfully. However, interesting patterns arise in subgroup analysis by firm size.

5.2 Heterogeneous effects by firm size

There are several rationales to predict different levels of compliance for different firm sizes. Smaller firms might be less likely to have a dedicated accountant and navigate

Table 11 Robustness check: reported income compliance of private sector employees, controlling for firm size, based on expenditures on food Dependent variable: $\ln(\text{food consumption})$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β_1 : Wage in the public sector (visible income)	0.87*** (0.02)	0.88*** (0.02)	0.81*** (0.02)	0.80*** (0.02)	0.79*** (0.02)	0.80*** (0.02)	0.80*** (0.02)	0.81*** (0.03)	0.90*** (0.03)
k_1 : Wage in the private sector	1.08*** (0.05)	1.07*** (0.05)	1.34*** (0.07)	1.23*** (0.06)	1.18*** (0.06)	1.17*** (0.06)	1.15*** (0.06)	1.13*** (0.05)	1.05*** (0.04)
ρ_1 : Private sector employee	2,169.14*** (166.14)	2,176.91*** (164.08)	3,747.75*** (302.60)	3,381.39*** (277.30)	2,940.56*** (269.74)	2,755.44*** (251.31)	2,585.25*** (252.73)	2,465.11*** (226.05)	1,667.91*** (122.91)
Age	0.01*** (0.00)	0.01*** (0.00)	0.00 (0.00)	0.00* (0.00)	0.00*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.00 (0.00)
Woman	0.19*** (0.02)	0.19*** (0.02)	0.10*** (0.01)	0.10*** (0.01)	0.11*** (0.01)	0.12*** (0.01)	0.12*** (0.01)	0.14*** (0.02)	0.20*** (0.02)
Married	0.17*** (0.01)	0.17*** (0.01)	0.14*** (0.01)	0.15*** (0.01)	0.16*** (0.01)	0.16*** (0.01)	0.17*** (0.01)	0.17*** (0.01)	0.15*** (0.02)
Finished high school	0.34*** (0.05)	0.34*** (0.05)	0.23*** (0.04)	0.26*** (0.04)	0.30*** (0.04)	0.31*** (0.04)	0.32*** (0.04)	0.33*** (0.05)	0.32*** (0.05)
Finished college	0.62*** (0.08)	0.62*** (0.08)	0.54*** (0.05)	0.58*** (0.05)	0.61*** (0.06)	0.61*** (0.06)	0.62*** (0.06)	0.62*** (0.07)	0.58*** (0.08)
Firm size (# employees)		-0.00 (0.00)							
Ratio of private sector compliance	0.93	0.93	0.75	0.81	0.85	0.86	0.87	0.88	0.95
Wald test $H_0 : k = 1$									
p-value	0.11	0.16	0.00	0.00	0.00	0.01	0.01	0.01	0.21

Table 11 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public employ-ees	All	All	All	All	All	All	All	All	All
Private employ-ees	All	All	<i>Emp.</i> ≤ 3	<i>Emp.</i> ≤ 5	<i>Emp.</i> ≤ 10	<i>Emp.</i> ≤ 15	<i>Emp.</i> ≤ 25	<i>Emp.</i> ≤ 50	<i>Emp.</i> > 50
Canton of residency F.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,798,204	1,797,725	570,530	619,230	719,555	780,691	862,221	969,466	1,342,846

The dependent variable is the natural logarithm of the expenditure on food reported to the e-billing system. The consumption variable is calculated using the third-party reported purchases made by the taxpayer to a company that reports to the e-billing system. The sample includes only individuals who purchased from a company with an e-billing system and worked in the formal sector as an employee. Public sector and private sector wages are third-party reported by the employer (government, or private entity). The firm size is the number of employees reported in the year. Ecuador does not allow for the possibility of part-time workers, and all employees are counted irrespective of how much of the year they worked. All columns include all the public employees. Columns (3) to (8) include employees in private firms with less than 3, 5, 10, 15, 25, and 50 employees, respectively. Column (9) includes private employees who work for firms with at least 51 employees. All monetary values are in dollars. Canton of residency is defined as the canton where the taxpayer had the largest portion of their total reported consumption. Standard errors are in parentheses and are clustered at the canton level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12 Robustness check: reported income compliance of private sector employees, controlling for firm size, based on expenditures in all categories Dependent variable: $\ln(\text{total consumption})$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β_1 : wage in the public sector (visible income)	0.99*** (0.02)	0.99*** (0.02)	0.85*** (0.01)	0.83*** (0.01)	0.83*** (0.01)	0.84*** (0.02)	0.85*** (0.02)	0.87*** (0.02)	1.03*** (0.02)
k_1 : wage in the private sector	1.10*** (0.02)	1.13*** (0.02)	1.66*** (0.07)	1.40*** (0.06)	1.26*** (0.06)	1.22*** (0.05)	1.18*** (0.04)	1.15*** (0.04)	1.03*** (0.02)
ρ_1 : private sector employee	1,192.51*** (154.43)	1,213.93*** (159.18)	2,111.42*** (246.17)	1,786.01*** (209.93)	1,372.57*** (177.10)	1,267.72*** (169.73)	1,210.53*** (165.31)	1,186.87*** (166.78)	895.79*** (108.14)
Age	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.00*** (0.00)
Woman	0.11** (0.04)	0.11*** (0.04)	-0.10*** (0.03)	-0.09*** (0.03)	-0.06** (0.03)	-0.04 (0.03)	-0.02 (0.03)	0.01 (0.03)	0.10** (0.04)
Married	0.19*** (0.02)	0.19*** (0.02)	0.11*** (0.01)	0.13*** (0.01)	0.16*** (0.01)	0.17*** (0.02)	0.18*** (0.02)	0.18*** (0.02)	0.15*** (0.02)
Finished high school	0.71*** (0.04)	0.71*** (0.04)	0.63*** (0.02)	0.69*** (0.02)	0.74*** (0.02)	0.75*** (0.02)	0.75*** (0.02)	0.75*** (0.02)	0.67*** (0.05)
Finished college	1.20*** (0.05)	1.21*** (0.05)	1.02*** (0.03)	1.10*** (0.03)	1.19*** (0.03)	1.22*** (0.03)	1.23*** (0.03)	1.24*** (0.04)	1.09*** (0.06)
Firm size (# employees)		0.00*** (0.00)							
Ratio of private sector compliance	0.91	0.88	0.60	0.71	0.79	0.82	0.85	0.87	0.97
Wald test $H_0 : k = 1$									
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14

Table 12 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Public employ-ees	All	All	All	All	All	All	All	All	All
Private employ-ees	All	All	<i>Emp.</i> ≤ 3	<i>Emp.</i> ≤ 5	<i>Emp.</i> ≤ 10	<i>Emp.</i> ≤ 15	<i>Emp.</i> ≤ 25	<i>Emp.</i> ≤ 50	<i>Emp.</i> > 50
Canton of residency F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,706,427	2,705,818	760,143	841,086	1,010,489	1,112,700	1,247,831	1,423,025	1,956,502

The dependent variable is the natural logarithm of the expenditure reported to the e-billing system. The consumption variable is calculated using the third-party reported purchases made by the taxpayer to a company that reports to the e-billing system. The sample includes only individuals who purchased from a company with an e-billing system and worked in the formal sector as an employee. Public sector and private sector wages are third-party reported by the employer (government, or private entity). The firm size is the number of employees reported in the year. Ecuador does not allow for the possibility of part-time workers, and all employees are counted irrespective of how much of the year they worked. All columns include all the public employees. Columns (3) to (8) include employees in private firms with less than 3, 5, 10, 15, 25, and 50 employees, respectively. Column (9) includes private employees who work for firms with at least 51 employees. All monetary values are in dollars. Canton of residency is defined as the canton where the taxpayer had the largest portion of their total reported consumption. Standard errors are in parentheses and are clustered at the canton level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

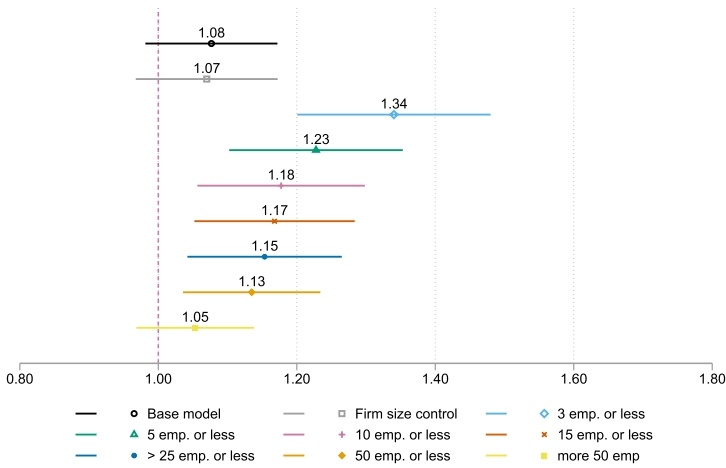


Fig. 4 k Coefficient for the estimation of the evasion gap between public and private sector employees using consumption of food. The k coefficient from Eq. (1) can be understood as the constant that the private wage will be multiplied by so the reported wage should be consistent with the food consumption and the pattern of consumption of the public sector employee. The null hypothesis is that k is equal to one. If k were equal to one, there would be no evidence of differences of consumption between public and private sector employees. When k is larger than one, the private sector employees of the corresponding firm size is underreporting their income as compared with the public sector employees. For each dollar that a public sector employee reports, the private sector reports $\frac{1}{k}$ dollars

the tax system correctly. Also, contracts that include envelope wages might be more difficult to keep confidential as the number of employees increases and the number of people who need to be coordinated is higher. These rationales apply in the private sector but not the public sector. Hence, we construct groups that include all public sector employees and only the employees of small firms. We create seven groups, including private sector employees in firms with up to 3 (i.e., 1 to 3 employees), up to 5, up to 10, up to 15, up to 25, up to 50 employees, and those with more than 50 employees (See Table 17). For each group, we estimate Eq. (1) using food consumption and total consumption (Tables 11 and 12, respectively, show the results). Using food consumption, we find a reporting gap ranging from 0.75 to 0.88. We find that the smaller the firm, the larger the gap is. For instance, for each dollar that public sector employees report, employees of firms with three or fewer employees report 75 cents. If we increase the sample to firms with 25 employees or less, the difference is 87 cents on the dollar. We do not find significant differences if we compare only employees of large firms (more than 50 employees). We find similar patterns with the total consumption estimation.

Notice that we find a gradient between the number of employees and the size of the evasion gap (See Fig. 4). So even if our identifying assumption does not hold (and the pattern of consumption of public and private sector employees are truly different), as long as those differences in consumption are not also correlated to the firm size, we can be confident that evasion is more likely at smaller firms (Fig. 5).

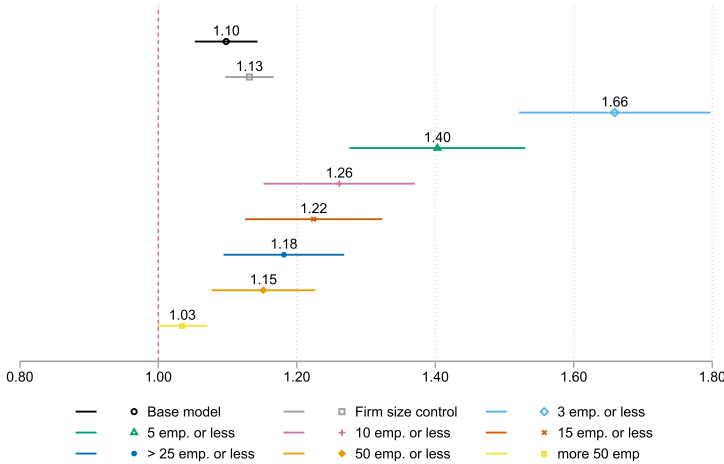


Fig. 5 k Coefficient for the estimation of the evasion gap between public and private sector employees using total consumption. The k coefficient from Eq. (1) can be understood as the constant that the private wage will be multiplied by so the reported wage should be consistent with the food consumption and the pattern of consumption of the public sector employee. The null hypothesis is that k is equal to one. If k were equal to one, there would be no evidence of differences of consumption between public and private sector employees. When k is larger than one, the private sector employees of the corresponding firm size is underreporting their income as compared with the public sector employees. For each dollar that a public sector employee reports, the private sector reports $\frac{1}{k}$ dollars

5.3 Robustness checks

In this subsection, we explore the shortcomings of our estimation. There is no joint filing in Ecuador, so it might be the case that the household’s primary breadwinner is different from the person who makes the household purchases. This could bias our estimation if there are systematic differences in the household composition of public and private sector employees. Also, there might be systematic differences in consumption for the same income level between public and private sector employees if their savings patterns are different depending on the relative stability in their sector.

5.3.1 Breadwinner versus primary spender

The Ecuadorian tax code does not allow joint filing, so each household income earner files their taxes independently. Imagine there is a household with two members, Chris and Pat. Chris earns the largest salary, but Pat makes all the household purchases. If that is the case, it would appear that Pat is overspending and Chris is saving. In principle, our estimation is unbiased as long as there is no correlation between this household consumption structure and being a public or private employee. Unfortunately, we do not have information on the household composition, nor can we identify the members of each household to construct income and consumption at a household level.

However, we can observe marital status,⁸ so to address this concern, we repeat the main estimation using only single individuals, and we do not observe differences from our main estimation. This result indicates that the household composition affects private and public sector employees similarly, so our estimation is not biased.

5.3.2 Job tenure

There is a possibility of systematic differences in consumption between public and private sector employees. The concern is that public sector jobs might be more stable; therefore, bureaucrats might have less precautionary savings than private sector employees because they are less worried about losing their jobs. In general, there is an expectation that public sector jobs might pay less but are more stable; if that were the case, our estimation would be biased downward because the consumption of public sector employees would be higher for all levels of income. We do not have a good way to construct a permanent income for each individual because we have access to consumption and tax return information for only one year; however, we can test the extent to which the job each individual has is stable, and compare private and public sector employees with the same tenure with the same employer.⁹ We are able to calculate the number of months each individual has been working for their current employer.¹⁰ The oldest reliable records we have access to go back to January 2005, so our maximum sample number of months is 156 (See Table 18). As the tenure increases, we observe that k becomes lower than one, indicating that public sector employees systematically consume a larger share than private sector employees. This is consistent with the idea that career bureaucrats will have very stable jobs. This means that the share of food consumption and total consumption for private sector employees should be larger than the proportion for public sector employees, not smaller; therefore, our estimation is conservative, and we estimate a lower bound of the underreporting of wages in the private sector.

6 Discussion and conclusion

This study analyzed the underreporting of income by private sector employees using a novel data source, electronic billing data on consumption matched to income tax records. The estimated underreporting of income is between 7 and 9 cents for each dollar of reported income of private sector employees. This result suggests that not only the self-employed exhibit underreported labor income and questions the prevalent practice of regarding tax authority income records of employees as the ‘gold standard.’ The estimated underreporting of private sector employee income translates to an estimated 3% of unregistered GDP from this source. For social security, underreporting has significant implications, reducing contributions by about 10 percent. Beyond the overall picture of underreporting, we detect substantial heterogeneities, notably a clear gradient of underreporting with respect to firm size. For example, in small firms of up

⁸ We get the marital status information for the National Registry. All civil marriages are performed and recorded by the National Registry.

⁹ We cannot tell if someone has held the same position, only if they have worked for the same employer.

¹⁰ If someone has more than one job, we take the tenure of the job with the largest wage.

to three employees, underreporting reaches 40 cents per dollar reported. A firm size gradient is in line with different risks and administrative costs of envelope wages in small versus large firms. Our results come from a middle-income country context with institutional weaknesses, but innovative uses of technology for tax administration and accounting that made this research possible.

The key assumption is that similar public and private sector employees have similar consumption patterns, particularly of food, independently of their source of income. Our robustness analyses suggest potential confounders, such as different household consumption structures or differential propensity to appear in the electronic billing system between public and private sector employees, are not biasing the results. If public sector employees have more stable jobs and consequently less precautionary savings than private sector employees, our underreporting estimates are biased downwards. The fact that the estimated reporting gap decreases with tenure is consistent with this notion.

The main limitation of our study is that we could not measure long-term income, because we used only one year of electronic billing information and tax return for data availability. However, our findings and methodology raise interesting policy questions and trade-offs. First, our data matching and methodology might enable tax and social security authorities to increase compliance and revenues. Second, given the underreporting gradient, it may seem as if tax authorities might want to audit more small businesses. However, the fixed costs of audits and small expected additional revenues from small firms put a limit on that implication. In fact, because small firms tend to be more economically vulnerable, non-enforcement of liabilities may be a cost-efficient way of flexibly supporting small businesses. Third, there are additional reasons for curbing envelope wages: full formalization has positive externalities and might also bring benefits to an individual small firm; enforcement might shift economic activity to more productive sectors and level the playing field among non-compliant and compliant firms, and envelope wages affect not only income tax, but also business tax and, importantly, social security and employee benefits.

Hence, neither the status quo of leaving envelope wages and quasi-informality unaddressed nor massive enforcement based on informative data likely constitute an optimal policy. An information campaign for small firms and their employees might be a more cost-effective strategy as smaller firms are unlikely to be fully aware of all tax regulations and how they can be compliant at limited administrative costs. In addition, disrupting incentives for paying envelope wages is key. The difference between the estimated income tax loss and the social security contribution loss, and experiences in other Latin American countries, suggests that changes in social security might be a better tool to decrease envelope wages. For example, in Mexico and Uruguay, underreporting of wages responded to changes in social security contributions and benefits (Bergolo and Cruces, 2014; Kumler et al., 2020).

In Ecuador, because retirement benefits are roughly calculated based on the five years with the highest contribution, employees do not have incentives to permanently receive their full wages on the books, especially at the beginning of their careers. A social security reform that links pensions more continuously to contributions could strengthen incentives to report wages truthfully. The positive effects of an incentive reform could be enhanced with a complementary information campaign.

Future research may investigate the effects of incentive reforms and information campaigns on envelope wages and related underreporting and evasion practices. Moreover, in light of advances in data availability and technology, the benefits and limitations of the methodology for measuring the shadow economy and tax and social security administration policy are interesting research areas.

Appendix

See Tables 13, 14, 15, 16, 17, and 18.

Table 13 Robustness check: reported income compliance of private sector employees—only single individuals based on expenditures on food Dependent variable: $\ln(\text{food consumption})$

	(1)	(2)	(3)	(4)
β_i : wage in the public sector (visible income)	0.93*** (0.05)	0.94*** (0.05)	0.86*** (0.03)	0.87*** (0.03)
k_i : wage in the private sector	0.98*** (0.04)	1.00*** (0.04)	1.07*** (0.05)	1.10*** (0.05)
ρ_i : private sector employee	1,645.20*** (127.14)	1,725.91*** (139.45)	1,909.20*** (145.36)	2,008.69*** (159.04)
Age			0.01*** (0.00)	0.01*** (0.00)
Woman			0.30*** (0.02)	0.30*** (0.02)
Finished high school			0.32*** (0.06)	0.32*** (0.06)
Finished college			0.57*** (0.09)	0.56*** (0.10)
Ratio of private sector compliance			0.93	0.91
Wald test $H_0 : k = 1$				
p -value	0.62	1	0.16	0.05
Sample	Everybody	Single sector	Everybody	Single sector
Canton of residency F.E	No	No	Yes	Yes
Controls	No	No	Yes	Yes
Observations	1,028,917	997,194	1,028,917	997,194

The dependent variable is the natural logarithm of the expenditure on food reported to the e-billing system. The consumption variable is calculated using the third-party reported purchases made by the taxpayer to a company that reports to the e-billing system. The sample includes only individuals who purchased from a company with an e-billing system and worked in the formal sector as an employee. Public sector and private sector wages are third-party reported by the employer (government, or private entity). Columns (1) and (3) include all individuals. Columns (2) and (4) include only individuals who worked solely in the private or public sector (i.e., excluding individuals who worked in both). All monetary values are in dollars. Canton of residency is defined as the canton where the taxpayer had the largest portion of their total reported consumption. Standard errors are in parentheses and are clustered at the canton level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14 Robustness check: reported income compliance of private sector employees - only single individuals based on expenditures in all categories Dependent variable: $\ln(\text{total consumption})$

	(1)	(2)	(3)	(4)
β_1 : Wage in the public sector (Visible income)	1.22*** (0.04)	1.25*** (0.04)	0.98*** (0.02)	1.00*** (0.02)
k_1 : Wage in the private sector	0.78*** (0.01)	0.80*** (0.01)	1.06*** (0.02)	1.07*** (0.02)
ρ_1 : Private sector employee	932.03*** (80.62)	1,000.76*** (91.62)	983.96*** (109.78)	1,053.59*** (120.52)
Age			0.01*** (0.00)	0.01*** (0.00)
Woman			0.23*** (0.03)	0.24*** (0.03)
Finished high school			0.68*** (0.05)	0.68*** (0.05)
Finished college			1.17*** (0.06)	1.16*** (0.06)
Ratio of private sector compliance			0.95	0.93
Wald test $H_0 : k = 1$				
p -value	0.00	0.00	0.00	0.00
Sample	Everybody	Single sector	Everybody	Single sector
Canton of residency F.E	No	No	Yes	Yes
Controls	No	No	Yes	Yes
Observations	1,614,325	1,575,931	1,614,325	1,575,931

The dependent variable is the natural logarithm of the expenditure reported to the e-billing system. The consumption variable is calculated using the third-party reported purchases made by the taxpayer to a company that reports to the e-billing system. The sample includes only individuals who purchased from a company with an e-billing system and worked in the formal sector as an employee. Public sector and private sector wages are third-party reported by the employer (government, or private entity). Columns (1) and (3) include all individuals. Columns (2) and (4) include only individuals who worked solely in the private or public sector (i.e., excluding individuals who worked in both). All monetary values are in dollars. Canton of residency is defined as the canton where the taxpayer had the largest portion of their total reported consumption. Standard errors are in parentheses and are clustered at the canton level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15 Robustness check: reported income compliance of private sector employees, controlling for tenure on the job, based on expenditures on food Dependent variable: $\ln(\text{food consumption})$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
β_1 : wage in the public sector (visible income)	0.87*** (0.02)	0.89*** (0.02)	0.84*** (0.03)	1.05*** (0.02)	1.04*** (0.02)	1.04*** (0.02)	1.04*** (0.02)
λ_1 : wage in the private sector	1.08*** (0.05)	1.03*** (0.05)	1.08*** (0.06)	0.84*** (0.05)	0.85*** (0.05)	0.85*** (0.05)	0.86*** (0.05)
ρ_1 : private sector employee	2,169.14*** (166.14)	2,377.94*** (148.28)	2,065.03*** (174.23)	3,841.63*** (190.38)	3,874.74*** (224.10)	3,828.45*** (253.52)	3,766.56*** (280.38)
Age	0.01*** (0.00)	0.00*** (0.00)	0.01*** (0.00)	0.00*** (0.00)	0.00*** (0.00)	0.00*** (0.00)	0.00*** (0.00)
Woman	0.19*** (0.02)	0.19*** (0.02)	0.18*** (0.02)	0.21*** (0.02)	0.21*** (0.03)	0.21*** (0.03)	0.21*** (0.03)
Married	0.17*** (0.01)	0.16*** (0.01)	0.18*** (0.01)	0.15*** (0.02)	0.15*** (0.02)	0.14*** (0.02)	0.14*** (0.02)
Finished high school	0.34*** (0.05)	0.35*** (0.05)	0.33*** (0.05)	0.37*** (0.06)	0.39*** (0.05)	0.39*** (0.05)	0.39*** (0.05)
Finished college	0.62*** (0.08)	0.62*** (0.09)	0.60*** (0.09)	0.65*** (0.10)	0.67*** (0.10)	0.68*** (0.09)	0.67*** (0.09)
Job tenure (months)		0.00* (0.00)					
Ratio of private sector compliance	0.93	0.97	0.93	1.19	1.18	1.17	1.16
Wald test $H_0 : k = 1$							
p-value	0.11	0.55	0.18	0.00	0.00	0.00	0.01

Table 15 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Restriction by job tenure	None	None	$0 < m \leq 48$	$48 < m \leq 72$	$48 < m \leq 96$	$48 < m \leq 120$	$48 < m \leq 144$
Canton of residency F.I.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,798,204	1,712,210	1,001,430	234,656	395,804	510,445	573,229

The dependent variable is the natural logarithm of the expenditure on food reported to the e-billing system. The consumption variable is calculated using the third-party reported purchases made by the taxpayer to a company that reports to the e-billing system. The sample includes only individuals who purchased from a company with an e-billing system and worked in the formal sector as an employee. Public sector and private sector wages are third-party reported by the employer (government or private entity). Job tenure is the number of months that the employee worked in the job reported in 2017, going back to 2005 (maximum number of months: 156). If the person has more than one job, the longest tenure is considered. All monetary values are in dollars. Canton of residency is defined as the canton where the taxpayer had the largest portion of their total reported consumption. Standard errors are in parentheses and are clustered at the canton level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16 Robustness check: reported income compliance of private sector employees, controlling for tenure on the job, based on expenditures in all categories Dependent variable: $\ln(\text{total consumption})$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
β_1 : wage in the public sector (visible income)	0.99*** (0.02)	1.00*** (0.02)	0.90*** (0.02)	1.17*** (0.02)	1.16*** (0.02)	1.18*** (0.02)	1.18*** (0.02)
λ_1 : wage in the private sector	1.10*** (0.02)	1.09*** (0.02)	1.17*** (0.03)	0.93*** (0.04)	0.93*** (0.04)	0.92*** (0.03)	0.92*** (0.03)
ρ_1 : private sector employee	1,192.51*** (154.43)	1,449.75*** (190.72)	1,119.70*** (165.30)	2,484.86*** (289.59)	2,427.90*** (302.28)	2,424.26*** (298.02)	2,442.36*** (308.63)
Age	0.01*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.00*** (0.00)	0.00*** (0.00)	0.00*** (0.00)	0.00*** (0.00)
Woman	0.11** (0.04)	0.11*** (0.04)	0.11*** (0.04)	0.12*** (0.04)	0.13*** (0.04)	0.14*** (0.05)	0.14*** (0.05)
Married	0.19*** (0.02)	0.17*** (0.02)	0.17*** (0.02)	0.17*** (0.02)	0.18*** (0.02)	0.18*** (0.02)	0.18*** (0.02)
Finished high school	0.71*** (0.04)	0.71*** (0.04)	0.66*** (0.04)	0.77*** (0.04)	0.79*** (0.03)	0.80*** (0.04)	0.80*** (0.04)
Finished college	1.20*** (0.05)	1.22*** (0.05)	1.21*** (0.06)	1.25*** (0.05)	1.28*** (0.05)	1.29*** (0.05)	1.28*** (0.05)
Job tenure (months)		0.00*** (0.00)					
Ratio of private sector compliance	0.91	0.91	0.86	1.07	1.08	1.09	1.09
Wald test $H_0 : k = 1$							
p-value	0.00	0.00	0.00	0.08	0.08	0.01	0.01

Table 16 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Restriction by job tenure	None	None	$0 < m \leq 48$	$48 < m \leq 72$	$48 < m \leq 96$	$48 < m \leq 120$	$48 < m \leq 144$
Canton of residency F.I.E	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,706,427	2,579,673	1,580,194	330,938	561,578	728,791	814,913

The dependent variable is the natural logarithm of the expenditure reported to the e-billing system. The consumption variable is calculated using the third-party reported purchases made by the taxpayer to a company that reports to the e-billing system. The sample includes only individuals who purchased from a company with an e-billing system and worked in the formal sector as an employee. Public sector and private sector wages are third-party reported by the employer (government or private entity). Job tenure is the number of months that the employee worked in the job reported in 2017, going back to 2005 (maximum number of months: 156). If the person has more than one job, the longest tenure is considered. All monetary values are in dollars. Canton of residency is defined as the canton where the taxpayer had the largest portion of their total reported consumption. Standard errors are in parentheses and are clustered at the canton level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17 Number of employees

	Mean	P25	P50	P75	P90	Obs
Public firms	271.92	8	22	183	568	2,699
Private firms	9.90	1	2	5	12	326,457
All firms	12.04	1	2	5	12	329,156

The number of employees is the total number of employees reported in 2007. If a person had more than one job, she is counted by all of their employers. The first column shows the mean; the second to the fifth the 0.25, 0.50, 0.75, and 0.9 percentiles, respectively; and the last column the number of firms in that category

Table 18 Job tenure in 2017 going back to 2005

	Mean	P25	P50	P75	P90	Obs
Public employees	52.45	23	41	64	131	579,731
Private employees	50.48	13	38	77	120	2,198,629
All employees	50.89	15	40	75	121	2,778,360

Job tenure is the number of months worked in the job, reported by the employee in 2017 and going back to 2005. If a person had more than one job, the one with largest salary is considered. The first column shows the mean; the second to the fifth the 0.25, 0.50, 0.75 and 0.9 percentiles, respectively; and the last column the number of observations on that category

Conceptual framework

We consider a standard tax evasion model similar to the version of the Allingham and Sandmo model presented by Kleven et al. (2011) that includes third-party reporting in the structure of the probability of detection. We use the model to guide the intuition of our estimation. Assume that the taxpayer is risk neutral and she has a real income of \bar{y} and a reported income of y ; the underreported income e is the difference between them ($e \equiv \bar{y} - y$). The probability of detection increases based on evasion (e) and on firm size (κ); $p = p(e, \kappa)$, where $\frac{\partial p(e, \kappa)}{\partial e} > 0$, $\frac{\partial p(e, \kappa)}{\partial \kappa} > 0$ and $\frac{\partial^2 p(e, \kappa)}{\partial e \partial \kappa} > 0$. The larger a firm, the more likely it is to report its withholdings to the tax authority correctly, and the less likely the rise of informal contracts where the employee is paid “envelope wages.” Several rationales can support this assumption. Assume an individual is willing to underreport income; there is no reason for him to choose to be an employee over being self-employed unless the former can provide a higher pay-off. When larger firms are more productive and provide a productivity boost to their employees, there might be a trade-off between productivity and the opportunity for evasion. Some individuals may choose to work for a larger and more productive firm even when there are no opportunities to evade; others might choose to be in a less productive, smaller operation but compensate for part of their loss of productivity by underreporting their income (Barth and Ognedal, 2018). Alternatively, if part of the contract is an “envelope wage” that goes unreported, there is always the chance that some employee will be a whistle-blower.

The probability of someone reporting the envelope wage will increase with the amount of underreported work (Kleven et al., 2016; Barth and Ognedal, 2018). We assume a linear tax (t) levied on the reported income. The penalty for evading is proportional to the evaded tax and is given by $\theta > 1$. The risk-neutral taxpayer chooses the level of evasion e to maximize:

$$\mathbb{E}(U) = \underbrace{[1 - p(e, \kappa)] [(\bar{y})(1 - t) + et]}_{\text{Expected return of not being audited}} + \underbrace{p(e, \kappa) [(\bar{y})(1 - t) - e\theta t]}_{\text{Expected return of being audited}} \quad (3)$$

The corresponding first-order condition after some manipulation is:

$$\left[p(e, \kappa) + \frac{\partial p(e, \kappa)}{\partial e} e \right] [1 + \theta] = 1 \quad (4)$$

The left-hand side of Eq. (4) represents the marginal cost for hiding an extra dollar, and the right-hand side is the marginal benefit. Since $\frac{\partial p(e, \kappa)}{\partial \kappa} > 0$ and $\frac{\partial^2 p(e, \kappa)}{\partial e \partial \kappa} > 0$, it is easy to see that the marginal cost increases with firm size but the marginal benefit is constant; hence, employees of larger firms will evade to a lesser degree.

Acknowledgements We are grateful to the Research Center for Fiscal Studies of the Ecuadorian Tax Authority staff for their outstanding collaboration. We are thankful for the valuable comments from participants at the 2022 Annual Congress of the International Institute of Public Finance. Our appreciation also extends to the Scientific Colombia Program-EFI Alliance for promoting programs and policies for a formal economy. The views and opinions presented herein are solely those of the authors and do not represent the official positions of the institutions they are affiliated with.

Funding Open Access funding provided by Colombia Consortium.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3–4), 323–338.
- Artavanis, N., Morse, A., & Tsoutsoura, M. (2016). Measuring income tax evasion using bank credit: Evidence from Greece. *Quarterly Journal of Economics*, 131(2), 739–798.
- Barth, E., & Ognedal, T. (2018). Tax Evasion in Firms. *LABOUR*, 32(1), 23–44.
- Barth, E., & Ognedal, T. (2018). Tax Evasion in Firms. *Labour*, 32(1), 23–44.
- Bergolo, M., & Cruces, G. (2014). Work and tax evasion incentive effects of social insurance programs: Evidence from an employment-based benefit extension. *Journal of Public Economics*, 117, 211–228.

- Cabral, A. C. G., Gemmell, N., & Alinaghi, N. (2021). Are survey-based selfemployment income under-reporting estimates biased? New evidence from matched register and survey data. *International Tax and Public Finance*, 28(2), 284–322.
- Cabral, A. C. G., Myles, G., & Kotsogiannis, C. (2015). Self-employment underreporting in Great Britain: Who and how much? en. In 14th journées LouisAndré Gérard-Varet.
- Carrillo, P., Pomeranz, D., & Singhal, M. (2017). Dodging the Taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*, 9(2), 144–164.
- Di Nola, A., Kocharkov, G., & Vasilev, A. (2019). Envelope wages, hidden production and labor productivity. *The BE Journal of Macroeconomics*, 19(2), 1–30.
- Domínguez-Barrero, F., López-Laborda, J., & Rodrigo-Sauco, F. (2017). Tax evasion in Spanish personal income tax by income sources, 2005–2008: From the synthetic to the dual tax. *European Journal of Law and Economics*, 44(1), 47–65.
- Ekici, T., & Besim, M. (2016). A measure of the shadow economy in a small economy: Evidence from household-level expenditure patterns. *Review of Income and Wealth*, 62(1), 145–160.
- Engström, P., & Holmlund, B. (2009). Tax evasion and self-employment in a high-tax country: Evidence from Sweden. *Applied Economics*, 41(19), 2419–2430.
- Feldman, N. E., & Slemrod, J. (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *The Economic Journal*, 117(518), 327–352.
- Dunbar, G. R., & Fu, C. (2015). Sheltered income: Estimating income under-reporting in Canada, 1998 and 2004. In Staff Working Papers. Bank of Canada.
- Gorodnichenko, Y., Martínez-Vázquez, J., & Peter Sabirianova, K. (2009). Myth and reality of flat tax reform: Micro estimates of tax evasion response and welfare effects in Russia. *Journal of Political Economy*, 117(3), 504–554.
- Instituto Ecuatoriano de Estadísticas y Censos. (2017). Reporte de Economía Laboral. Tech. rep. Quito: INEC.
- International Labour Office. (2013). The measurement of informality: Statistical manual on the informal sector.
- Jensen, A. (2022). Employment structure and the rise of the modern tax system. *American Economic Review*, 112(1), 213–34.
- Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–692.
- Kleven, H. J., Kreiner, C. T., & Saez, E. (2016). Why can modern governments tax so much? An agency model of firms as fiscal intermediaries. *Economica*, 83(330), 219–246.
- Kumler, T., Verhoogen, E., & Frias, J. (2020). Enlisting employees in improving payroll tax compliance: Evidence from Mexico. *Review of Economics and Statistics*, 102(5), 881–896.
- Ohnsorge, F., & Yu, S. (2021). The long shadow of informality: Challenges and policies. Tech. rep.
- Paulus, A. (2015). Income underreporting based on income-expenditure gaps: Survey versus tax records. ISER Working Paper Series 2015-15. Institute for Social and Economic Research.
- Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39(1), 17–32.
- Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, 57(4), 904–54.
- Torregrosa-Hetland, S. (2020). Inequality in tax evasion: The case of the Spanish income tax. *Applied Economic Analysis*, 28(83), 89–109.
- Williams, C. C., & Horodnic, I. A. (2017). Evaluating the illegal employer practice of under-reporting employees' salaries. *British Journal of Industrial Relations*, 55(1), 83–111.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.