Tax evasion and productivity

By Hans Martinez*

Draft: July 17, 2022

Although tax evasion is a widespread phenomenon, it has not been addressed in the estimation of firm-level productivity. Firms, nevertheless, have incentives to under-report profits to evade taxes by either consistently over-reporting expenses or under-reporting sales. Consequently, productivity estimates are likely to be downward biased. The literature has coped with tax evasion by treating it as a measurement error problem. However, this is inconsistent with economic intuition about firms' profit-maximizing incentives. In particular, systematic misreporting due to tax evasion is not likely to be mean zero and uncorrelated with firms' characteristics. In this paper, I jointly model tax evasion and productivity, allowing for classical measurement error, and estimate firm-level productivity for manufacturing firms in Mexico and Colombia. The key identifying assumption is that tax evasion depends on the expected return to cheating, which, in turn, depends on firms' characteristics.

JEL: C12, C14, C60, C72, D22, D43, L13

Keywords: Tax evasion, Productivity

I. Prospectus

Important dates:

- July 31: Prospectus document;

* Martinez: Western University, hmarti33@uwo.ca.

- Final prospectus document: 7 days before oral presentation;
- Last two weeks of Sept: Oral defense (presentation).

A. Prospectus Objective

Demonstrate that:

1) Clear and original question
2) The question is interesting (it addresses a nontrivial problem) and important (it's of interest to scholars in the student's field)
3) Feasible solution and within student's reach during the remainder of Ph.D.
4) Resulting research is publishable in a good field journal

B. Prospectus document

Outline thesis with particular focus on *planned* JMP. Document must convey research *plan* satisfies criteria:

1) clarity
2) interest and importance
3) feasibility
4) quality

Present and discuss:

1) JMP's idea
2) Motivation
3) Originality
4) Strategy

Empirical JMP

1) Outline identification and estimation strategy

2) Explain the data that will be used

3) Table of summary statistics or preliminary results

Format details:

- about 10 pages long.
- Up to two pages describing other components if they are known.
- List names of three or more members of thesis committee

C. RAP

RAP1

A: Firm-level productivity estimates using gross-production functions are downward biased when firms over-report expenses of the flexible input to reduce their tax obligations and increase their profits.

R: Are our firm-level productivity estimates downward biased because of firms' tax evasion?

P: Although tax evasion is a widespread phenomenon, it has been ignored in the estimation of firm-level productivity.

RAP2

A: Firm-level productivity estimates using gross-production functions are less spread out when accounting for systematic under-reporting of expenses in the flexible input to evade taxes.

R: Does accounting for the firm's tax evasion help us explain the great variation of firm-level productivity estimates?

P: Tax evasion has been ignored as a possible cause for the wide variation of productivity across firms.

RAP3

A: Low productivity firms in Mexico are not as low as believed.

R: Why are low-productivity firms in Mexico surviving in the market?

P: This paper is the first to consider tax evasion as a possible explanation for the productivity puzzle in Mexico, i.e., why are low productivity firms surviving in the market?

Introduction, or clarity, interest, and importance

Although tax evasion is a widespread phenomenon[1], it has not been addressed in the estimation of firm-level productivity. Firms, however, have incentives to reduce their fiscal obligations and maximize after-tax profits. With this goal, firms might systematically misreport profits. This is important because, then, productivity estimates at the firm level are likely to be downward biased.

Intuitively, misreporting due to tax evasion biases how we measure firm-level productivity —as the residual of a production function, where the output is a function of the inputs— by distorting the observed relationship between inputs and outputs with respect to the true unobserved one. If firms are taxed on their profits —a common in many countries—, then, firms have incentives to underreport true profits to the authority to reduce their tax liabilities. To reduce their declared profits, firms might underreport sales or over-report expenses. Now, consider estimating productivity by a gross-output production. In this approach, gross output is a function of the inputs and a residual, productivity. A key assumption is that input demand is strictly monotonic on the productivity (Gandhi, Navarro, and Rivers 2020; Ackerberg, Caves, and Frazer 2015; Levinsohn and Petrin 2003). That is, more productive —technically efficient— firms will use fewer inputs and produce more output. Therefore, if firms underreport their output to reduce sales, or if firms overreport inputs to increase expenses, then the productivity estimates would be biassed downward. That is, certain firms —the

[1] Tax evasion has been a long-standing concern for developing countries, but since the 2008 economic crisis, it has also been of increasing importance for developed countries (**Slemrod2019?**).

ones with higher incentives to not comply— could be more technically efficient than we think they are. Note: Does not belong here, move to productivity gap

Why do we care about the productivity measurement bias due to tax evasion? First, it might help explain part of the productivity gap puzzle. Economists have found *enormous and consistent* productivity differences across producers (Syverson 2011). We still don't know how much of this productivity gap —the difference between the lower and higher percentiles of the distribution— tax evasion can explain. Second, it should be considered in the design of public policies aiming at reducing resource misallocations. Aggregate productivity —and, thus, the economic growth of a country and the welfare of its citizens— depends on the efficient allocation of its inputs among its firms. Ideally, resources should be allocated to the most productive firms. Additionally, it is not necessarily the case that firms with the highest incentives to evade taxes are always the lowest productive. The incentives depend on the tax system. For example, in countries where a profit threshold determines different tax rates, the firms near the threshold are the ones with the highest incentives to misreport, but the threshold might be completely unrelated to productivity. Therefore, firm-level productivity estimates adjusted by tax evasion misreporting might inform better the design of public policies aiming at an efficient reallocation of resources to boost economic growth.

Note: here

Moreover, accounting for tax evasion in estimating and measuring firm-level productivity is particularly significant for developing countries. This is the case because tax evasion is likely to be higher in low- and middle-income countries[2], and because the productivity gap in these countries is wider —wider productivity gaps are associated with higher resource misallocations. For example, recent estimates of non-detected tax evasion are up to $10 billion USD per year in Mexico

[2]Informality —the extreme form of tax evasion— is larger in developing countries [(**Loayza2006?**); LaPorta2014].

(Zumaya et al. 2021), approximately 1% of the country's GDP. Furthermore, some studies argue that input misallocation —implied by a wider productivity gap— explains a significant part of the differences in income per capita between developed and developing countries (Syverson 2011; Levy 2018). For instance, evidence from Colombia suggests that labor and financial policy reforms during the 1990s aimed at reallocating away from low- and towards high-productivity firms —effectively reducing the productivity gap— resulted in a higher aggregate productivity (Eslava et al. 2004).

The main challenge of dealing with tax evasion is that it is hidden by nature, for that reason, the literature has coped with it by treating it as a measurement error problem (e.g., Blalock and Gertler 2004, 204). The argument is that some firms under-report and others over-report so that the misreporting does not bias the production function estimates. In other words, it is assumed that tax-evasion misreporting has a mean zero in expectation and is independent of firm attributes.

However, the previous assumption posed by the literature to minimize the effect of tax-evasion bias is inconsistent with economic intuition on firms' profit-maximization incentives. To start, systematic misreporting due to tax evasion goes in one direction only. Put differently, there's no economic incentive for firms to artificially increase their profits for tax purposes by either overreporting sales or underreporting input expenses. Therefore, it is unlikely that systematic underreporting due to tax evasion is mean zero. Additionally, the degree of misreporting likely varies across firms depending on the firm's cost of misreporting, the probability to get caught, and the firm's risk aversion. Indeed, it is likely that the cost of cheating is higher for larger firms than for smaller ones simply because of the more people involved in tracking the double accounting of inputs. Also, for a large firm, the risk of an anonymous whistle-blower employee is higher. As a result, the firm's characteristics might be correlated with the intensive and extensive margins of tax evasion.

Note: Part 2

Note: Bridge

So, then, in this paper, I ask,

> can we recover unbiased productivity estimates at the firm level using a gross-output production function in the presence of systematic misreporting due to tax evasion? if so, what is the magnitude of this bias, in particular for developing countries? given that different tax systems —rates, rules and enforcement procedures— generate different tax evasion incentives, how does the producitivity bias vary according to different tax systems? how much of the productivity gap can tax evasion explain within a country and across countries?

To answer these questions, this paper studies productivity at the establishment level for Mexican and Colombian firms accounting for tax evasion. I employ data from a sample of anonymized firms' tax filings and INEGI's EAIM annual production survey, in the case of Mexico. For Colombia, I use the EAM annual survey on manufacturing firms. In the estimation framework, I jointly model the firm-level productivity and the tax evasion, allowing for classical measurement error. The key identifying assumption is that while tax evasion depends on the expected return to cheating, productivity follows a Markov process and measurement error is uncorrelated with inputs and across time. In turn, the expected returns to cheating depend on the firm's characteristics[3]. In particular, larger firms are less likely to underreport inputs because of their higher cheating cost, their higher probability to be anonymously denounced by an employee, and their access to complex tools to legally *avoid* taxes[4].

By jointly modeling tax evasion and productivity, this paper departs from the literature in estimating a latent variable that follows an incentive constraint (IC) —tax compliance—, and in focusing on the productivity bias —rather than on

[3]Tax evasion might also depend on the firm's market characteristics, e.g., market concentration, end-consumer vs. intermediate goods market, etc.

[4]I follow the literature by referring to evasion as illegal actions to reduce tax liability, while avoidance refers to legal actions to reduce tax.

an input's coefficient. Recent work has studied the effect of measurement error in capital inputs on production function estimation. Their approach has been using an IV (Collard-Wexler and Loecker 2020) or a centering condition *à la* Hu and Schennach (2008) (Charlotte's JMP). However, systematic misreporting is different from measurement error in that it goes only one way and it follows an IC, as discussed above. Likewise, because capital is not a flexible input[5], the focus of these papers is on the effect of the marginal return to capital coefficient rather than on the effect on productivity estimates.

Moreover, I focus on gross-output production functions, as opposed to value-added, to reduce the noise that might be introduced by firms cheating also on prices. Besides the fact that prices are an equilibrium object and might be affected by demand also —and not only by supply—, firms might also artificially increase the price of inputs or reduce the price of sales to reduce declared profits. Then, in that sense, the tax-evasion bias magnitude of productivity estimates obtained through a value-added production function would be greater than through the gross-output production function. In the latter, only the price of the flexible input —intermediates—, and the output sale price enter through the first-order conditions (FOC) of the profit maximization problem of the firm[6]. In contrast, the prices of the non-flexible inputs, —capital and labor—, might introduce noise in the case of the value-added production function. Note: tax evasion under-reporting sales: use firms that sell to other firms because of opposite incentives in contrast to firm-end consumer sales. Unskilled labor might be flexible but it's the least productive, so less damage.

II. Relation to the literature, or originality

Note: Unfinished section

Having discussed the paper's idea and motivation in the Introduction, where I

[5]In the literature, a flexible input is neither dynamic —current input choice depends on its lagged value—, nor predetermined —completely defined in the past.

[6]If the cross-derivative of the flexible input with respect to the non-flexible inputs is not zero, then systematic misreporting in the non-flexible inputs will also affect productivity.

try to highlight the *clarity* of the question, and why is *interesting* and *important*, in this section, I highlight the *originality*. I, therefore, discuss how this paper's question is different from current literature and the paper's potential contribution to different research areas.

A. *Production functions and productivity*

Production functions and measurement error in capital [Charlotte's JMP; Collard-Wexler and Loecker (2020)]. Capital is a non-flexible input. Measurement error is different from systematic under/over-reporting due to tax evasion.

B. *Tax evasion*

Individuals' tax evasion linking tax records and survey data. Paulus (2015) jointly models tax evasion and measurement error.

Work studying a causal link between productivity and tax evasion. Dabla-Norris et al. (2019) argues that productivity improvements by firms lead to lower tax evasion. However, the authors use a self-reported percentage of unreported sales as tax evasion from surveys, and productivity is measured as the log of sales per worker.

C. *Measurement error*

Measurement error (Hu and Schennach 2008; Hu and Yao 2022; Schennach 2021) is different from systematic under/over-reporting from after-tax profit-maximizing incentives because the latter follows an incentive constraint and goes only in one direction.

D. *Lit Review Notes*

For personal reference, to be removed from the final version.

Productivity improvements by firms can causally lead to lower tax evasion, i.e., underreporting sales for tax purposes. In other words, an increase in firm productivity leads to a higher share of sales reported for tax purposes. The authors use the self-reported share of declared income as a proxy for tax evasion.

Limitations: Using the World Bank Enterprise (47 emerging and developing countries) survey question to get the percentage of sales that are reported for tax purposes, which is used to measure tax evasion. In addition, productivity is measured as the log of sales per worker. No access to quantities data, they use LP but with sales and workers.

Undisputed (?) facts: 1) Tax evasion is higher in poor countries; 2) Tax evading firms tend to be less productive (?).

Productivity gaps between firms that comply with taxes and regulations and those which do not are significant, (between 25 to 50 %) (Amin et al. 2019; Fajnzylber, 2011; Busso et al., 2012). These large productivity gaps can translate into low-economy wide productivity and growth: 1) if, by not complying with existing taxes and regulations, firms enjoy a potentially large implicit subsidy that allows them to stay in business despite low productivity, or 2) to expand their market share at the expense of more productive firms (Farrell, 2004; Bobbio, 2016).

Almeida and Carneiro (2009): Tax evasion may also increase firms' efficiency (productivity?). Avoiding onerous taxes might confer firms more flexibility in their employment and production decisions.

Kleven et. al (2016): Larger firms are less likely to avoid taxes. firms' use of business records increases as it hires more employees. They show that the government may be able to enforce higher tax compliance even in the presence of a low threat of audit due to the increased ease of whistle-blowing as the firm grows.

tax evasion is hidden so that it is difficult to obtain precise measures (Kundt et al., 2017; Slemrod and Weber, 2012)

Loayza and Rigolini (2006) and La Porta and Shleifer (2014): informality is larger in lower GDPpC countries

Tax evasion and informality are associated with lower income levels and productivity (Dabla-Norris and Feltenstein, 2003; Loayza, 1996; Sarte, 2000)

@Paulus2015

The author models discrepancies between tax records and income survey data as individual tax evasion and measurement error jointly. Using data for Estonia, the author finds that people in the bottom and the top part of earnings distribution evade much more and about 12

The main constraint for empirical research on tax evasion is the lack of suitable data. This is worse for developing countries. A measure of undeclared income for individuals is needed. This kind of data is usually unreliable and very difficult/or expensive to obtain.

Baldini et al. (2009) use linked income data with tax records at the individual level to estimate tax evasion. However, the author assumes people report their true income in the survey to obtain a measure of non-reporting. Furthermore, in the measurement error literature, several studies assume the opposite, e.g. Krueger (1991) and Bollinger(1998); survey of lit Chen et al. (2011) and Bound et al. (2001). Linked data is referred to as Validation data in the measurement error literature. Later studies have started to consider matching errors (Kapteyn and Ypma, 2007; Meijer et al., 2012) and errors in register data (Abowd and Stinson, 2013).

Pissarides and Weber (1989) pioneered a methodology to study underreported income. They assume that the underreported income for one population group is inferred from a comparison with a reference group assumed to have negligible non-compliance but to be similar in other respects. These types of studies assume implicitly that the underreporting of income in a survey corresponds to the underreporting of income to the tax authority.

@Slemrod2019

Tax-evasion policies have been a greater priority in developed countries since the 2008 great recession (IMF, 2015). For example, the UK announced efforts to combat tax avoidance, tax evasion, and 'imbalances' in the tax system that would bring in £5 billion in additional revenue each year.

Tax evasion matters because it affects both the resource cost of raising taxes and the distribution of the tax burden, i.e. core interests of public economics. Policy question? Does curbing evasion improve the equity and efficiency of the public finances, given the cost of doing so? If so, what's the best way to go about it?

Tax evasion creates horizontal inequity because people evade taxes by different proportionate amounts because of differences in personal characteristics —risk aversion, tax system, and honesty— and because of different opportunities and potential rewards for evasion. Likewise, efforts to combat tax evasion might create vertical inequity. For example, going after lower-income households such as the IRS fighting fraud related to the earned income tax credit (EITC) than going after unreported foreign accounts that are more likely to be held by high-income households.

Tax evasion creates efficiency costs, e.g., individual's cost of implementing and hiding non-compliance, tax authority's cost of combating non-compliance, etc. Tax evasion also creates socially inefficient incentives to engage in activities that facilitate tax evasion. For example, tax-noncompliance opportunities attract people to self-employment who would otherwise be employees.

Tax evasion creates misallocations. For example, it can cause companies that otherwise would not find it attractive to do so to set up operations in a tax haven to facilitate or camouflage evasion.

Curtailing tax evasion cost shapes optimal policy —it is not optimal to completely eradicate tax evasion because of the cost of doing so. The presence of evasion also alters the optimal setting of tax rates —because it affects the marginal efficiency cost of so doing— and the choice of the tax base —because different tax bases are more or less susceptible to evasion.

Slemrod and Weber (2012) review macroeconomic measures of the informal economy.

The standard tax evasion framework is the model of tax deterrence by Allingham and Sandmo (1972), who adapted Becker's (1968) model of criminal behavior to tax evasion. A relatively recent simplified framework is where the maximand is $y(1-t)+te-c(e,\alpha)$, y is income, e is the understated tax liability, t is a linear tax rate, and c is the private cost of evasion which includes the utility cost of bearing risk and the expected value of punishment. The private cost may depend on the amount of attempted evasion and α, the vector of enforcement instruments, such as the extent of auditing.

The model predicts that a risk-averse agent will evade taxes if the expected return is positive. Additionally, an increase in either the probability of detection or the penalty if detected will reduce evasion. Importantly, the effect of the tax rate is less clear, for even its sign depends on the penalty function (Yitzhaki, 1974).

What matters about p is the perception of taxpayers. Much work is focused on what alters the taxpayer's perception of the enforcement environment.

This framework naturally applies to individuals, and small, single-owned firms. Large firms might be risk-neutral rather than risk-averse.

The challenge of measuring tax evasion is the incentive of evaders to conceal their behavior. Many taxpayers are unwilling to respond accurately maybe because of the threat of punishment or social shame. For this reason, even the most credible empirical studies of evasion don't actually have a reliable measure of evasion.

Tax gap studies by auditing (real differences between countries' tax gaps or differences in methodologies?):

- U.S. Overall gross tax-gap average over 2008-10 was $458 billion USD, 18.3% of estimated tax liability (IRS, 2016). IRS National Research Program (NRP).

- UK: overall tax gap from 2013-2014 6.4% of true liability: 6.4% for the corporation income tax 5.0% individual income tax, and 11.1% value-added tax (VAT). Small- and medium-sized enterprises account for over half of the overall tax gap (HMRC, 2015). Her Majesty's Revenue and Customs (HMRC, 2015) Toro et a. (2013).

- Denmark overall tax gap is only 2.2 percent of net income (Kleven et al., 2011).

- OECD(2015) other countries.

Sharp variations in the compliance rate by the extent of information reporting to the tax authority provide compelling evidence that the probability of successful evasion play a key role in the deterrence of tax evasion. For example, in the US, the estimated non-compliance rate when there is little to no third-party-reported information (e.g., self-employment income) is 63%, compared to 19% and 7% for 'some' and 'substantial' information reporting, respectively. Small businesses represent a large portion of the tax gap in individual income, approximately 47% of underreporting. In contrast, only 1

Administrative data: Canada, UK, Brazil, Chile, China, Costa Rica, Ecuador, India, Pakistan, Rwanda, Tunisia, and Uganda. US access https://www.irs.gov/pub/irs-soi/16jsrpprojects.pdf

RCT's review: Hallsworth (2014)

Bérgolo et al. (2018) RCT with firms using letters. Value-added-tax compliance, 20,000 Uruguayan small and medium-sized firms. The authors find that adding to the baseline letter a paragraph with statistics about the probability of being audited and the penalty rates increased tax compliance by about 6.3 percent (about one-fourth of its current level), and adding a paragraph that informs firms that evading taxes increased the probability of being audited increased tax compliance by about 7.4 percent.

Ortega and Scartascini (2018) Colombia, individuals. Method of communication. The authors find sizable differences across delivery methods. Personal visits by a tax inspector are more effective than the impersonal methods; they are, alas, also much more expensive. In a follow-up study, Ortega and Scartascini (2015) find that the effect of phone calls falls between those of the impersonal methods and the personal visits.

Almunia and Lopez-Rodriguez (2018) study the behavior of Spanish firms in response to a notch in enforcement intensity that arises because the Spanish Large Taxpayer's Unit (LTU) monitors firms with revenues above €6 million for all major taxes. Sure enough, there is significant bunching of firms just below the threshold. The bunching is more pronounced for firms that produce intermediate goods, which makes sense because their transactions create more of a paper trail than firms that sell to final consumers, and thus the discontinuous increase in enforcement intensity affects these firms more than retailers.

Examining data from Ecuador, Carrillo, Emran, and Rivadeneira (2012) find evidence of bunching in reported tax liability just above the withholding threshold, suggesting that some firms manipulate their self-reported tax liability and possibly real economic choices to minimize expected tax payments subject to the discontinuity in the audit probability. Third-party data on sales and intermediate input costs reported by large firms who act as withholding agents indicate that bunching is indeed associated with tax evasion: self-reported sales are smaller than third-party reports for more than 10 percent of firms. "withholding", large firms withhold and remit to the tax authority a fixed share of their purchases from small firms (and individuals), who can then apply the withheld amount as a credit against their self-reported tax liability. While withholding does not change the firms' true tax liability, there is typically a discontinuity in the audit probability at the withholding rate; firms seeking tax refunds (because self-reported tax liability is lower than the withheld amount) are audited at a higher rate than firms making additional tax remittances.

In Chile, firms can only claim tax credits for inputs bought from tax-compliant suppliers, the invoice–credit VAT system has a built-in (albeit imperfect) self-enforcement mechanism for firm-to-firm transactions. Pomeranz (2015) tests this hypothesis by mailing increased-audit-threat letters to over 100,000 randomly selected Chilean firms, using a sample of over 300,000 firms receiving no letter as the control group. Consistent with theoretical predictions on the self-enforcement mechanism, the increase in VAT receipts (and therefore the inferred level of evasion) induced by the letters is concentrated at the level of sales from firms to final consumers, for which there is no paper trail.

In Ecuador, Carrillo, Pomeranz, and Singhal (2017) examine the effect of a change in the tax authority's use of third-party information on reported corporate tax revenues. In the authors' study, the Ecuadorian tax authority informed some firms of the discrepancy between the two reports and offered them the opportunity to file an amended return. The authors compare the reporting behavior of firms before and after the notification and find that 24 percent of firms underreported revenue in the years when the government did not use the third-party-verified information. In the three rounds of the experiment, 11–19 percent of notified firms filed an amended return.

In amended returns, firms correctly reported their revenues but also increased their reported costs almost one-for-one with the increase in reported revenues (96 cents for each dollar!), nearly eliminating any impact on apparent tax liability. This study also reveals that reported costs were lower than third-party information on costs, which at first blush seems to be at odds with a model of firms that seek to maximize after-tax profits. The authors propose that this behavior is consistent with firms who may believe that the probability of an audit is a function of firm size and profits. To appear small, and lower the odds of being audited, some firms may underreport both revenues and costs.

Kumler, Verhoogen, and Frías (2013) study the effects on evasion of a 1997 pension reform in Mexico that tied younger workers' retirement benefits more closely to the wage reported by employers. This reform provided a new incentive for this group of workers to ensure that their employers accurately reported their wages, which in turn might lower payroll tax evasion by firms. To examine the impact of this initiative, the authors combine two sources of data on wages: administrative data from the Mexican Social Security Insitute (IMSS) and household survey data from the Encuesta Nacional de Empleo Urbano (ENEU). The authors find that, consistent with a decline in evasion, the gap in median or mean wage within a cell between the ENEU data and the IMSS data fell for younger age groups after the pension reform but, as predicted, for older workers not affected by this reform there was no decrease in the gap between the two income reports.

In Pakistan, corporations either pay a tax on profits or turnover, depending on which liability is greater. This effectively implies that at a profit rate lower than the ratio of the turnover tax rate to the profit tax rate, firms cannot deduct costs. Because in Pakistan a large portion of evasion is through misreporting of costs, this tax regime trades off loss in production efficiency (as firms move from a neutral profit tax to a distorting output tax regime) for a gain in revenue collection. Best et al. (2015) use administrative data on the universe of corporations in Pakistan to estimate the elasticity of taxable income using the bunch-ing of firms below the threshold profit rate. They find clear evidence of such bunching, whose location shifts along with changes in tax rates that move the threshold. Using the analysis-of-bunching methodology, they estimate that turnover taxes reduce evasion by between 60–70 percent.

Relevance of tax evasion Widespreadness of tax evasion Seminal model of tax evasion Real effects of tax evasion firm-to-firm sales less likely to evade taxes small firms more likely to evade taxes evidence of misreporting of costs (Pakistan) Chile and Colombia tax evasion evidence

III. Tax evasion and productivity in Mexico and Colombia

Note: Unfinished section

To lay out the context of the empirical approach and justify some assumptions during the estimation, this section provides a general overview of the corporate tax system in both Mexico and Colombia, and firm-level productivity estimates from previous work. The objective is to highlight the *relevance* of tax evasion in both countries and how it relates to the *critical* questions regarding productivity.

A. The corporate tax system and tax evasion

In Mexico, firms have incentives to stay small or declare profits below 2 million Mexican pesos (MXN) (Levy 2018).

B. Productivity

Productivity has stagnated in Mexico during the last decades despite macroeconomic stability and significant growth in human capital (Levy 2018). Are the firm-level productivity measures reliable considering the potentially high tax evasion by firms?

IV. The model

To draw intuition on the mechanisms and consequences of tax evasion on productivity measurement, I set up a simplified model. I use then the model to guide Monte Carlo simulations and explore the results of a counterfactual policy aiming to reduce the productivity gap by a reallocation of resources from low to high-productivity firms. I contrast the result of the policy using a naive measure of productivity —ignoring the effect of tax evasion— with the results of using the true measure.

The simulations show that there is a reshuffling of firms in the naive measure of the productivity distribution; tax-evading firms move to the lower end. In addition, the mismeasured productivity distribution is more spread out than the underlying true productivity. The mismeasured productivity distribution is higher spread apart because the productivity of firms at the low end of the distribution is underestimated up to the 80th percentile, while the top end of the distribution is overestimated.

The reshuffling of firms in the mismeasured productivity distribution is relevant for policies; using the wrong measure will lead to failure of achieving their target. Tax evading firms appear to be less productive than they are, but their likelihood to evade taxes is independent of true productivity. Therefore, if a policy uses a naive productivity measurement to reduce the productivity gap by targeting the lower quartile of the distribution, the result is that the true productivity gap will not be significantly reduced. A secondary result is that the tax gap —the ratio of the tax revenue the government is collecting to what it *should* collect if firms were not evading taxes— decreases because the share of firms that are likely to evade shrinks. If the policy uses the true measure, the productivity gap is effectively reduced but the tax gap is unaffected.

In the model, price-taking firms maximize after-tax profits in a two-stage game. At the start, firm $i \in \mathcal{I}$ gets a random draw of productivity ω_i with probability distribution function (pdf) f_ω and size s_i with pdf f_s.

In the first stage, firms maximize before-tax profits given input and output prices, ρ and P, the production function $G(\cdot)$, and their productivity ω_i.

$$\max_{x_i \in \mathcal{X}} PG(x_i)e^{\omega_i} - \rho x_i$$

Their optimal input choice is x_i^*. In other words, let $x_i^* = argmax_x PG(x_i)e^{\omega_i} - \rho x_i$. Then, they produce output $y_i = G(x_i^*)e^{\omega_i}$.

In the second stage, firms choose how much inputs to overreport to maximize their after-tax profits, given the tax rate τ, their optimal input choice x_i^* of the first stage, their size s_i, and the tax-evasion cost. The cost function of the tax-evasion decision is increasing and convex on the overreporting —the greater the difference from the true input the higher the cost— and increasing in the firm's size.

$$\max_{x \in \mathcal{X}} (1-\tau)Py_i - \rho(x_i^* - \tau x_i) - C(x_i - x_i^*, s_i)$$

Likewise, the optimal overreporting input choice is $\tilde{x}_i$. That is, $\tilde{x}_i = argmax_x (1-\tau)Py_i - \rho(x_i^* - \tau x_i) - C(x_i - x_i^*, s_i)$

Finally, firms choose to overreport if the after-tax profits, $V^E = (1-\tau)Py_i - \rho(x_i^* - \tau\tilde{x}_i) - C(\tilde{x}_i - x_i^*, s_i)$, are greater than reporting the truth, $V^T = (1-\tau)Py_i - \rho x_i^*$. In summary,

$$\max\{V^T, V^E\}$$

$$x_i - x_i^* = \begin{cases} \geq 0 & , V^E \geq V^T \\ 0 & , \text{otherwise} \end{cases}$$

The first-order conditions of the first stage are well-known, $G_x(x_i^*)e^{\omega_i} = \frac{\rho}{P}$. Firms choose inputs such that the slope of the production function equals the input/output price ratio. From this FOC, we can solve for the optimal input choice $x_i^* = G_x^{-1}(\frac{\rho}{P}e^{-\omega})$. Here is easy to observe why input demand is strictly monotonic in productivity; the higher the productivity the lower the quantity demanded by firms.

On the second stage, the FOC, $\tau\rho = C_x(\tilde{x}_i - x_i^*, s_i)$, reveal that overrerporting is increasing in the tax rate and input prices, but decreasing in firm size, $\tilde{x}_i - x_i^* = C_x^{-1}(\tau\rho, s_i)$. Productivity does not directly influence tax-evasion overreporting, but only through the optimal input choice x_i^*.

A. Monte Carlo simulations

To run the Monte Carlo simulations, I choose the following functional forms for the production function and the tax-evading cost:

$$G(x_i)e^{\omega_i} = x_i^{\beta} e^{\omega_i}$$
$$C(x_i - x_i^*, s_i) = (x_i - x_i^*)\gamma s_i + \alpha^{(x_i - x_i^*)}$$

With these functional forms, the FOC of the first and second stage become

$$x_i^* = \left(\frac{\beta P e^{\omega_i}}{\rho}\right)^{\frac{1}{1-\beta}} \tag{1}$$

$$\tilde{x}_i = \log \frac{\left(\frac{\tau\rho - \gamma s_i}{\log \alpha}\right)}{\log \alpha} + x_i^* \tag{2}$$

The parameters are set as indicated in the next table.

Parameter	Value
β	0.4
γ	0.0059
α	1.05
τ	0.30
ρ	1.00
p	2.00

Finally, $\omega_i \sim N(0,1)$ and $s_i \sim U[1,50]$. I generate 1,000 firms and simulate their decisions.

Distribution of true and reported profits and inputs

Measuring productivity

Table 2—Recovering productivity-Observed inputs

	OLS	IV	2SIV	FOC
β	0.151***	0.215***	-0.064***	0.633***
	(0.018)	(0.020)	(0.024)	(0.031)
N	1000	1000	1000	1000
R^2	0.056	0.105	-0.002	-0.624

Different estimates for the marginal output of input of the production function using the observed values of input.

Table 3—Recovering productivity-True inputs

	OLS	IV	2SIV	FOC
β	0.971***	0.997***	0.423***	0.400***
	(0.004)	(0.001)	(0.138)	(3.016e-18)
N	1000	1000	1000	1000
R^2	0.983	0.999	2.236e-04	1.000

Different estimates for the marginal output of input of the production function using the true unobserved values of input.

Policy counterfactuals

V. Empirical framework and data, or feasibility

In this section, I argue that it is *feasible* to answer the question regarding the estimation of firm-level productivity in the presence of tax evasion. First, I outline the *identification* and *estimation* strategy. Namely, I explain how we can build upon the proxy variable literature and recent developments in the measurement error literature to account for unobserved systematic deviations from the truth by employing an incentive constraint. Moreover, I sustain that our identification assumption —that large firms are less likely to overreport intermediate inputs— is not unreasonable. Finally, I describe the *data* that I use in the estimation.

A. Empirical framework

Assuming momentarily that firms do not cheat on sales or prices for the sake of exposition, firms will, then, overreport inputs if the expected profits of cheating is greater than the profits of reporting the truth. In other words, firms cheat if the incentive constraint is binding, otherwise they don't. Let $\Pi(X_{it}) = P_t Y_{it} - \rho_t X_{it}$ be the price-taking firm's profits. Therefore, the input quantities we observe in data are as follows:

$$X_{it} = X_{it}^* + \varepsilon_{it}^X(S_i) \tag{3}$$

$$\varepsilon_{it}^X(S_i) = \begin{cases} 0, & (1-\tau)\Pi(x_{it}^*) \geq \Pi(x_{it}^*) - \tau\Pi(x_{it}) - C(x_{it} - x_{it}^*, s_i) \\ (0, \tilde{x}], & \text{otherwise} \end{cases} \tag{4}$$

X_{it} is the observed input quantities in data, while X_{it}^* is the true input quantity consumed by the firm, and τ is the tax on firms' profits. Once firms reach zero profits, firms don not have incentives to over-report more Note: Firms get tax benefits for (past) losses? Alternatively, upper bound could be inf, then the probability of being audited increase $C(\cdot)$ goes up.. Consequently, $\Pi(\tilde{x}) = 0$. Lastly, $C(\cdot)$ is the cost function of cheating. This cost function might be decomposed into $C(\cdot) = \kappa Pr(a) + c(X_{it} - X_{it}^*, S_i)$, where κ is the economic penalty for cheating, $Pr(a)$ is the probability of being caught cheating, and $c(\cdot)$ is a convex cost function that depends on how much the firm departs from the truth $X_{it} - X_{it}^*$ and its size S_i.

Now, we're interested to estimate the gross output production function,

$$Y_{it} = G(X^*_{it})e^{\omega_{it}+\varepsilon^Y_{it}} \tag{5}$$

Y_{it} is the gross output for firm i at time t. ω_{it} is the scalar productivity —the object of interest— and ε^Y_{it} is the classical measurement error. The output is a function $G(\cdot)$ of the true unobserved input X^*_{it}, measured in quantities.

The properties of this functional form of the production function have been well studied by the proxy variable literature (Ackerberg, Caves, and Frazer 2015; Gandhi, Navarro, and Rivers 2020). In particular, there exists a simultaneity problem between the inputs and productivity. We follow this literature and assume that productivity ω_{it} is scalar and follows a Markov process, input demand is strictly monotonic in ω_{it}, and that the classical measurement error ε^Y_{it} is uncorrelated with the inputs and across time.

A simple way to see how tax evasion biases downward the productivity ω_{it} is to assume a log-linear production function like the Cobb-Douglas (CD). Then,

$$\begin{aligned}
\mathbb{E}[\log Y_{it}|x] &= \mathbb{E}[\beta \log X_{it}|x] + \mathbb{E}[\omega_{it}|x] \\
\mathbb{E}[\log Y_{it}|x] &= \mathbb{E}[\beta \log(X^*_{it} + \varepsilon^X_{it})|x] + \mathbb{E}[\omega_{it}|x] \\
\mathbb{E}[\omega_{it}|x] &= \mathbb{E}[\log Y_{it}|x] - \beta\mathbb{E}[\log(X^*_{it} + \varepsilon^X_{it})|x]
\end{aligned}$$

Therefore, in the case of the CD production function, the bias is

$$
\begin{aligned}
\Delta_\omega &= -\beta\left(\mathbb{E}[\log(X_{it}^* + \varepsilon_{it}^X) - \log(X_{it}^*)|x]\right) \\
&= -\beta\left(\mathbb{E}[\log\left(\frac{X_{it}^* + \varepsilon_{it}^X}{X_{it}^*}\right)|x]\right) \\
&\leq \beta\left(\log(\mathbb{E}\left[\frac{X_{it}^* + \varepsilon_{it}^X}{X_{it}^*}|x\right])\right) \\
&\leq 0
\end{aligned}
$$

by Jensen's inequality and because $\varepsilon_{it}^X \geq 0$ by **??**.

In this paper, the key identification assumption is that larger firms do not overreport inputs. For larger firms, it is costlier to keep the double accounting because more people need to get involved. Likewise, the larger the firm the higher the probability to be denounced to the authority by an anonymous employee just because more eyes are watching. Finally, larger firms have access to other legal tools to reduce their tax liabilities, which might be more profitable than tax evasion. Let $L(X_{it}^*, S_i) = \tau[\Pi(X_{it}^*) - \Pi(X_{it})] - C(X_{it} - X_{it}^*, S_i)$. Therefore, the location condition for large firms $S_i = s$ is

$$
\mathbb{E}[L(X_{it}^*, S_i)|X_{it}^*, S_i = s] = 0 \tag{6}
$$

In other words, the expected return of cheating conditional on being a large firm is zero.

In addition, I follow the proxy variable literature in using the first-order conditions from the firm's profit-maximization problem and the flexible input assumption. In particular, the firm's optimization problem is the following:

$$
\max_{X_{it}^*} P_t \mathbb{E}[G(x_{it}^*) e^{\omega_{it} + \varepsilon_{it}^Y}] - \rho_t X_{it}^* \tag{7}
$$

From the first-order conditions of **??**, equation **??**, multiplying by X_{it}^*/P_t, taking logs, and rearranging, we get the share equation,

$$\ln\left(\frac{\rho_t X_{it}^*}{P_t Y_{it}}\right) = \ln\left(\frac{G_x(x_{it}^*) X_{it}^*}{G(x_{it}^*)}\right) - \varepsilon_{it}^Y \tag{8}$$

B. Estimation strategy

We can state our problem as recovering the joint distribution $f(y_{it}, x_{it}, p_t, s_i) = f(y_{it}, x_{it}, x_{it}^*, p_t, \omega_{it}, \omega_{it-1}, s_i)$ from the data $\mathcal{O} = \{Y_{it}, X_{it}, P_t, S_i\}_{i \in I, t \in T}$. Notice x_{it}^* and ω_{it} are unobserved. Additionally, we can rewrite our joint distribution in terms of the conditional and marginal probabilities densities.

Note that so far we have implicitly state before the following assumptions that we ennumerate below.

ASSUMPTION 1: *The production function from equation* **??** *satisfies* $f_1(y_{it}|x_{it}, x_{it}^*, \omega_{it}, \omega_{it-1}, s_i, p_t) = f_1(y_{it}|x_{it}^*, \omega_{it})$.

ASSUMPTION 2: *The share function from equation* **??** *satisfies* $f_2(p_t|x_{it}, x_{it}^*, \omega_{it}, \omega_{it-1}, s_i) = f_2(p_t|x_{it}^*)$.

ASSUMPTION 3: *The Markov process of the scalar productivity satisfies* $f_3(\omega_{it}|x_{it}, x_{it}^*, \omega_{it-1}, s_i, p_t) = f_3(\omega_{it}, |\omega_{it-1})$.

ASSUMPTION 4: *The systemaric underreporting from tax evasion satisfies* $f_4(x_{it}|x_{it}^*, \omega_{it}, \omega_{it-1}, s_i, p_t) = f_4(x_{it}|x_{it}^*, s_i)$.

Therefore, we can redefine our joint distribution as follows:

$$\begin{aligned}
&f(y_{it}, x_{it}, p_t, s_i) \\
&\quad = f(y_{it}, x_{it}, p_t, \omega_{it}, \omega_{it-1}, x_{it}^*, s_i) \\
&\quad = f_1(y_{it}|x_{it}^*, \omega_{it}) f_2(p_{it}|x_{it}^*, \omega_{it}) f_3(\omega_{it}|\omega_{it-1}) f_4(x_{it}|x_{it}^*, s_i) f_5(x_{it}^*|s_i) f(s_i, \omega_{it-1})
\end{aligned}$$

Note that $f(\omega_{it-1})f(x^*_{it}, s_i)$ follows from the Note: flexible input assumption: Don't need it.

Let $\mathcal{A} = (\Theta, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \mathcal{F}_4)$ be the parameter space of the functional forms of equations **??**, **??**, **??**, and **??** and of the conditional distributions $f_1, f_2, f_3,$ and f_4. Assuming $f(y_{it}, x_{it}, p_t, s_i)$ is time invariant, we can form the quasi likelihood function as follows:

$$\sum_{i \in I} \left(\sum_{t \in T} \log f(y_{it}, x_{it}, p_t, s_i) \right) \equiv \sum_{i \in I} l(D_i; \alpha)$$

in which $D_i = (Y_{it}, X_{it}, P_t, S_i)$ and

$$\begin{aligned}
l(D_i : \alpha) \equiv\ & l(D_i; \theta, f_1, f_2, f_3, f_4) \\
&= \sum_{t \in T} \log \Bigg\{ \int f_1(y_{it} - g(x^*_{it}) + \omega_{it}) \\
& f_2 \left(- \ln \left(\frac{\rho_t X^*_{it}}{P_t Y_{it}} \right) + \ln \left(\frac{G_x(x^*_{it}) X^*_{it}}{G(x^*_{it})} \right) \right) \\
& f_3(\omega_{it} - h(\omega_{it-1})) \\
& f_4(0|s_i)^{\mathbb{1}[L(x^*_{it}, s_i) \geq 0]} f_4(x_{it} - x^*_{it}|s_i)^{\mathbb{1}[L(x^*_{it}, s_i) < 0]} \\
& f_5(x^*_{it}|s_i) dx^* \Bigg\} \\
& + \sum_{t \in T} \log f(\omega_{it-1}, s_i)
\end{aligned}$$

Let $\mathbb{E}[\cdot]$ be the expectation from the true data generating process for D_i. Then,

$$\alpha_0 = \arg\ \sup_{\alpha \in \mathcal{A}} \mathbb{E}[l(D_i; \alpha)]$$

We can use a sequence of finite-dimensional sieves spaces[7] $\mathcal{A}_N = (\Theta, \mathcal{F}^N_1, \mathcal{F}^N_2, \mathcal{F}^N_3, \mathcal{F}^N_4)$ to approximate the functional space $\mathcal{A} = (\Theta, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \mathcal{F}_4)$. The semiparametric sieve MLE $\hat{\alpha} \in \mathcal{A}_N$ for $\alpha_0 \in \mathcal{A}$ is defined as:

[7] Hu and Yao (2022) use Hermite orthogonal polynomials because they work well with just a few sieve terms.

$$\hat{\alpha}_N = \text{argmax}_{\alpha \in \mathcal{A}_N} \sum_{i \in I} l(D_i; \alpha)$$

C. Data

I use anonymized annual tax declarations from a sample of Mexican firms and survey data at the establishment level for manufacturing firms in Mexico and Colombia collected annually. The main challenge is that in the Mexican data, although I observe sales value and production value, I do not observe input prices and quantities for intermediate inputs in the survey data —only the value—, and the tax records cannot be linked to the survey records. However, Colombian data on firms does include input prices.

The anonymized records come from the tax agency in Mexico *Secretaria de Administración Tributaria* (SAT), while the annual survey of Mexican manufacturing firms (*Encuesta Anual de la Industria Manufacturera*, EAIM) is collected by the National Institute of Statistics and Geography (INEGI) of Mexico. Although the tax records and the survey datasets cannot be linked, the tax records can be informative about the assumptions we make during the estimation.

The tax records include the annual tax declarations of a sample of 520,000-720,000 firms, between 2010-2015, depending on the year. The data includes information about the total income, profit/losses before taxes and before paying the share of profits to employees (PTU), taxes on profits (ISR), and credits and reductions for different reasons, e.g., reductions for Maquiladoras, R&D credits, among others. It also includes some firm characteristics like country region, broad industry classification, e.g, service, commerce, manufacturing, etc., and classification by —*what seems to be*— the number of employees Note: Not sure about this one, though. Description only says Range_date. Table **??** describe the summary statistics for this data.

The Manufacturing Industry Annual Survey (INEGI-EAIM) data contains information at the establishment level on manufacturing firms. The survey classifies

Table 4—Tax records summary statistics

Variable	N	Mean	S.D.	Min	Max
Total income	56712	63.966	303.858	0	5695.213
Total deductions	47871	71.983	310.741	0	6130.882
Utility/Losses (before PTU)	54945	6.874	822.602	-885.261	192681
Losses from previous years	11680	2.266	25.251	0	2185.51
Tax on rents (ISR)	56712	1.016	7.651	0	394.477
Range (# of employees?)	56718				
... 1 a 5	8752	15.4%			
... 11 a 20	24453	43.1%			
... 21 a 25	4421	7.8%			
... 6 a 10	16241	28.6%			
... Mayor a 25	2851	5%			
Geographic zone	56718				
... CENTRO NORTE	8062	14.2%			
... CENTRO OCCIDENTE	27158	47.9%			
... NORTE	13909	24.5%			
... SUR ORIENTE	7589	13.4%			

Annual tax declarations of mexican manufacturing firms in 2014, millions of Mexican Pesos —MXN. Source: SAT Mexico.

the manufacturing sector according to the North American Industrial Classification System (NAICS, 2013). The survey covers the years 2013-2020 and it is representative at the national level by industry sector, disaggregated at the subsector, branch, and economic activity class according to the 2013 NAICS. The data can only be accessed by remote estimation or on-site at the INEGI data laboratories in Mexico City.

The INEGI-EAIM includes the following variables: sales value, production value, intermediate inputs —raw materials, packaging, fuel, energy— value, labor quantity and total wage bill, capital value, and stock of raw materials, work-in-process, and production at the end of the year. Table **??** shows the summary statistics for a small sample of observations publicly available to test programming code before submitting it to remote processing at INEGI.

Table 5—EAIM sample summary statistics

Variable	N	Mean	S.D.	Min	Max
Sales value	15	2.069	0.898	0.878	3.824
Production value	15	2.077	0.892	0.878	3.824
Intermediate inputs —raw materials	15	1.394	0.703	0.764	2.791
Labor quantity	15	0.188	0.231	0.002	0.663
Wage bill	15	0.072	0.096	0.005	0.298
Capital value	15	0.39	0.343	0.194	1.142
Stock of raw materials	15	-0.03	0.111	-0.43	0.012
Work-in-process stock	15	0	0	-0.001	0.001
Production stock	15	-0.007	0.027	-0.099	0.019

EAIM survey sample of Mexican manufacturing firms in the animal food industry (311110), 2016-2018, millions of Mexican Pesos —MXN. Source: INEGI Mexico.

The Colombian Annual Manufacturer Survey (*Encuesta Anual de Manufactura*, EAM) recollects information on manufacturing plants in Colombia with a minimum income of 500 million Colombian pesos (COP) or 10 employees. The EAM includes the production value, sales value, raw materials consumed and its value, employees and payroll, as well as the book value of capital. The table **??** present the summary statistics of the 2018 survey.

Table 6—EAM 2018 summary statistics

Variable	N	Mean	S.D.	Min	Max
Production value	7911	32.363	159.813	0.003	5466.184
Sales value	7911	31.819	158.873	-1.735	5474.678
Raw materials consumed	7911	18.589	132.207	0	5191.503
Raw materials value	7911	18.371	130.224	0	5060.732
Employees	7911	0	0	0	0.002
Payroll	7911	1.621	3.623	0	69.758
Capital in book value	7911	21.365	170.869	0	6588.104

EAM survey of Colombian manufacturing firms, 2018, millions of Colombian Pesos —COP.
Source: DANE Colombia.

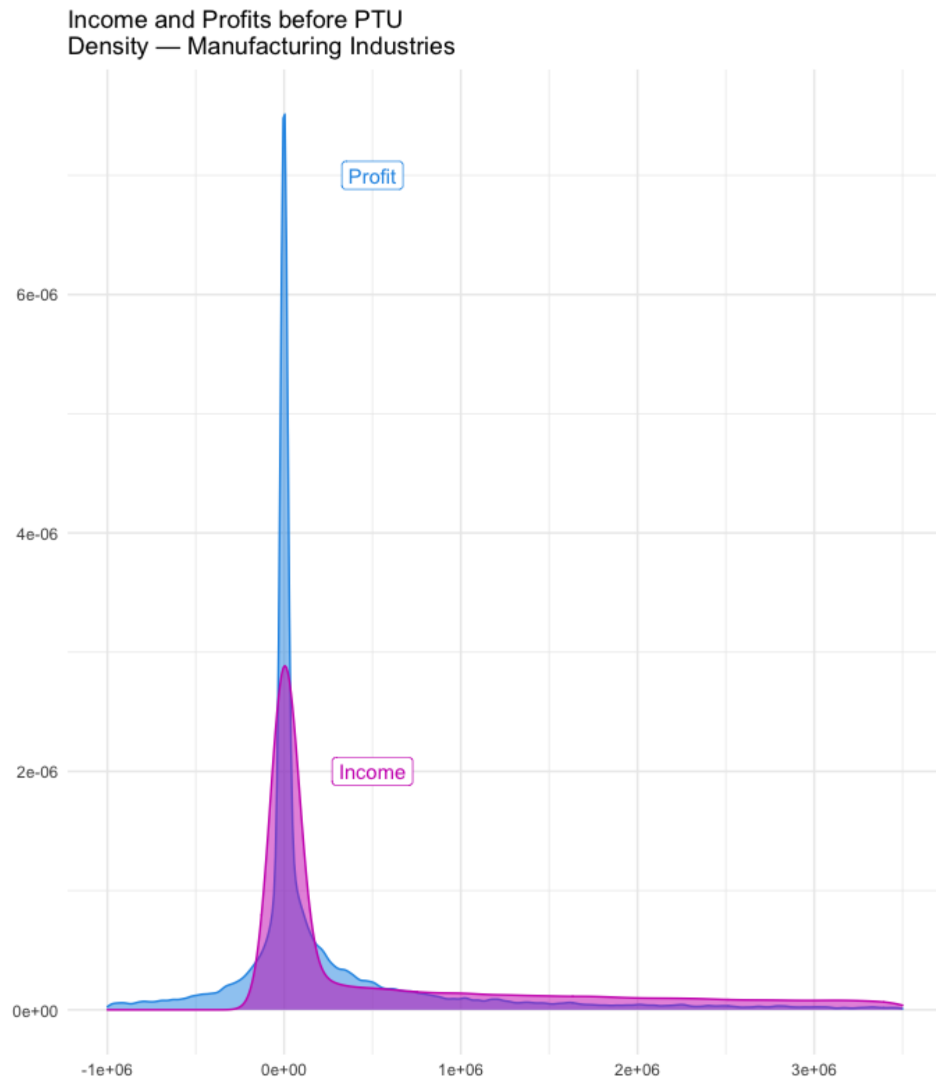
D. Exploratory analysis

During a preliminary exploratory analysis of the tax records, I find highly-concentrated profit and income distributions at zero (figure **??**). However, the profit distribution is more concentrated than the income. Moreover, the concentration is higher as the firm size decreases (figure **??**). It is highest for 1-to-5 employee firms than for firms with more than 25 employees.

E. Expected challenges

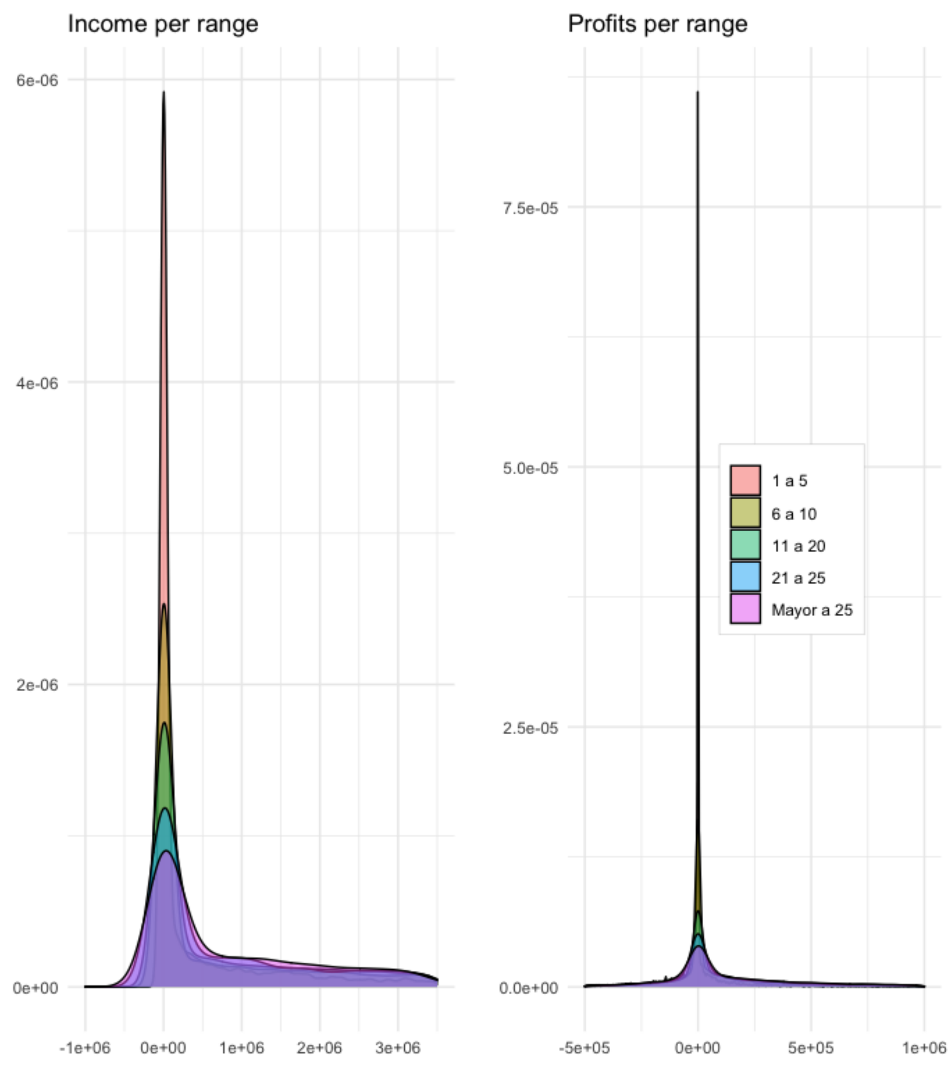
- Firms might cheat on sales. Possible solution: Tax evasion in sales needs collusion between selling and buying parties. I can focus on business-to-business (B2B) sales firms instead of business-to-consumer (B2C) ones. Counter-incentives will likely eliminate collusion incentives decreasing tax evasion noise —buying firm wants to over-report expenses, selling firm wants to under-report sales. Whereas in the latter, consumers might be willing

FIGURE 1. INCOME AND PROFIT DENSITIES



Income and profit densities of a sample of manufacturing firms in Mexico in millions of Mexican pesos —MXN. Source: SAT Mexico.

FIGURE 2. INCOME AND PROFIT DENSITIES BY RANGE



Income and profit densities of a sample of manufacturing firms in Mexico by range (number of employees?) in millions of Mexican pesos —MXN. Source: SAT Mexico.

to collude with firms by getting a lower price if they request no invoice, for example.

- Firms might cheat on input prices. Possible solution: for the base case, I can assume no cheating on prices and argue it's a lower bound for tax evasion bias. For robustness, I can model cheating on prices for the case in which data includes input prices.

- Firms might also report losses and use past losses to reduce the tax burden of the current period. Counter-argument: if firms keep reporting losses or the reported loss is too high, then the probability of getting audited and caught cheating increases, increasing the cheating cost. So, the probability of over-reporting expenses such that there are losses instead of zero profits is low. Data shows a high concentration around zero.

- No input prices on Mexican data. Possible solution: because firms are price-taking, then a time-dummy can take care of it if firms do not cheat on prices.

- In Mexico, if firms are classified as Small or Medium, they are subject to a lower tax rate. Solution: tax rate should adjust depending on the firms' classification. I think classification depends on the number of employees and sales level.

- Colombia: what's the tax evasion evidence? what's the tax scheme? is there data on firms' taxes?

References

Appendix

List of Corrections

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "IDENTIFICATION

PROPERTIES OF RECENT PRODUCTION FUNCTION ESTIMATORS" 83: 2411–51. https://doi.org/10.3982/ECTA.

Blalock, Garrick, and Paul J. Gertler. 2004. "Learning from Exporting Revisited in a Less Developed Setting." *Journal of Development Economics* 75: 397–416. https://doi.org/10.1016/j.jdeveco.2004.06.004.

Collard-Wexler, Allan, and Jan De Loecker. 2020. "Production Function Estimation and Capital Measurement Error." NBER. http://www.nber.org/papers/w22437.

Dabla-Norris, Era, Mark Gradstein, Fedor Miryugin, and Florian Misch. 2019. "Productivity and Tax Evasion." CESifo. www.RePEc.org.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler. 2004. "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia." *Journal of Development Economics* 75 (December): 333–71. https://doi.org/10.1016/J.JDEVECO.2004.06.002.

Gandhi, Amit, Salvador Navarro, and David A Rivers. 2020. "On the Identification of Gross Output Production Functions."

Hu, Yingyao, and Susanne M Schennach. 2008. "INSTRUMENTAL VARIABLE TREATMENT OF NONCLASSICAL MEASUREMENT ERROR MODELS." *Econometrica*. Vol. 76. http://www.bls.census.gov/cps/cpsmain.htm.

Hu, Yingyao, and Jiaxiong Yao. 2022. "Illuminating Economic Growth." *Journal of Econometrics* 228 (June): 359–78. https://doi.org/10.1016/J.JECONOM.2021.05.007.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies* 70 (April): 317–41. https://doi.org/10.1111/1467-937X.00246.

Levy, Santiago. 2018. *Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico | Publications*. IADB. https://publications.iadb.org/publications/english/document/Under-Rewarded_Efforts_The_Elusive_

Quest_for_Prosperity_in_Mexico.pdf.

Paulus, Alari. 2015. "Tax Evasion and Measurement Error: An Econometric Analysis of Survey Data Linked with Tax Records." University of Essex, Institute for Social; Economic Research (ISER). http://hdl.handle.net/10419/126489.

Schennach, Susanne. 2021. "Measurement Systems."

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49 (June): 326–65. https://doi.org/10.1257/JEL.49.2.326.

Zumaya, Martin, Rita Guerrero, Eduardo Islas, Omar Pineda, Carlos Gershenson, Gerardo Iñiguez, and Carlos Pineda. 2021. "Identifying Tax Evasion in Mexico with Tools from Network Science and Machine Learning." http://arxiv.org/abs/2104.13353.