Tax Evasion and Productivity

Hans Martinez*
Department of Economics, University of Western Ontario

July 9, 2025

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.

Updates

- Tax Evasion and Productivity: Using PF to identify tax evasion through input overreporting
  - CD
    * Preliminary results of Tax Evasion: Moments, MLE (Truncated Normal and LogNormal)
    * Identification of PF parameters when h AR(1)
    * **New** Discussion: Variance of the error term and Choosing intermediates in GNR-CD.
    * **New** Preliminary results of PF parameters
    * **New** Productivity results
  - TransLog
    * Identification strategy for tax evasion
- Leveraging Tax Policy change to identify changes in tax evasion
  - Why? Relax common technology assumption
  - Triple difference identification strategy (with CD)

Next Steps

- Tax Evasion and Productivity
  - Back of Envelope estimates for Gov Losses in Tax Revenue
  - App of Productivity: Aggregate Productivity and Resource Misallocations
  - Translog Tax Evasion results
  - Translog PF identification strategy

    - Translog PF results
    - Translog Productivity results
- DiD
    - Relaxing parallel trends: Linear trends that in the absence of policy change would have continued

1 TODO

- Using ELVIS to allow for more flexible functional forms of h the markov process of productivity
- Non-parametric Deconvolution
- Non-parametric PF

2 A parsimonious model of tax evasion through input overreporting

Price-taking firms maximize expected after-tax profits. Firms choose the flexible input M_{it} to produce output Y_{it} given output and input prices $\{P_t, \rho_t\}$, a common technology, the production function (Equation **??**), and their productivity ω_{it}.

$$Y_{it} = G(M_{it}) \exp(\omega_{it} + \varepsilon_{it}) \tag{1}$$

As standard in the literature, productivity ω_{it} is known to firms when they make input decisions. This is the well-known endogeneity problem of simultaneity. On the other hand, firms face output shocks. The output shock ε_{it} is not part of the firms' information set.

The model departs from the literature by allowing firms to overreport their inputs e_{it} to reduce their tax burden and optimize after-tax profits. Firms, then, consider in their optimization problem the profit tax τ, the evasion penalty/cost $\kappa(e)$, and the probability of detection $q(e_{it}|\theta_{it})$.

Firms solve Equation **??**

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

The probability of detection $q(e_{it}|\theta_{it})$ is monotonically increasing in the amount evaded e_{it}, conditional on the type of the firm θ_{it}. Intuitively, for a given type, firms that evade more are more likely to get caught.

The type of the firm θ_{it} might be discrete, like the type of juridical organization, or continuous, like the level of revenue[1]. Some types might be more likely to be detected if the firm engages in tax evasion. For example, in contrast to other types of juridical organizations in Colombia, corporations are closely supervised and are required to have an auditor. That is, for a given level of tax evasion e_0 and two different types $\theta' \neq \theta \in \Theta$, then $q(e_0|\theta') \geq q(e_0|\theta)$.

If the type θ is continuous, it might be a function of inputs; for example, level of revenue. Firms will then affect their probability of detection $q(e|\theta)$ in two ways: directly, by choosing how much they evade e; and indirectly, when choosing inputs M.

The optimal decision of the firm will depend on the fiscal environment $\Gamma = \{\tau, \kappa, q\}$, namely

[1]Level of revenue is a common measure for fiscal authorities to determine a firm's taxes and/or level of scrutiny, e.g., Mexico, Spain, Colombia, and Ecuador.

the tax rates, the penalty/cost of detection, and the probability of detection.

The firms' problem (Equation **??**) can be rewritten as follows,

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

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.

Assuming well-behaved functions and no corner solutions, the first-order conditions lead to the following system of differential equations,

$$G_M(M_{it})\exp(\omega_{it})\mathcal{E} - \frac{\rho_t}{P_t} = \frac{1}{(1-\tau)}\frac{\partial q(e_{it}|\theta_{it})}{\partial\theta_{it}}\frac{\partial\theta_{it}}{\partial M}\left[\frac{\rho_t}{P_t}e_{it}\tau + \kappa(e_{it})\right] \tag{4}$$

$$[1-q(e_{it}|\theta_{it})]\frac{\rho_t}{P_t}\tau - q(e_{it}|\theta_{it})\kappa'(e_{it}) = q'(e_{it}|\theta_{it})\left[\frac{\rho_t}{P_t}\tau e_{it} + \kappa(e_{it})\right] \tag{5}$$

where $E = \mathbb{E}[\exp(\varepsilon_{it})]$. The type of firms is continuous and increasing on the input. The probability of detection is increasing in the type continuum. In particular, $\frac{\partial q(e_{it}|\theta_{it})}{\partial\theta_{it}}\frac{\partial\theta_{it}}{\partial M} \geq 0$.

The left-hand side of Equation **??** is the familiar marginal output of inputs and the price ratio. In the absence of incentives' distortions induced by the fiscal environment, they are equal. But now, the equality holds no more. There's a wedge arising from the fiscal environment. The right-hand side of the equation is positive by the assumptions of the model.

Equation **??** solves the optimal evasion decision. The left-hand side is the marginal benefit net of the marginal cost of evasion. The right-hand side is the rate of change of the probability of detection due to a change in evasion weighted by the benefit and cost of evading.

2.1 Case 1 (Independence): $q(e|\theta) = q(e)$ and $\kappa(e) = \kappa_0$

Consider the case when the probability of detection is independent of type, $q(e|\theta) = q(e)$. This could be the case if the type is the juridical organization of the firm. Hence, the type of the firm, and thus the probability of detection, does not change with the firm's input decisions, $\frac{\partial q(e_{it}|\theta_{it})}{\partial \theta_{it}} \frac{\partial \theta_{it}}{\partial M} = 0$. In addition, assume the evasion cost is constant, $\kappa(e) = \kappa_0$, for simplicity.

In this case, the first-order conditions of Equation **??** with respect to the input M_{it} and the tax evasion e_{it} yield the following

$$G_M(M_{it}) \exp(\omega_{it}) \mathcal{E} = \frac{\rho_t}{P_t} \tag{6}$$

$$e_{it} = \frac{1 - q(e_{it})}{q'(e_{it})} - \frac{\kappa_0}{\frac{\rho_t}{P_t}\tau} \tag{7}$$

Equation **??**, the well-known optimality condition, says that the price ratio is equal to the marginal product of the inputs.

Likewise, Equation **??** reveals the firms' optimal tax evasion decision decreases if the probability of detection $q(e_{it})$ or the penalty of evading κ 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 τ 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 τ 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[2].

2.2 Case 2 (Spain): Discrete increase in the probability of detection after a certain threshold of revenue

In Spain, the Large Taxpayers Unit (LTU) of the tax authority focuses exclusively on firms with total operating revenue above 6 million euros. The LTU has more auditors per taxpayer than the rest of the tax authority, and these auditors are on average more experienced and better trained to deal with the most complex taxpayers. This LTU creates a discontinuity

[2] 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.

in the monitoring effort of the tax authority. Consequently, at this arbitrary revenue level, the probability of detection increases discretely (**?**).

In this scenario, depending on the productivity shock, the firm might be better off choosing not to produce past the revenue threshold. Indeed, for a relevant range of productivity draws $\Omega^B = [\omega^L, \omega^H]$, the firms will not choose to grow past the revenue threshold if the expected after-tax profits of staying small are greater than the expected after-tax profits of growing.

In the model, there is now a threshold of revenue θ^L after which the probability of detection increases discretely. 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.

Formally, let $\Theta_L = \{\theta_i : \theta_i < \theta^L\}$ and $\Theta_H = \{\theta_i : \theta_i \geq \theta^L\}$, then for all e_0 and $\theta_i' \neq \theta_i$, $q(e_0|\theta_i \in \Theta_k) = q(e_0|\theta_i' \in \Theta_k)$ with $k = \{L, H\}$, but $q(e_0|\theta_i' \in \Theta_H) \geq q(e_0|\theta_i \in \Theta_L)$.

Firms' revenue with productivity draw ω^L corresponds exactly to the enforcement threshold θ^L. Production and reporting decisions of firms with productivity draws below ω^L are not affected by the change in the probability of detection. Firms choose their inputs according to Equation **??** and their evasion decision according to Equation **??**. Firms with productivity draws above ω^U

Firms with productivity $\omega_i \in \Omega^B$ will choose the input level $\tilde{M}_i$ resulting in an expected revenue below the threshold $\theta_i < \theta^L$, if the expected after-tax profit of staying small are greater than growing, $\mathbb{E}[\pi_i|\Theta_L, \Omega^B] - \mathbb{E}[\pi_i|\Theta_H, \Omega^B] \geq 0$.

The optimal input choice M_i^* for firms with productivity $\omega_i \in \Omega^B$ implies an expected revenue greater than or equal to the threshold $\theta_i^* \geq \theta^L$. Let the expected profits given M_i^* and the optimal tax evasion in the range of size θ_l, e_{it}^*, is $\pi_l \equiv \mathbb{E}[\pi(M_{it}^*, e_{it}^*)|\theta_l]$. Let $\tilde{M}_{it}$ be the input level such that the expected revenue is below the threshold $\tilde{s}_{it} < \theta^L$ and $\tilde{e}_{it}$ be the optimal tax evasion in the range of size θ_s. Let also the expected profits of staying small are $\pi_s \equiv \mathbb{E}[\pi(\tilde{M}_{it}, \tilde{e}_{it})|\theta_s]$.

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 **??** no longer holds, hence, the marginal product of the input is now greater than the relative prices. Therefore, according to Equation **??**, bunching firms would compensate for their *higher* costs by increasing overreporting.

2.2.1 Discussion

What is new in this paper relative to the literature is that it focuses on the production function framework using public data whereas **?** and other papers use a bunching estimator with government administrative data which is difficult to access. Second, the paper focuses on input overreporting rather than on revenue underreporting, which is the relevant margin of evasion for manufacturing firms. More on this point in the revenue underreporting section. Finally, in contrast to **?** where the authors conclude that misreporting does not imply real losses in production but only fictitious reduction of the real sales, firms might optimally forgo higher revenue levels if the expected profits of staying small and evade taxes by misreporting are greater than the expected profits of growing and avoid misreporting.

2.3 Case 3 (Colombia & Mexico): Discrete increase in the tax rate after a revenue threshold

2.3.1 Colombia, Individual Proprietorships

In Colombia between 1981 and 1991, individual firm proprietors were subject to the individual income tax schedule. Individuals had incentives to not form juridical organizations to avoid double taxation. The tax authority suffered from severe limitations and inefficiencies at the time.

In this case, after the revenue threshold, the tax rate increases discretely but the probability of detection does not. The jump in the tax rate generates the incentive to increase evasion. However, a higher level of evasion increases the cost of evading by increasing the probability of detection. If the cost of an increased evasion outweighs the benefits of growing past the revenue threshold, the firms would bunch below the cutoff.

2.3.2 Mexico, Irreversible Change in Tax Regime after a Revenue Threshold

In Mexico, firms with annual revenues below 2 million pesos are taxed under the REPECO (*Regime de Pequeños Contribuyentes*) regime of small contributors at 2 percent of annual revenues, while firms above that threshold are taxed under the general regime at 30 percent. Firms must transition to the general regime if revenues increase beyond the threshold. Once in the general regime, firms cannot revert to the REPECO regime.

Firms' decision is now dynamic. Firms will maximize the sum of current and future after-tax profits. The discrete jump in the tax rate will lead to a bunching below the threshold. Moreover, the bunching will be exacerbated because firms will choose to grow past the cutoff only if the future productivity shocks allow the firm to continue to be profitable.

2.4 Case 4 (Colombia): Firms first choose type, input decisions do not affect the probability of detection

In Colombia between 1981 and 1991, Corporations were closely supervised by the Superintendent of Corporations and were required to have an auditor. All other firms were subject to the regular monitoring efforts of the tax authority, which suffered from severe limitations and inefficiencies at the time.

In the model, firms first choose their type. Input decisions do not affect the probability of detection. However, if the type is *Corporation* the probability of detection is higher than *Partnership*. Firms maximize the sum of their expected profits. In their optimization problem, firms will consider the sum of expected productivity shocks and their corresponding probability of detection. High-productivity firms will self-select into *Corporations*.

3 Colombia 1981-1991

3.1 Colombian Corporate Tax System

The relevant corporate taxes for input overreporting in Colombia during this period are the Corporate Income Tax (CIT) and the Sales Tax. The Sales Tax gradually transformed into a kind of Value-Added Tax (VAT). Also relevant for the CIT are the different juridical organizations that exist in Colombia.

This period was characterized by high levels of overall tax evasion (**?**). The fiscal rules had a system of penalties and interest that encouraged false and delinquent returns (**?**). The fiscal authority was characterized by having an inefficient auditing system, being overburdened, and legal loopholes (**?**).

3.1.1 Juridical Organizations

In Colombia, there are five types of juridical organizations: Corporations, Partnerships, Limited Liability Companies, and Individual Proprietorships.

Corporations (*sociedad anónima*) are the typical associations of capital. They are the counterpart of the US corporation. The capital of a corporation is provided by the shareholders (no less than 5) and is divided into tradable shares of equal value. Shareholders' liability is limited to the capital contributed. Corporations are subject to the Superintendent of Corporations and are closely supervised, being required to have an auditor.

Joint Stock Companies (*sociedad en comandita por acciones*) comprises two or more managing partners who are jointly and severally liable, and five or more limited partners whose liability is limited to their respective contributions. Joint Stock Companies are taxed as Corporations. Its shares are tradable, like the shares of Corporations.

Partnerships are associations of two or more persons. Partners are jointly and severally liable for the partnership's operations. Partnerships include general partnerships (*sociedad colectiva*), de Facto partnerships (*sociedades de hecho*), and ordinary limited partnerships (*sociedad en comandita simple*).

A limited liability company (*sociedad de responsabilidad limitada*) is an association of two or more persons —not exceeding 20 (Fiscal Survey) or 25 (1992 *EAM* survey documents)—, whose shares cannot be traded. The personal liability of the partners is limited to the capital contributed. The Limited Liability Company is quite important in Colombia (Fiscal Survey).

Finally, proprietorships are individuals (natural persons) who allocate part of their assets to conduct commercial activities.

There are other juridical organizations in the data that will be excluded from the final analysis. These organizations are non-profit, like cooperatives and community enterprises, or state industrial enterprises, the proceeds of which come from taxes, fees, or special contributions.

3.1.2 Corporate Income tax

The juridical organizations were subject to different Corporate Income Tax rates. Corporations were taxed at a fixed rate of 40%, while Partnerships and Limited Liability companies at 20%. Individual proprietors were subject to the graduated Individual Tax Schedule consisting of 56 rates, ranging from 0.50 to 51 percent.

Corporations were taxed on their distributed dividends, while partnerships and limited liability companies were taxed on their profits, whether or not distributed. Owners of juridical organizations were double taxed, at the firm and the individual level, whereas the income of proprietorships was taxed only once, at the individual level.

Since 1974, individuals and juridical organizations, except for limited liability companies, were subject to the minimum presumptive income. Rent (income and profits) was presumed to be no less than 8 percent of net wealth (assets less depreciation, real estate, livestock, securities).

Certain industries like airlines, publishing, and reforestation sectors, and various activities in selected regions (primarily "frontier" and other less developed ones) had their income tax exempted, limited, or reduced.

Table 1: Juridical Organizations in Colombia (1980s), A Summary

Organization	Corporate Income Tax	Liability	Capital	Owners
Corporation	40% (on distributed dividends)	Limited to capital participation	Tradable capital shares	$N \geq 5$
Limited Co.	20% (on profits)	Limited to capital participation	Non-tradable capital shares	$2 \leq N \leq 20(25)$
Partnership	20% (on profits)	Full	Not a capital association	$N \geq 2$
Proprietorship	Individual Income Tax	Full	Owner	$N = 1$

3.1.3 Sales taxes

Sales taxes were originally targeted at the manufacturing sector on finished goods and imports. Since 1974, manufacturers were allowed to credit taxes paid on any purchase made by the firm, except the acquisition of capital goods (**?**). The credits worked through a system of refunds. Consequently, the tax became a kind of value-added tax (VAT).

The basic rate was 15 percent. There was also a preferential rate of 6 percent for certain industries like clothing, footwear, and major inputs used for building popular housing, and a rate of 35 percent for luxury goods. Exports, foodstuff, drugs, and textbooks were excluded from the beginning. Also excluded were inputs, transportation equipment, agricultural machinery, and equipment.

3.1.4 Discussion

From Colombia's tax system, we can conclude that corporations are the least likely to evade taxes by misreporting because of at least three reasons. One, the Superintendent of Corporations closely monitored corporations by requiring them to have an auditor. In the model, this implies that the probability of detection is higher for them. Second, free tradable shares impose an incentive to be profitable because the market value of the shares might be negatively affected otherwise. In other words, if a corporation fakely reports lower profits, the value of its shares would likely decrease scaring away shareholders and potential investors. Joint stock companies have freely tradable shares too. Three, corporations pay CIT on distributed dividends, not on profits as Partnerships and LLCs did. Corporations have an additional margin regarding the income tax they pay because they decide when to pay dividends. This might generate other types of incentives that might be influenced by the corporation's policy regarding their dividends and the demands of their shareholders.

On the other hand, Proprietorships and LLCs are subject to the incentive to evade CIT by artificially reducing their profits.

Moreover, Proprietorships, Partnerships, and Limited Companies had incentives to overreport inputs to evade VAT and CIT. The incentives to evade varied across industries because the sales tax rate differed between industries. The incentives to evade also varied within industry sectors because juridical organizations within an industry were subject to different CIT rates. There were additional sources of variation depending on the firm's location due to local exemptions and sales composition (inputs to other firms, to consumers, to the foreign market).

Lastly, Individual proprietorships were likely to bunch at the individual income thresholds because they were subject to individual income tax which was increasing by brackets.

3.2 Fiscal Reforms

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

3.2.1 1983

The 1983 reform tried to alleviate double taxation by introducing a tax credit of 10% of dividends received by shareholders of corporations.

In addition, Law 9 of 1983 instituted a new measure of presumptive income equal to 2 percent of gross receipts to supplement the measure based on net wealth. This reform was aimed specifically at the commerce and service sectors; the former was thought to evade the wealth-based presumptive tax by systematically understating inventories. Under the original presumptive income, a juridical entity cannot declare income less than 8% of its capital (wealth).

In addition, the 1983 reform extended the presumptive income tax to limited liability companies. Before this reform, all juridical organizations were subject to the presumptive income tax except for limited liability companies.

In 1983, the value-added tax (VAT) was extended to the retail level, with a *simplified system* being made available to small retailers to ease compliance costs and the administrative burden.

The 1983 reform relatively unified the value-added tax (VAT) rates by combining previously taxed goods at 6% and 15 percent into 10%. The number of the products and services that were levied expanded.

In 1984, VAT exemptions for agricultural machinery, transportation equipment, and certain other goods were eliminated.

3.2.2 1986

The 1986 reforms unified the taxation of corporations and limited liability companies by taxing both at a rate of 30%. The top tax rate applied to individual income was reduced from 49 to 30%.

Double taxation was eliminated. The reform exempted corporate dividends and participation in profits of limited liability companies from tax at the individual shareholder/owner level.

Lastly, the 1986 reform relocated the tax collection and reception of tax reports to the banking system, among other things.

3.2.3 1990

The 1990 reform increased the VAT from 10% to 12%.

Table 2: Income Tax Reforms in Colombia (1980-1986)

Reform Year	J.O. Affected	Income Tax Change
1983	Individuals	8% increase in most scales; Max tax rate was reduced from 56 to 49%
	Ltd. Co.	Reduction from 20 to 18%; Now subject to presumptive income
1986	Individuals	Max tax rate applied was reduced from 49 to 30%
	Ltd. Co.	Increased from 18 to 30%
	Corporations	Decreased from 40 to 30%

In addition, the individual income obtained from the sale of shares through the stock market was exempted from taxation. Income tax was waived for investment funds, mutual funds, and securities, and the rates for remittances and income for foreign investment were reduced. Tax amnesties were granted, and the sanitation tax was reduced to encourage the repatriation of capital.

3.2.4 Discussion

Increases in the VAT would increase the incentives to evade. Decreases in CIT would decrease them. Intuitively, we expect higher levels of tax evasion if tax rates increase.

The CIT homogenization between Corporations and LLCs in 1986 would have motivated LLCs to incorporate. Likewise, the elimination of double taxation also in 1986 would have motivated proprietorships to become LLCs, Corporations, or Partnerships.

On the other hand, reporting more information to the tax authority — like the banking system being responsible for the collection and reception of tax reports in 1986— would decrease tax evasion.

Table 3: Sales Tax Reforms in Colombia (1980-1986)

P&T (1990)			
Industry Description	Sales Tax Change	SIC	Industry
Beverages and Tobacco	- to 35;10	313	Beverage Industries
	- to 35;10	314	Tobacco Manufactures
Textiles	6 to 10	321	Textiles
Paper	15 to 10	341	Paper and Paper Products
Other Chemical Products	15 to 10	351	Industrial Chemicals
Soap	6;15 to 10	352	Other Chemical Products
Oil and Coal Derivatives	10 to 14	354	Miscellaneous Products of Petroleum and C
Plastics	15 to 10	356	Plastic Products Not Elsewhere Classified
Iron and Steel; Nickel Smelting	6;15 to 10	371	Iron and Steel Basic Industries
Equipment and Machinery	6 to 10	382	Machinery Except Electrical
	6 to 10	383	Electrical Machinery Apparatus, Appliances
Transportation	6 to 10	384	Transport Equipment

4 Data

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

Besides the information on output, intermediates, capital, and labor, the dataset includes the type of juridical organization and the sales taxes. Table **??** offers some summary statistics.

4.1 Top Industries by Revenue in Colombia

4.2 Corporations by Industry

Looking into industry sectors.

4.3 The Fiscal Reforms

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 **??**). The dataset does not capture any changes after the 1990 reform, although there is only one more year of data.

As a validation exercise, we can see that the VAT changes induced by the three fiscal reforms are captured in the dataset. Figure **??** 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 changes in overreporting, Figure **??** shows that sales, for instance, were not exactly following

Table 4: Summary Statistics, Manufacturing Firms in Colombia (1981-1991)

	Missing (%)	Mean	SD	Q1	Median	Q3
Share of Revenues						
Sales Taxes	0	0.066	0.051	0.006	0.069	0.100
Skilled Labor (Wages)	0	0.043	1.620	0.009	0.021	0.046
Unskilled Labor (Wages)	0	0.086	0.106	0.019	0.051	0.119
Capital	0	0.495	7.555	0.126	0.261	0.499
Intermediates						
Materials (M)	0	0.457	0.191	0.333	0.457	0.582
Electricity (E)	0	0.018	0.028	0.005	0.010	0.020
Fuels (F)	0	0.011	0.025	0.001	0.003	0.010
Repair & Maintenance (R&M)	0	0.007	0.013	0.001	0.003	0.009
Services (S)	0	0.103	0.167	0.052	0.084	0.130
Deductible Inter. (M+E+F+R&M)	0	0.493	0.180	0.371	0.488	0.610
Non-Deductible Inter. (S)	0	0.103	0.167	0.052	0.084	0.130
J. Org.	**N**	**%**				
Proprietorship	5343	12.877				
Ltd. Co.	25457	61.353				
Corporation	8761	21.114				
Partnership	1425	3.434				

Table 5: Top 10 industries in Colombia by revenues

description	sic_3	n_sic	n_Corp	Market Shar
Food Manufacturing	311	912	150	23.
Beverage Industries	313	130	74	9.
Other Chemical Products	352	301	101	8.
Industrial Chemicals	351	117	69	7.
Textiles	321	476	71	7.
Transport Equipment	384	238	42	6.
Food Manufacturing	312	193	41	5.
Paper and Paper Products	341	153	41	5.
Fabricated Metal Products, Except Machinery and Equipment	381	633	85	3.
Electrical Machinery Apparatus, Appliances and Supplies	383	213	57	3.

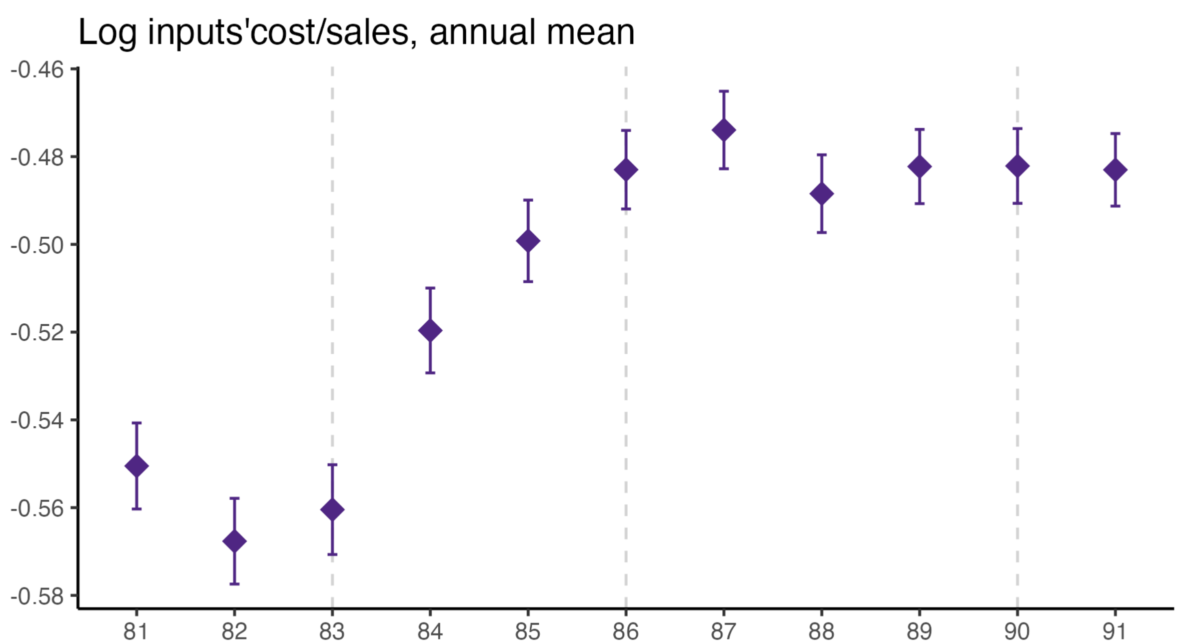


Figure 1: Input's cost share of sales, average by year of the logs.

Table 6: Corporations by Industry. Top 20 Industries in Colombia by number of firms.

Industry
Food Manufacturing
Wearing Apparel, Except Footwear
Fabricated Metal Products, Except Machinery and Equipment
Textiles
Machinery Except Electrical
Printing, Publishing and Allied Industries
Other Chemical Products
Other Non-Metallic Mineral Products
Plastic Products Not Elsewhere Classified
Transport Equipment
Electrical Machinery Apparatus, Appliances and Supplies
Food Manufacturing
Furniture and Fixtures, Except Primarily of Metal
Footwear, Except Vulcanized or Moulded Rubber or Plastic Footwear
Other Manufacturing Industries
Wood and Wood and Cork Products, Except Furniture
Paper and Paper Products
Beverage Industries
Industrial Chemicals
Leather and Products of Leather, Leather Substitutes and Fur, Except Footwear and Wearing Apparel

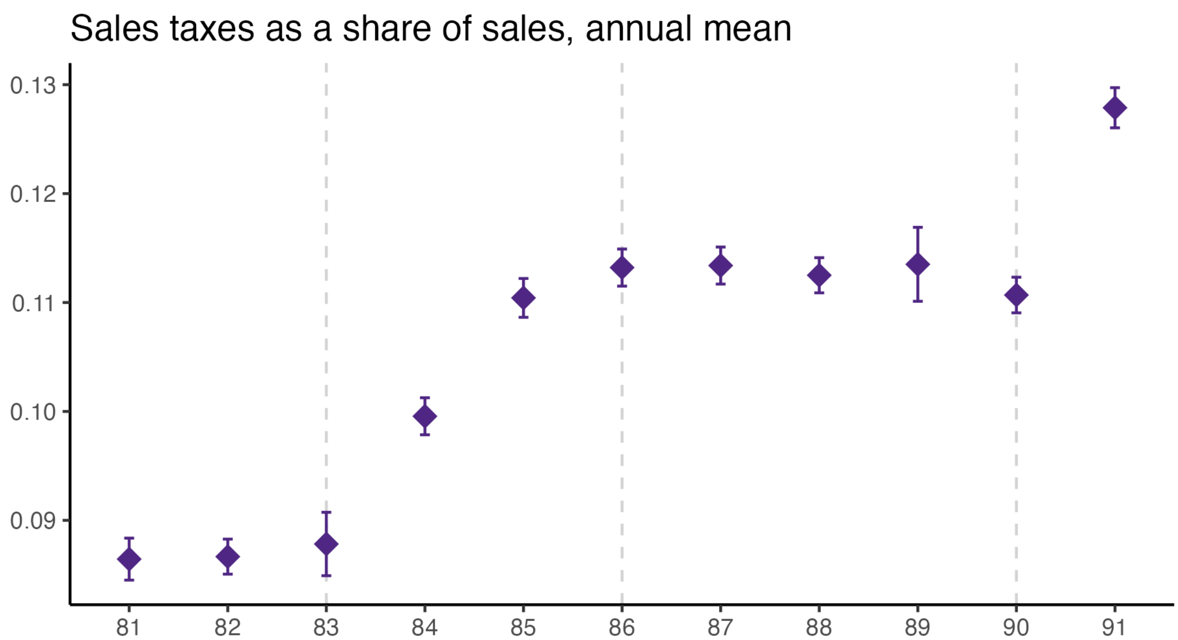


Figure 2: Sale taxes paid as share of sales, average by year.

the changes in fiscal policy. Sales started to grow during 1983, the year of the reform, whereas the cost share of sales started to grow the year after. Likewise, sales fell in 1986, while the cost share seems to reduce its growth after 1986.

Finally, in a preliminary empirical assessment, I observe that the sales tax rate is a significant determinant of the log share of revenue and that non-corporations consistently use 13-17 percent more intermediates than Corporations for a rich set of controls Table **??**. The results were estimated following Equation **??**

$$log(s_{it}) = \alpha_1 Tax_{it} + \beta_1' JurOrg_i + \beta_2' JurOrg_i \times \gamma_t + \gamma_t + \gamma_{ind} + \gamma_{metro} + \beta_3' Z + \varepsilon_{it} \quad (8)$$

Although this is not deterministic evidence, it does not contradict the hypothesis that firms other than corporations have incentives to overreport intermediates to evade taxes and that the higher the taxes the higher the incentives to evade by misreporting.

Table 7: Effect of the Juridical Organization Type and Sales Tax on the Log Share of Intermediate Inputs.

Dependent Variable:		Interm.	
Model:	(1)	(2)	(3)
Variables			
Sales Tax Rate	-0.6387***	-0.5235***	-0.5521***
	(0.1419)	(0.1027)	(0.1028)
Proprietorship	0.1654***	0.1269***	0.1237***
	(0.0275)	(0.0180)	(0.0168)
LLC	0.1717***	0.1333***	0.1280***
	(0.0286)	(0.0182)	(0.0180)
Partnership	0.1607***	0.1248***	0.1234***
	(0.0295)	(0.0199)	(0.0191)
Fixed-effects			
Industry		Yes	Yes
Metro Area			Yes
Fit statistics			
Observations	41,467	41,467	41,467
R^2	0.60315	0.62848	0.63387
Within R^2		0.54989	0.55348

Clustered (Industry & Year) standard-errors in parentheses

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

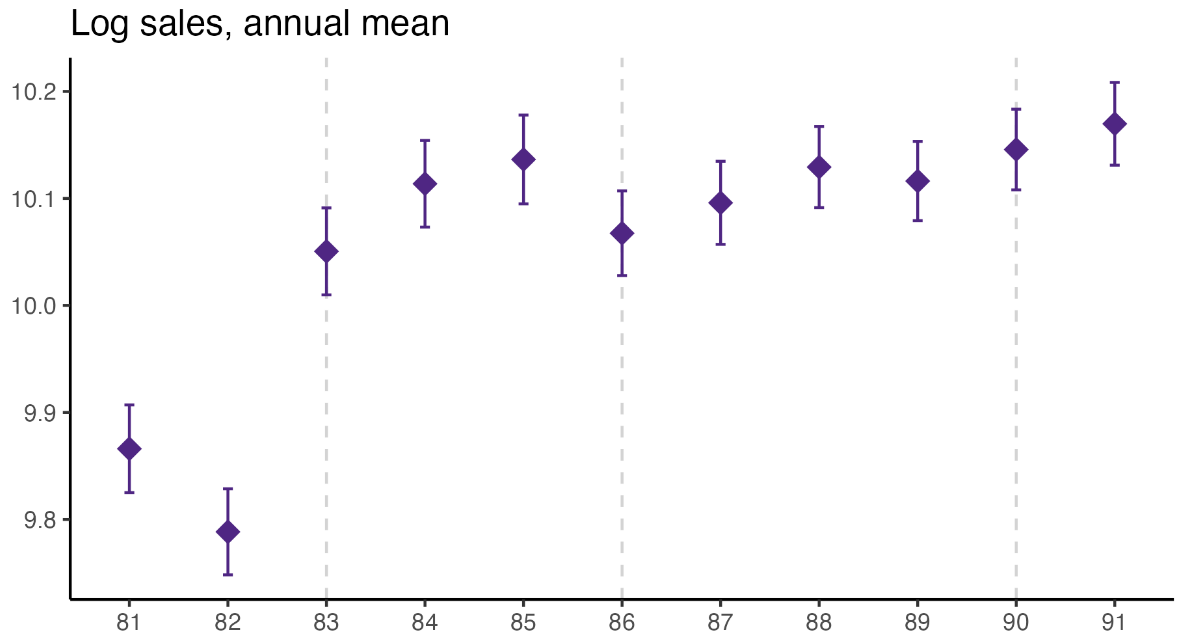


Figure 3: Sales in logs, annual mean.

4.4 The Fiscal Reform of 1983 by Industry

5 Empirical Application

Given the fiscal reforms of 1983 and 1986 in Colombia, it is natural to think in a difference-in-difference empirical application to test if the fiscal reforms induced any change in the tax evasion behavior of the firms. We can expect that an increase in either the sales tax or corporate income tax rate would lead to higher levels of evasion.

Among other changes, the fiscal law of 1983 tried to homogenize the sales taxes of the manufacturing industry. The reform reduced from 15 to 10% for some industries; for others, it increased the sales tax from 6 to 10%. Still, some others, like the Food Products industry were exempt and certain others were not affected. See Table **??** for a description of the changes documented in **?**.

Table 8: Fiscal reform of 1983. Sales tax share of revenue, mean by industry year. Top 20 industries with more firms.

SIC	Industry
311	Food Manufacturing
312	Food Manufacturing
322	Wearing Apparel, Except Footwear
381	Fabricated Metal Products, Except Machinery and Equipment
321	Textiles
382	Machinery Except Electrical
384	Transport Equipment
383	Electrical Machinery Apparatus, Appliances and Supplies
313	Beverage Industries
341	Paper and Paper Products
324	Footwear, Except Vulcanized or Moulded Rubber or Plastic Footwear
342	Printing, Publishing and Allied Industries
369	Other Non-Metallic Mineral Products
332	Furniture and Fixtures, Except Primarily of Metal
390	Other Manufacturing Industries
351	Industrial Chemicals
352	Other Chemical Products
356	Plastic Products Not Elsewhere Classified

The 1983 reform also adjusted the income tax rates for limited liability companies and individuals. For individuals, the income tax rate increased by 8% in most scales, while the maximum was reduced from 56 to 49%. For limited liability companies, the CIT was reduced from 10 to 18%.

To evaluate the change in tax evasion by input cost overreporting due to the change in the sales tax, I apply a triple difference approach. I use corporations in the industries exempted from sales taxes the year before the policy change as the control group.

Formally, non-corporations in industry k, which might have received an increment or decrement in their sales tax rate,

$$s^k_{1,j,t} = \lambda^k_t + \mu^k_1 + e^{VAT}_{j,t} + e^{CIT}_{j,t} + \varepsilon_{jt}$$

Corporations in industry k,

$$s^k_{0,j,t} = \lambda^k_t + \mu^k_0 + \varepsilon_{jt}$$

Likewise, Non-corporations and Corporations in an industry exempt from sales taxes

$$s^E_{1,j,t} = \lambda^E_t + \mu^E_1 + e^{CIT}_{j,t} + \varepsilon_{jt}$$

$$s^E_{0,j,t} = \lambda^E_t + \mu^E_0 + \varepsilon_{jt}$$

Taking the difference between time t' and t in industry k for both, corporations and non-corporations,

$$\mathbb{E}[s^k_{1,j,t'}] - \mathbb{E}[s^k_{1,j,t}] = \Delta^k_\lambda + \Delta^{VAT}_e + \Delta^{CIT}_e$$

$$\mathbb{E}[s^k_{0,j,t'}] - \mathbb{E}[s^k_{0,j,t}] = \Delta^k_\lambda$$

The diff-in-diff method will recover the joint effect of both policy changes,

$$\mathbb{E}[s^k_{1,j,t'}] - \mathbb{E}[s^k_{1,j,t}] - \Big(\mathbb{E}[s^k_{0,j,t'}] - \mathbb{E}[s^k_{0,j,t}]\Big) = \Delta_e^{VAT} + \Delta_e^{CIT}$$

The joint effect might be ambiguous because an increase in the sales tax rate will increase the incentive to overreport inputs cost but a decrease in the CIT might decrease the incentive.

To recover the effect of the change in the sales tax rate, we can use the firms of the industries that are exempted from the sales tax. Intuitively, exempted firms would not react to the change in the sales tax —which is industry-specific—, but only to the CIT —which affects all industries.

So we have,

$$\begin{aligned} &\mathbb{E}[s^k_{1,j,t'}] - \mathbb{E}[s^k_{1,j,t}] - \Big(\mathbb{E}[s^k_{0,j,t'}] - \mathbb{E}[s^k_{0,j,t}]\Big) \\ &\qquad - \Big[\mathbb{E}[s^E_{1,j,t'}] - \mathbb{E}[s^E_{1,j,t}] - \Big(\mathbb{E}[s^E_{0,j,t'}] - \mathbb{E}[s^E_{0,j,t}]\Big)\Big] = \Delta_e^{VAT} \end{aligned}$$

In regression form,

$$s_{jt} = \alpha \left[\mathbb{1}\{t = t'\} \times \mathbb{1}\{\text{treat} = \text{Non-Corp}\} \times \mathbb{1}\{k \neq E\}\right] + \beta_Z' Z_{jt} + \gamma_j + \gamma_t + \varepsilon_{jt}$$

5.1 Changes in Composition of Total Expenditure

Depending on the industry, firms might be adjusting different margins, raw materials or other expenses. For example, firms in the non-metallic mineral products industry might not fake raw materials but they can adjust deductible expenses. Another example might be the textile industry. Although the available records and data might not allow for separating expenditures precisely, the evidence shows that firms are significantly adjusting these margins.

Table 9: Log of Inputs Cost Share of Revenue by Industry. Triple diff-in-diff. The reference group is Corporations in industries exempted from the Tax Rate the year before the Reform of 1983 (1982).

Dependent Variables:	log_mats_share	log_energy_share	log_fuels_share	log_
Industry		Wearing Apparel, Except Footwe...		
Tax Change		Increased		
Model:	(1)	(2)	(3)	
Variables				
1984 × Non-Corp. × treat_3tax_treat	0.0842**	-0.0708	-0.0235	
	(0.0315)	(0.0496)	(0.0571)	
1985 × Non-Corp. × treat_3tax_treat	0.3449***	0.0248	0.2256**	
	(0.0628)	(0.0621)	(0.0871)	
1986 × Non-Corp. × treat_3tax_treat	0.1842***	0.0385	0.1604*	
	(0.0495)	(0.0733)	(0.0802)	
Fixed-effects				
plant	Yes	Yes	Yes	
Year	Yes	Yes	Yes	
Fit statistics				
Observations	10,399	10,410	9,320	
R^2	0.77685	0.85498	0.83131	
Within R^2	0.01625	0.01232	0.01154	

Clustered (Year) standard-errors in parentheses

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

? defines services as general expenditures minus machinery rental and interest payments. The Colombian survey uses industrial expenditure to calculate intermediate consumption. I identified potentially deductible expenses with information from **?**.

5.2 Parallel Trends

6 Empirical Appendix

6.1 Changes in Sales Tax

7 Changes in Relative Prices

An alternative hypothesis is that changes in the sales taxes change the relative prices of inputs. It could be that the regressions are capturing not the changes in evasion by overreporting but the changes in relative prices.

Suppose we have two flexible inputs M and L. M is deductible but L is not. Equation **??** becomes then,

$$\begin{aligned} \max_{M_{it}, L_{it}, e_{it} \in [0,\infty),} & [1 - q(e_{it}|\theta_{it})] \left[P_t \mathbb{E}[Y_{it}] - \rho_t M_{it} - w_t L_{it} - \tau \left(P_t \mathbb{E}[Y_{it}] - \rho_t (M_{it} + e_{it})\right)\right] \\ & + q(e_{it}|\theta_{it}) \left[P_t \mathbb{E}[Y_{it}] - \rho_t M_{it} - w_t L_{it} - \tau \left(P_t \mathbb{E}[Y_{it}] - \rho_t M_{it}\right) - \kappa(e)\right] \\ \text{s.t.} \quad & Y_{it} = G(M_{it}, L_{it}) \exp(\omega_{it} + \varepsilon_{it}) \end{aligned} \tag{9}$$

In the case of Colombia, θ is their juridical organization. Assume firms choose their θ before the start of the period. I'll come to this decision later, but for now, it follows that $q(e|\theta) = q(e)$. In other words, firms cannot affect their detection probability when choosing inputs.

Table 10: Classifications of Expenditure.

Expenditure	Code	Services	Indust
Purchases of accessories and replacement parts of less than one year duration	c1		$+
Purchases of fuels and lubricants consumed by the establishment	c2		$+
Payments for industrial work by other establishments	c3		$+
Payment of domestic workers	c4		$+
Payments of third parties for repairs and maintenance	c5		$+
Purchases of raw materials and goods sold without transformation	c6		$+
TOTAL Industrial Expenditures (c1:c6)	**c7**		
Rent of fixed property	c8	$+$	
Payments for professional services	c9	$+$	
Machinery rental	c10		
Insurance, excl. employe benefits	c11	$+$	
Water, mail, telephone, etc.	c12	$+$	
Publicity and advertising	c13	$+$	
Interest payments	c14		
Royalty payments	c15	$+$	
Other expenditures	c16	$+$	
TOTAL General Expenditures (c8:c16)	**c17**		
TOTAL Expenditure (c7+c17)			

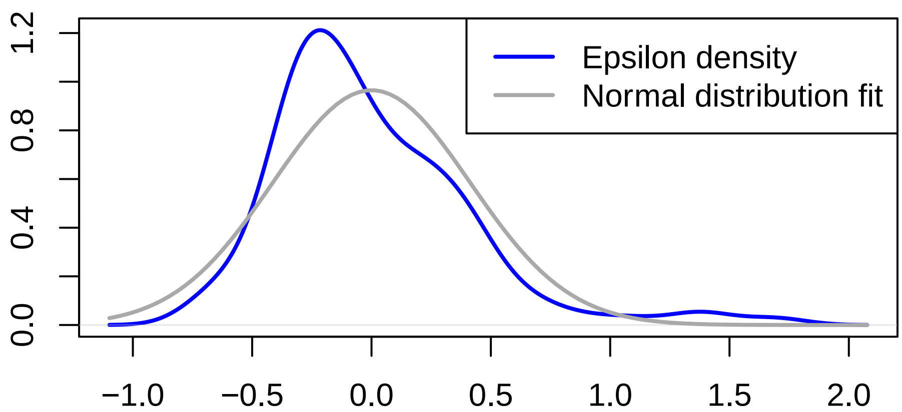
Table 11: Log of Types of Expenses Share of Total Expenses by Industry. Triple diff-in-diff. The reference group is Corporations in industries exempted from the Tax Rate the year before the Reform of 1983 (1982).

Dependent Variables:	log_services_share	log_energy_share	log_fuels_share	lo
Industry		322–Wearing Apparel, Except Fo...		
Tax Change		Increased		
Model:	(1)	(2)	(3)	
Variables				
1984 × Non-Corp. × treat_3tax_treat	-0.1164**	-0.0641	-0.0235	
	(0.0469)	(0.0498)	(0.0571)	
1985 × Non-Corp. × treat_3tax_treat	0.0751	0.0530	0.2256**	
	(0.0560)	(0.0641)	(0.0871)	
1986 × Non-Corp. × treat_3tax_treat	0.0201	0.0737	0.1604*	
	(0.0503)	(0.0746)	(0.0802)	
Fixed-effects				
plant	Yes	Yes	Yes	
Year	Yes	Yes	Yes	
Fit statistics				
Observations	10,508	10,417	9,320	
R^2	0.72910	0.75102	0.83131	
Within R^2	0.01639	0.01122	0.01154	

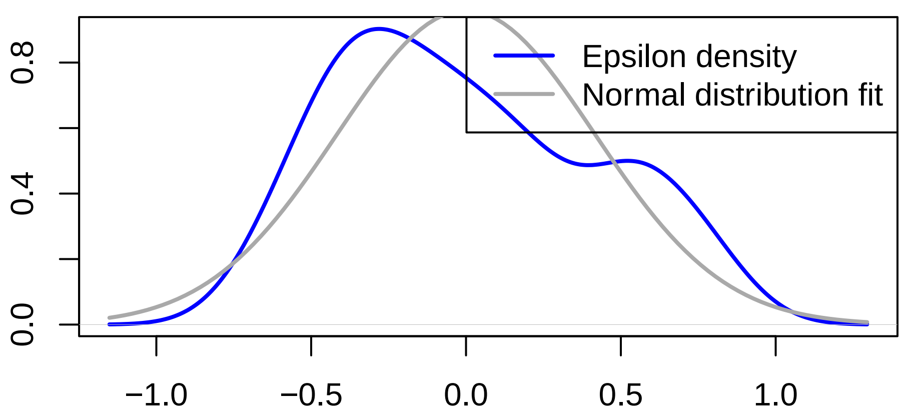
Clustered (Year) standard-errors in parentheses

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

Industry 322



Industry 331



Industry 321

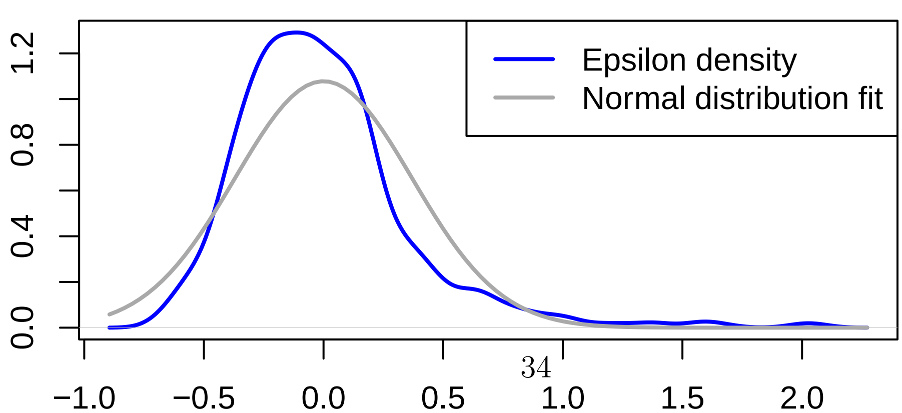


Table 12: DiD on Sales Taxes by Industry. The Control Group is Corporations in 1983 (1982).

Dependent Variable:	Sales Tax Rate			
Industry	311	312	322	381
Tax Change	exempt		increased	
Model:	(1)	(2)	(3)	(4)
Variables				
1984	-0.0001	-0.0007***	0.0351***	-0.0018**
	(8.27×10^{-5})	(5.57×10^{-5})	(0.0012)	(0.0007)
1985	-0.0005***	-0.0017***	0.0408***	-0.0014*
	(0.0002)	(0.0002)	(0.0017)	(0.0007)
1986	-0.0005**	-0.0026***	0.0382***	0.0017
	(0.0002)	(0.0002)	(0.0020)	(0.0011)
Non-Corp.	0.0015	-0.0020***	0.0009	0.0076
	(0.0008)	(0.0006)	(0.0026)	(0.0056)
1984 × Non-Corp.	-0.0004**	-0.0001	-0.0038***	0.0020**
	(0.0001)	(0.0002)	(0.0012)	(0.0007)
1985 × Non-Corp.	-0.0003	0.0018***	-0.0021	0.0101***
	(0.0003)	(0.0003)	(0.0018)	(0.0006)
1986 × Non-Corp.	8.2×10^{-5}	0.0036***	0.0013	0.0083***
	(0.0002)	(0.0001)	(0.0021)	(0.0011)
Fixed-effects				
plant	Yes	Yes	Yes	Yes
Fit statistics				
Observations	6,074	1,207	4,445	3,678

The first-order conditions now yield the following. For deductible flexible inputs, the optimality condition remains the same as Equation **??**,

$$G_M(M_{it}, L_{it}) \exp(\omega_{it})\mathcal{E} = \frac{\rho_t}{P_t} \tag{10}$$

However, for non-deductible flexible inputs, the sales tax induces a distortion in the optimality condition.

$$G_L(M_{it}, L_{it}) \exp(\omega_{it})\mathcal{E} = \frac{w_t}{(1-\tau)P_t} \tag{11}$$

Finally, the optimality condition for tax evasion becomes,

$$[1 - q(e_{it})]\frac{\rho_t}{P_t}\tau - q(e_{it})\kappa'(e_{it}) = q'(e_{it})\left[\frac{\rho_t}{P_t}\tau e_{it} + \kappa(e_{it})\right] \tag{12}$$

What can we conclude? First, if the production function is a Cobb-Douglas, then we can independently solve for each input. In this case, changes in VAT or CIT would not affect the optimality condition of deductible flexible inputs because it would not affect the relative prices. Therefore, changes in the consumption of deductible flexible inputs due to changes in VAT or CIT can only be explained by an increase in the incentives to evade taxes by overreporting.

On the other hand, an increase in the VAT or CIT would increase the relative prices of non-deductible flexible inputs, leading to a decrease in their consumption.

Alternatively, if non-deductible flexible inputs become deductible the distortion in the optimality condition induced by the VAT and CIT would be eliminated. Eliminating the distortion would lead to a reduction of the relative prices inducing a higher consumption of these inputs.

Second, if the production function is not Cobb-Douglas, then the firm has to solve the system of equations. Theoretically, the changes in the relative prices of the non-deductible flexible inputs might affect the consumption of the deductible flexible inputs depending on whether these inputs are complements or substitutes.

Empirically, however, if I can observe deductible inputs separately from non-deductible inputs, I can still be able to run the share regression to recover the input elasticity of output. The observed consumed non-deductible inputs will capture the changes in relative prices.

If I want to run the share regression using the non-deductible flexible inputs, I have to be careful to account for the distortion induced by the changes in taxes. If I observe the sales taxes, I can still be able to control for changes in sales taxes and run this regression.

Theoretically, we would have,

$$\ln\left(\frac{\rho_t M_{it}}{P_t Y_{it}}\right) = \ln D^{\mathcal{E}}_{M}(M_{it}, L_{it}) - \varepsilon_{it}$$
$$\ln\left(\frac{w_t L_{it}}{(1-\tau) P_t Y_{it}}\right) = \ln D^{\mathcal{E}}_{L}(M_{it}, L_{it}) - \varepsilon_{it}$$

Note that while we can use the gross revenue of sales for deductible flexible inputs, we would have to use the net of tax sales revenues for non-deductible flexible inputs.

More imporantly, note that the practicioner observes only $\ln M^* = \ln M + e$ but not M. In section Section **??**, I show how to recover these functions.

8 Do Corporations in Colombia have different technologies than other juridical organizations?

An implicit assumption in the identification strategy is that the subset of non-evaders has the same common technology as evaders. In the case of Colombia, this implies that Corporations have the same common technology as Proprietorships, Limited Companies, and Partnerships. However, we can think that firms with better technology are self-selecting into corporations, and thus, firms with low-performance technology will be mislabeled as evaders.

The key question is how firms are choosing their juridical organization, and in particular, how firms are choosing to become corporations. I argue that choosing the juridical organization is a result of the capital needs of the firm and it is affected by the preferences of the owners over their corporate liability and their social connections.

From the definitions of the juridical organizations, we can expect corporations to be the largests in terms of capital, followed by LLCs and Partnerships, and Proprietorships to be smallest. The reason is simply that more people can participate by contributing their capital to the firm. This ordering in turn would suggest a growth path for the firms.

Turning to the data, I find some firms change their juridical organization type. I build a juridical organization transition matrix from the previous year to the next one at the firm level. Table **??** shows the transition matrix. Although the transition matrix suggests that firms mostly stay in their juridical organization, a growth trend emerges. It looks like proprietorships turn into LLCs, and LLCs into Corporations.

Looking into the capital distributions of proprietorships, LLCs, and corporations we find

Table 13: Transition Matrix

	Proprietorship	Partnership	Ltd. Co.	Corporation	Total
Proprietorship	94.8	0.7	4.5	0.1	4404
Partnership	0.5	92.7	5.6	1.3	1241
Ltd. Co.	0.2	0.1	98.6	1.0	22154
Corporation	0.0	0.1	0.5	99.4	7805

that corporations are the largest, followed by LLCs, and proprietorships are the smallest Figure **??**. However, there are considerable overlaps. The same is true for revenues (in logs) Figure **??**. This is not surprising, as the need for larger capital increases, the more people would need to participate.

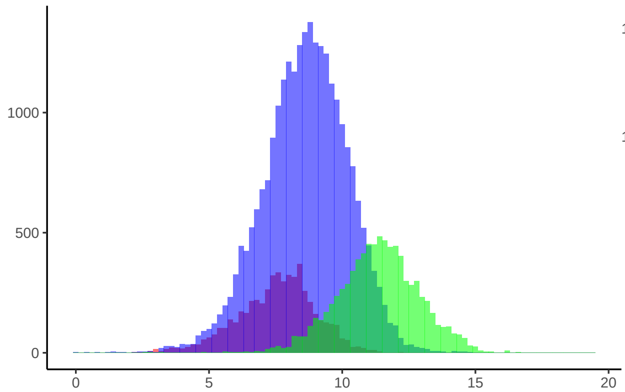
However, when we look at the frequency distributions of the capital over revenue ratio in logs, the distributions perfectly overlap Figure **??**. If corporations had different technologies we would expect to see two different distributions. In particular, the distribution of corporations would have been farther to the right than the distributions of proprietorships and LLCs.

These patterns remain even after controlling for industry, metropolitan area, and year Table **??**.

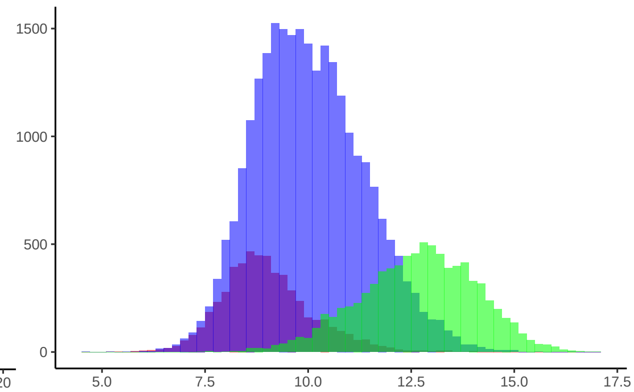
8.0.1 Model of firm (capital) growth (sketch)

A limited liability company looking to acquire more capital had three options.

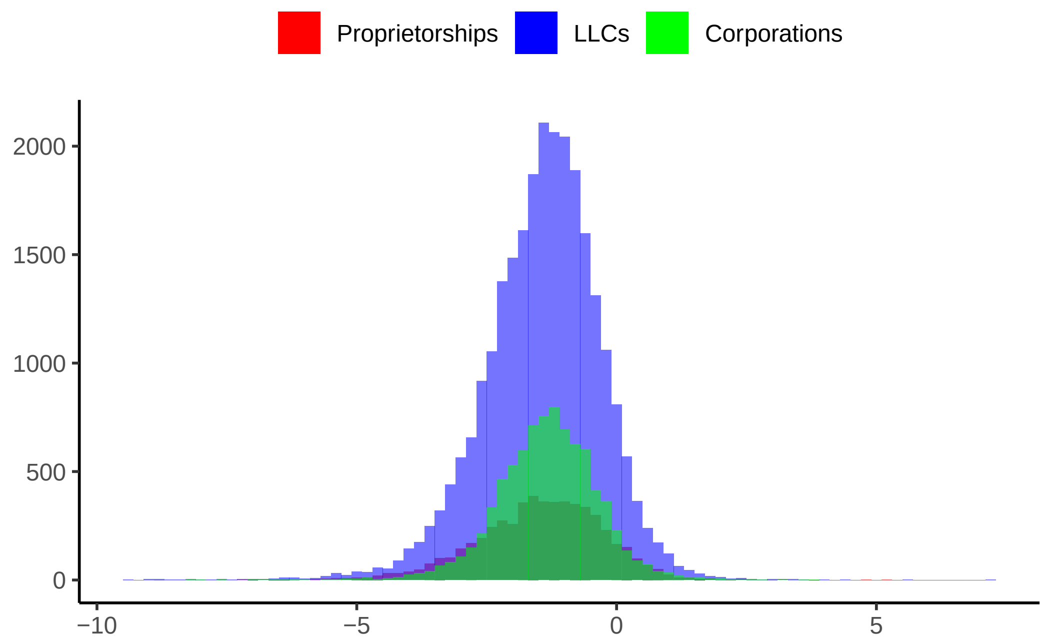
First, partners can increase their capital participation, using personal wealth or through a bank credit. This option however will increase the partners' liability. The more capital I bring the greater my liability is.



(a) Capital



(b) Revenue



(c) Capital/Revenue

Figure 4: Frequency historgrams of capital, revenue and capital/revenue (in logs) of Corporations, LLCs, and Proprietorships.

Table 14: Testing Differences in the log of Capital, Revenue, and Capital/Revenue for Corporations and Other Juridical Organizations

Dependent Variables:	log(capital)	log(sales)	log(capital/sales)
Model:	(1)	(2)	(3)
Variables			
LLC	-2.446***	-2.424***	-0.0212
	(0.0743)	(0.0639)	(0.0601)
Proprietorship	-3.346***	-3.372***	0.0265
	(0.1672)	(0.1220)	(0.0897)
Fixed-effects			
Industry	Yes	Yes	Yes
Metro Area	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	39,628	39,628	39,628
R^2	0.44574	0.51147	0.12003
Within R^2	0.31430	0.39972	0.00025

Clustered (Industry & Year) standard-errors in parentheses

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

Second, the LLC might increase its capital by inviting additional partners. As the firm increases its capital, inviting more partners will become more difficult. The funding partners risk losing control of their firm, as more people participate in the firm's management and decision-making.

Third, the firm can incorporate. This will bring capital to the firm. Anyone in the market for the shares of the firm can participate. Shares can be traded. A CEO reporting to the shareholders might be appointed. Liability is limited to shares.

The decision to incorporate is a decision on how to acquire capital. Firms still need to consider the stream of future productivity shocks, but the decision also depends on other unobservables such as risk-aversion (increasing the partners' liability), the partners' ability to convince more partners to join the LLC, and preferences over the control of the firm's management.

Assume there's an unobserved scalar ζ_i that captures all of the owners' heterogeneity in preferences over liability risk, access to networks of potential investors, and preferences over the control of the firm's management. Let a higher value of ζ capture less risk aversion, a wider network of potential investors, or a higher preference for retaining control over the firm. Consider the case in which two LLCs firms are looking to optimally increase their capital by the same quantity, but the owners are different in ζ. In particular, let the owner of firm 1 have a higher $\zeta_1 > \zeta_2$ than the owner of firm 2. The firms also have the same technology and productivity shocks. The owner of firm 1 ζ_1 would be able to convince additional investors to join their LLC, but the owner 2 ζ_2 would not. ζ_2 owner would have to choose to incorporate to get the capital for his firm or forgo growth.

The 20 pp difference in CIT rate between LLCs and Corporations might have deterred firm growth. The CIT was homogenized to 30% in 1986. What was the capital misallocation due to the tax rate differential? Are more firms incorporating after the reform?

8.0.2 Evasion vs. Production Technology

Production technology is industry-specific, but evasion technology is the same across industries.

Take the face value of the regressions and assume, without conceding, that corporations have different technologies. This will imply that corporations have technologies that are consistently 15% more productive than non-corporations regardless of the industry. In other words, it does not matter if the firms produce canned food, shoes or furniture, corporations have technologies consistently more productive. However, this is difficult to believe because there could be industries in which such an improvement in technology is difficult to imagine. For example, coffee grain mills (food products), steel foundries (steel basic products), plastic molding and extrusion (plastic products), or cement (non-metallic mineral products).

However, these cross-sectional differences across industries could be better explained by the evasion technology. We would expect the evasion technology to be the same across industries. It does not matter what I produce and what technology my firm employs, to overreport inputs all that is needed is a fake invoice.

8.0.3 Inputs Cost Share of Revenues

The evidence suggests that Corporations have similar technologies to the rest of the firms. I look at the share of revenues of different input costs. In doing so, I consider the incentives generated by the fiscal environment in Colombia. Firms had the lowest incentives to evade taxes by misreporting capital, consumed energy, and skilled labor. The cost share of revenue of these inputs suggests that Corporations and the rest of the firms had the same technologies.

8.0.3.1 Capital

Firms could not deduct capital goods from sales taxes. However, firms of capital-intensive industries might have had incentives to underreport capital because capital was used to set the minimum income tax. Income tax could not be less than 8 percent of the capital.

8.0.3.2 Energy

Before the electrical energy reform in 1994, the power market was almost exclusively supplied by public companies with negligible participation from private firms. The public companies were inefficient and suffered financial distress due to poor management and low tariffs set by political actors. Two major blackouts in 1983 and 1992-1993 led to the 1994 reform that opened the sector to private participation.

It is unlikely that firms overreported energy, as they have to purchase from a Corporation or a public company. Both of those organizations had no incentives to cooperate and provide fake invoices.

In the data, energy sold by corporations accounted on average for 73% of the total energy sold, but only 1.7% of the total energy purchased. In addition, corporations sold energy at 12 times the market price on average Table **??**. The high price might suggest corporations had market power in the electric energy market, however, this is unlikely to be relevant to affect any estimations as their overall market share is small.

8.0.3.3 Labor

Table 15: Electric energy market in Colombia (1981-1991).

	Sold/Purchased Price Ratio	Sold Energy (% of Total Sold Energy)	Sold Energy (%
Corporation	11.9	70.3	
Ltd. Co.	7.4	2.1	
Other	2.5	27.6	
Partnership	1.0	0.0	
Proprietorship	1.0	0.0	

Firms might have incentives to underreport labor to evade Payroll Taxes (PRT) if the expected benefits of evading the PRT outweigh the opportunity costs of evading Corporate Income Tax (CIT) by overreporting labor costs and the expected cost of evading PRT (**?**).

Firms are more likely to underreport unskilled rather than skilled labor. Skilled employees are less likely to cooperate with firms in underreporting their wages. Firms might offer employees cash compensations in order for employees to accept lower reported wages or report their wages at all. The cost for employees is that these payroll taxes provide them with social benefits such as social security or public health access. These benefits are not obvious in the short run. It is more likely that unskilled labor to be short-sighted and accept to cooperate with firms to underreport their wages. Skilled workers, on the other hand, are less likely to accept these conditions and to have outside options and move if the conditions are not favorable to them.

8.0.3.4 Services and other expenditures

The Colombian dataset allows, to some extent, separating expenditures, including services, into deductible and non-deductible expenses. In contrast to non-evading firms, evading firms should use a higher share of deductible expenses. The reason is that firms have incentives to overreport deductible expenditure to evade ICT and VAT.

The Colombian data separates firms' total expenditures into general and industrial expenditures. Industrial expenditure is defined as the indirect costs and expenditures incurred by the firm in order to perform its industrial activity. The data lists the purchase of accessories and replacement parts, fuels and lubricants consumed by the establishment, industrial work by other establishments, and third-party repairs and maintenance, among others, as industrial expenditures[3]. All other expenses, including insurance (excluding employee benefits) and machinery rentals, are considered general expenses.

Most services were excluded from sales taxes. Some non-excluded services in this period were insurance premiums (excluding life insurance), repair and maintenance, national and international telegrams, telex, and telephone, and rental of goods and chattels, including financial leasing.

I classify as deductible expenses the following industrial expenditures: purchase of fuels and lubricants, payments to third parties for repair and maintenance, purchase of raw materials and goods sold without transformation; and the following general expenditures: machinery rental, insurance excluding employee benefits, and water, mail and telephone expenses.

8.0.3.5 Technical note

According to the notes on the dataset, the quantity of energy consumed by firms is estimated as the difference between purchased plus generated energy and sold energy. In contrast, the value of consumed energy is the difference between the value of purchased and sold energy.

[3]After 1992, industrial expenditure included rent of machinery and property, payment for professional services, insurance (excluding employee benefits), water, mail, and telephone.

Consequently, using the calculated value over the calculated quantity of consumed energy ignores the quantity of generated energy and the increase in the price of sold energy by corporations.

In **?**, services are defined as general expenditures minus machinery rental and interest payments. However, this approach does not include industrial expenditures which are closely linked to the production process. In the Colombian data, the intermediate consumption is defined as raw materials, purchased electric energy, and industrial expenditure. This definition is close to what is commonly defined as intermediate inputs.

8.0.3.6 Results

Table **??** shows that in terms of the capital and consumed energy share of revenue, corporations are not different from other types of juridical organizations. Capital and consumed energy are reliable indicators that Corporations have similar technologies to other firms because firms have no incentives to overreport capital or consumed energy. On the other hand, the raw materials share of revenues is 3 percent higher for non-corporations. Firms have incentives to overreport raw materials to evade sales taxes and CIT. Hence, materials are not a good reference to compare technology between Corporations and other types of firms.

Table **??** shows that the skilled labor share of revenue for corporations is not statistically different from other types of firms. On the other hand, unskilled labor is significantly lower for non-corporations. Unskilled workers are more likely to cooperate with non-corporations to underreport their employment/wages. An alternative explanation is that corporations employ more unskilled workers because they are bigger. Therefore, skilled labor is a more reliable indicator of technology and shows that corporations have similar technology to other firms.

Table **??** also shows that the total expenditure share of revenue is 8 percent lower for non-corporations on average. On one hand, this might be due to size. Corporations tend to be larger than non-corporations. However, the composition of the expenditure matters.

?@tbl-reg-shares-3 shows that the deductible expenses share of total expenditure is 6 percent significantly higher for non-corporations. This is not the case if we classify expenditures by industrial and general expenses. Services as defined in GNR, that is general expenses minus machinery rental and interest payments, as a share of total expenditure is 5 percent higher for non-corporations. Services include deductible expenses such as insurance excluding employee benefits, and telegrams and telephone services.

In summary, looking at the inputs that firms are less likely to misreport due to tax evasion incentives, the evidence suggests that Corporations have similar technologies to the other types of firms. In addition, I find that the cost share of inputs that are likely to be misreported due to tax evasion incentives are significantly different from corporations and in the expected direction. In particular, it looks like firms overreport materials and deductible expenses, and underreport unskilled labor.

9 Intermediate Inputs

What are intermediate inputs? Intermediate inputs are flexible and in our case firms should have incentives to overreport them. Raw materials, electricity, fuels, and services are what researchers have in mind when they refer to intermediates. In the context of tax evasion in Colombia, materials are the input of interest for several reasons.

Table 16: Shares of Revenue for different inputs by Corporations and Non-Corporation.

Dependent Variables:	Capital	energy_share	Materials	fuels_share
Model:	(1)	(2)	(3)	(4)
Variables				
Non-Corporation	0.1524	0.0010	0.0173**	-5.62×10^{-5}
	(0.1063)	(0.0018)	(0.0072)	(0.0011)
Fixed-effects				
Industry	Yes	Yes	Yes	Yes
Metro Area	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	42,145	42,145	42,145	42,145
R^2	0.00180	0.10438	0.23606	0.22140
Within R^2	5.74×10^{-5}	0.00022	0.00147	8.62×10^{-7}

Clustered (Industry & Year) standard-errors in parentheses

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

Table 17: Shares of Revenue for different inputs by Corporations and Non-Corporation.

Dependent Variables:	Skilled Labor	Unskilled Labor	services_share	repair_maint_share
Model:	(1)	(2)	(3)	(4)
Variables				
Non-Corporation	0.0134	0.0261***	-0.0324***	-0.0011
	(0.0201)	(0.0062)	(0.0048)	(0.0007)
Fixed-effects				
Industry	Yes	Yes	Yes	Yes
Metro Area	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	42,145	42,145	42,145	42,145
R^2	0.00192	0.36301	0.02607	0.07357
Within R^2	9.65×10^{-6}	0.01266	0.00539	0.00095

Clustered (Industry & Year) standard-errors in parentheses

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

Study	Intermediates
Levinsohn and Petrin (2003)	Materials, Electricity, and Fuels
Gandhi, Navarro, & Rivers (2020)	Materials, Energy (Electricity and Fuels), and Services

First, firms have incentives to overreport production expenses to evade CIT, but only VAT deductible intermediates reduce their VAT burden. In fact, Colombia used system of refunds to companies with VAT tax balances in their favor. Raw materials are VAT deductible during the whole period of out data, 1981-1991. In addition, materials are the greater share of the firm's expenses among the other intermediate inputs.

Electricity was mainly provided by government companies during this period. Even though there were private participation in the market, the data shows that corporations sold most of the private electricity. In our paper, we argue that Corporations are the firms with the lowest incentives to misreport their commercial activity. Therefore, it could have been relatively riskier to fake an electric bill than other inputs. Why? Because any electricity bill not coming from a Corporation or a Government company might look suspicious. In addition, it is unclear that firms could deduct from their taxes their electric bill. Finally, electricity is a small share of the firm's expenses in the data.

Fuels consumed in production were deductible however these represent a non-significant share of the firms' expenses.

Finally, only repair and maintenance services were taxed from the beginning during this period. The authority gradually introduced services they could control easily like telephone and telecommunication, insurances, airline tickets, and parking, among others. The rest of the services look more like fixed costs and not like intermediate inputs for production, or at least they do not necessarily increase with production. Take for instance, telephone bills. They might increase if an increase of output comes from handling more clients, but they might keep the same and production might increase if current clients increase their volume of demanded products.

10 Industries

Given that I'll focus on materials, the industries I pick would have to make sense. Meaning, that firms should have nontrivial incentives to overreport raw materials to evade taxes and raw materials should represent a nontrivial share of their total production expenses.

To start, unprocessed primary goods such as mining, agricultural, fishing, and forestry were exempt of sales taxes. If the main raw materials of an industry is a primary good, then firms in that industry would have no incentive to overreport their raw materials because they are not paying sales taxes on their raw material that could be credited towards the taxes they have to pay for their sales. Industries like concrete and clay used in construction (369-Non-Metallic Minerals) and leather products (323-Leather tanning and dyeing, Fur skins) might fall into this category.

For other industries, few specialized suppliers provide their main raw materials. These firms might have to import a non-trivial share of their materials. It would be highly suspicious to fake invoices for raw materials that come from few known local and international suppliers. For example, plastic pellets the main raw material in injection molding and plastic products (356) are supplied internationally by a handful of suppliers located in the US and Europe.

Likewise, industries exempt of sales tax would have relative lower incentives to evade than non-exempt. In Colombia, firms producing foodstuffs (311 and 312), drugs, textbooks, transportation equipment (384), agricultural machinery (382-Non-Electrical Machinery including Farm Machinery), and equipment (383-Electrical Equipment) were exempt of sales taxes since 1974. Exemption to Agricultural machinery and transportation equipment was eliminated in 1984. Almost half-way through the data.

(In addition, exporters were exempted from sales taxes.)

Lastly, it would be better to avoid industries selling a significant share of their output directly to consumers and retailers. Tax evasion through underreporting of sales severely affected Colombia during this time (**?**). It has been documented elsewhere that revenue underreporting is most significant for firms selling directly consumers. Why? Firms might offer consumers a discount if they do not request an invoice. Since the consumer is the final payer of the sales tax, they are more like to cooperate. On the other hand, firms selling to other firms face counterincentives. The firm buying the product is interested in having an invoice to report the sales tax pay to reduce the balance of the sales tax they collect when selling. In addition, during this period there was a tendency to create fictitious distributors to reduce the tax base.

A report based on a sample from that period in **?** reveals that industries 322 (Wearing Apparel — 67.5%), 324 (Footwear — 49%), 342 (Publishing — 53.9%), 369 (Non-Metallic — 50%), and 390 (Other Manufacturing — 61.3%) sell half or more of their output to retailers and directly to consumers.

Table 19: Distribution of Sales from Manufacturing Firms, by type of Purchaser (in percentage) (**?**)

SIC	Total no. of firms	% to retailers	% to public	% retailers and public combined	% government	% wholesalers
312	25	24.16	15.72	39.88	1.20	58.92
313	3	-	29.33	29.33	0	70.67
321	8	15.06	0.11	15.17	1.17	83.67
322	2	52.92	14.58	67.50	0	32.50
324	5	45.00	4.00	49.00	0	51.00
332	8	10.85	20.15	31.00	5.92	63.08
341	6	5.00	0	5.00	0	95.00
342	9	19.44	34.44	53.89	7.22	38.89
351	5	15.20	6.00	21.20	2.80	76.00
352	8	37.81	5.50	43.31	4.47	52.22
353	3	5.67	5.00	10.67	0	89.33
355	5	21.00	0	21.00	0	79.00
362	1	10.00	0	10.00	0	90.00
369	8	21.25	28.75	50.00	12.88	37.13
372	3	10.00	0	10.00	1.67	88.33
381	22	12.41	14.73	27.14	2.50	70.36
382	20	11.20	30.00	41.20	9.00	49.80
383	15	13.87	7.13	21.00	9.07	69.93
384	7	0	34.57	34.57	0	65.43
390	3	58.00	3.33	61.33	3.33	35.33

SIC	Total no. of firms	% to retailers	% to public	% retailers and public combined	% government	% wholesalers
356	11	15.55	6.00	21.55	3.18	75.27

11 Identification Strategy

Because the firms' optimization decisions depend on the fiscal environment, the identification strategy should be motivated by the fiscal environment Γ. In particular, the identification strategy will be as good as how well we can tell apart a subset of firms that have the highest incentive to not evade. For example, in the case of Spain, the firms above the revenue LTU threshold. In the case of Colombia, the corporations.

Assumption 11.1: Non-Evaders

Based on the fiscal environment Γ, the researcher can identify a subset of firms $\theta_i \in \Theta^{NE} \subset \Theta$ that does not evade taxes by overreporting inputs.

For those firms, then $\mathbb{E}[e_{it}|\theta_i \in \Theta^{NE}] = 0$

In addition, I impose the following timing assumption.

Assumption 11.2: Independence

Firms choose overreporting e_{it} *before* the output shock ε_{it}

Assumption 11.2 implies that input overreporting is independent of the current period output shock, $e_{it} \perp \varepsilon_{it}$. In the literature is not rare to assume that the output shock is not part of the information set of the firms, $\varepsilon_{it} \notin \mathcal{I}_t$ (**?**). Timing and information set assumptions are not uncommon for identification strategies in production functions and demand estimation (**??**).

11.1 Identifying the production function parameters

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

The econometrician observes output, capital, labor, and prices of intermedite inputs and output $\{Y_{it}, K_{it}, L_{it}, \rho_t, P_t\}$, but she only observes overreported intermediate inputs, $M_{it}^* = M_{it}\exp(e_{it})$[4].

Then, for the case of Colombia, equation **??** applies since the type of firms is the juridical organization and the non-evaders are Corporations. By multiplying both sides by the intermediate inputs and dividing by the output, we get

$$\begin{aligned} \ln\left(\frac{\rho_t M_{it}}{P_t Y_{it}}\right) &= \ln\beta + \ln\mathcal{E} - \varepsilon_{it} \\ &\equiv \ln D^{\mathcal{E}} - \varepsilon_{it} \end{aligned} \tag{13}$$

$$\ln(M_{it}^*) = \ln(M_{it}) + e_{it}$$

$$\ln\left(\frac{\rho_t M_{it}^*}{P_t Y_{it}}\right) = \ln\beta + \ln\mathcal{E} - \varepsilon_{it} + e_{it}$$

where, $\mathcal{E} \equiv \mathbb{E}[\exp(\varepsilon_{it})|\mathcal{I}_{it}] = \mathbb{E}[\exp(\varepsilon_{it})]$.

We can use Equation **??** and assumption 11.1 to recover the production function parameter β

$$\mathbb{E}\left[\ln\left(\frac{\rho_t M_{it}^*}{P_t Y_{it}}\right)\middle|\Theta^{NE}\right] = \ln D^{\mathcal{E}} \tag{14}$$

[4]Note we can always rewrite $M^* + e = M^*\exp\{a\}$, then $\exp\{a\} = \frac{e}{M^*} + 1$.

The constant $\mathcal{E}$ is also identified (?),

$$\mathcal{E} = \mathbb{E}\left[\exp\left(\ln D^{\mathcal{E}} - \ln\left(\frac{\rho_t M_{it}^*}{P_t Y_{it}}\right)\right)|\theta^{NE}\right] = \mathbb{E}\left[\exp(\varepsilon_{it})|\theta^{NE}\right] = \mathbb{E}[\exp(\varepsilon_{it})] \tag{15}$$

and, thus, the output elasticity of input, β, is also identified,

$$\beta = \exp\left(\ln D^{\mathcal{E}} - \ln \mathcal{E}\right). \tag{16}$$

11.2 Identifying Tax Evasion

Having recovered both the flexible input elasticity, β, and the constant $\mathcal{E}$, for all firms, I can form the following variable using observed data.

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

Tax evasion therefore can be recovered up to an independent random variable. We can do better however. By assumption 11.2, the tax evasion, ε_{it} ,is independent of e_{it}. Note that, from Equation **??**, we also recovered $f_\varepsilon(\varepsilon)$ the distribution of ε. By using deconvolution methods, I can recover the tax evasion apart from the output shock.

In particular, from probability theory,

Definition 11.1.

Definition: Convolution

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

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

where f is the density of $\mathcal{Z}$ (Meister, 2009)

11.3 Identifying Productivity

Here, I show how to recover the rest of the parameters of the production function, including productivity and its Markov process. We can do it in several ways, depending on our object of interest.

In the literature, it is not uncommon to assume that productivity follows a Markov process. That is,

$$\omega_{it} = h(\omega_{it-1}) + \eta_{it} \tag{17}$$

Let $\ln X_{it} = x_{it}$. With Equation **??**, we can form the following observed variable,

$$\begin{aligned} \mathcal{W}_{it} &\equiv y_{it} - \beta(m^*_{it} - \mathcal{V}_{it}) \\ &= y_{it} - \beta(m_{it} + e_{it} - e_{it} + \varepsilon_{it}) \\ &= y_{it} - \beta m_{it} - \beta\varepsilon_{it} \\ &= \alpha_K k_{it} + \alpha_L l_{it} + \omega_{it} + (1 - \beta)\varepsilon_{it} \end{aligned} \tag{18}$$

Here, intuitively, we are trading unobservables, the unobserved tax evasion for the output shock, just to make our lives easier. We could have use use directly m^* since we know the distribution of tax evasion e, which is equivalent to the distribution of $f_{m^*|m}(m^*|m)$. Also note that is it not necessary to learn tax evasion to recover productivity. This is useful if a practicioner is only interested in productivity and not on tax evasion.

By replacing ω_{it} from Equation **??** in Equation **??**, we get

$$\mathcal{W}_{it} = (1-\beta)\varepsilon_{it} + \alpha_K k_{it} + \alpha_L l_{it} + h(\mathcal{W}_{it-1} - (1-\beta)\varepsilon_{it-1} - \alpha_K k_{it-1} - \alpha_L l_{it-1}) + \eta_{it} \quad (19)$$

If $h(\xi) = \delta_0 + \delta_1 \xi$, to learn α and h, we need to find instruments Z such that,

$$\mathbb{E}[\eta_{it} + (1-\beta)\varepsilon_{it} + \delta_1(1-\beta)\varepsilon_{it-1}|Z] = 0. \quad (20)$$

Table **??** displays candidates for Z

Table 20: Orthogonality by Residuals

	Orthogonality
η_{it}	$k_{it}, l_{it}, \varepsilon_{it},\ k_{it-1}, l_{it-1}, \varepsilon_{it-1}, \omega_{it-1}, \mathcal{W}_{it-1}, m_{it-1}, m^*_{it-1}$
ε_{it}	$k_{it}, l_{it}, \omega_{it}, m_{it}, m^*_{it}, \eta_{it},\ k_{it-1}, l_{it-1}, \varepsilon_{it-1}, \omega_{it-1}, \mathcal{W}_{it-1}, m_{it-1}, m^*_{it-1}$
ε_{it-1}	$k_{it}, l_{it}, \omega_{it}, \varepsilon_{it}, m_{it}, m^*_{it}, \eta_{it},\ k_{it-1}, l_{it-1}, \omega_{it-1}, \mathcal{W}_{it}, m_{it-1}, m^*_{it-1}$

The possible instruments are $Z_{it} = \{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m^*_{it-1}, \mathcal{W}_{it-2}\}$.

During estimation, I use labor and capital as instruments for themselves $Z_{it} = \{k_{it}, l_{it}\}$, and the constant and the lag of observed overreported intermediates as instrument for the AR(1) productivity Markov process $h(\cdot)$.

Once the paramaters of the production function are identified, we can use the following observed variable to recover ω_{it}

$$
\begin{aligned}
\widetilde{\mathcal{W}_{it}} &\equiv \mathcal{W}_{it} - \alpha_K k_{it} - \alpha_L l_{it} \\
&= \omega_{it} + (1-\beta)\varepsilon_{it}
\end{aligned} \tag{21}
$$

11.4 Identification with two Flexible Inputs

Suppose now that M_{it} and L_{it} are flexible inputs, but K_{it} is not. Furthermore, M_{it} is deductible and L_{it} is not. Under a Cobb-Douglas production function, Equation **??** and Equation **??** become,

$$
\begin{aligned}
\ln\left(\frac{\rho_t M_{it}}{P_t Y_{it}}\right) &= \ln\beta + \ln\mathcal{E} - \varepsilon_{it} \\
&= \ln D_M^{\mathcal{E}} - \varepsilon_{it}
\end{aligned} \tag{22}
$$

$$
\begin{aligned}
\ln\left(\frac{w_t L_{it}}{(1-\tau) P_t Y_{it}}\right) &= \ln\alpha_L + \ln\mathcal{E} - \varepsilon_{it} \\
&= \ln D_L^{\mathcal{E}} - \varepsilon_{it}
\end{aligned} \tag{23}
$$

Note that firms do not have incentives to overreport non-deductible flexible inputs. So the observed non-deductible flexible inputs are the true inputs consumed in production. For the deductible flexible inputes, we still observe the overreported inputs, $M_{it}^* = M_{it}\exp(e_{it})$.

So, Equation **??** becomes,

$$
\ln\left(\frac{\rho_t M_{it}^*}{P_t Y_{it}}\right) = \ln D_M^{\mathcal{E}} - \varepsilon_{it} + e_{it} \tag{24}
$$

The constant $\ln D_L^{\mathcal{E}}$ is identified by a linear regression from Equation **??**. Likewise, using assumption 11.1, we can identify the constant $\ln D_M^{\mathcal{E}}$ by running a linear regression of Equation **??** for the non-evaders.

To identify the elasticities from the constant $\mathcal{E}$, let $\hat{\varepsilon}_{it}^M$ and $\hat{\varepsilon}_{it}^L$ be the residuals of the regressions of Equation **??** and Equation **??**, respectively. Then, we can form the following observed variable,

$$\varepsilon_{it} = \begin{cases} \lambda_{it}\hat{\varepsilon}_{it}^M + (1-\lambda_{it})\hat{\varepsilon}_{it}^L & \text{if } i \text{ is a Non-evader} \\ \\ \hat{\varepsilon}_{it}^L & \text{otherwise} \end{cases} \tag{25}$$

where λ_{it} is a weight that can be chosen by the practicioner. For example, it can be set to the share of the cost of both flexible inputs in the total cost of production, $\lambda_{it} = \frac{\rho_t M_{it}^*}{\rho_t M_{it}^* + w_t L_{it}}$. In this way, we would have a time-firm specific weight that captures the relative importance of each flexible input in the production process. alternatively, we can can have time or firm fixed weights, $\lambda_t = N^{-1}\sum_i \lambda_{it}$ or $\lambda_i = T^{-1}\sum_t \lambda_{it}$, respectively.

Another option is to use the optimal weight to minimize the variance of ε_{it},

$$\lambda = \arg\min_{\lambda} \mathbb{E}\left[\varepsilon^2|\Theta^{NE}\right]$$

$$\begin{aligned} \mathbb{E}\left[\varepsilon^2|\Theta^{NE}\right] =& \lambda^2\mathbb{E}\left[\left(\hat{\varepsilon}_{it}^M\right)^2|\Theta^{NE}\right] + (1-\lambda)^2\mathbb{E}\left[\left(\hat{\varepsilon}_{it}^L\right)^2|\Theta^{NE}\right] \\ &+ 2\lambda(1-\lambda)\mathbb{E}\left[\hat{\varepsilon}_{it}^M\hat{\varepsilon}_{it}^L|\Theta^{NE}\right] \end{aligned}$$

Deriving the first order condition, we get

$$2\lambda\mathbb{E}\left[\left(\hat{\varepsilon}_{it}^M\right)^2|\Theta^{NE}\right] + 2(1-2\lambda)\mathbb{E}\left[\hat{\varepsilon}_{it}^M\hat{\varepsilon}_{it}^L|\Theta^{NE}\right] - 2(1-\lambda)\mathbb{E}\left[\left(\hat{\varepsilon}_{it}^L\right)^2|\Theta^{NE}\right] = 0$$

Finally, solving for λ gives us the optimal weight,

$$\lambda^* = \frac{\mathbb{E}\left[\left(\hat{\varepsilon}_{it}^L\right)^2 |\Theta^{NE}\right] - \mathbb{E}\left[\hat{\varepsilon}_{it}^M \hat{\varepsilon}_{it}^L |\Theta^{NE}\right]}{\mathbb{E}\left[\left(\hat{\varepsilon}_{it}^M\right)^2 |\Theta^{NE}\right] + \mathbb{E}\left[\left(\hat{\varepsilon}_{it}^L\right)^2 |\Theta^{NE}\right] - 2\mathbb{E}\left[\hat{\varepsilon}_{it}^M \hat{\varepsilon}_{it}^L |\Theta^{NE}\right]}. \tag{26}$$

This might work only if the covariance of the residuals is less than the variance of any residual. In theory, the covariance cannot be larger than the variance of either residua. In that case, it means the both residuals are good estimates of the output shock ε_{it} and we can use either of them to recover the output shock ε_{it}.

Then, we can recover $\mathcal{E} = \mathbb{E}[\exp(\varepsilon_{it})]$ and the elasticities $\beta = \exp(\ln D_M^{\mathcal{E}} - \ln \mathcal{E})$ and $\alpha_L = \exp(\ln D_L^{\mathcal{E}} - \ln \mathcal{E})$.

11.5 Identifying Tax Evasion with Two Flexible Inputs

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

In the case of the two flexible inputs, we can form a firm-time specific variable that captures the tax evasion by overreporting the deductible flexible input, $\mathcal{V}_{it} = e_{it}$.

11.6 Identifying Productivity with Two Flexible Inputs

To recover α_K, we can now form the following observed variable,

$$\begin{aligned}
\mathcal{Y}_{it} \equiv &y_{it} - \beta(m_{it}^* - e_{it}) - \alpha_L l_{it} - \varepsilon_{it} \\
= &y_{it} - \beta m_{it} - \alpha_L l_{it} - \varepsilon_{it} \\
= &\alpha_K k_{it} + \omega_{it}
\end{aligned}$$

Then, using ortogonality of η_{it},

$$\mathbb{E}[\mathcal{Y}_{it}|Z_{it}] = \alpha_K k_{it} + \mathbb{E}[h(\mathcal{Y}_{it-1} - \alpha_K k_{it-1})|Z_{it}]$$

where $Z_{it} = \{k_{it}, k_{it-1}, \mathcal{Y}_{it-1}\}$.

11.7 Translog Production Function

To identify tax evasion and relaxing the assumption of a CD production function, we would need a flexible input for which firms have no incentives to overreport. Firms might face flexible inputs that they cannot deduct from their VAT or CIT, for example. If this is the case, then firms would have no incentives to overreport non-deductible flexible inputs.

Assume now that L_{it} is a non-deductible flexible input and, without loss of generality, there are only two inputs (L_{it}, M_{it}). Let's assume the production function is now *translog*. We have,

$$\ln G(l, m) = \beta_0 m_{it} + \beta_1 m_{it} l_{it} + \beta_2 m_{it} m_{it} + \beta_3 l_{it} + \beta_4 l_{it} l_{it}$$

Then, equation Equation **??** becomes

$$\begin{aligned} s_{it}^L &= \ln\left(\beta_3 + 2\beta_4 l_{it} + \beta_1(m_{it}^* - e_{it})\right) + \ln \mathcal{E} - \varepsilon_{it} \\ &\equiv \ln D^{\mathcal{E}}(l_{it}, m_{it}^* - e_{it}) - \varepsilon_{it} \end{aligned} \tag{27}$$

where $s_{it}^L \equiv \ln\left(\frac{w_t L_{it}}{(1-\tau)P_t Y_{it}}\right)$.

Note that by assumption 11.1, $D^{\mathcal{E}}$ and the density of ε are still identified.

11.7.1 Testing for Tax Evasion

We can design a test for tax evasion by forming the analog of Equation **??**,

$$
\begin{aligned}
\mathcal{V}_{it}^{L} &= \ln D^{\mathcal{E}}(l_{it}, m_{it}^{*}) - s_{it}^{L} \\
&= \ln D^{\mathcal{E}}(l_{it}, m_{it} + e_{it}) - s_{it}^{L} \\
&= \ln D^{\mathcal{E}}(l_{it}, m_{it} + e_{it}) - \ln D^{\mathcal{E}}(l_{it}, m_{it}) + \varepsilon_{it} \\
&= \ln \left(\frac{D^{\mathcal{E}}(l_{it}, m_{it} + e_{it})}{D^{\mathcal{E}}(l_{it}, m_{it})} \right) + \varepsilon_{it}
\end{aligned}
$$

Note that $\ln \left(\frac{D^{\mathcal{E}}(l,m+e)}{D^{\mathcal{E}}(l,m)} \right) \geq 0$ is always positive because $e_{it} \geq 0$ and $D^{\mathcal{E}}$ is strictly increasing in m. Thus, $\frac{D^{\mathcal{E}}(l,m+e)}{D^{\mathcal{E}}(l,m)} \geq 1$.

11.7.2 Identifying Tax Evasion

Once we recovered the parameters of Equation **??**, we can invert it to solve for the true inputs m^{*} as a function of observed data and the output. Then, the difference between the observed and true inputs is by definition the tax evasion by overreporting.

$$
\begin{aligned}
m_{it}^{*} &= m_{it} + e_{it} \\
&= D^{-1}(\exp(s_{it}^{L} - \ln \mathcal{E} + \varepsilon_{it}), l_{it}) + e_{it}
\end{aligned}
$$

Again, having recovered the distribution of the output shock $f_{\varepsilon}(\varepsilon)$, we can use deconvolution techniques to learn the distribution of e. In particular, to learn the p.d.f of tax evasion $f_e(e)$, we can deconvolute it by

$$
\int f_e(m_{it}^{*} - D^{-1}(\exp(s_{it}^{L} - \ln \mathcal{E} - \varepsilon_{it}), l_{it})) f_{\varepsilon}(\varepsilon) d\varepsilon
$$

11.8 Non-Parametric Identification of Tax Evasion

The previous result also suggests a non-parametric identification strategy, as long as $D^{\mathcal{E}}$ is monotonic in m. This identification strategy is analogous to **?**, where the authors also require monotonicity and independence to recover a nonparametric function of m^* with nonclassical measurement error.

In our case, intuitively, if we know the density of ε and the function $D^{\mathcal{E}}$, the variation left is due to e, which can be recovered as long as we can vary $D^{\mathcal{E}}$ by moving e.

To see why the non-deductible flexible input is needed to identify tax evasion consider the following. Suppose that only the input M is flexible and deductible.

$$\ln\left(\frac{\rho M}{PY}\right) = \ln D^{\mathcal{E}}(K, L, M) - \varepsilon$$

$D^{\mathcal{E}}(K, L, M)$ is still identified by assumption 11.1, however, when we form the analogous of Equation **??**, we now have

$$
\begin{aligned}
\ln\left(\frac{\rho M^*}{PY}\right) - \ln D^{\mathcal{E}}(K,L,M^*) = & \\
& \ln\left(\frac{\rho(M+e)}{PY}\right) - \ln D^{\mathcal{E}}(K,L,M+e) \\
& = \ln\left(\frac{\rho(M+e)}{PY}\right) - \ln D^{\mathcal{E}}(K,L,M+e) \\
& \quad + \left[\ln\left(\frac{\rho M}{PY}\right) - \ln D^{\mathcal{E}}(K,L,M)\right] \\
& \quad - \left[\ln\left(\frac{\rho M}{PY}\right) - \ln D^{\mathcal{E}}(K,L,M)\right] \\
& = \ln\left(\frac{\rho M}{PY}\right) - \ln D^{\mathcal{E}}(K,L,M) \\
& \quad + \left[\ln\left(\frac{\rho(M+e)}{PY}\right) - \ln\left(\frac{\rho(M)}{PY}\right)\right] \\
& \quad - \left[\ln D^{\mathcal{E}}(K,L,M+e) - \ln D^{\mathcal{E}}(K,L,M)\right] \\
& = -\varepsilon \\
& \quad + \left[\ln D^{\mathcal{E}}(K,L,M+e) - \varepsilon - \ln D^{\mathcal{E}}(K,L,M) + \varepsilon\right] \\
& \quad - \left[\ln D^{\mathcal{E}}(K,L,M+e) - \ln D^{\mathcal{E}}(K,L,M)\right] \\
& = -\varepsilon(K,L,M)
\end{aligned}
$$

Now, we are not able to separate the variation of ε from e.

12 Implementation

We are interested in the distribution of tax evasion e but it cannot be observed. What is observed is the contaminated version $\mathcal{V}$ Equation **??**. Evasion e and the output shock ε are independent [11.2] with probability density distributions f_e and f_ε. Then, from Definition **??**

$$f_{\mathcal{V}}(\mathcal{V}) = \int f_e(\mathcal{V}+\varepsilon) f_\varepsilon(\varepsilon)\mathrm{d}\varepsilon$$

where $f_{\mathcal{V}}$ denotes the density of $\mathcal{V}$.

12.1 Parametric Deconvolution

Assume a functional form for $f_\varepsilon(\cdot;\gamma)$ that depends on known parameters γ. Assume a known functional form for the density $f_e(\cdot;\lambda)$ that depends on unknown parameters λ. We can estimate parameters λ by

$$\hat{\lambda} = \arg\max_{\lambda} \sum_{i=1}^{n} \log\left(\int f_e(\mathcal{V}_i + \varepsilon;\lambda) f_\varepsilon(\varepsilon;\gamma)\mathrm{d}\varepsilon\right) \tag{28}$$

Properties of MLE with unobserved scalar heterogeneity have been derived elsewhere before (**??**).

12.2 Non-Parametric Deconvolution by Splines

Following **?**, consider the following logspline model:

$$f_{e|\Theta}(e) = \exp(s(e;\theta) - C(\theta)), \quad L < e < R$$

where, $[L, R]$ define the support of f_e,

$$s(e;\theta) = \sum_{j=1}^{K-1} \theta_j B_j(e),$$

$\{B_j(E), j = 1, 2, \ldots, K\}$ is a sequence of cubic B-spline basis functions, and

$$C(\theta) = \log\left(\int_L^R \exp(s(e;\theta))\mathrm{d}e\right) < \infty.$$

Then $f_{e|\Theta}(e)$ is a positive density function on (L, R). The logspline model assumes that f_e is a twice continuously differentiable function. The last basis function is not used for identifiability purposes, given that by the property of B-splines, $\sum_{j=1}^{K} B_j(e) = 1$ for all $e \in (L, R)$.

The log likelihood of the observed variable $\mathcal{V}$ is

$$
\begin{aligned}
l_{\mathcal{V}}(\theta) &= \sum_{i=1}^{n} \log \left(\int f_{\varepsilon}(e - \mathcal{V}_{\rangle}) f_{e|\theta}(e) \mathrm{d}e \right) \\
&= \sum_{i=1}^{n} \log \left(\int f_{\varepsilon}(e - \mathcal{V}_{\rangle}) \exp(s(e; \theta) - C(\theta)) \mathrm{d}e \right) \\
&= \sum_{i=1}^{n} \log \left(\int f_{\varepsilon}(e - \mathcal{V}_{\rangle}) \exp(s(e; \theta)) \mathrm{d}e \right) - nC(\theta)
\end{aligned}
$$

The usual maximum likelihood estimate $\hat{\theta}$ is the maximizer of $l_{\mathcal{V}}(\theta)$.

To avoid overfitting, I add a penalty term to the log-likelihood function. The penalty term is a linear combination of the squared second derivatives of the B-spline basis functions. The second derivative of the B-splines is approximated by finite differences (**?**).

The penalty term is given by

$$
\text{Penalty}(\theta) = \lambda \cdot \|\mathbf{D}\theta\|^2
$$

Intuitively, we want to penalize excessive curvature in the log-density $\log f_e(e) \approx s(e; \theta) = \sum_j \theta_j B_j(e)$, where B_j are B-spline basis functions.

The second derivative of a smooth function is a measure of curvature. Since we don't have an analytical form for $s''(e)$, we can approximate it using finite differences on the coefficients θ.

Given a vector of spline coefficients $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$, define the second-order differences as:

$$\Delta^2 \theta_j = \theta_j - 2\theta_{j+1} + \theta_{j+2}$$

These values approximate the second derivative of the spline function at positions along the domain.

We can write all second differences as a matrix operation:

Let $D \in \mathbb{R}^{(k-2)\times k}$, then:

$$D\theta = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & -2 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & -2 & 1 & \ldots & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & \ldots & 0 & 1 & -2 & 1 & \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_k \end{bmatrix}$$

So $D\theta$ is a vector of second differences:

$$[\theta_1 - 2\theta_2 + \theta_3, \theta_2 - 2\theta_3 + \theta_4, \ldots]$$

The roughness penalty becomes:

$$\text{Penalty} = \lambda \cdot \|D\theta\|^2 = \lambda \cdot \sum_{j=1}^{k-2} (\Delta^2 \theta_j)^2$$

where λ is a smoothing parameter that can be choosen based on BIC/AIC or cross-validation.

The final log-likelihood function to be maximized is:

$$l_{\mathcal{V}}(\theta) = \sum_{i=1}^{n} \log \left(\int f_{\varepsilon}(e - \mathcal{V}_{\rangle}) \exp(s(e;\theta)) \mathrm{d}e \right) - nC(\theta) - \lambda \cdot \|D\theta\|^2$$

12.2.1 Simulations

Before deconvoluting with the actual data, I run some simulations to check the performance of my implmenetation in R.

I simulated a thousand observations of $\varepsilon \sim \mathcal{N}(0, 0.35)$ and $\log(e) \sim \mathcal{N}(-1.8, 0.95)$, and then I generated $\mathcal{V} = e - \varepsilon$. Figure **??** shows the histogram of the simulated data.

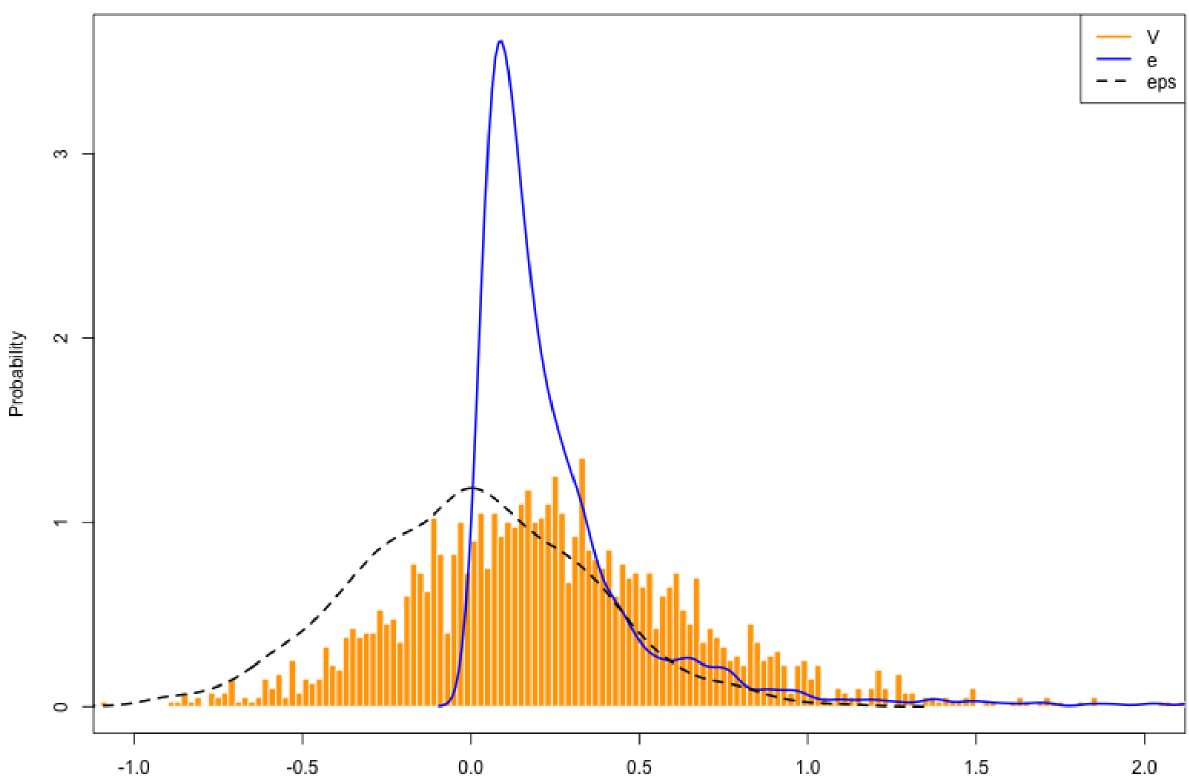


Figure 5: Simulated data

I then used the logspline model to estimate the parameters of the distribution of e. I used a B-spline basis with 10 knots. The knots where placed at equally spaced percentiles of the non-negative V observations. For the integral, I used Gauss Legendre quadrature. To ensure good coverage of the high probability region of the density, I used a recursive function to form a grid placing a higher number of nodes in the high probability region. I set the penalty parameter λ to 0.1. The estimated density is shown in Figure **??**.

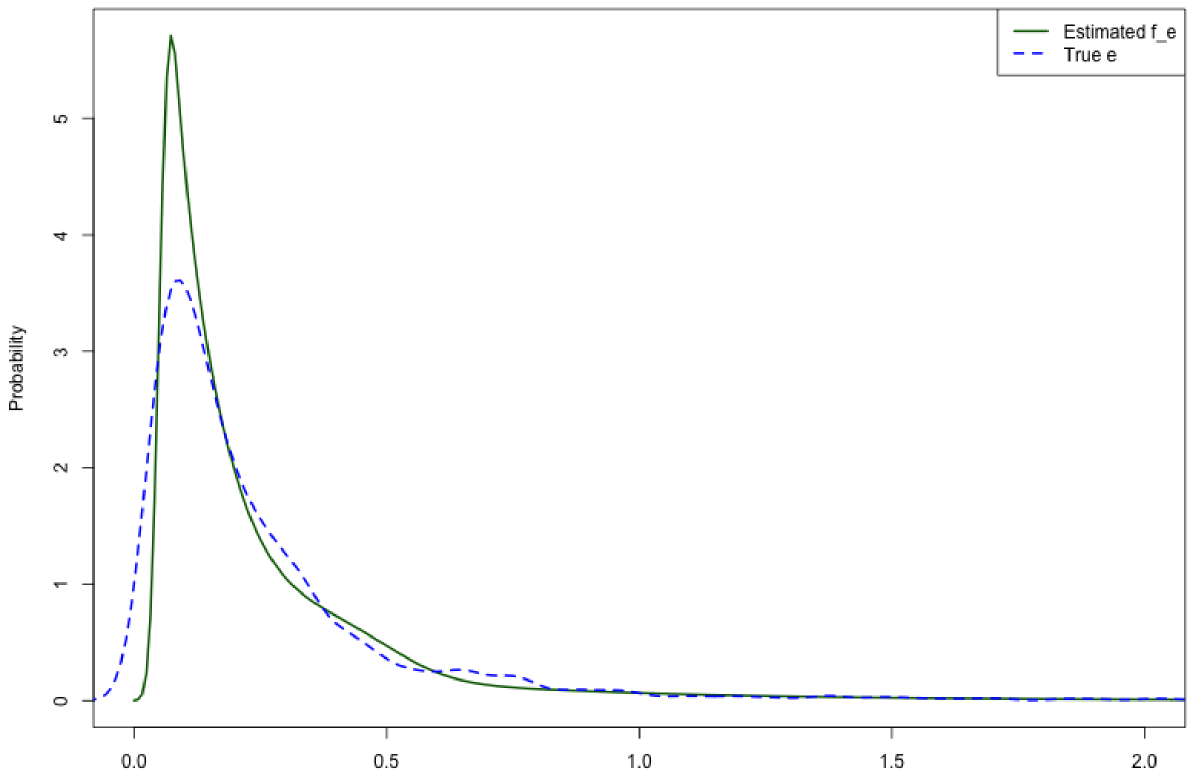


Figure 6: LogSpline Deconvolution

13 Results

14 Getting the Story Straight

Firms face incentives to evade sales taxes by overreporting their input costs. Firms overreport their cost by acquiring fake invoices to claim additional deductions of their sales taxes, maximizing after-tax profits. The higher the sales tax, the greater the incentive to evade. However, the probability of detection and the threat of penalties limit how much firms overreport.

The key incentive for firms is the **effective sales tax rate** — the amount firms ultimately pay or receive from the tax authority. The effective tax rate is calculated as the difference between the total sales tax charged by the firm on its sales and the total sales tax paid by the firm on its inputs, divided by total sales. A positive difference indicates what the firm should pay to the government. If the difference is negative, the firm receives a refund from the tax authority.

$$\tau = \frac{\tau_0 P_t Y_{it} - \tau_1 \rho_t M_{it}}{P_t Y_{it}}$$

In Colombia, the total amount charged by the firm as sales tax depends on the statutory tax rate of each product sold domestically; **exports are exempt** from sales tax. Low statutory sales tax rates and exempt products result in lower sales tax collected by the firm and thus, a lower effective tax rate. Consequently, firms whose products are exempt from sales tax or taxed with **low statutory rates** have low incentives to evade. Likewise, as the share of their exports increases, exporter firms bear decreasing incentives to evade taxes because a larger portion of their sales is exempt from sales tax, reducing the effective sales tax rate.

On the other hand, the total amount paid by the firm as sales tax depends on the statutory tax rate of each of its inputs, including imports. **Imports**, in contrast to exports, are **not exempt**. Hence, the composition of the foreign and domestic inputs does not affect the effective sales tax rate, nor the incentives to evade. However, the tax system imposed a zero sales tax rate on **unprocessed primary inputs**, such as forestry, mining, fishing, and agriculture. Therefore, firms consuming a high share of primary inputs face a higher effective sales tax rate and thus, greater incentives to evade.

Note that the effective sales tax rate might naturally be **negative** for firms in certain industries. Before the 1983 reform in Colombia, there were four different statutory sales tax rates: 4, 10, 15 and 25 percent. Because the sales tax was applied at the product level, *the total amount paid as sales tax by the firms on its inputs could be larger than the total amount charged by the firm as sales tax on its products.* This situation would result in a negative effective sales tax rate. Firms in this situation would receive a refund from the tax authority.

Would firms with negative effective tax rates have incentives to overreport inputs to increase their refunds? Unlikely. Although the refunds had to be made within 30 days, *the procedure was complicated, and the tax authority delayed, on average, one year to process the payment.* Suspicion of fraudulent activities might further delay the process. Therefore, firms in this situation did not have incentives to overreport inputs to increase their refunds.

(On the contrary, firms with negative effective tax rates would have incentives to *underreport* inputs to reduce the risk of being audited due to the refunds. These firms would be *better off reporting enough inputs to offset the sales tax charged on their products, thus avoiding the complications of refunds and potential audits.* This might explain the negative coefficients found in the statistical test in the empirical section of this paper. Why? Because corporations have incentives to report truthfully their inputs, not only because they are subject to harder scrutiny by the tax authority, but also because they would not like to reduce their profits artificially; profits influence the value of their shares, and thus, the wealth of their shareholders.

Underreporting inputs is the only reasonable explanation of the significantly negative coefficients found in the statistical test of tax evasion. Neither market power nor technology differences could explain the negative coefficients. Both market power on either side and technology differences would result in a smaller input cost share of revenues.

Market power in the sales market implies that firms can sell at higher prices, increasing the denominator of the share, resulting in a smaller share. Market power in the buying market implies firms can buy inputs at lower prices, decreasing the numerator of the share, leading to smaller shares, too. Because corporations are more likely to have market power, their smaller share would lead to a positive coefficient in the statistical test and smaller output elasticity of inputs.

Likewise, firms with better technology are more likely to be more efficient in using inputs, resulting in smaller input costs relative to revenues. Because corporations are more likely to have better technology, their smaller share would lead to a positive coefficient in the statistical test and smaller output elasticity of inputs. Therefore, the negative coefficients found in the statistical test cannot be explained by market power or technology differences, but rather by firms with negative effective tax rates underreporting their inputs to avoid complications with refunds and potential audits.)

Whether firms in certain industries have incentives to overreport their inputs to evade sales taxes might **not** be **obvious** *ex-ante*. For example, if the available data does not allow for estimating the average effective tax rate by type of firm in each industry.

A first contribution of this paper is to provide practitioners and policymakers a way to identify industries where tax evasion through cost-overreporting might be non-trivial and can be identified from data. The simple model employed in this paper to investigate the mechanisms of tax evasion provides conditions that can be tested with data. As a result, I propose a statistical test where the null hypothesis states that firms in an industry of interest do not evade sales taxes by overreporting; the alternative states that firms do commit tax evasion Section **??**.

14.1 Showing the Evidence

31 Food & Beverages:

- 311 Food Products (Meat, Dairy, Sugar, Bakery, etc.)
- 312 Food Products (Animal feeds, Others)
- 313 Beverages (Soft drinks, Alcoholic beverages, etc.)

In contrast to 311 and 312, 313 is not exempt from the sales tax. The evidence of tax evasion is significant as shown by the positive coefficient of the statistical test. Also, the output elasticity of intermediates is significantly smaller for non-evaders (Corporations) than for evaders (non-Corporations).

32 Textile, Apparel & Leather:

- 321 Textiles
- 322 Wearing apparel
- 323 Leather Products
- 324 Footwear

Table 21: Intermediate elasticities for Corporations (Non-Evaders) and non-Corporations (Evaders) for different industries. Estimates were obtained using GNR (2020) method assuming a CD functional form. Intermediates are defined as raw materials. ε is defined as **measurement error**. The Tax Evasion Test displays the coefficient of the statistical test, where the null hypothesis states that there is no tax evasion through cost overreporting. Among the industry characteristics, Sales is the industry's market share of the total sales in Colombia between 1981 and 1991. Sales Tax Rate is the average sales tax reported rate. Exports is the average share of exports of the industry total sales, and Exporters is the share of firms exporting 10 percent or more of their total sales.

	Elasticities		Test	Inds. Char		
SIC	Corps.	Others	Tax Ev.	Sales (Mkt %)	Sales Tax Rate	Expo
31 Food & Beverages						
311	0.6	0.61	-0.0085 (0.0124)	23.5	0.2	
	[0.57, 0.62]	[0.59, 0.62]				
312	0.58	0.57	-0.1493 (0.0257)	5.3		
	[0.49, 0.65]	[0.52, 0.61]				
313	0.29	0.35	0.0659 (0.0123)***	9.6	8.4	
	[0.26, 0.3]	[0.32, 0.39]				
32 Textile, Apparel & Leather						
321	0.39	0.44	0.0842 (0.0111)***	7.4	7.9	
	[0.35, 0.41]	[0.42, 0.45]				
322	0.35	0.42	1.1582 (0.0145)***	2.6	8.0	
	[0.3, 0.4]	[0.4, 0.43]				
323	0.56	0.47	-0.1731 (0.0179)	1.1	8.1	
	[0.5, 0.58]	[0.44, 0.5]				
324	0.43	0.46	0.0331 (0.0188)**	1.0	8.0	
	[0.36, 0.47]	[0.45, 0.47]				

Table ?? also shows evidence for tax evasion by overreporting in industries 321, 322, and 324, but not in 323. Corporations in industries 321, 322, and 324 display intermediate elasticities statistically smaller than the rest of the firms and the tax evasion test display a significantly positive coefficient. Not, for industry 323.

This makes sense because industry 323 displays by far the largest share of exporters and firms exports represent the largest average share of sales. Because exports are zero-rated for sales tax, firms in this industry have low incentives to evade taxes by overreporting their inputs.

33 Wood & Wood Products:

- 331 Wood Products except furniture (Sawmills, Containers, etc.)
- 332 Wood Furniture (Wooden furniture and fixtures)

Both industries sell their products locally (low shares of exports) and one third of their joint consumption comes from forestry, which is exempt of sales tax (IO-Matrix 1985 Colombia). For industry 331, the large share of primary goods probably leads to a high effective sales tax rate because we expect there might not be many more other inputs needed in the production process. Hence, firms in this industry might face high incentives to overreport their non-primary inputs to evade sales taxes.

However, for firms in industry 332, even though they also consume a large share of primary goods, they also consume a more diverse set of inputs to produce the wood furniture, such as textiles, chemicals, foam cushions, and others. Because of this wide range of inputs, the sales tax paid for their inputs might be high leading to a low effective sales rate and hence, low incentives to evade taxes.

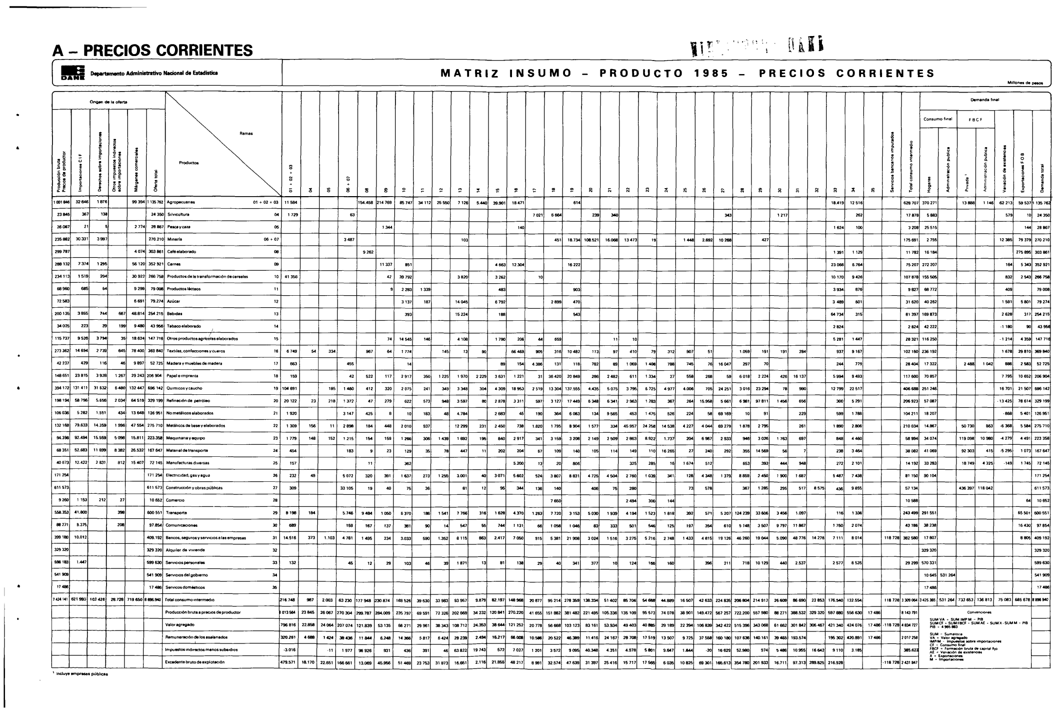


Figure 7: IO-Matrix 1985 Colombia

Figure 8

14.2 Production Function

Table **??** displays the results of estimating the production function parameters following the identification strategy described above. I use an AR(1) model for the Markov process of productivity. I employed $Z = \{k_{it}, l_{it}\}$ as instruments for themselves. One additional instrument to control for the endogeneity between ε_{it-1} and $\mathcal{W}_{it-1}$ is needed in the first stage of 'TE+GNR'. The instrument used is specified in the table Table **??**.

To compare, I also include the uncorrected estimates using GNR and a simple OLS. In the case of GNR estimation method, the Markov process of productivity is a third degree polynomial, and capital and labor, the instruments.

I focus on the top 5 industries in terms of revenue: 1) 311 - Food Products, 2) 313 Beverages, 3) 352 Other Chemical Products. 4) 351 Industrial Chemicals, and 5) 321 Textiles.

From Table **??**, we can conclude that m^*_{t-1} is a better than $\mathcal{W}_{t-2}$, k_{t-1} and l_{t-1}. The following discussion will be focusing on the results relating using m^*_{t-1} as the instrument.

First, because tax evasion artificially inflates the intermediates's cost share of revenues, lower output elasticities of intermediates than traditional methods are expected. In table Table **??** shows that correcting for tax evasion results in smaller intermediates' output elasticity with respect with OLS. With respect to GNR, all estimates are statistically lower except one. The output elasticity of intermediates of the Textile industry (321) is significantly greater than the GNR result.

Three of the five intermediate elasticities resulted as not statistically different from zero. These zeros are not exclusive to the method correcting by tax evasion, as we can see that GNR for Cobb-Douglas approach zero for some industries. These zeros could be the result of the functional form. A Cobb-Douglas specification leaves out the cross-elasticities which

Table 22: Cobb-Douglas production function parameters estimates, correcting intermediates —raw materials— for tax evasion vs. naive estimation and OLS. h is an AR(1) process for 'Tax Evasion + GNR' but a third degree polynomial for 'GNR'. $Z = \{k_{it}, l_{it}\}$ for GNR. One additional instrument to GNR's Z control for the endogeneity between ε_{it-1} and $\mathcal{W}_{it-1}$ in the first stage of 'TE+GNR'. The instrument used is specified in the table.

		Tax Evasion + GNR					
Inds.	Input	k_{it-1}	l_{it-1}	m^*_{it-1}	$\mathcal{W}\_{it-2}$	GNR	OLS
322	m	0.3123	0.3123	0.3123	0.3123	0.3683	0.5669
	k	0.0911	0.0540	0.1698	0.1875	0.1921	0.0477
	l	0.3125	0.2986	0.4076	0.4106	0.3608	0.4119
331	m	0.2937	0.2937	0.2937	0.2937	0.3485	0.5593
	k	0.0388	0.0000	0.1156	0.0721	0.1043	0.0636
	l	0.5819	0.3383	0.6179	0.5809	0.5063	0.4311
321	m	0.3544	0.3544	0.3544	0.3544	0.3869	0.6616
	k	0.0697	0.0000	0.3389	0.3214	0.2604	0.0966
	l	0.1324	0.2780	0.2987	0.3383	0.3412	0.2796
342	m	0.3096	0.3096	0.3096	0.3096	0.3350	0.5789
	k	0.0414	0.0916	0.2836	0.3016	0.2426	0.0716
	l	0.2016	0.2105	0.3703	0.4034	0.4698	0.4147
313	m	0.2701	0.2701	0.2701	0.2701	0.2826	0.7497
	k	0.0160	1.0000	0.4610	0.0000	0.3395	0.1171
	l	0.0283	0.1592	0.1229	0.2157	0.2943	0.2702
311	m	0.5530	0.5530	0.5530	0.5530	0.5630	0.7907
	k	0.2196	0.2768	0.2661	0.2716	0.2443	0.0559
	l	0.2393	0.0990	0.1963	0.1688	0.2084	0.1750
381	m	0.3743	0.3743	0.3743	0.3743	0.3719	0.6449
	k	0.0641	[illegible]	0.2450	0.2430	[illegible]	[illegible]

Table 23: Cobb-Douglas production function parameters estimates, correcting intermediates —raw materials— for tax evasion vs. naive estimation and OLS. h is an AR(1) process for 'Tax Evasion + GNR' but a third degree polynomial for 'GNR'. $Z = \{k_{it}, l_{it}\}$ for GNR. One additional instrument to GNR's Z control for the endogeneity between ε_{it-1} and $\mathcal{W}_{it-1}$ in the first stage of 'TE+GNR'. The instrument used is specified in the table. ε_{it-1} is defined as measurement error.

		Tax Evasion + GNR					
Inds.	Input	k_{it-1}	l_{it-1}	m^*_{it-1}	$\mathcal{W}_{it-2}$	GNR	OLS
322	m	0.3469	0.3469	0.3469	0.3469	0.4218	0.5669
	k	0.0865	0.0520	0.1612	0.1781	0.1759	0.0477
	l	0.2968	0.2837	0.3871	0.3900	0.3302	0.4119
331	m	0.3209	0.3209	0.3209	0.3209	0.3989	0.5593
	k	0.0373	0.0000	0.1112	0.0693	0.0962	0.0636
	l	0.5596	0.3253	0.5941	0.5587	0.4671	0.4311
321	m	0.3869	0.3869	0.3869	0.3869	0.4316	0.6616
	k	0.0662	0.0000	0.3218	0.3052	0.2414	0.0966
	l	0.1257	0.2640	0.2838	0.3213	0.3163	0.2796
342	m	0.3431	0.3431	0.3431	0.3431	0.3642	0.5789
	k	0.0394	0.0854	0.2699	0.2870	0.2319	0.0716
	l	0.1919	0.2007	0.3523	0.3839	0.4492	0.4147
313	m	0.2894	0.2894	0.2894	0.2894	0.3091	0.7497
	k	0.0155	1.0000	0.4489	0.0000	0.3270	0.1171
	l	0.0276	0.1531	0.1197	0.2100	0.2834	0.2702
311	m	0.5991	0.5991	0.5991	0.5991	0.6108	0.7907
	k	0.1971	0.2483	0.2386	0.2436	0.2176	0.0559
	l	0.2144	0.0888	0.1761	0.1514	0.1856	0.1750
381		0.4986	0.4986	0.4986	0.4986	0.4965	0.6449

Table 24: First stage 'Weak Instruments' F-test of the 2SLS-IV estimation. The null hypothesis is that the coefficients are equal to zero, in other words, uncorrelated with the endogenous variable $\mathcal{W}_{it-1}$.

Ins.	Inds.	df1	df2	statistic	p-value	stars
k_{t-1}	322	1	3581	1419.5262	0.0000	***
	331	1	785	121.4029	0.0000	***
	321	1	2339	3714.9184	0.0000	***
	342	1	1707	4800.6672	0.0000	***
	313	1	985	2758.2637	0.0000	***
	311	1	5102	5.4289	0.0198	**
	381	1	3060	3212.6117	0.0000	***
l_{t-1}	322	1	3581	3840.2193	0.0000	***
	331	1	785	778.9743	0.0000	***
	321	1	2339	4176.4138	0.0000	***
	342	1	1707	7325.5613	0.0000	***
	313	1	985	435.1505	0.0000	***
	311	1	5102	213.1251	0.0000	***
	381	1	3060	7087.4688	0.0000	***
m^*_{t-1}	322	1	3581	4963.1390	0.0000	***
	331	1	785	416.9415	0.0000	***
	321	1	2339	420.5456	0.0000	***
	342	1	1707	844.1156	0.0000	***
	313	1	985	458.2628	0.0000	***
	311	1	5102	1344.3363	0.0000	***
	381	1	3060	1275.1243	0.0000	***
$\mathcal{W}_{t-2}$	322	1	2949	8673.2398	0.0000	***
	331	1	636	1061.0846	0.0000	***

could lead to omitted variables bias. In other words, we effectively ignore complementarity or substitutability between the inputs. This CD omitted variable bias could be accentutated because we defined raw materials as intermediates leaving out others, such as energy, fuels, and services.

The elasticities of capital and labor vary across industries, with some being lower, some higher, and others quite similar. This is expected. The direction of the bias in the elasticities of capital and labor depends on whether thse inputs are complements or subsitutes of intermediate inputs. This relatioship will vary by industry. Put differently, for different industries, the cross-elasticity between intermediates and capital, and intermediates and labor will be different in direction and magnitude. Since these cross-elasticities are not included in the CD production function, the elasticities of capital and labor should reflect different degrees of these variations once intermediates are corrected for overreporting.

14.3 Productivity

Once, the production functio parameters are known, we can recover ω by deconvolution using Equation **??**. Here I assume ω follows a normal distribution.

?@tbl-boot-omega displays the estimates of the distribution parameters of ω corresponding to the instrument used in the previous section and the parametric specification of an AR(1).

To compare these productivity estimates to uncorrected ones, I simulate productivity using the parameters estimated in **?@tbl-boot-omega**. I use 1000 independent draws from ω and ε. Then, I computed productivity, $\varphi_{it}^S = \exp\{\omega_{it}^S + \varepsilon_{it}^S\}$.

Table **??** compares summary statistics the simulated productivity using the estimated parameters after correcting for tax evasion with the summary statistics of the productivity

Table 25: Estimates Omega Distribution by Industry after Deconvoluting from the Output Shock.

	μ	σ
322	5.43	0.58
342	4.25	0.47
313	3.29	0.45
351	3.47	0.65
331	4.04	0.38

estimates of GNR.

Industries 321 Textiles, and 351 Industrial Chemicals, and 313 Beverages have means lower or equal to the uncorrected ones for both instruments considered. Industry 311 Food Products display a higher mean for both estimates, while industry 352 Other Chemicals displays a higher or lower mean depending on the instrument.

With respect to standard deviations, industries 321 Textiles, 351 Industrial Chemicals, and 352 Other Chemicals display lower magnitudes, while industries 311 Food Products display significantly higher magnitudes. Lastly, industry 313 Beverages standard deviation is lower or higher depending on the instrument.

14.4 Testing for the Presence of Tax Evasion Through Overreporting

Equation **??** suggest a way to test the presence of tax evasion through cost overreporting. Let $\mathbb{E}[\mathcal{V}_{it}] \equiv \mu_{\mathcal{V}}$. Define the null hypothesis as the absence of cost overreporting, $H_0 : \mu_{\mathcal{V}} = 0$, and

Table 26: Summary statistics of the productivity estimates using GNR correcting and not correcting for tax evasion. Tax-evasion corrected estimates were simulated using the parameters of **?@tbl-boot-omega**, with 1000 draws.

inds	Method	Mean	SD	Q1	Median	Q3
313	GNR	70.6	44.7	45.1	59.0	82.7
	TE: m^*_{t-1}	55.8	39.9	29.3	47.0	70.2
321	GNR	23.6	11.0	17.4	21.5	27.3
	TE: m^*_{t-1}	21.6	13.1	12.6	18.2	26.7
322	GNR	41.6	14.3	32.2	39.1	48.6
	TE: m^*_{t-1}	83.1	54.5	47.2	68.1	106.0
331	GNR	61.1	23.5	46.3	56.5	70.3
	TE: m^*_{t-1}	73.8	44.5	42.9	63.7	93.0
342	GNR	23.1	8.3	18.1	21.7	26.3
	TE: m^*_{t-1}	32.1	18.4	18.5	27.7	40.7

the alternative hypothesis as the presence of cost overreporting, $H_1 : \mu_{\mathcal{V}} > 0$. Consequently, we can use a one-sided t-test to verify for the presence of tax evasion by overreporting.

Under the null hypothesis, there is no tax evasion. Then,

$$
\begin{aligned}
z = \frac{\bar{\mathcal{V}}}{\hat{\sigma}_{\mathcal{V}}/\sqrt{N}} &= \frac{\mathbb{E}[\mathcal{V}]}{\sqrt{\mathbb{V}\partial\diagdown[\mathcal{V}]/N}} \\
&= \frac{-\mathbb{E}[\varepsilon] + \mathbb{E}[e]}{\sqrt{\mathbb{V}\partial\diagdown[\varepsilon] + \mathbb{V}\partial\diagdown[e]/N}} \\
&= \frac{-\mathbb{E}[\varepsilon]}{\sqrt{\mathbb{V}\partial\diagdown[\varepsilon]/N}} \sim N(0,1)
\end{aligned}
$$

I test for the presence of tax evasion for different classifications of intermediates. Intermediates include raw materials, energy and services. Deductibles include raw materials and deductible expenses. Materials include only raw materials. The items included as deductible expenses or services are detailed in Table **??**

Table **??** shows that the null hypothesis of no tax evasion is rejected at the 1% significance level for twelve of the top twenty manufacturing industries.

In particular, there is no evidence of tax evasion for most industries in which products or raw materials are exempt of sales taxes, such as 312 (Other Food Products), 382 (Non-Electrical Machinery), 384 (Transport Equipment), 323 (Leather Products), and 369 (Non-Metallic Mineral Products); there is also no evidence of materials overreporting for 356 (Plastic Products)industry, whose main raw materials are likely to be specialized and supplied by few local and international suppliers.

Among the industries with exempted products, 311 (Food Products) and 383 (Electrical Machinery) there is evidence of tax evasion but as we'll see later the average overreporting is low compared to other industries.

As expected, the evidence is stronger particularly for the other industries that do not fall

Table 27: Tax Evasion Through Cost-Overreporting One-Side t-Test by Industry in Colombia. Under the null hypothesis, there is no tax evasion. Values of the statistic were computed from Equation ?? for different intermediate inputs. Standard errors shown in parethesis. Stars indicate significance level at the 1% (***), 5% (**), and 10% (*).

SIC	Category	Deductibles	Materials	Electricity	Fuels
311	1 Exempt Product	0.0203 (0.0056)***	-0.0085 (0.0124)	0.2154 (0.0137)***	0.0741 (
312	1 Exempt Product	0.0117 (0.0127)	-0.1493 (0.0257)	0.1471 (0.0373)***	0.4202 (
383	1 Exempt Product	0.0519 (0.0105)***	0.0413 (0.0113)***	0.1495 (0.0238)***	0.301 (0
322	4 Direct Customer Sales	0.582 (0.009)***	1.1582 (0.0145)***	-0.1566 (0.0131)	-0.398 (
342	4 Direct Customer Sales	0.0618 (0.0084)***	0.0747 (0.0097)***	0.1848 (0.016)***	-0.0632
390	4 Direct Customer Sales	0.082 (0.0132)***	0.061 (0.0146)***	0.476 (0.0338)***	0.7123 (
321		0.0752 (0.0079)***	0.0842 (0.0111)***	-0.3619 (0.0201)	-0.6187
313		0.0377 (0.0111)***	0.0659 (0.0123)***	-0.3124 (0.0235)	-0.2885
351		0.0144 (0.0148)	0.0417 (0.026)*	-0.1065 (0.0486)	0.1192 (
352		0.0441 (0.008)***	0.0322 (0.0098)***	0.1309 (0.0191)***	0.2939 (

Table 28: Tax Evasion Through Cost-Overreporting One-Side t-Test by Industry in Colombia. Under the null hypothesis, there is no tax evasion. Values of the statistic were computed from Equation ??. Standard errors shown in parethesis. Stars indicate significance level at the 1% (***), 5% (**), and 10% (*).

sic_3	sic_3_lab
311	Food Manufacturing
313	Beverage Industries
352	Other Chemical Products
351	Industrial Chemicals
321	Textiles
384	Transport Equipment
312	Food Manufacturing
341	Paper and Paper Products
381	Fabricated Metal Products, Except Machinery and Equipment
383	Electrical Machinery Apparatus, Appliances and Supplies
369	Other Non-Metallic Mineral Products
356	Plastic Products Not Elsewhere Classified
342	Printing, Publishing and Allied Industries
322	Wearing Apparel, Except Footwear
382	Machinery Except Electrical
323	Leather and Products of Leather, Leather Substitutes and Fur, Except Footwear and Wearing A
324	Footwear, Except Vulcanized or Moulded Rubber or Plastic Footwear
390	Other Manufacturing Industries
331	Wood and Wood and Cork Products, Except Furniture
332	Furniture and Fixtures, Except Primarily of Metal

in the previous categories. Namely, there is evidence of tax evasion for industries 313 (Beverages); for 321 (Textiles), 322 (Wearing Apparel), and 324 (Footwear); 331 (Wood Products except Furniture) and 332 (Wood Furniture); for industry 341 (Paper) and 342 (Publishing); 351 (Industrial Chemicals) and 352 (Other Chemicals); 381 (Metal Products Except Machinery), and 390 (Other Manufacturing Industries).

14.5 Deconvoluting Tax Evasion Using Moments

One simple way to start with the deconvolution is using moments. In particular, for every n-th moment $\mathbb{E}[\varepsilon_{it}^n|\Theta^{NE}] = \mathbb{E}[\varepsilon_{it}^n|t] = \mathbb{E}[\varepsilon_{it}^n]$

Therefore, any moment of the tax evasion e_{it} distribution $\forall t \in T$ can be estimated in theory.

Namely, from Equation **??**, we can estimate the average tax evasion by

$$\begin{aligned}\mathbb{E}[e_{it}|t] &= \mathbb{E}[\mathcal{V}_{it}|t] + \mathbb{E}[\varepsilon_{it}] \\ &= \mathbb{E}[\mathcal{V}_{it}|t] + \mu_{\varepsilon}\end{aligned}$$

Note that we learned the distribution $f_{\varepsilon}(\varepsilon)$ of ε from the first stage, so μ_{ε} and σ_{ε} are known.

Table **??** displays the estimated average of the log fraction that firms increase their costs of raw materials to evade taxes by claiming a greater deductible amount of their owed sales taxes.

The top five tax evading industries, 322 (Wearing Apparel), 342 (Publishing), 313 (Beverages), 351 (Industrial Chemicals), and 331 (Wood Products), display an average tax evasion e greater than 12%, which is non-trivial.

Although deconvolution by moments is the simplest method, it displays the undesirable characteristic that estimate of variances frequently result with a negative sign. In the next section, I address this problem by using parametric MLE.

Table 29: Average tax evasion by Industry. Estimates show the average tax evasion from the output shock in Equation ??. LCI and UCI are the bias-corrected bootsrap confidence intervals at the 10% significance level with 250 bootstrap replicates. Intermediates are defined as raw materials.

SIC	Inter	Mean	LCI	UCI
322	Deductibles	0.1589	0.0373	0.2199
	Materials	0.1955	0.0910	0.2895
331	Deductibles	0.1205	0.0140	0.2458
	Materials	0.2177	0.0475	0.3956
321	Deductibles	0.0885	0.0530	0.1465
	Materials	0.1094	0.0721	0.1782
342	Deductibles	0.0436	-0.0255	0.1162
	Materials	0.0597	0.0005	0.1572
313	Deductibles	0.0377	0.0058	0.0759
	Materials	0.0659	0.0320	0.1077
311	Deductibles	0.0071	-0.0360	0.0242
	Materials	0.0193	-0.0074	0.0634
381	Deductibles	-0.0035	-0.0460	0.0425
	Materials	-0.0053	-0.0494	0.0457

14.6 Deconvoluting by Parametric MLE

I use parametric MLE to obtain better estimates of the variances using equation Equation **??**. I assume that the error term ε follows a normal distribution.

For tax-evasion, theory suggests that firms only have incentives to overreport costs, not to underreport them. Therefore, as explained elsewhere, it might be expected that overreporting $e \geq 0$ is non-negative. In addition, it might also be expected that most firms overreport a little and a few firms overreport greater amounts. Therefore, a lognormal or a truncated normal distribution might be more appropriate.

By definition, if a random variable U is log-normal distributed, then $log(U) \sim N(\mu, \sigma)$. Thus, we cannot directly compare the parameters of the log-normal distribution to our previous estimates. We can however, used the parameters to compute any moment of the log-normally distributed variable U by

$$E[U^n] = e^{n\mu + \frac{1}{2}n^2\sigma^2}$$

In particular, the first moment and the variance are computed as follows,

$$E[U] = e^{\mu + \frac{1}{2}\sigma^2}$$

$$Var[U] = E[U^2] - E[U]^2 = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

In addition, the mode and the mean by

$$\text{Mode}[U] = e^{\mu - \sigma^2}$$

$$\text{Med}[U] = e^{\mu}$$

Likewise, the probability density function of random variable U with a normal distribution with parameters $\tilde{\mu}$ and $\tilde{\sigma}$ and truncated from below at zero is

$$f(u; \tilde{\mu}, \tilde{\sigma}) = \frac{\varphi(\frac{u-\tilde{\mu}}{\tilde{\sigma}})}{\tilde{\sigma}(1 - \Phi(\alpha))}$$

where

$$\varphi(\xi) = \frac{1}{2\pi} \exp(-\frac{1}{2}\xi^2)$$

is the probability density function of the standard normal distribution, $\Phi(.)$ is its cumulative distribution function, and $\alpha = -\tilde{\mu}/\tilde{\sigma}$

Then, the mean becomes

$$E[U|U > 0] = \tilde{\mu} + \tilde{\sigma}\frac{\varphi(\alpha)}{1 - \Phi(\alpha)}$$

and the variance, median, and mode become

$$\begin{aligned} Var[U|U > 0] &= \sigma^2[1 + \alpha\varphi(\alpha)/(1 - \Phi(\alpha)) - (\varphi(\alpha)/[1 - \Phi(\alpha)])^2] \\ \text{Median}[U|U > 0] &= \tilde{\mu} + \Phi^{-1}\left(\frac{\Phi(\alpha) + 1}{2}\right)\tilde{\sigma} \\ \text{Mode}[U|U > 0] &= \tilde{\mu} \end{aligned}$$

Table **??** displays the estimates of the parameters for both the log-normal and truncated normal distributions using an MLE approach. Other moments of the densities are computed as explained previously and shown for reference.

Table 30: Estimates of Deconvoluting Tax Evasion from the Output Shock in Equation **??** by Parametric MLE Using A Log-Normal and a Truncated-Normal Distribution. μ and σ are the parameters of the densities. Moments displayed are estimated using the distribution paramters as explained above. Intermediates are defined as raw materials.

sic_3	dist	mu	sigma	mean	sd	mode	median
311	lognormal	-1.7460	0.9181	0.2659	0.3059	0.0751	0.1745
	truncated normal	0.2280	0.0894	0.2294	0.0876	0.2280	0.2286
313	lognormal	-1.4006	0.9855	0.4005	0.5131	0.0933	0.2465
	truncated normal	0.3285	0.1647	0.3376	0.1549	0.3285	0.3332
321	lognormal	-1.4011	0.8883	0.3655	0.4006	0.1119	0.2463
	truncated normal	0.3183	0.1302	0.3209	0.1269	0.3183	0.3195
352	lognormal	-1.5406	0.9517	0.3370	0.4091	0.0866	0.2143
	truncated normal	0.2847	0.1218	0.2879	0.1180	0.2847	0.2862
383	lognormal	-1.5748	0.9288	0.3187	0.3730	0.0874	0.2071
	truncated normal	0.2718	0.1230	0.2761	0.1180	0.2718	0.2739

Both the log-normal and the truncated normal distributions point to similar means. The differences between the other moments of the distributions can be explained by the differences in the shapes of their probability density functions. The standard deviation is larger in the log-normal distribution than in the truncated normal distribution which can be explained by the asymmetry of the log normal distribution which is skewed to the right. With respect to the mode, it is the same to the mean in the case of the truncated normal distribution, while it is lower than the median for the log-normal. Finally, the median is lower than the mean in the case of the log-normal, but higher than the mean in the case of the truncated normal because of the truncation from below.

Both distributions show higher estimates of the average overreporting in logs than using only moments. Now it looks like firms evade taxes by inflating their cost of their materials by 40 percent or more.

Table **??** shows the bias corrected bootstrap confidence intervals for the mean of each distribution by industry using 200 replicates.

Table 31: Estimates of Deconvoluting Tax Evasion from the Output Shock in Equation **??** by Parametric MLE Using A Log-Normal and a Truncated-Normal Distribution. LCI and UCI are the bias corrected bootstrap confidence intervals at the 5 percent significance level using 250 replicates.

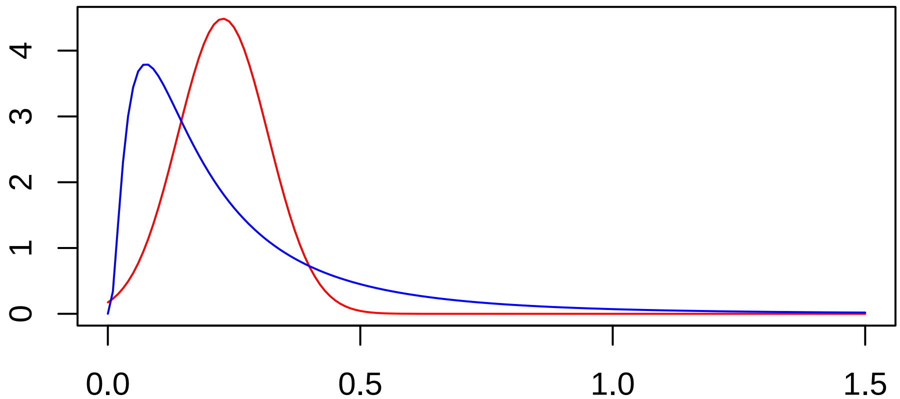
SIC	Density	Mean	SD
311	lognormal	0.2659	0.3059
		[0.2285, 0.2999]	[0.2069, 0.3536]
	truncated normal	0.2294	0.0876
		[0.2116, 0.2597]	[0.0791, 0.0991]
313	lognormal	0.4005	0.5131
		[0.3407, 0.468]	[0.3279, 0.6767]
	truncated normal	0.3376	0.1549
		[0.2965, 0.3831]	[0.1343, 0.1791]

352	lognormal	0.337	0.4091
		[0.2718, 0.3804]	[0.2442, 0.4747]
	truncated normal	0.2879	0.118
		[0.2509, 0.3287]	[0.1046, 0.134]
321	lognormal	0.3655	0.4006
		[0.3176, 0.4179]	[0.2702, 0.4448]
	truncated normal	0.3209	0.1269
		[0.2909, 0.3751]	[0.1183, 0.1443]
383	lognormal	0.3187	0.373
		[0.2613, 0.3727]	[0.2391, 0.4416]
	truncated normal	0.2761	0.118
		[0.226, 0.3368]	[0.1082, 0.1351]

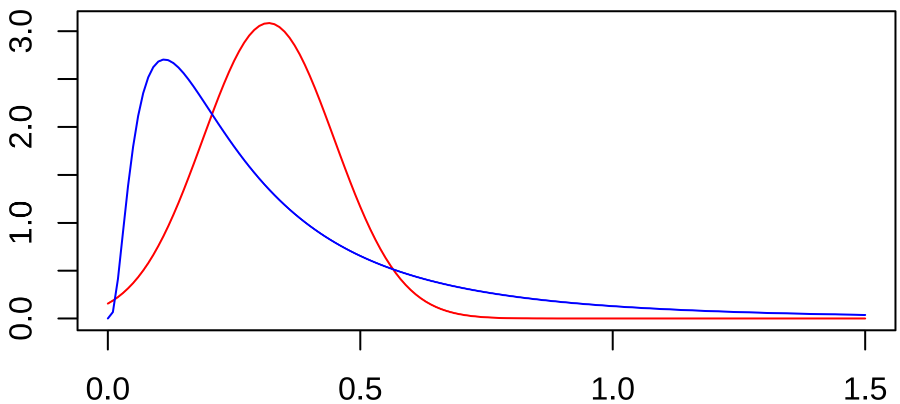
The difference between our previous estimates using moments can be explained by the fact that the log-normal and truncated distribution are restricted to positive values while we did not apply the restriction when using moments. Were we to take the average of only the positive values, the mean of the moments' method would be higher and closer to the log-normal and truncated normal distributions.

From the perspective of theory, firms would have incentives to overreport their materials. Therefore, using moments to estimate overreporting without restricting to positive values like in the case of the MLE method using a log-normal or truncated distribution might underestimate tax evasion through cost overreporting.

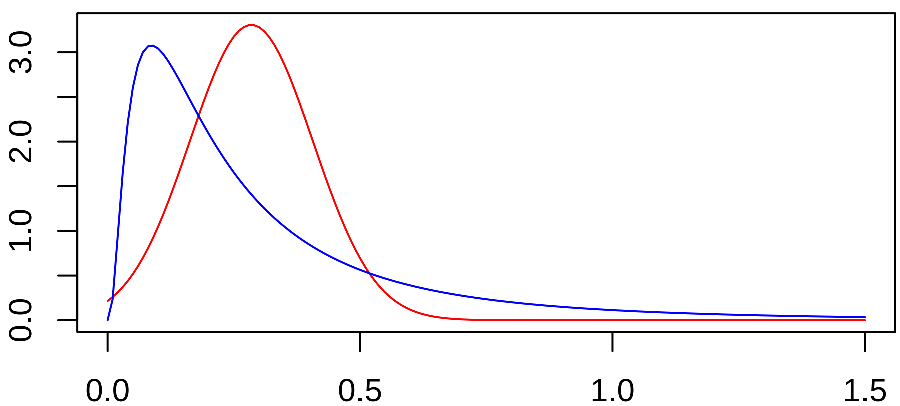
311 Food Manufacturing ...



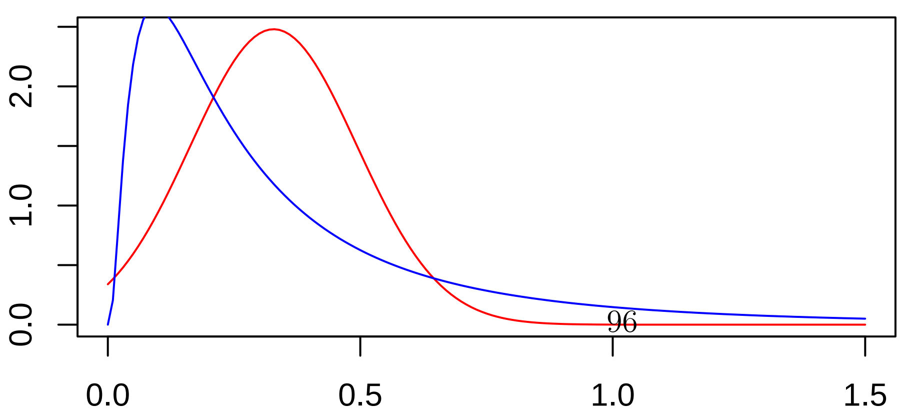
321 Textiles ...



352 Other Chemical Products ...



313 Beverage Industries ...



383 Electrical Machinery Apparatus ...

Table 32: Semiparametric Deconvolution of Tax Evasion using Penalized B-Splines. The error density function is assumed to be normally distributed with mean and variance obtained from the residuals of the first stage conditional on firms beign Non-Evaders.

sic_3	mean	sd
322	0.199	0.074
331	0.226	0.110
321	0.113	0.053
342	0.058	0.013
313	0.072	0.037
311	0.021	0.004
381	0.015	0.002

14.7 Non-Parametric Deconvolution using Penalized B-Splines

15 Tax Evasion and Resource Missallocation

A popular way to decompose total productivity introduced by **?** is the following :

$$\varphi_t = \bar{\varphi}_t + \sum_{i=1}^{I} (s_{it} - \bar{s}_t)(\varphi_{it} - \bar{\varphi}_t)$$

where $\varphi_{it} = \exp\{\omega_{it} + \varepsilon_{it}\}$ is the firm-level productivity, s_{it} is firm' i share of output at time t. $\bar{\varphi}_t$ and $\bar{s}_t$ represent the unweighted mean productivity and unweighted mean share.

Intuitively, the more output is allocated to high productivity firms the higher the total productivity will be. Alternatively, the smaller the shares allocated to low productivity firms the greater the total productivity.

Ideally, we would like resources to be allocated to high productivity firms. Hence, the link between the decomposition of total aggregate productivity and missallocation.

In the tax evasion context, the productivity of misreporting firms would look lower than it really is. How would this affect our measures of total productivity? If misreporting firms have small shares, their corrected productivity measure might be closer to the mean reducing their contribution to total productivity, leading to lower measured levels of total productivity.

In other words, tax evasion through cost overreporting might lead to biased measures of total aggregate productivity.

However, if we think about the marginal firm that is struggling to stay in the market, the additional income gained through misreporting might help her survive longer. Therefore, tax evasion increases missallocation in the sense that a share of the output is allocated to firms that otherwise would not be in the market.

There are alternative ways to decompose total productivity that account for the entry and exit of firms (Foster et al. 2008). In addition, alternative measures to aggregate productivity have been used elsewhere (e.g. aggregate demand shocks in **?**).

For example, we can decompose aggregate productivity like this

$$\Delta\varphi_t = \sum_{i\in C} \bar{s}_i \Delta\varphi_{it} + \sum_{i\in C} (\bar{\varphi}_i - \bar{\varphi})\Delta s_{it} + \sum_{i\in N} s_{it} \left(\varphi_{it} - \bar{\varphi}\right) - \sum_{i\in X} s_{it-1} \left(\varphi_{it-1} - \bar{\varphi}\right)$$

where Δ is the change from $t-1$ to t, and the sets C,N, and X represent the continuing, entering, and exiting firms.

In the case of corporate tax evasion a relevant measure could be aggregate profitability. Low productivity firms are surviving not because they are technologically efficient but because their profitability is marginally higher when they evade taxes. Decomposing an aggregate profitability measure by market shares and firm dynamics would indicate how exiting firms are contributing to the aggregate measure. Intuitively, again higher levels of aggregate profitability correspond to larger shares of the market being allocated to high profitability firms.

In the absence of tax evasion, firms overreporting costs would have lower after-tax profits. The aggregate profitability level would be higher if these firms have small market shares and remain in the sample. On the other hand, the aggregate profitability level would decrease if the small-share and low-profitability firms were to exit the market.

Therefore, in the case of tax evasion, the key aspect of aggregate profitability is the counterfactual decision of low-profitability firms to exit the market in addition to the corrected profits if firms were not to overreport their costs.

Consequently, the economic model should account for firm entry and exit along with tax evasion.

Table 33: Fortran implementation of the Non-Parametric GNR(2020) method to estimate production functions. Comparison with Stata code and replication data from GNR(2020), and the working data of this paper and my own Fortran implementation code.

	NP GNR								
	Replication Data			Working Data					
	Stata			Stata			Fortran		
	m	k	l	m	k	l	m	k	l
311	0.67	0.12	0.22	0.67	0.12	0.22	0.67	0.12	0.22
321	0.54	0.16	0.32	0.54	0.16	0.32	0.54	0.16	0.32
322	0.52	0.05	0.42	0.52	0.05	0.42	0.52	0.05	0.42
331	0.51	0.04	0.44	0.51	0.04	0.44	0.51	0.04	0.44
381	0.53	0.10	0.43	0.53	0.10	0.43	0.53	0.10	0.42

TABLE 2
AVERAGE INPUT ELASTICITIES OF OUTPUT
(Structural vs. Uncorrected OLS Estimates)

	Industry (ISIC Code)											
	Food Products (311)		Textiles (321)		Apparel (322)		Wood Products (331)		Fabricated Metals (381)		All	
	GNR	OLS	GNR	OLS	GNR	OLS	GNR	OLS	GNR	OLS	GNR	OLS
Colombia:												
Labor	.22	.15	.32	.21	.42	.32	.44	.32	.43	.29	.35	.26
	(.02)	(.01)	(.03)	(.02)	(.02)	(.01)	(.05)	(.03)	(.02)	(.02)	(.01)	(.01)
Capital	.12	.04	.16	.06	.05	.01	.04	.03	.10	.03	.14	.06
	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)	(.01)	(.00)
Intermediates	.67	.82	.54	.76	.52	.68	.51	.65	.53	.73	.54	.72
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.00)	(.00)
Sum	1.01	1.01	1.01	1.03	.99	1.01	.99	1.00	1.06	1.05	1.04	1.04
	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)	(.04)	(.02)	(.01)	(.01)	(.00)	(.00)
Mean (capital)/mean (labor)	.55	.27	.49	.27	.12	.04	.08	.08	.23	.11	.40	.23
	(.08)	(.07)	(.09)	(.06)	(.04)	(.02)	(.05)	(.05)	(.04)	(.04)	(.03)	(.01)
Chile:												
Labor	.28	.17	.45	.26	.45	.29	.40	.20	.52	.32	.38	.20
	(.01)	(.01)	(.03)	(.02)	(.02)	(.02)	(.02)	(.01)	(.03)	(.02)	(.01)	(.01)
Capital	.11	.05	.11	.06	.06	.03	.07	.02	.13	.07	.16	.09
	(.01)	(.00)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.00)	(.00)
Intermediates	.67	.83	.54	.75	.56	.74	.59	.81	.50	.71	.55	.77
	(.00)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.00)	(.00)
Sum	1.05	1.05	1.10	1.06	1.08	1.06	1.06	1.04	1.15	1.10	1.09	1.06
	(.01)	(.00)	(.02)	(.01)	(.02)	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.00)
Mean (capital)/mean (labor)	.39	.28	.24	.22	.14	.12	.18	.12	.25	.21	.43	.42
	(.03)	(.03)	(.04)	(.04)	(.03)	(.03)	(.03)	(.05)	(.03)	(.04)	(.02)	(.02)

NOTE.—Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates. For each industry, the numbers in the first column are based on a gross output specification using a complete polynomial series of degree two for each of the two nonparametric functions ($\mathcal{D}$ and $\mathcal{C}$) of our approach (GNR). The numbers in the second column are also based on a gross output specification and are estimated using a complete polynomial series of degree two with OLS. Since the input elasticities are heterogeneous across firms, we report the average input elasticities within each given industry. The "Sum" row reports the sum of the average labor, capital, and intermediate-input elasticities, and the "Mean (capital)/mean (labor)" row reports the ratio of the average capital elasticity to the average labor elasticity.

Figure 10: **?** Table 2 (GNR2020)

Table 34: Fortran implementation of the Non-Parametric GNR(2020) method to estimate production functions. Comparison of the error standard deviation and $\mathcal{E}$ between the Stata code and replication data from GNR(2020), and the working data of this paper and my own Fortran implementation code.

NP GNR						
	Replication Data		Working Data			
	Stata		Stata		Fortran	
	$\mathcal{E}$	err sd	$\mathcal{E}$	err sd	$\mathcal{E}$	err sd
311	1.05	0.24	1.05	0.24	1.05	0.24
321	1.03	0.23	1.03	0.23	1.03	0.23
322	1.02	0.22	1.02	0.22	1.02	0.22
331	1.04	0.25	1.04	0.25	1.04	0.25
381	1.02	0.20	1.02	0.20	1.02	0.20

Table 35: Fortran implementation of the Cobb-Douglas GNR(2020) method to estimate production functions. Comparison with Stata code and replication data from GNR(2020), and the working data of this paper and my own Fortran implementation code.

CD GNR									
	Replication Data			Working Data					
	Stata			Stata			Fortran		
	m	k	l	m	k	l	m	k	l
311	0.62	0.20	0.20	0.62	0.20	0.20	0.62	0.20	0.20
321	0.52	0.21	0.27	0.52	0.21	0.27	0.52	0.21	0.27
322	0.44	0.19	0.29	0.44	0.19	0.29	0.44	0.19	0.29
331	0.46	0.10	0.44	0.46	0.10	0.44	0.46	0.10	0.44
381	0.51	0.18	0.36	0.51	0.18	0.36	0.51	0.18	0.36

Table 36: Fortran implementation of the Cobb-Douglas GNR(2020) method to estimate production functions. Comparison of the error standard deviation and $\mathcal{E}$ between the Stata code and replication data from GNR(2020), and the working data of this paper and my own Fortran implementation code.

CD GNR						
	Replication Data		Working Data			
	Stata		Stata		Fortran	
	$\mathcal{E}$	err sd	$\mathcal{E}$	err sd	$\mathcal{E}$	err sd
311	1.11	0.33	1.11	0.33	1.11	0.33
321	1.06	0.30	1.06	0.30	1.06	0.30
322	1.16	0.43	1.16	0.43	1.16	0.43
331	1.10	0.39	1.10	0.39	1.10	0.39
381	1.05	0.29	1.05	0.29	1.05	0.29

Table 37: First State of Cobb-Douglas production function intermediate elasticities estimates using GNR Stata estimator and R estimator with trimmed data at 0.05. **All firms**.

		Working Data					
		Stata			R		
Industry	Inter.	m	$\mathcal{E}$	err sd	m	$\mathcal{E}$	err sd
311	Intermediates	0.65	1.07	0.31	0.65	1.07	0.31
	Materials	0.56	1.08	0.35	0.56	1.08	0.35
	Deductibles	0.56	1.12	0.39	0.56	1.12	0.39
321	Intermediates	0.52	1.06	0.30	0.52	1.06	0.3
	Materials	0.39	1.12	0.41	0.39	1.12	0.41
	Deductibles	0.43	1.09	0.37	0.43	1.09	0.37
322	Intermediates	0.45	1.13	0.41	0.45	1.13	0.41
	Materials	0.37	1.15	0.45	0.37	1.15	0.45
	Deductibles	0.37	1.15	0.45	0.37	1.15	0.45
331	Intermediates	0.47	1.09	0.38	0.47	1.09	0.38
	Materials	0.35	1.14	0.48	0.35	1.14	0.48
	Deductibles	0.39	1.12	0.43	0.39	1.12	0.43
381	Intermediates	0.51	1.05	0.29	0.51	1.05	0.29
	Materials	0.37	1.09	0.39	0.37	1.09	0.39
	Deductibles	0.41	1.07	0.35	0.41	1.07	0.35

16 References

17 Fortran GNR

17.1 Non-Parametric Production Function

17.2 Cobb-Douglas Production Function

17.3 CD GNR + trimming

17.4 CD GNR + Trimming + Corporations

17.5 CD GNR + Trimming + Measurement Error

17.6 Industry Characteristics

18 CD GNR Intermediates

Observations with value of zero for some intermediates were driving down the estimates of the output elasticity of intermediates. This is a common problem when using data from surveys. To avoid this problem, I trimmed the observations with a share of intermediates below 0.05.

These observations increase the variance of the error term, which in turn increases the value of $\mathcal{E}$. The higher the value of $\mathcal{E}$, the lower the value of the elasticity of intermediates.

Figure **??** show that at when trimming observations with a share of intermediates below 0.05, the estimates of $\mathcal{E}$ and the output elasticity of intermediates start to stabilize. That is the changes in the values of these variables are small. In addition, Figure **??** shows that at

Table 38: First State of Cobb-Douglas production function intermediate elasticities estimates using GNR Stata estimator and R estimator with trimmed data at 0.05. **Only Corporations**.

		Working Data					
		Stata.			R		
Industry	Inter.	m	$\mathcal{E}$	err sd	m	$\mathcal{E}$	err sd
311	Intermediates	0.67	1.06	0.28	0.67	1.06	0.28
	Materials	0.55	1.08	0.35	0.55	1.08	0.35
	Deductibles	0.57	1.09	0.35	0.57	1.09	0.35
321	Intermediates	0.51	1.03	0.24	0.51	1.03	0.24
	Materials	0.35	1.09	0.37	0.35	1.09	0.37
	Deductibles	0.40	1.07	0.33	0.4	1.07	0.33
322	Intermediates	0.25	1.59	0.82	0.25	1.59	0.82
	Materials	0.31	1.11	0.41	0.31	1.11	0.41
	Deductibles	0.34	1.09	0.38	0.34	1.09	0.38
331	Intermediates	0.50	1.03	0.24	0.5	1.03	0.24
	Materials	0.29	1.09	0.42	0.29	1.09	0.42
	Deductibles	0.37	1.05	0.30	0.37	1.05	0.3
381	Intermediates	0.54	1.04	0.27	0.54	1.04	0.27
	Materials	0.37	1.09	0.40	0.37	1.09	0.4
	Deductibles	0.41	1.08	0.37	0.41	1.08	0.37

Table 39: First State of Cobb-Douglas production function intermediate elasticities estimates using GNR Stata estimator and R estimator with trimmed data at 0.05. **Corporations vs. All**. ε is the **measurement error**.

		Working Data (R)					
		Corps.			All		
Industry	Inter.	m	$\mathcal{E}$	err sd	m	$\mathcal{E}$	err sd
311	Intermediates	0.71	1	0.28	0.69	1	0.31
	Materials	0.6	1	0.35	0.61	1	0.35
	Deductibles	0.62	1	0.35	0.63	1	0.39
321	Intermediates	0.53	1	0.24	0.55	1	0.3
	Materials	0.39	1	0.37	0.43	1	0.41
	Deductibles	0.43	1	0.33	0.47	1	0.37
322	Intermediates	0.4	1	0.82	0.51	1	0.41
	Materials	0.35	1	0.41	0.42	1	0.45
	Deductibles	0.37	1	0.38	0.43	1	0.45
331	Intermediates	0.51	1	0.24	0.51	1	0.38
	Materials	0.32	1	0.42	0.4	1	0.48
	Deductibles	0.39	1	0.3	0.44	1	0.43
381	Intermediates	0.56	1	0.27	0.53	1	0.29
	Materials	0.41	1	0.4	0.41	1	0.39
	Deductibles	0.44	1	0.37	0.44	1	0.35

Table 40: First State of Cobb-Douglas production function intermediate elasticities estimates using GNR Stata estimator and R estimator with trimmed data at 0.05. **Corporations vs. All**. ε is the **measurement error**. Top evading industries.

		Working Data (R)					
		Corps.			All		
Industry	Inter.	m	$\mathcal{E}$	err sd	m	$\mathcal{E}$	err sd
322	Intermediates	0.4	1	0.82	0.51	1	0.41
	Materials	0.35	1	0.41	0.42	1	0.45
	Deductibles	0.37	1	0.38	0.43	1	0.45
331	Intermediates	0.51	1	0.24	0.51	1	0.38
	Materials	0.32	1	0.42	0.4	1	0.48
	Deductibles	0.39	1	0.3	0.44	1	0.43
321	Intermediates	0.53	1	0.24	0.55	1	0.3
	Materials	0.39	1	0.37	0.43	1	0.41
	Deductibles	0.43	1	0.33	0.47	1	0.37
342	Intermediates	0.54	1	0.25	0.5	1	0.27
	Materials	0.34	1	0.41	0.36	1	0.37
	Deductibles	0.37	1	0.37	0.39	1	0.35
313	Intermediates	0.46	1	0.29	0.46	1	0.35
	Materials	0.29	1	0.36	0.31	1	0.41
	Deductibles	0.33	1	0.32	0.35	1	0.37

Table 41: Intermediate elasticities for different groups of firms using GNR method assuming a CD functional form. Data trimmed at 0.05 of the intermediates share cost of gross output. ε is defined as **measurement error**.

		m			
Inds.	Intermediate	Corps.	Exp.	Imp.	All
322	Intermediates	0.4	0.5	0.56	0.51
	Materials	0.35	0.41	0.42	0.42
	Deductibles	0.37	0.4	0.44	0.43
331	Intermediates	0.51	0.5	0.57	0.51
	Materials	0.32	0.38	0.42	0.4
	Deductibles	0.39	0.41	0.45	0.44
321	Intermediates	0.53	0.55	0.54	0.55
	Materials	0.39	0.45	0.41	0.43
	Deductibles	0.43	0.48	0.45	0.47
342	Intermediates	0.54	0.53	0.52	0.5
	Materials	0.34	0.38	0.34	0.36
	Deductibles	0.37	0.41	0.37	0.39
313	Intermediates	0.46	0.29	0.43	0.46
	Materials	0.29	0.23	0.27	0.31
	Deductibles	0.33	0.25	0.32	0.35
311	Intermediates	0.71	0.71	0.77	0.69
	Materials	0.6	0.64	0.67	0.61
	Deductibles	0.62	0.66	0.7	0.63
381	Intermediates	0.56	0.58	0.53	0.53
	Materials	0.41	0.42	0.4	0.41
	Deductibles	0.44	0.45	0.43	0.44

Table 42: One side test of tax evasion through inputs overreporting. Different non-evaders groups used as reference.

	Corporations		Exporters		
Inds.	Materials	Deductibles	Materials	Deductibles	Materials
322	1.1582 (0.0145)***	0.582 (0.009)***	0.1215 (0.0145)***	0.1103 (0.009)***	-0.1187 (0.0
331	0.1851 (0.0198)***	0.11 (0.0148)***	0.086 (0.0198)***	0.1038 (0.0148)***	-0.0852 (0.0
321	0.0842 (0.0111)***	0.0752 (0.0079)***	0.0604 (0.0111)***	0.115 (0.0079)***	0.016 (0.011
342	0.0747 (0.0097)***	0.0618 (0.0084)***	-0.0731 (0.0097)	-0.0611 (0.0084)	0.0396 (0.00
313	0.0659 (0.0123)***	0.0377 (0.0111)***	0.2794 (0.0123)***	0.3439 (0.0111)***	0.1218 (0.01
311	-0.0085 (0.0124)	0.0203 (0.0056)***	-0.1489 (0.0124)	-0.0416 (0.0056)	-0.2007 (0.0
381	-0.0041 (0.0066)	-4e-04 (0.0058)	-0.0246 (0.0066)	-0.0169 (0.0058)	0.0254 (0.00

this threshold, the change in percentage of observations dropped is also small.

A more detailed picture of this is displayed in **?@tbl-cd-gnr-inter-trim**. The table compares the estimates of the output elasticity of intermediates and other statistics of the first stage of GNR(2020) assuming a Cobb-Douglas Production Function for different definitions of intermediates and different thresholds for trimming observations.

19 Density

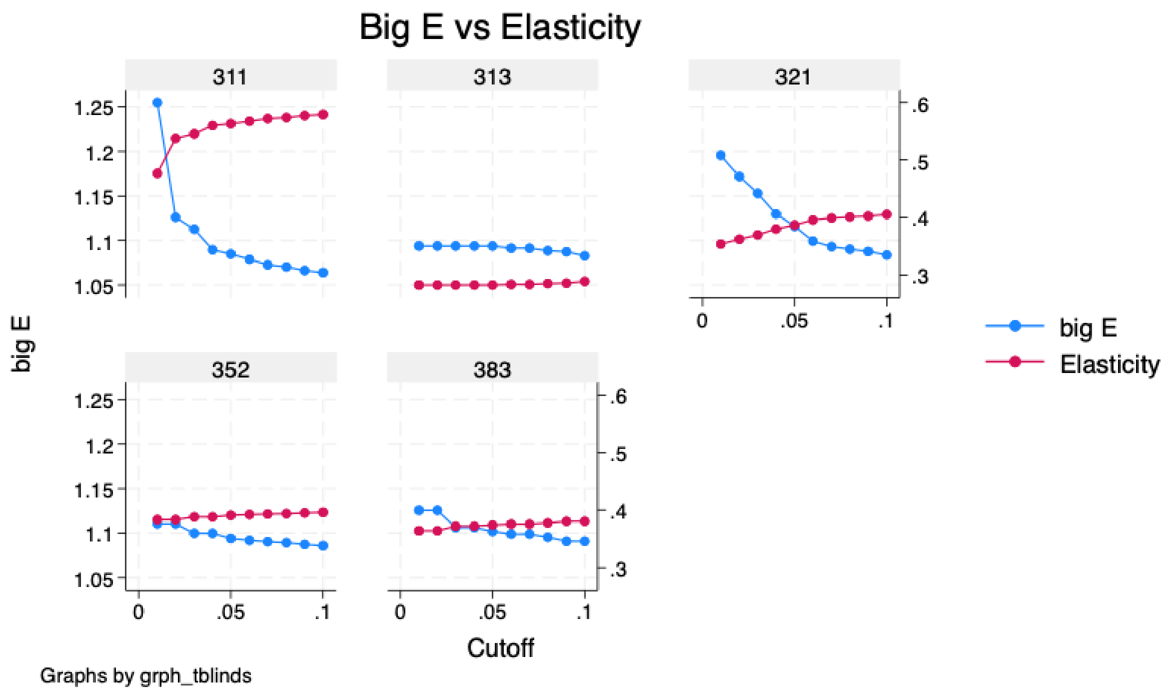


Figure 11: Estimates of $\mathcal{E}$ and the output elasticity of intermediates (raw materials) trimming observations with a share of intermediates below different thresholds going from 0.01 to 0.1 . The elasticities of intermediates were estimated assuming a Cobb-Douglas production function using the first stage of GNR(2020).

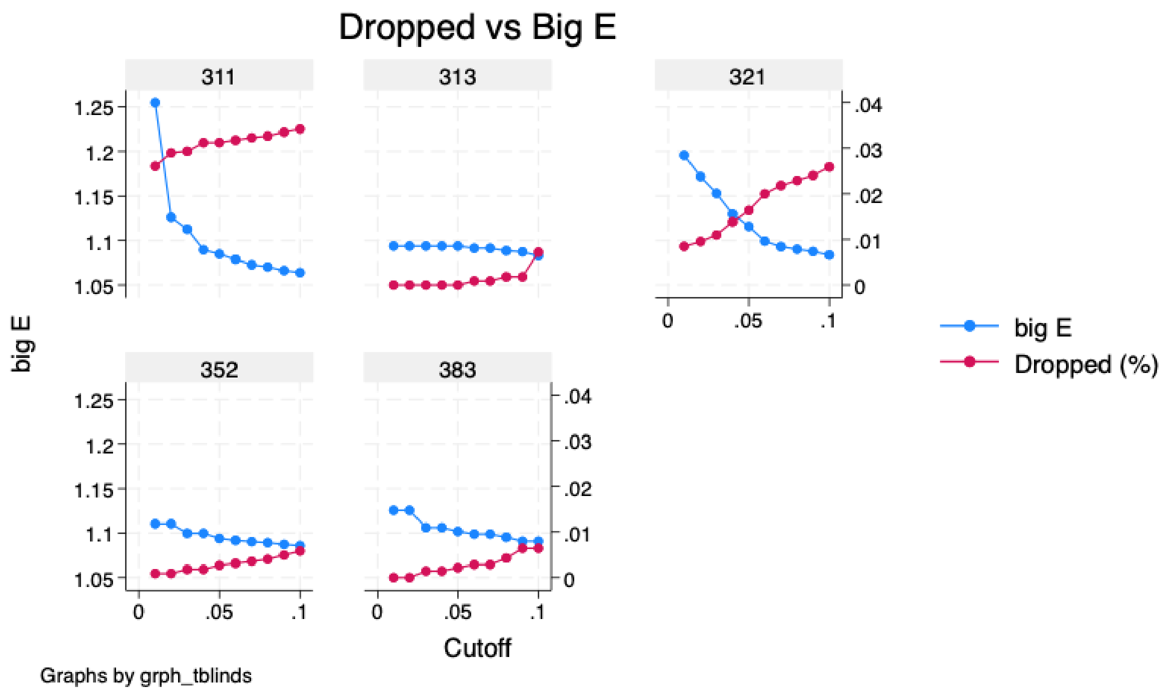
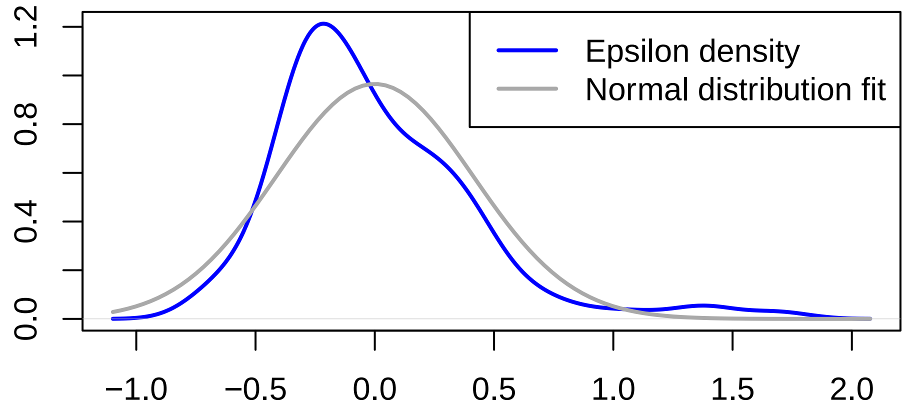


Figure 12: Estimates of $\mathcal{E}$ and percentage of observations dropped by trimming observations with a share of intermediates (raw materials) below different thresholds going from 0.01 to 0.1 . The elasticities of intermediates were estimated assuming a Cobb-Douglas production function using the first stage of GNR(2020).

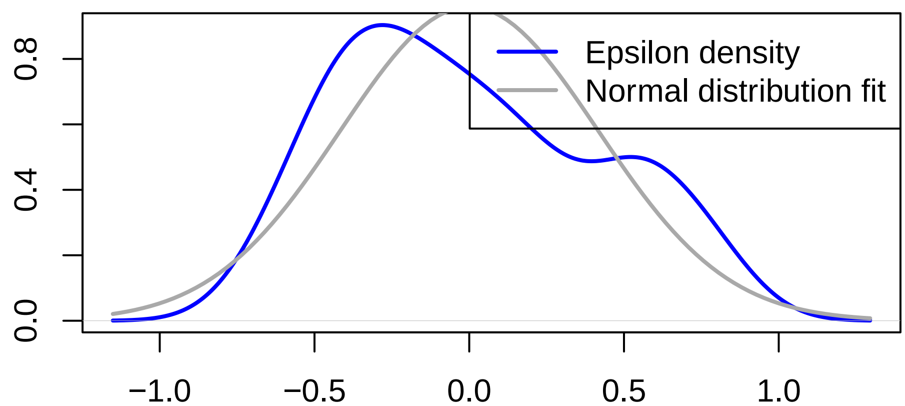
Table 43: Estimates of Cobb-Douglas Production Function parameters using GNR(2020) and statistics of the first stage for different definitions of intermediates. *m s e* stands for raw materials, services, and energy, *mats* for raw materials, and *ded* for deductible intermediates (raw materials, electricity, fuels, and repair and maintenance services). Observations were trimmed below a share of intermediates of 0.05.

Ind.	Inter.	m	l	k	bigE	si_mean	err_mean	err_sd
311	m_s_e	0.6488	0.1827	0.1832	1.0652	0.7219	0	0.3091
	mats	0.5630	0.2084	0.2443	1.0848	0.6416	0	0.3470
	ded	0.5567	0.2194	0.2337	1.1232	0.6631	0	0.3948
321	m_s_e	0.5183	0.2658	0.2088	1.0557	0.5689	0	0.2996
	mats	0.3869	0.3412	0.2604	1.1158	0.4633	0	0.4136
	ded	0.4273	0.3237	0.2470	1.0934	0.4949	0	0.3720
352	m_s_e	0.5749	0.2133	0.2415	1.0437	0.6201	0	0.2718
	mats	0.3909	0.2858	0.3527	1.0940	0.4574	0	0.3922
	ded	0.4114	0.2797	0.3360	1.0877	0.4756	0	0.3746
313	m_s_e	0.4324	0.1658	0.3086	1.0721	0.4893	0	0.3479
	mats	0.2826	0.2943	0.3395	1.0938	0.3347	0	0.4104
	ded	0.3216	0.2912	0.3203	1.0767	0.3699	0	0.3726
383	m_s_e	0.5261	0.3666	0.1892	1.0445	0.5694	0	0.2786
	mats	0.3739	0.4889	0.2523	1.1018	0.4427	0	0.4069
	ded	0.4015	0.4532	0.2427	1.0891	0.4655	0	0.3798

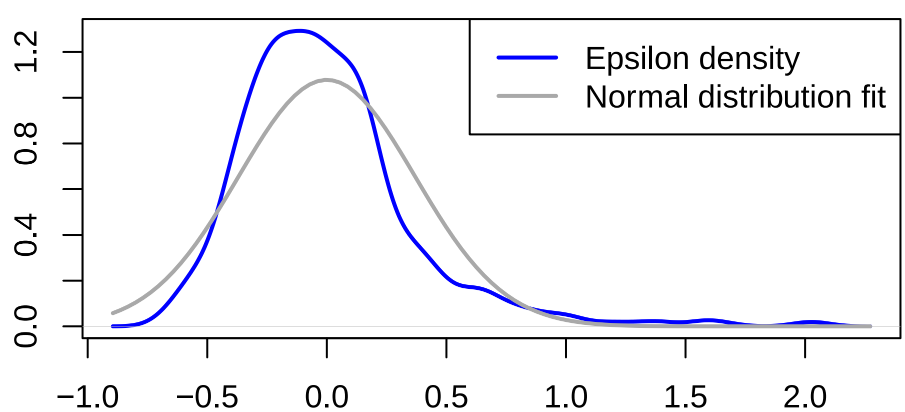
Industry 322



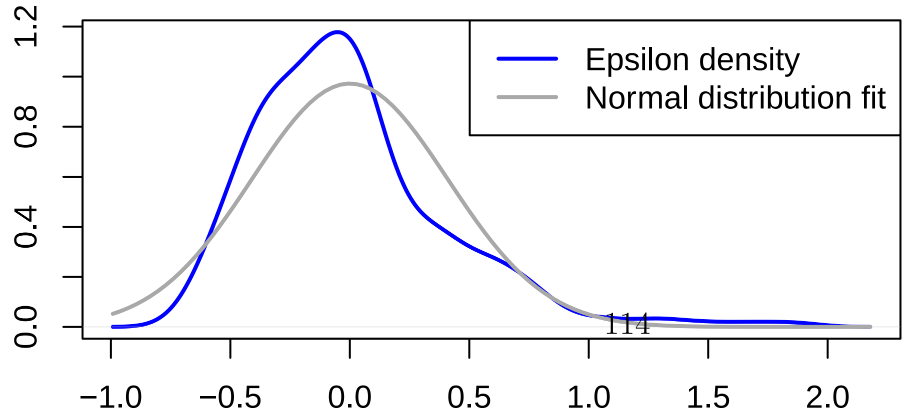
Industry 331



Industry 321



Industry 342



Industry 313