

# Tax evasion and productivity

Hans Martinez\*

Department of Economics, University of Western Ontario

August 1, 2023

## Abstract

Corporate tax evasion through cost overreporting spreads internationally causing governments significant tax revenue losses. Detecting and measuring the magnitude of tax evasion remains a challenge, even for the few studies on overreporting where researchers can exploit administrative data. Moreover, if this evasion strategy accounts for economic losses as large as reported, then cost overreporting might bias estimates of production functions, especially productivity. This paper addresses both issues. I first provide a novel strategy to estimate cost overreporting using commonly available firm-level data. I then formally show that ignoring cost overreporting leads to downward biased productivity estimates. Finally, I demonstrate how to recover productivity in the presence of tax evasion.

*Keywords:* Tax Evasion, Cost Overreporting, Production Function Estimation, Productivity

---

\*email: hmarti33@uwo.ca. I thank my supervisors Salvador Navarro, David Rivers, and Victor Aguiar for their guidance.

# Introduction

Corporate tax evasion through cost overreporting globally spreads causing governments significant tax revenue losses. Detecting and measuring the magnitude of cost overreporting remains a challenge, even for governments with administrative data. Moreover, if this evasion strategy accounts for economic losses as large as reported, then cost overreporting might bias estimates of production functions and, especially, productivity. This paper addresses both issues. I first provide a novel strategy to estimate cost overreporting using commonly available firm-level data. I then formally show that ignoring cost overreporting leads to downward biased productivity estimates. Finally, I demonstrate how to recover productivity in the presence of tax evasion.

Cost overreporting arises when firms acquire false invoices to claim additional tax deductions on value-added and corporate income taxes. According to the OECD’s document *Technology Tools to Tackle Tax Evasion and Tax Fraud* (2017), cost overreporting — also known as “fake invoicing”, “ghost firms”, “invoice mills”, or “missing traders”— permeates internationally reaching Latin America, Eastern Europe, Asia, and Africa. Reports claim cost overreporting generated annual tax revenue losses amounting to 5.6% of Poland’s GDP in 2016 (Poland’s Minister of Finance, 2018); 0.2% of Chile’s GDP in 2004 (Gonzalez and Velasquez, 2013; Jorrat, 2001; CIAT, 2008); 0.2% of Colombia’s GDP (Portafolio, 2019); and 0.03% of Mexico’s GDP in 2018 (Senado de la Republica, 2019).

Recently, evidence from Ecuador (Carrillo et al. 2022) shows that cost overreporting disseminates across all types of firms; constitutes a quantitatively large share of firms’ total deductions; and occurs mostly in the intermediate inputs. Contrary to the literature consensus, evasion by overreporting is not limited to small, semi-formal firms, but large and

formal firms overreport too. Likewise, firms were found to overreport up to 14.1% of the value of their purchase deductions. Finally, when challenged by the tax authority, firms adjusted mostly intermediate inputs.

The evidence from Ecuador also shows that big firms do not overreport inputs. This is unsurprising for several reasons. First, a large firm arguably draws more attention from the tax authority. Given its limited resources, the government optimizes its expected revenues by targeting the firms with the higher potential tax recovery, the few big ones. Second, the cost of being caught cheating is potentially higher for big firms. Large firms are likely to participate in international markets. A tax evasion scandal in Colombia, for example, might affect US sales. Finally, big firms potentially have more sophisticated strategies for tax evasion (e.g., profit shifting [Bustos et al. 2022](#)).

Despite its relevance, cost overreporting has been mostly overlooked by the literature. On one hand, the few studies on this evasion strategy exploit detailed administrative data ([Zumaya et al. 2021](#), [Carrillo et al. 2022](#)). Government tax authorities restrict access to administrative data because of firms' confidentiality concerns. On the other hand, to the best of my knowledge, no study has attempted to structurally identify cost overreporting. Unlike the case of individuals ([Pissarides & Weber 1989](#), [Paulus 2015](#)), when it comes to corporate tax evasion, researchers have to account for an additional source of unobserved heterogeneity, productivity. Why? Because cost overreporting might be naively quantified as low productivity. Intuitively, for a given output level, high input utilization by a firm could be explained by either the amount of input the firm overreports to evade taxes or by a low productivity shock.

To address this gap in the literature, first I formally show that ignoring tax evasion leads to downwardly biased productivity estimates. I then provide a new estimation strategy,

requiring only commonly available firm-level data, to jointly recover the densities of tax evasion and productivity. The intuition works as follows. In the absence of tax evasion, the first-order conditions of the firms' cost-minimization problem inform about a common technology, the production function. Consequently, in the presence of cost overreporting, deviations from this common technology identify tax evasion up to the current-period output shock. Then, from the subset of non-overreporting firms, the strategy identifies the production function parameters and the density of the output shock. Finally, using deconvolution techniques, I can jointly recover the distributions of tax evasion and productivity.

In a preliminary application, I show how to use the method to learn about tax evasion even when we do not know the subset of non-evading firms, but there has been a change in fiscal policy that incentivizes cost overreporting. Using firm-level data from Colombia between 1981 and 1991, I show that tax evasion through cost overreporting increased after the fiscal reform of 1983 between 8 and 9% in 1985 and 1986. This result stands at odds with previous studies that indicate that the evasion of income tax and VAT declined during this period (Sanchez & Gutierrez, 1994)

# **1 A parsimonious model of tax evasion through input overreporting**

Price-taking firms maximize expected after-tax profits. Firms choose input  $M_{it}$  and tax evasion  $e_{it}$  given the prices —output and input prices  $\{P_t, \rho_t\}$ —, and expected output —firms take the expectation of the production function and their current period productivity shock over the output shock  $\varepsilon_{it}$ , i.e.,

$$\mathbb{E}[Y_{it}] = \mathbb{E}[G(M_{it}) \exp(\omega_{it} + \varepsilon_{it})] \quad (1)$$

In their optimization problem, firms also consider the profit tax  $\tau$ , the penalty for evading taxes  $\kappa$ , and their perceived probability of detection  $q_{it}(e_{it}|s_{it})$ .

Firms, then, solve Equation 2

$$\begin{aligned} \max_{M_{it}, e_{it} \in [0, \infty)} & [1 - q_{it}(e_{it}|s_{it})] [(P_t \mathbb{E}[Y_{it}] - \rho_t M_{it}) - \tau (P_t \mathbb{E}[Y_{it}] - \rho_t (M_{it} + e_{it}))] \\ & + q_{it}(e_{it}|s_{it}) [(1 - \tau)(P_t \mathbb{E}[Y_{it}] - \rho_t M_{it}) - \kappa] \\ \text{s.t. } & Y_{it} = G(M_{it}) \exp(\omega_{it} + \varepsilon_{it}) \end{aligned} \quad (2)$$

Intuitively, for a given size, firms that evade more are more likely to get caught. Formally, the firm's  $i$  perceived probability of detection  $q_{it}(e_{it}|s_{it})$  is monotonically increasing in the amount evaded  $e_{it}$ , conditional on the size of the firm  $s_{it}$ . Likewise, larger firms might be more likely to get caught. That is, for a given level of tax evasion  $e_0$  and two different sizes  $s' > s \in \mathcal{S}$ , then  $q(e_0|s') \geq q(e_0|s)$ , where  $s$  can be continuous or discrete for a range of expected revenues.

Expected revenue determines the size of firms. The fiscal authority in the countries I examined in the empirical application defines the tax rules in terms of firms' revenue, i.e., Mexico, Colombia, Ecuador, and Chile (?). In these countries, firm revenues represent the main input for VAT and corporate profit taxes.

Because size  $s$  is measured in expected revenues, a function of inputs, firms affect their perceived probability of detection in two ways. Firms affect the probability  $q(e|s)$  directly, by choosing how much they evade  $e_{it}$ , and indirectly, by their implicit choice of size when choosing inputs  $M_{it}$ .

The firms' problem can be rewritten as follows,

$$\begin{aligned} \max_{M_{it}, e_{it}} \quad & \mathbb{E}[\pi|s_{it}] = (1 - \tau)(\mathbb{E}[Y_{it}] - \frac{\rho_t}{P_t}M_{it}) + [1 - q_{it}(e_{it}|s_{it})](\frac{\rho_t}{P_t}e_{it}\tau) - q_{it}(e_{it}|s_{it})\kappa \\ \text{s.t.} \quad & Y_{it} = G(M_{it})\exp(\omega_{it} + \varepsilon_{it}) \end{aligned} \quad (3)$$

Intuitively, if the firm overreports her inputs' cost, she will get the share of the value she overreported with probability  $(1 - q)$  and she will be penalized with probability  $q$ .

The optimal decision of the firm will depend on the fiscal environment  $\{\tau, \kappa\}$  and its perceived probability of detection.

### 1.1 Case 1: $q(e|s) = q(e)$

Consider the case when the probability of detection is independent of size,  $q(e|s) = q(e)$ .

This could be the case if, given a productivity shock, the firm perceives that the probability of detection is independent of her size within the range of its maximum optimal size.

The optimal input choice is independent, hence, of the evasion decision. In this case, the first-order conditions of Equation 3 with respect to the input  $M_{it}$  and the tax evasion  $e_{it}$  yield the following

$$G_M(M_{it})\exp(\omega_{it})\mathcal{E} = \frac{\rho_t}{P_t} \quad (4)$$

$$e_{it} = \frac{1 - q(e_{it})}{q'(e_{it})} - \frac{\kappa}{\frac{\rho_t}{P_t}\tau} \quad (5)$$

Equation 4, the well-known optimality condition, says that the price ratio is equal to the marginal product of the inputs, given the productivity and the expected value of the output shock,  $\mathcal{E} = \mathbb{E}[\exp(\varepsilon_{it})]$ .

Likewise, Equation 5 reveals the firms' optimal tax evasion decision decreases if the probability of detection  $q(e_{it})$  or the penalty of evading  $\kappa$  increases. Tax evasion also depends on how sensitive the probability of detection is to the level of evasion  $q'(e)$ . In particular, greater sensibility will result in lower levels of evasion.

Note that the net change of tax evasion due to an increase in the relative prices  $\frac{\rho_t}{P_t}$  or the tax rate  $\tau$  is not evident at first sight. The net effect will also depend on the change in the detection probability induced by the changes in the relative prices or the tax rate. In particular, an increase in relative prices  $\frac{\rho_t}{P_t}$  or the tax rate  $\tau$  will incentivize a higher tax evasion level, however, a higher tax evasion level will increase the probability of detection—depending on the shape of the probability as a function of  $e$ —, so it will deter higher levels of evasion. An increase in the tax rate, for instance, will only increase tax evasion if the change in the tax rates increases the incentives to evade more than the decrease in the incentives due to the changes in the detection probability.

Formally, suppose a firm increases its tax evasion,  $e_1 - e_0 > 0$  because of an increase in taxes  $\tau_1 > \tau_0$ . Then, it follows that

$$\left( \frac{\tau_1 - \tau_0}{\tau_1 \tau_0} \right) \frac{P\kappa}{\rho} > \left( \frac{1 - q(e_1)}{q'(e_1)} - \frac{1 - q(e_0)}{q'(e_0)} \right)$$

The change in the probability of detection weighted by the slope of the probability function should be less than the change in the tax rate weighted by the penalty of evading and the relative prices<sup>1</sup>.

---

<sup>1</sup>An analogous condition for an increase in relative prices leading to higher levels of tax evasion exists. Under this condition, the model is consistent with the literature that macroeconomic downturns lead to higher evasion.

## 1.2 Case 2: increase in the probability of detection after a certain threshold of size

Consider now the case where there is a threshold of size  $s_0$  after which the probability of detection increases significantly. To make things simpler, assume that before the threshold, the probability changes as a function of evasion but does not vary conditional on size. After the threshold, the probability increases for every level of evasion but does not vary conditional on size.

This could be the case in which there are different fiscal regimes for small and medium-large firms, and firms believe that the probability of detection in the medium-large firms' regime is higher. The medium-large regime might imply submitting more information that can be traced by a third party, like in Mexico and Colombia. Another possibility could be that the tax authority increases auditing efforts significantly after a certain size threshold, like in Spain.

In this scenario, depending on the productivity shock, the firm might be better off choosing not to grow. Indeed, for a small enough productivity shock, the firm will not choose to grow past the size threshold if the expected profits of staying small are greater than the expected profits of growing.

Formally, let  $S_s = \{s_i : s_i < s_0\}$  and  $S_l = \{s_i : s_i \geq s_0\}$ , then for any  $e_0$  and  $s'_i > s_i$ ,  $q(e_0|s_i \in S_k) = q(e_0|s'_i \in S_k)$  with  $k = \{s, l\}$ , but  $q(e_0|s'_i \in S_l) \geq q(e_0|s_i \in S_s)$  for all  $e_0$ .

Suppose the realized productivity shock of the current period  $\omega_{it}$  is such that, taking the probabilities as constant, the optimal input  $M_{it}^*$  implies an expected revenue greater than threshold  $s_{it}^* \geq s_0$ . Let the expected profits given  $M_{it}^*$  and the optimal tax evasion in the range of size  $S_l$ ,  $e_{it}^*$ , is  $\pi_l \equiv \mathbb{E}[\pi(M_{it}^*, e_{it}^*)|S_l]$ . Let  $\tilde{M}_{it}$  be the input level such that the



expected revenue is below the threshold  $\tilde{s}_{it} < s_0$  and  $\tilde{e}_{it}$  be the optimal tax evasion in the range of size  $S_s$ . Let also the expected profits of staying small are  $\pi_s \equiv \mathbb{E}[\pi(\tilde{M}_{it}, \tilde{e}_{it})|S_s]$ .

In this second case, therefore, firms might optimally choose to remain small if, for a low productivity shock, the expected profits of not growing are greater than the expected profits of growing  $\pi_l < \pi_s$ . Firms choosing to remain small will lead to a bunching below the threshold in the size distribution of firms.

Besides the higher levels of evasion before the threshold —simply because of the higher probability of detection—, we can also expect bunching firms to evade more than their similar-sized peers. At  $\tilde{M}_{it}$ , the optimization condition of Equation 4 no longer holds, hence, the marginal product of the input is now greater than the relative prices. Therefore, according to Equation 5, bunching firms would compensate for their *higher* costs by increasing overreporting.

## 2 The setting

Suppose we have access to panel data, where we observe output  $Y_{it}$ , intermediate inputs  $M_{it}$ , capital  $K_{it}$ , labor  $L_{it}$ , and output  $P_t$  and intermediate input prices  $\rho_t$  for  $I$  firms over  $T$  periods. Then our set of observations is  $\mathcal{O} = \{Y_{it}, M_{it}, K_{it}, L_{it}, P_t, \rho_t\}_{i \in I, t \in T}$ . As is standard in the literature, firms are price-takers and the intermediates are flexible.

The objects of interest are the production function (PF),  $Y_{it} = G(M_{it}, K_{it}, L_{it})e^{\omega_{it} + \varepsilon_{it}^Y}$ , and productivity  $\omega_{it}$ .  $\varepsilon_{it}^Y$  is the current period output shock. We are also interested in the Markov process of productivity, which we assume is AR(1),  $\omega_{it} = h(\omega_{it-1}) + \eta_{it}$ , with  $\mathbb{E}[\eta_{it}|\omega_{it-1}] = 0$ .

## 2.1 Tax evasion and the productivity bias

Firms overreport their true intermediate inputs  $M_{it}^*$  by  $e^{\varepsilon_{it}^M}$  to evade taxes. Then, reported inputs are

$$M_{it} = M_{it}^* e^{\varepsilon_{it}^M}$$

with  $\varepsilon_{it}^M \geq 0$  and  $\varepsilon_{it}^M \not\perp M_{it}^*$ . A simple model of tax evasion through input overreporting can be found in the appendix.

It is fairly easy to see that the productivity bias, the difference between the naively estimated  $\tilde{\omega}_{it}$  and true productivity  $\omega_{it}$  is as follows:

$$\mathbb{E}[\tilde{\omega}_{it} | \mathcal{J}_{it}] - \mathbb{E}[\omega_{it} | \mathcal{J}_{it}] \leq \ln \mathbb{E} \left[ \frac{G(M_{it}^*, K_{it}, L_{it})}{G(M_{it}^* e^{\varepsilon_{it}^M}, K_{it}, L_{it})} \middle| \mathcal{J}_{it} \right] \leq 0$$

Where  $\mathcal{J}$  stands for the information set of firm  $i$  in time  $t$ .

The previous result holds because of Jensen's inequality and because  $G(\cdot)$  is monotonically increasing in its arguments.

## 3 Identification strategy

My identification assumption is motivated by the empirical evidence showing that very large firms do not evade taxes by input overreporting ([Carrillo et al. 2022](#)). Hence, I assume I know the quantile of the size distribution of compliant firms.

**Assumption 1** (Non-evaders). *Firms above percentile  $s$  do not overreport inputs*

$$\mathbb{E}[\varepsilon_{it}^M | S_{it} > s] = 0$$

In addition, I impose the following timing assumption.

**Assumption 2.** *Firms choose overreporting  $\varepsilon_{it}^M$  before the output shock  $\varepsilon_{it}^Y$*

Assumption 2 implies that input overreporting is independent of the current period output shock,  $\varepsilon_{it}^M \perp \varepsilon_{it}^Y$ . In the literature is not rare to assume that the output shock is not part of the information set of the firms,  $\varepsilon_{it}^Y \notin \mathcal{J}_t$  (Gandhi et al. 2020). Timing and information set assumptions are not uncommon for identification strategies in production functions and demand estimation (Akerberg et al. 2021, Akerberg 2019).

### 3.1 Identifying the production function parameters

Suppose the production function is Cobb-Douglas,  $G(M_{it}^*, K_{it}, L_{it})e^{\omega_{it} + \varepsilon_{it}^Y} = M_{it}^{*\beta} K_{it}^{\alpha_K} L_{it}^{\alpha_L} e^{\omega_{it} + \varepsilon_{it}^Y}$ .

Then, we can use the first-order conditions of the firms' cost minimization problem on the flexible input, as in Gandhi et al. (2020), and assumption 1 to recover the production function parameter  $\beta$ .

$$\begin{aligned} \ln \left( \frac{\rho_t M_{it}^*}{P_t Y_{it}} \right) + \varepsilon_{it}^M &= \ln \beta + \ln \mathcal{E} - \varepsilon_{it}^Y \\ \mathbb{E} \left[ \ln \left( \frac{\rho_t M_{it}^*}{P_t Y_{it}} \right) \middle| S_{it} > s \right] &= \ln \beta \end{aligned} \tag{6}$$

$$\mathcal{E} = \mathbb{E}[e^{\varepsilon_{it}^Y} | \mathcal{J}_{it}]$$

Note that this is the case because  $\mathbb{E}[\varepsilon_{it}^Y | S_{it} > s] = \mathbb{E}[\varepsilon_{it}^Y] = 0$ . That is, there is no selection on the output shock. Therefore, as long as our measure of size is not structurally related to

Equation 6 — like output or revenue in the same period—, the output elasticity of input,  $\beta$ , can be recovered.

### 3.2 Identifying the Markov Process of Productivity

If we were to follow [Gandhi et al. \(2020\)](#) to recover the Markov process of productivity, we could only have estimates for firms with size  $S_{it} > s$ .

$$y_{it} \equiv \ln Y_{it} - \varepsilon_{it}^Y - \beta \ln M_{it} - \alpha_K \ln K_{it} - \alpha_L \ln L_{it} = \omega_{it}$$

$$\mathbb{E}[y_{it}|y_{it-1}, S_{it} > s] = \mathbb{E}[\omega_{it}|\omega_{it-1}, S_{it} > s]$$

However, it is likely that there would be selection on  $S$ , i.e.,  $\mathbb{E}[\omega_{it}|S_{it} > s] \neq \mathbb{E}[\omega_{it}]$ . In other words, larger firms will have a higher probability of getting a higher draw of productivity. Therefore, the parameters of the AR(1) process of productivity would be biased.

However, for every firm, I can form the following variable using observed variables and the production function parameter I recovered from the first stage.

$$\begin{aligned} \mathcal{V}_{it} &\equiv \ln \left( \frac{\rho_t M_{it}}{P_t Y_{it}} \right) - \ln \beta \\ &= \ln \left( \frac{\rho_t M_{it}^*}{P_t Y_{it}} \right) - \ln \beta + \varepsilon_{it}^M \\ &= -\varepsilon_{it}^Y + \varepsilon_{it}^M \end{aligned}$$

This is useful because I can recover the tax evasion up to a random variable,  $\varepsilon_{it}^Y$ , with mean zero and, by assumption 2, independent of  $\varepsilon_{it}^M$ .

Likewise, I can recover the true inputs  $M_{it}^*$  and productivity up to  $\varepsilon_{it}^Y$ .

$$\begin{aligned}
\mathcal{M}_{it} &\equiv \ln M_{it} - \mathcal{V}_{it} \\
&= \ln M_{it}^* + \varepsilon_{it}^M - (\varepsilon_{it}^M - \varepsilon_{it}^Y) \\
&= \ln M_{it}^* + \varepsilon_{it}^Y
\end{aligned}$$

$$\begin{aligned}
\mathcal{W}_{it} &\equiv \ln Y_{it} - \beta \mathcal{M}_{it} - \alpha_K \ln K_{it} - \alpha_L \ln L_{it} \\
&= \beta \ln M_{it}^* + \alpha_K \ln K_{it} + \alpha_L \ln L_{it} + \omega_{it} + \varepsilon_{it}^Y - \beta \ln M_{it}^* - \beta \varepsilon_{it}^Y - \alpha_K \ln K_{it} - \alpha_L \ln L_{it} \\
&= \omega_{it} + (1 - \beta) \varepsilon_{it}^Y
\end{aligned}$$

I can use  $\mathcal{W}_{it}$  to recover  $\delta = [\delta_0, \delta_1]'$ .

$$\begin{aligned}
\mathbb{E}[\mathcal{W}_{it} | \mathcal{W}_{it-1}] &= \\
\mathbb{E}[\omega_{it} | \mathcal{W}_{it-1}] &= \delta_0 + \delta_1 \mathbb{E}[\omega_{it-1} + (1 - \beta) \varepsilon_{it-1}^Y | \mathcal{W}_{it-1}] \\
&\quad - \mathbb{E}[(1 - \beta) \varepsilon_{it}^Y | \mathcal{W}_{it-1}] + \mathbb{E}[\eta_{it} | \mathcal{W}_{it-1}] \\
&= \delta_0 + \delta_1 \mathbb{E}[\omega_{it-1} | \mathcal{W}_{it-1}]
\end{aligned}$$

### 3.3 Deconvoluting

But I can go further and recover the densities of tax evasion and productivity at every period. The key insight is that from the 1st stage, I also learn the distribution  $m$  of  $\varepsilon_{it}^Y$ . Hence, I can use non-parametric density deconvolution methods.

For completeness, I reproduce a result from probability theory that allows me to pursue these techniques.

**Definition 3.1.** The density of the sum of two *independent* random variables is equal to the *convolution* of the densities of both addends; hence

$$h^d = f^d * m = \int f^d(\mathcal{Z}^d - \varepsilon^Y) m(\varepsilon^Y) d\varepsilon^Y$$

where  $h^d$  is the density of  $\mathcal{Z}^d \in \{\mathcal{E}, \mathcal{X}, \mathcal{W}\}$  (Meister, 2009)

The result holds by definition for productivity and true inputs, and by assumption 2 for tax evasion.

Intuitively, for every  $n$ -th moment  $\mathbb{E}[(\varepsilon_{it}^Y)^n | S_{it} > s] = \mathbb{E}[(\varepsilon_{it}^Y)^n | t] = \mathbb{E}[(\varepsilon_{it}^Y)^n]$ . Therefore, I can get any moment of the tax evasion  $\varepsilon_{it}^M$  distribution  $\forall t \in T$ . For example,

$$\mathbb{E}[\mathcal{E}_{it} | t] = \mathbb{E}[\varepsilon_{it}^M | t] - \mathbb{E}[\varepsilon_{it}^Y] = \mathbb{E}[\varepsilon_{it}^M | t]$$

$$\mathbb{V}\text{ar}[\mathcal{E}_{it} | t] = \mathbb{V}\text{ar}[\varepsilon_{it}^M | t] + \underbrace{\mathbb{V}\text{ar}[\varepsilon_{it}^Y]}_{\text{known}}$$

Therefore, I can recover the distribution of tax evasion and how it has changed over time.

Furthermore, I can do the same with true inputs  $M_{it}^*$  and productivity  $\omega_{it}$ .

Moreover, I can get any mixed moment  $\mathbb{E}[(\mathcal{Z}^c, \mathcal{Z}^d)^n | t]$  and learn how tax evasion changes with productivity, for instance, and how this relationship has evolved.

$$\mathbb{C}\text{ov}[\mathcal{E}_{it}, \mathcal{W}_{it} | t] = \mathbb{E}[\varepsilon_{it}^M \omega_{it} | t] - \underbrace{(1 - \beta) \mathbb{E}[(\varepsilon_{it}^Y)^2] - \mathbb{E}[\varepsilon_{it}^M | t] \mathbb{E}[\omega_{it} | t]}_{\text{known}}$$

## 4 Colombia 1981-1991: an application

During this period, Colombia went through three major fiscal reforms (1983, 1986, 1990).

The 1983 reform increased the value-added tax (VAT) from 6% to 10% and expanded the number of the products and services that were levied. In addition, the 1983 fiscal reform

required firms with profits over COL\$ 200,000 and owned by individuals to pay profit taxes. The 1986 reform relocated the tax collection and reception of tax reports to the banking system, among other things. The 1990 reform increased the VAT from 10% to 12%. See section the appendix for more details on the fiscal reforms.

Intuitively, we expect higher levels of tax evasion if tax rates increase. On the other hand, reporting more information to the authority —like the firms required to report and pay profit taxes in 1983—, or having a third party reporting on your private information —like the banking system being responsible for the collection and reception of tax reports in 1986— would decrease tax evasion.

## 4.1 Data

The data is a well-known data set that has been used in the estimation of production functions and productivity before. The dataset comes from the Annual Survey of Manufacturing (EAM) and contains information about manufacturing firms with more than 10 employees from 1981 to 1991.

A simple graphical analysis shows that the average (of the log) intermediates cost share of sales started growing after 1983 and that it stabilized in 1988 after the policy changes of the 1986 reform settled in (Figure 1). The dataset does not capture any changes after the 1990 reform, although there is only one more year of data.

Furthermore, we can appreciate in Figure 2 two different patterns before and after the threshold where firms start being required to report and pay profit taxes. We can observe that the intermediate’s cost share decreases abruptly by the 20th percentile. Then, it reaches its minimum by the percentile 25th and starts growing up till the end of the distribution.

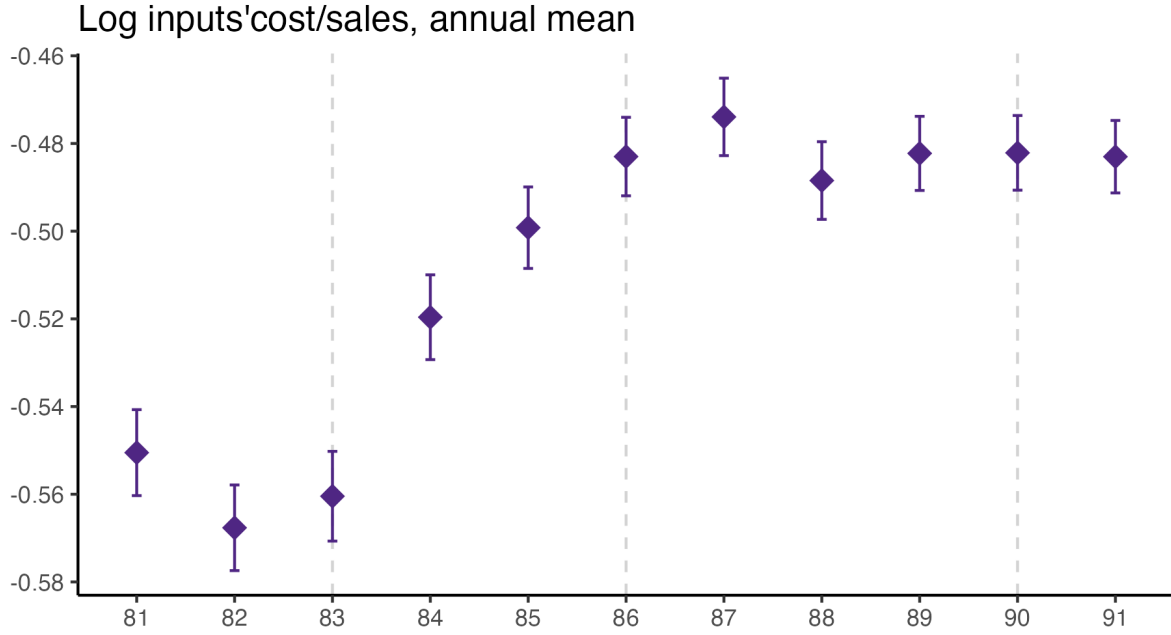


Figure 1: Input's cost share of sales, average by year of the logs.

Before the 25th percentile, firms only file and pay VAT, starting in the 25th percentile, firms report and pay in addition profit taxes. In other words, the information they have to provide to the authority is greater. The bunching before the 25th percentile is also informative about firms' beliefs of the probability of getting caught evading, or the difficulty for firms in evading as they report more information to the authority.

In addition, Colombia has a progressive profit tax schedule. Firms with higher profits pay a higher tax rate. Higher tax rates lead to increased incentives to evade them; a plausible explanation of why the increasing pattern.

As a validation exercise, we can see that the VAT changes induced by the three fiscal reforms are captured in the dataset. Figure 3 shows that the sales tax increased to 10% after the 1983 reform, and then around 12% after the 1990 reform.

Just as an exercise to see if other economic changes in this period were driving the apparent



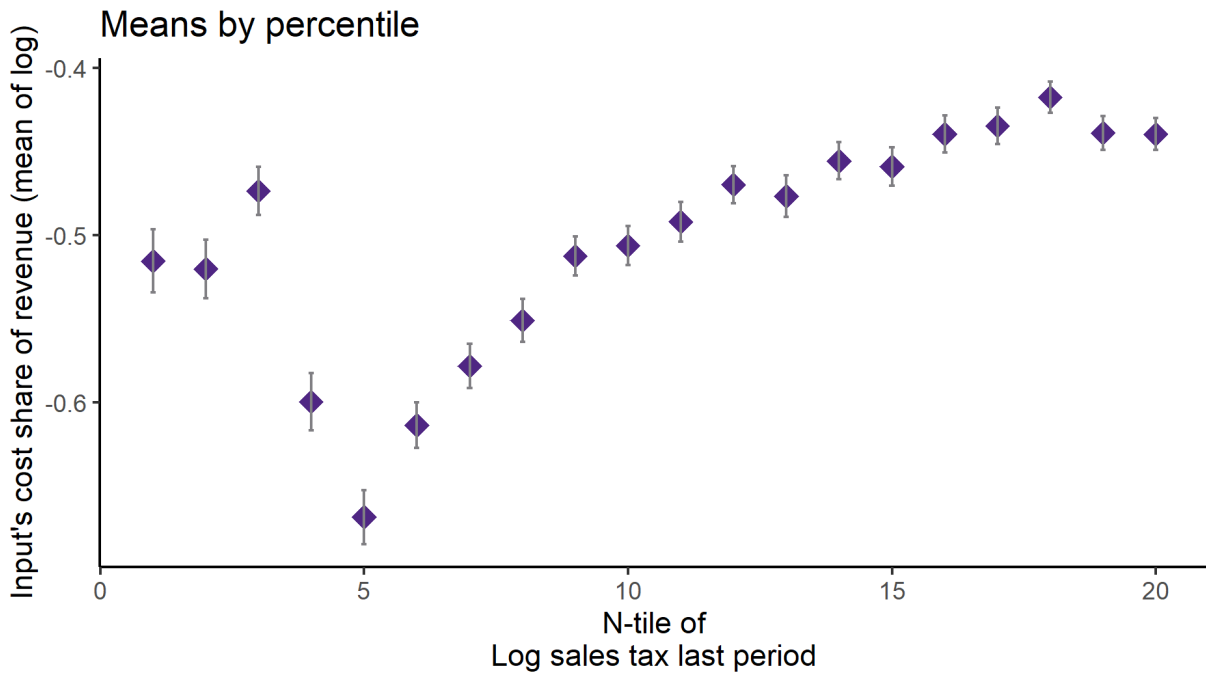


Figure 2: Input's cost share of sales, average by percentile of sales taxes paid the previous year (in logs, groupings of 5 percentiles).

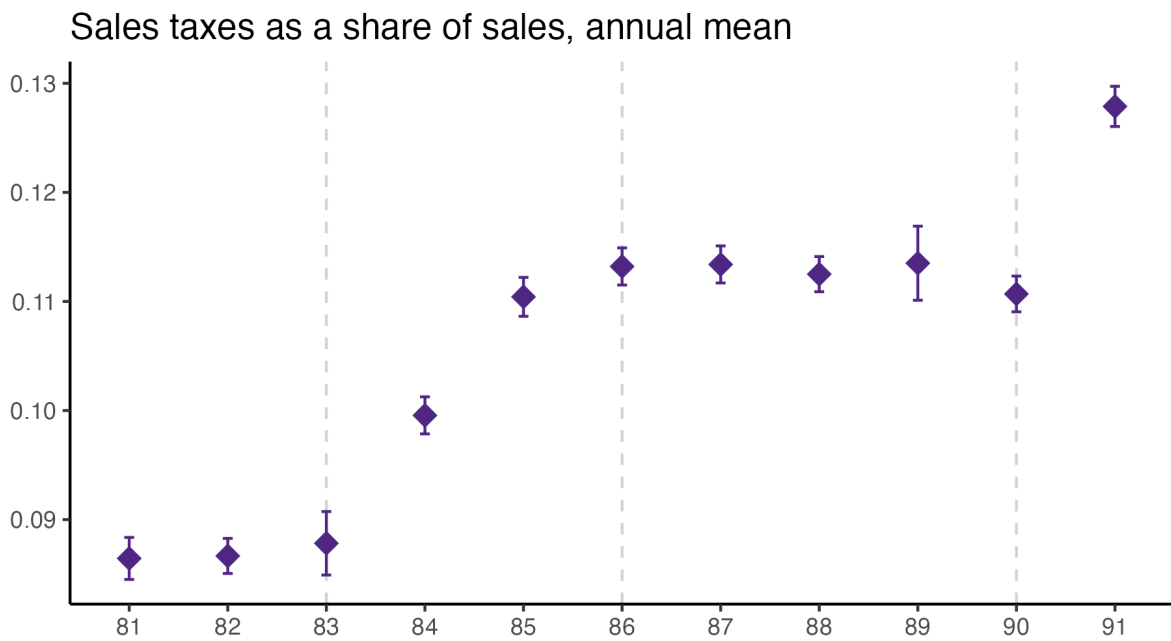


Figure 3: Sale taxes paid as share of sales, average by year.

changes in overreporting, Figure 4 shows that sales, for instance, were not exactly following the changes in fiscal policy. Sales start to grow during 1983, the year of the reform, whereas the cost share of sales started to grow the year after. Likewise, sales fall in 1986, while the cost share seems to reduce its growth after 1986.

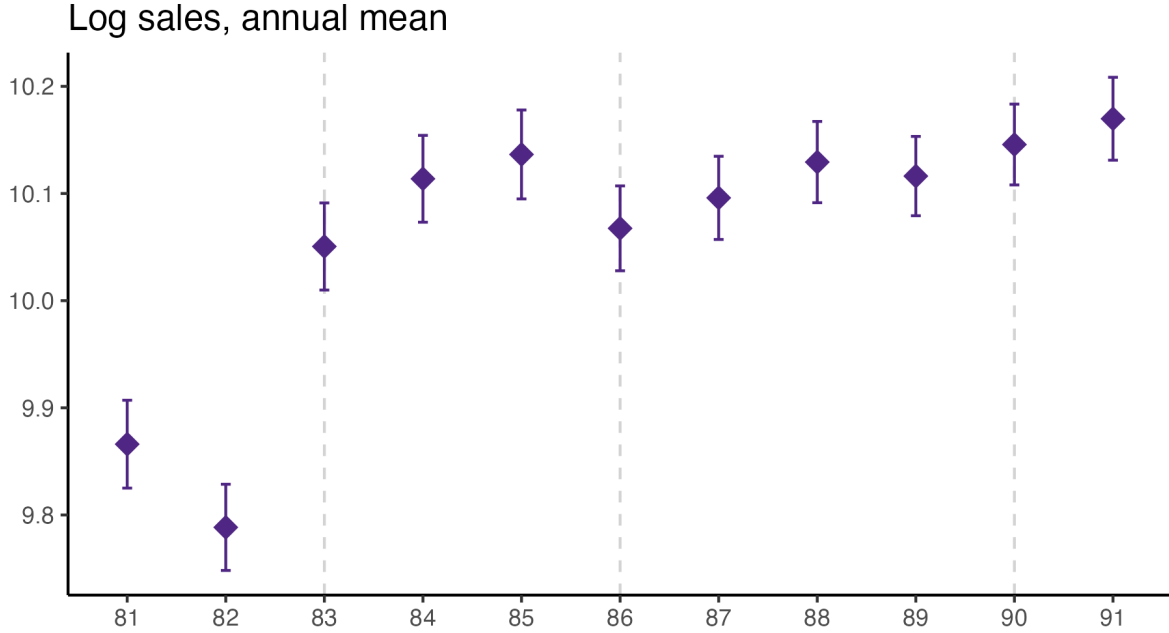


Figure 4: Sales in logs, annual mean.

## 5 Empirical evidence

To inform the empirical application, I analyze Figure 2 by year, focusing on the years around the policy change of 1983. Figure 5 shows the mean of the intermediate's cost share of sales (mean log share, henceforth) by percentile of the sales taxes paid the previous year and by year, only for 1982 to 1984.

The first observation that jumps into view is that the firms that in 1982 only paid VAT and were required to pay profit taxes in 1983 seem to decrease their intermediates' cost

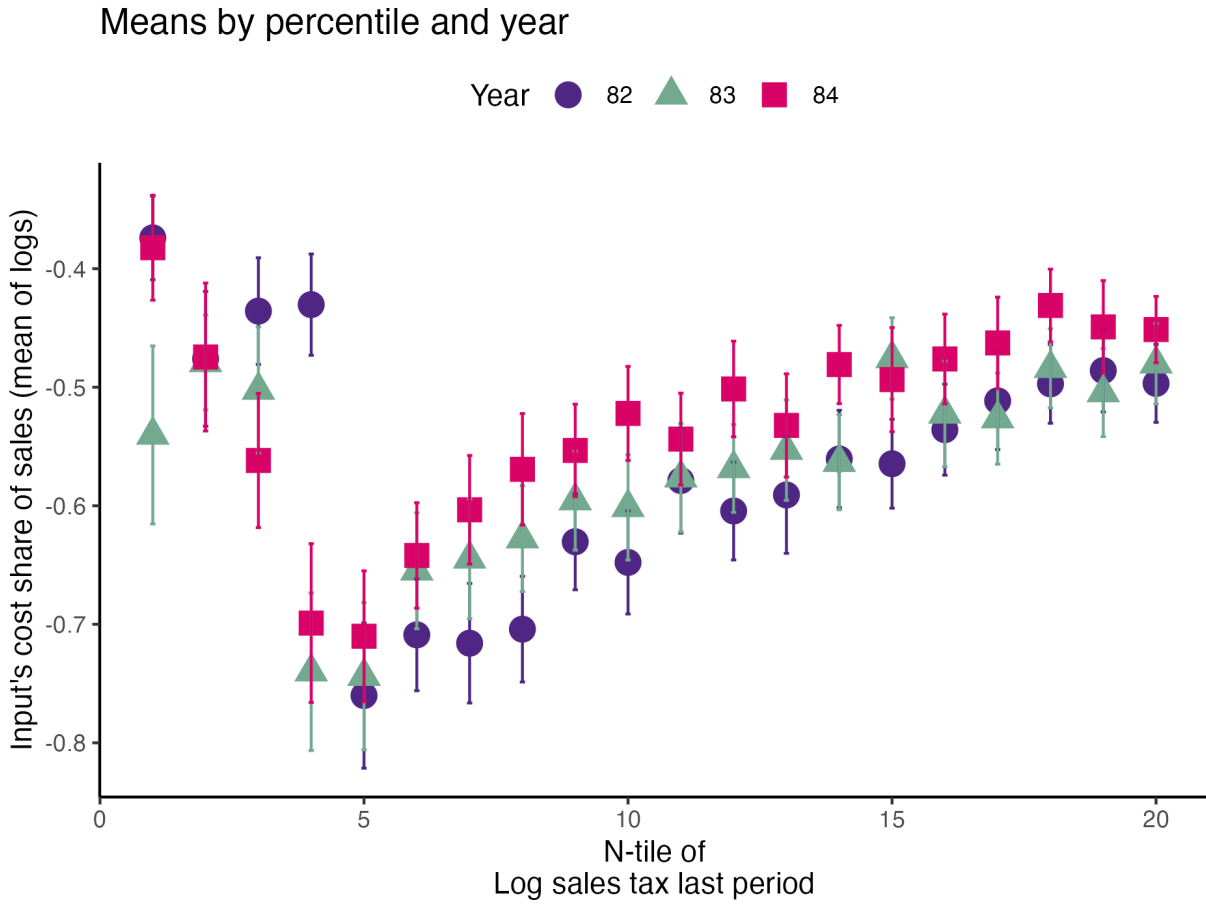


Figure 5: Input's cost share of sales, average by percentile of sales taxes paid the previous year and by year (in logs).

share of sales drastically in 1983, precisely. Please note how the mean log share of the 20th percentile (4/20 in Figure 5) drops significantly from 1982 to 1983.

The second consideration is that there is a generalized increase in the mean log share from 1983 to 1984 across almost all the distribution. Although the increases differ across the percentiles of the distribution, the highest 10% of the distribution increased the least. This makes sense if the tax authority scrutinizes with the finest detail the firms paying the greatest tax bill.

Based on this graphical analysis, I investigate with a simple empirical model 1) how the firms required to pay profit taxes by the 1983 reform adjusted their tax evasion and 2) the change in overreporting induced by the 1983 and 1986 reforms in subsequent years.

On one hand, as Figure 5 suggests, I expect firms required to report and pay profit taxes to reduce their tax evasion. This might seem counterintuitive as higher taxes will increase the incentive to evade. These firms however have now to report more information to the tax authority increasing the difficulty to evade. In addition, firms had to adjust their beliefs about the probability of being caught evading taxes under the additional required information. On the other hand, Figure 1 and Figure 5 suggest that tax evasion increased after the 1983 reform and it ceased to grow after the 1986 reform.

Of course, we would like to control for other factors that might have affected the log share during these years. We would need to control also for industry and regional variation. Consequently, to test the change in overreporting by firms required to pay profit taxes in 1983 (20th percentile in Figure 5), I use as the control group the firms at the 10 and 95 percentile of the distribution. The intuition is that these firms were not affected by this fiscal policy change. Furthermore, firms in the 10th percentile are more similar to

the affected firms than the ones at the 95th percentile. However, the firms at the 95th percentile receive the most attention from the tax authority, consequently, they are less likely to adjust their log share because of tax evasion motives. Likewise, I use the 95th percentile firms as the control group to test the change in overreporting induced by the 1983 and 1986 reforms. For these two specifications, the base year is 1983 and 1986 respectively.

The results of the estimations in Table 1 show that firms that were required to report and pay profit taxes in 1983 because of the change in the threshold reduced their tax evasion through cost overreporting by 14%. Table 2 shows that tax evasion grew between 8% and 9% in 1985 and 1986, following the increased value-added tax rates of the 1983 fiscal reform. Finally, Table 3 indicates that the tax evasion growth came to a halt after 1986, as the subsequent years do not show a significant increase with respect 1986.

## Conclusions

I provide a new estimation strategy to jointly recover tax evasion through cost overreporting and productivity that only requires commonly available firm-level data.

Studying manufacturing firms from Colombia, I show that the method can be used to investigate tax evasion using surveys and fiscal policy changes. The 1983 reform required firms with profits above COL\$ 200,000 or more to report income tax in 1983 in addition to the value-added tax. These firms reduced their tax evasion by cost overreporting by 14% likely due to the additional information firms needed to report and the adjustment in their beliefs on the probability of getting caught evading.

Likewise, the evidence suggests that this type of tax evasion increased after the reform of 1983 between 8 and 9% in 1985 and 1986. The increase in value-added tax likely amplified

Table 1: 1983 Reform, Change in Threshold

Dependent Variable:	Inter. cost share of sales (logs)	
Model:	(1)	(2)
<i>Variables</i>		
Affected firms	0.0003 (0.0309)	0.0061 (0.0309)
Year of 1983	-0.0090 (0.0212)	-0.0091 (0.0208)
Affected firms $\times$ Year of 1983	-0.1531*** (0.0456)	-0.1543*** (0.0457)
<i>Fixed-effects</i>		
Industry	Yes	Yes
Metropolitan area		Yes
<i>Fit statistics</i>		
Observations	1,714	1,714
R <sup>2</sup>	0.17053	0.17614
Within R <sup>2</sup>	0.01889	0.01825

*Heteroskedasticity-robust standard-errors in parentheses*

*Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1*

Table 2: 1983 Fiscal Reform

Dependent Variable:	Inter. cost share of sales (logs)	
Model:	(1)	(2)
<i>Variables</i>		
1-15% $\times$ Year of 1984	0.0039 (0.0281)	0.0052 (0.0280)
20-50% $\times$ Year of 1984	0.0226 (0.0217)	0.0210 (0.0216)
55-75% $\times$ Year of 1984	0.0099 (0.0213)	0.0098 (0.0212)
80-95% $\times$ Year of 1984	0.0284 (0.0219)	0.0294 (0.0219)
1-15% $\times$ Year of 1985	0.0536* (0.0275)	0.0534* (0.0275)
20-50% $\times$ Year of 1985	0.0762*** (0.0219)	0.0737*** (0.0218)
55-75% $\times$ Year of 1985	0.0753*** (0.0215)	0.0751*** (0.0214)
80-95% $\times$ Year of 1985	0.0801*** (0.0225)	0.0796*** (0.0224)
1-15% $\times$ Year of 1986	-0.0271 (0.0267)	-0.0249 (0.0267)
20-50% $\times$ Year of 1986	0.0681*** (0.0211)	0.0662*** (0.0210)
55-75% $\times$ Year of 1986	0.0827***	0.0829***

Table 3: 1986 Fiscal Reform

Dependent Variable:	Inter. cost share of sales (logs)	
Model:	(1)	(2)
<i>Variables</i>		
1-15% $\times$ Year of 1987	0.0107 (0.0249)	0.0108 (0.0248)
20-50% $\times$ Year of 1987	-0.0031 (0.0198)	-0.0029 (0.0196)
55-75% $\times$ Year of 1987	-0.0102 (0.0197)	-0.0095 (0.0196)
80-95% $\times$ Year of 1987	0.0181 (0.0201)	0.0178 (0.0199)
1-15% $\times$ Year of 1988	0.0019 (0.0263)	-0.0053 (0.0262)
20-50% $\times$ Year of 1988	0.0198 (0.0194)	0.0125 (0.0192)
55-75% $\times$ Year of 1988	0.0184 (0.0195)	0.0116 (0.0194)
80-95% $\times$ Year of 1988	0.0187 (0.0199)	0.0115 (0.0198)
1-15% $\times$ Year of 1989	0.0321 (0.0248)	0.0226 (0.0247)
20-50% $\times$ Year of 1989	0.0190 (0.0192)	0.0113 (0.0191)
55-75% $\times$ Year of 1989	-0.0065	-0.0140



the incentives to evade. This result stands at odds with previous studies that indicate that the evasion of income tax and VAT declined during this period (Sanchez & Gutierrez, 1994). Finally, evidence suggests that tax evasion growth came to a halt after the 1986 reform where the banking system took over the tax collection and reception of tax reports from the tax authority.

## References

Ackerberg, D. A. (2019), ‘Timing assumptions and efficiency: Empirical evidence in a production function context’.

Ackerberg, D., Frazer, G., Luo, Y. & Su, Y. (2021), ‘Under-identification of structural models based on timing and information set assumptions’.

**URL:** <https://ssrn.com/abstract=3717757>

Bustos, S., Pomeranz, D., Carlos, J., Serrato, S., Vila-Belda, J. & Zucman, G. (2022), ‘The race between tax enforcement and tax planning: Evidence from a natural experiment in chile the race between tax enforcement and tax planning: Evidence from a natural experiment in chile’.

**URL:** <http://www.nber.org/papers/w30114>

Carrillo, P., Donaldson, D., Pomeranz, D. & Singhal, M. (2022), ‘Ghosting the tax authority: Fake firms and tax fraud’.

**URL:** <http://www.nber.org/papers/w30242>

Gandhi, A., Navarro, S. & Rivers, D. A. (2020), ‘On the identification of gross output production functions’.

Paulus, A. (2015), ‘Tax evasion and measurement error: An econometric analysis of survey

data linked with tax records’.

**URL:** <http://hdl.handle.net/10419/126489>

Pissarides, C. A. & Weber, G. (1989), ‘An expenditure-based estimate of britain’s black economy’, *Journal of Public Economics* **39**, 17–32.

*Technology Tools to Tackle Tax Evasion and Tax Fraud* (2017).

**URL:** [www.oecd.org](http://www.oecd.org)

Zumaya, M., Guerrero, R., Islas, E., Pineda, O., Gershenson, C., Iñiguez, G. & Pineda, C. (2021), ‘Identifying tax evasion in mexico with tools from network science and machine learning’.

**URL:** <http://arxiv.org/abs/2104.13353>