Tax evasion and productivity

Hans Martinez*
Department of Economics, University of Western Ontario

April 14, 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 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.

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 **?**, cost overreporting — also known as "fake invoicing", "ghost firms", "invoice mills", or "missing traders"— permeates internationally reaching Latin America, Eastern Europe, Asia, and Africa. Cost overreporting reportedly generated annual tax revenue losses amounting to non-trivial shares of GDP in countries such as Poland, 6%; Chile, 0.2%; Colombia, 0.2%; and Mexico, 0.03%. Recent evidence from Ecuador (**?**) shows that, contrary to current belief, cost overreporting is widespread across firms —not limited to small, semi-formal firms— and quantitatively large.

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 ?).

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 (??). 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 (??), 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 productivity estimates that are downwardly biased. 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 jointly recover the distributions of tax evasion and productivity.

I also estimate how tax evasion evolves with productivity, in other words, their joint distribution.

Using firm-level data from Ecuador, Colombia, Chile, and Mexico, the estimates show that . . .

1 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 ρ_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 ω_{it}. ε_{it}^Y is the current period output shock. We are also interested in the Markov process of productivity, which we assume is AR(1), $\omega_{it} = \delta_0 + \delta_1\omega_{it-1} + \eta_{it}$, with $\mathbb{E}[\eta_{it}] = 0$ and $\eta_{it} \perp \omega_{it}$.

1.1 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

$$M_{it} = M_{it}^* e^{\varepsilon_{it}^M} \tag{1}$$

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 ω_{it} is as follows:

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

Where $\mathcal{I}$ 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.

2 Identification strategy

My identification assumption is motivated by the empirical evidence showing that very large firms do not evade taxes by input overreporting (**?**). Hence, I assume I know the quantile of the size distribution of compliant firms.

Assumption 1 (Compliers)**.** *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 ε_{it}^{M}* after *productivity shock ω_{it} is realized and true inputs M_{it}^{*} are chosen but* before *the output shock ε_{it}^{Y}*

Assumption **??** 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, $\varepsilon_{it}^{Y} \notin \mathcal{I}_{t}$ (**?**). Timing assumptions for identification are common in production functions and demand estimation (**??**).

2.1 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 **?**, and assumption **??** to recover the production function parameter β.

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

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 output, the output elasticity of input, β, can be recovered.

2.2 Identifying the Markov Process of Productivity

If we were to follow **?** to recover the Markov process of productivity, we could only have estimates for firms with size $S_{it} > s$.

$$\mathcal{Y}_{it} \equiv \ln Y_{it} - \varepsilon_{it}^Y - \beta\ln M_{it} - \alpha_K \ln K_{it} - \alpha_L \ln L_{it} = \omega_{it}$$
$$\mathbb{E}[\mathcal{Y}_{it}|\mathcal{Y}_{it-1}, S_{it} > s] = \mathbb{E}[\omega_{it}|\omega_{it-1}, S_{it} > s]$$

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.

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

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

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

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

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

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

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

2.3 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 ε_{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 2.1. 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 **??** 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 ε_{it}^{M} distribution $\forall t \in T$. For example,

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

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

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 ω_{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.

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

3 Preliminary results