Limited Tax Capacity and the Optimal Taxation of Firms

Marcelo Arbex* **Enlinson Mattos†** **Rebeca Regatieri‡**

July 29, 2021

Abstract

Limited tax capacity, inadequate tax administration, and poorly design tax systems are common features of most developing countries. We study a tax reform (*SIMPLES*) in Brazil, a developing economy with a large informal sector. Our empirical analysis reveals that the *State-SIMPLES* tax reform had a substantial effect on firms at different stages of production regarding their intensive margins (i.e., reported revenue and costs) to the *SIMPLES* massive tax reductions. For downstream eligible firms, we observed reductions in reported production costs vis-a-vis non-eligible firms. We also characterize the optimal tax instruments with limited tax capacity for a multistage, value-added, and turnover tax systems and show that the elasticities of misreported sales and purchase to policy instruments are behavioral statistics that substitute the traditional Diamond and Mirrlees (1971)'s efficiency effect of taxation. We evaluate numerically the optimal tax systems. Assuming similar misreporting costs to all firms, the optimal taxes on upstream is larger than zero and tax rebate smaller than one.

Keywords: Limited tax capacity, tax reform, optimal taxes, informal firms

JEL Classification: H21, H26, H32.

*Department of Economics, University of Windsor. arbex@uwindsor.ca.; †*Corresponding author.* São Paulo School of Economics, Fundação Getúlio Vargas. enlinson.mattos@fgv.br; ‡Ph.D. candidate, SÃ£o Paulo School of Economics Fundação Getúlio Vargas and Brazilian Treasury. regatieri.rebeca@gmail.com. This paper was developed during EM sabbatical at the University of Michigan (2019). We have benefited comments and suggestions from Joel Slemrod, James Hines Jr., Christian Trudeau, Tobias Hauck, Xuan Wang and all participants at the PFFLS 2019- University of Michigan. Any errors are our own.

1 Introduction

Limited tax capacity and tax evasion are common features of most developing countries. In these economies, the government limitations to collect taxes can be due to inadequate tax administration as well as poorly design tax systems. In fact, many poor and developing countries rely heavily on indirect taxation and have promoted a minimum tax scheme to raise tax revenues and fight evasion. Moreover, the understanding of how firms, in particular, small and potentially informal firms, are affected and respond on their intensive margins (i.e., reported revenue and costs) have important implications for the optimal design of tax systems and policy instruments in limited tax capacity economies. Our approach to this issue is twofold. We develop a positive analysis of a tax reform (*SIMPLES*) in Brazil, a developing economy with a large informal sector. The *SIMPLES* represented massive tax reduction for micro-enterprises and small firms and it was followed by States participation in the program that changed sales tax rates for small firms. Then, we propose a normative analysis of taxation in a limited tax capacity economy and alternative tax systems.

Our empirical analysis investigates tax reforms in the Brazilian States that followed the federal tax simplification program called *SIMPLES* (a multi-stage turn-over tax to replace a value-added and retail sales tax) and we consider a survey of upstream and downstream informal firms to analyze the effects of the tax reform on their reported revenues and production cost. Under the new system, often called *Federal-SIMPLES*, firms were allowed to remit a single (lower) tax on their total annual revenue. In addition to changes in the federal tax system, individual states were allowed to opt in the *SIMPLES* by imposing an additional tax rate to collect the correspondent value-added taxes (VAT) on eligible firms. By the end of 2003, not only individual states had started to participate in the SIMPLES tax program - either *Federal-SIMPLES* or *State-SIMPLES* - but also informal firms were surveyed in October of that year. These two facts create an interesting opportunity to investigate how the SIMPLES tax reform implemented in a few states in 2003 might have affected the reported expenses and revenues of upstream and downstream firms located in those states - a tax reform in a limited tax capacity environment. Our data source for small (potentially informal) firms is the *ECINF* (*Pesquisa de Economia Informal Urbana*), a national survey conducted by the

Brazilian Statistics Bureau (IBGE) in October of 2003.

Our identification strategy explores the state-level variation of the tax reform, separating firms between eligible and not eligible to join the *SIMPLES*, similar to de Paula and Scheinkman (2010) with a relevant difference. As we are interested in the intensive response margin, we compare firms in eligible sectors located in states that had a tax reform in the exact year of the survey (2003) with those without eligible firms in non-tax reformed states. We aim to capture short-run reported sales and costs responses to tax reform [1] For small upstream and downstream firms, the SIMPLES tax reform of 2003 represented a change from VAT system, in which both production stages were taxed at the same tax rate and downstream firms could claim the full amount of taxes remitted by upstream firms as part of their production costs, to a turnover tax (TOT) system - upstream and downstream firms can be taxed at different tax rates depending on their sizes and sector eligibility. We observe that informal firms identified by *ECINF* are mainly concentrated in the downstream production stage, their owners have less schooling and seem to manage less labor-intensive business in terms of the share of expenses. More relevant to our theoretical approach, these are small, mostly price-takers firms, with very few of them reporting to the tax authority. These features reinforce the need to control for observable characteristics at the firms level. The results of our empirical analysis reveals that the *State-SIMPLES* tax reform had a substantial effect on firms in different stages of production. Within the same sector, eligible upstream firms in states that implemented the *SIMPLES* tax reform in 2003 seem to have responded similarly (although with smaller size) than their downstream counterparts, with the last showing noticeable reductions in reported production costs *vis-à-vis* non-eligible firms. Young downstream firms show reduction in the reported sales, again similar to results we obtain for upstream firms.

In order to characterize an optimal tax system in a limited tax capacity economy, we develop a simple framework that highlights the main trade-offs faced by a government when optimally designing its policy instruments in such environment. There are four main players in our model economy: consumers, downstream firms, upstream firms and the government. Production occurs in two stages, upstream and downstream, and firms operate constant returns to scale technologies

[1] See Monteiro and Assuncao (2012) and Piza (2018) for the evaluation of the impact of the federal program on the decision of the firm to become formal.

(Diamond and Mirrlees, 1971a,b). An upstream firm produces an intermediate good using only labor as input and sells it to downstream firms (B2B transactions), which are subject to taxation. Combining the intermediate good and labor, downstream firms produce a final good that is sold to consumers (B2C transactions). As their production costs, along with labor and intermediate goods costs, downstream firms can also claim a fraction of the taxes remitted by the upstream firm. Hence, the tax authority has three policy instruments, namely, *ad valorem* taxes on the upstream (intermediate) good and the downstream (final) good, and a refund tax rate (the proportion of the tax paid by upstream firms that downstream firms are allowed to deduct from their production costs). However, tax administration is constrained by its limited tax capacity, which creates opportunities for firms to misreport their purchases and revenues. In line with our empirical findings, downstream and upstream firms decide how much of their sales and purchases they report to the tax authority with heterogeneous misreporting cost both at the revenue and purchase margins. Consumers work in both production stages and they consume the final good but not the intermediate good. We focus on homogeneous firms within an industry and homogeneous consumers.

Misreporting taxes is costly and directly associated with the government limited tax capacity. We focus on the main differences between business-to-business (B2B) and business-to-consumers (B2C) transactions to characterize the optimal linear taxes on firms. The government's limited tax capacity creates opportunities for upstream and downstream firms to misreport their sales and purchases for tax purposes. Our model allow for heterogeneous misreporting costs across upstream and downstream firms and evasion margins. The quantitative analysis also addresses the fact that upstream firms are as responsive to tax changes than downstream firms, i.e., moving from a VAT to a TOT system leads downstream firms to report less production costs with similar response by the upstream firms.

In our model, misreported gaps (the difference between true revenues and expenses and the amount firms choose to report to the tax authority) are interpreted as tax evasion margins due to the government's limited tax capacity. Using a parsimonious model, we discuss the normative conditions under which the tax authority may optimally implement a general tax system with the three above-mentioned tax instruments. We show that the optimal policies are determined by two - a behavioral and a mechanical - effects in both B2B and B2C transactions, a reinterpretation of the traditional optimal tax literature result (Diamond and Mirrlees, 1971a,b). If the misreporting elasticity is small relative to the mechanical effect in the production chain, then it is optimal for the government to increase tax rates on the upstream and downstream sales. Conversely, if the evasion elasticity is large relative to the real output elasticity, the revenue efficiency concern will be relatively stronger, making it optimal to reduce taxation of both stages of production. This, in turn, reduces evasion (measured by the sales and purchases misreported gaps) and tax distortions in the economy. We also show the conditions for an interior (neither zero nor one) optimal refund tax rate. That is, it optimal for the government to allow downstream firms to deduct a fraction of the taxes remitted by upstream firms as part of their production costs. Finally, we characterize the optimal tax instruments with limited tax capacity for the two tax systems considered in the *SIMPLES* tax reform (VAT, TOT). We show that the optimal taxation of upstream firms must take into account the behavior and mechanical responses of their own production as well as the effects on other firms along the production chain (i.e., downstream firms), even in a TOT system. Thus, we show that optimal taxes on intermediate and final goods can be potentially different, in part due their asymmetric effects on firms' misreporting (tax evasion) margins.

Related Literature. The Diamond and Mirrlees (1971a,b) model has guided the intuition about the optimal taxation of goods and services: optimal taxes are zero on all intermediate goods. While the Diamond-Mirrlees results have been amply investigated, exceptions to their benchmark have been noted, in particular, in the presence of unlimited tax capacity and costly monitoring; a branch of the literature that our paper belongs. In modern tax systems, firms remit are required to file tax information reports and remit taxes to the tax authorities. As Kopczuk and Slemrod (2006) emphasize, governments could take advantage of arm's length transactions (upstream, downstream) as well as firms' economies of scale when setting up the tax code and tax enforcement policies, features that must be included in the study of optimal taxation of intermediate and final goods. The authors explore the key role of a corporation as tax remitter in a theoretical framework that proposes a comparative evaluation of a uniform retail sales tax (RST) versus a VAT, both consumption taxes, as well as wage and firms' profit taxation. Administrative and enforcement issues are critical and may distort the theoretical foundation introduced in Diamond and Mirrlees (1971a,b), because firms give rise to easy-to-monitor transactions as well as to income-shifting tax responses across different bases.

The main contribution of our model is to highlight the role of each misreporting margin of firms along the production chain, summarized by behavioral and mechanical effects, on the optimal choice of tax instruments. Tax evasion brings an additional challenge to the optimal design of taxation along the production chain. If upstream and downstream can evade taxes on their sales and/or in the purchase of inputs, the government must account for the direct and indirect effects of its choice of optimal tax instruments, as well as tax policy changes, on the real and reported tax liabilities. Informality creates a fiscal as well as an efficiency problem. While some studies argue that the optimal tax system must aim to reduce tax evasion (e.g., Emran and Stiglitz (2005); Gordon and Li (2009)), others suggest that it may be optimal to leave hard-to-tax firms untaxed, e.g., Keen and Mintz (2004); Keen (2008); Dharmapala et al. (2011). Alternatively, Best et al. (2015) explore the government's trade-off between production efficiency versus (tax) revenue efficiency. In particular, the authors suggest that in a limited tax capacity economy and depending on the firm's tax liability, it might be optimal to either tax the firm's profits or to implement a TOT system.

The empirical implications of the government's limited tax capacity have also been investigated in the literature. Pomeranz (2015) and Carrillo et al. (2017), for instance, investigate how alternative auditing policies affect firms' tax base reporting behavior.[2] And, there are several papers that study the Brazilian economy using the same data as ours but with particular emphasis on the extensive margin decision, i.e., the probability of a firm to become formal. de Paula and Scheinkman (2010) test implications of a simple equilibrium model of informality using data from the ECINF to show that an increase in enforcement in a production stage increases downstream and upstream formalization. These chain effects are reduced with withholding taxes. Fajnzylber

[2]See Besley and Coate (1991), Cuff et al. (2017) for a normative model on tax-auditing with limited evasion. Gadenne et al. (2018) study the effect of audit policies on the production chain in India. Besley and Persson (2013) review the literature and evidence on the economic and political forces that shape the way that fiscal capacity is created and sustained.

et al. (2011) and Monteiro and Assuncao (2012) examine whether the Federal-Simples act to diminish informality of small firms in Brazil and they find mixed results. While Fajnzylber et al. (2011) suggests a reduction in the informality decision, Monteiro and Assuncao (2012) findings suggest that Federal-Simples only affects firms decision to become formal in the retail sector. Finally, Piza (2018) demonstrates that this policy is, in fact, not effective to induce the firm's formalization decision (extensive margin). Our paper focus on the firms intensive margin response to changes in tax rates at the state level, a consumption tax. We interpret that these tax changes are nothing but a change from a VAT to a TOT system, and we find that only downstream firms are sensitive to such tax changes without any effect on the decision to becoming formal.

The paper is divided as follows. Section 2 presents tax reform institutional details and a description our data set. In this section, we also discuss our empirical strategy and main empirical results. Section 3 introduces a stylized model to study firm's production decisions in a limited tax capacity environment. We characterize the optimal taxation of firms when misreporting taxes is costly at different margins. In Section 4, we evaluate numerically the optimal tax policies and tax systems and conduct several quantitative exercises. Section 5 concludes.

2 A Tax Reform in a Limited Tax Capacity Economy

We are interested in changes in the upstream and downstream firms' intensive margins (reported sales and reported costs) with respect to the State level tax reform, two relevant features of potential informal firms associated with a tax reform. First, smaller firms are harder to tax by the tax authority (limited tax capacity) and, second, the size of the firm, measured by the amount of sales (revenue), is an additional criterion for firms in eligible economic sector to opt in the *State-SIMPLES* tax system. Therefore, firms' reported revenues and costs at the moment of tax reforms seems to capture the relevant taxable income responses to tax changes.

2.1 Understanding Firms' Reported Responses to a Tax Reform

In this section we investigate a tax reform carried out in Brazil, a developing economy with a large informal sector. The *SIMPLES* (*Sistema Integrado de Pagamento de Impostos e Contribuições das Microempresas e Empresas de Pequeno Porte*) tax reform meant a significant tax reduction for micro-enterprises and small firms. It allowed Brazilian states to replace a value-added tax system at the state level by a multistage turnover tax system. Our first goal is to present a brief description of this tax reform and document how small and potentially informal firms are organized along the production chain and to provide empirical evidence (reduced form estimates) of the impact of this tax reform on firms' intensive margin responses, such as sales and non-labor costs. We classify a business as either an upstream firm or a downstream firm depending on the buyer of goods and services produced and sold by them. Our analysis highlights that these firms operate in a limited tax capacity economy, an environment characterized by the government's inability to fully tax business and consumers. On one hand, this feature of the Brazilian economy, as well as many other developing economies, can be attributed to an inefficient tax administration and auditing procedures. On the other hand, the government limited tax capacity can be due to firms' efforts to hide and misreport their output and input purchases. Of course, all these factors can be at play simultaneously.

The second goal of this section is to provide evidence that firms' misreporting behavior vary along the production chain. That is, upstream and downstream firms have different margins to misreport taxes to the government and they might react differently regarding the amount of purchases and sales, depending on their respective misreporting costs. Our empirical analysis reveals that upstream and downstream firms may respond differently regarding their revenues and production costs when they were exposed to the *SIMPLES* tax reform. The remainder of this section is organized as follows. We describe the main features of the tax reform and our dataset. Then, we present our identification strategy and discuss in details our main empirical results.

2.2 The *SIMPLES* Tax Reform

The Brazilian federal government launched the *SIMPLES* in December of 1996. Its primary goal was to reduce and to simplify the taxation of micro-enterprises and small firms. The *SIMPLES* works as a minimum alternative tax scheme, combining six different types of federal taxes and social contributions: (i) corporate income tax (*IRPJ*), (ii) contributions to employee's savings program (PIS/PASEP), (iii) social security contributions (*COFINS*), (iv) tax on industrialized products (*IPI*), (iv) contributions to the net profit (*CSLL*) and (vi) employer's social security contributions. Under the new tax system, often called Federal-SIMPLES, micro-enterprises (small firms) in eligible sectors are required to pay a single tax, which varies from 3 to 5% (5.4 - 8.6%) of their total annual revenue.[3] Eligible economic sectors include retail and trade, construction, transportation and manufacturing (see Table A.1, Appendix A.1 for more details). Although participation in the *SIMPLES* was not mandatory, there is evidence that firms would benefit by joining it. According to Monteiro and Assuncao (2012), eligible firms that opted into the program experienced a reduction in the overall tax burden of up to 8% of their annual revenue.

In additional to the changes at the Federal level, legislation was put forward to allow individual Brazilian States to augment the program at the State level. States that decided to join the Federal tax reform were allowed to impose an additional value-added tax rate on eligible firms. In practice, it meant that a micro-enterprise and a small firm in a participating State can be required to remit, in addition to the federal tax rate, a marginal tax rate of up to 2.5 and 1.1 percentage points, respectively. In fact, the *SIMPLES* tax reform consisted in an opportunity for each state to decide about an alternative tax scheme for VAT remittance under the States' responsibility. Participating States could either join the Federal-*SIMPLES* program or create their own version of the program. For small firms in eligible sectors, the alternative tax scheme reform could be described as a change from a value-added to a turnover tax system. Interestingly, States kept most of the Federal-*SIMPLES* eligibility criteria (firms' sector and size), with only few exceptions.[4] By the end of 2001, among the twenty-six States plus the Distrito Federal, fifteen States had joined the program. In 2003, five States - Pará, Rio Grande do Norte, Paraná, Ceará, and Paraíba - also promoted tax changes following the Federal-*SIMPLES* tax reform and, by then (2003) seven Brazilian States had not participated in this tax reform - Amazonas (AM), Mato Grosso (MT), Minas Gerais (MG), Piaui (PI), Rondonia (RO), Rio Grande do Sul (RS) and Tocantins (TO) - see Table A.2, Appendix A.1, for specific dates.

[3]Firms must still pay and remit municipal and state taxes. The tax reform we investigate focuses on states adhesion to the federal *SIMPLES* and, hence, consumption taxes are included in the firms' tax remittances. Micro-enterprises are defined as firms with annual revenue less than R$120,000; USD40.000) and small firms are those with annual revenue less than R$1.200.000; USD400.000).

[4]We were very rigorous comparing eligibility criteria of the States that implemented SIMPLES with those Federal rules and there was only small differences. This comparison is available upon request.

2.3 ECINF survey and the *SIMPLES* Tax Reform

Our empirical analysis relies on a national survey conducted by the Brazilian Statistics Bureau (*IBGE*) in October of 2003 to study small and potentially informal firms along the production chain in Brazil. The *ECINF* (*Pesquisa de Economia Informal Urbana*) surveyed a total of 48,701 urban entrepreneurs in all Brazilian states, mainly focused on business with less than five employees. Participants answered a series of questions regarding their business ranging from individual characteristics of the business owner, obstacles to formalization, main sources of revenue and costs.

Given a coincidence of dates - the *ECINF* was conducted in 2003 and several States changed their tax system as part of the *SIMPLES* tax reform - we believe the *ECINF* is an ideal dataset to investigate the firms' intensive margins responses in a limited tax capacity economy. More specifically, the *ECINF* was conducted in 2003 when the Brazilian States of Pará, Rio Grande do Norte, Paraná, Ceará, and Paraíba joined the *Federal-SIMPLES* tax reform and changed their tax code at the State level, which we call *State-SIMPLES* tax reform. This fact represents an interesting opportunity to investigate how small firms surveyed by the *ECINF* were affected by the tax reform mostly because we can investigate how tax system changes affected firms' revenues and production cost reported in the same year. This strategy eliminates any biases alternative policies might have had on eligible firms prior to the tax reform and the *ECINF* survey.[5]

However, there are some challenges that need to be addressed. Given that each individual State adopted a different tax code with extensive case-specific rules as well as contingencies, we had to obtain precise data on individual firm's eligibility and tax rates for each State firms are located. Next, we compare them with the federal eligibility rule to participate in the *Federal-SIMPLES*. Not surprisingly, the vast majority of States used the same sectors eligibility to implement their versions of alternative tax schemes. Moreover, as we aim to compare States with local reform versus those that do not implement any local tax reform, it seems natural to consider a unified eligibility criteria for both samples. Therefore, we consider the *Federal-SIMPLES* eligibility criteria in our

[5] A previous version of this dataset (*ECINF* 1997) was previously used by Monteiro and Assuncao (2012); Piza (2018); Fajnzylber et al. (2011); de Paula and Scheinkman (2010); Ulyssea (2018) to document firms' extensive margin responses (i.e., their formalization decision). Our empirical strategy builds on these papers, but we focus on intensive margin responses to the State-*SIMPLES* tax reform, focusing on the States that reformed their tax system in the year the *ECINF* survey was conducted, i.e., 2003.

main empirical strategy.[6]

We focus on eligible and non-eligible firms in the States that implemented the *State-SIMPLES* tax reform in 2003 versus eligible and non-eligible firms in the States without a reform. This reduces our sample to 23,087 firms. The main variables and descriptive statistics for all firms in our sample are summarized in Table A.3 (Appendix A.1). We compare all firms (54,905) surveyed in the *ECINF* 2003 versus firms (23,087) in States that either have adopted the *State-SIMPLES* in 2003 or have not implemented a tax reform at the State level. We observe that our sample reports similar magnitudes for almost all variables, although with smaller reported sales, costs and non-labor costs.

Of particular interest for our study is the businesses customers - i.e., whether a firm sells goods and services to other firms or sell them directly to final consumers or the government. We use this information (i.e., a firm main customer) to classify a firm either as an upstream firm or as a downstream firm. If a firm sells goods and services to individuals consumers or to the government, we define it as a downstream firm. Otherwise, it is classified as an upstream firm. In our sample, there are 20,312 downstream firms and 2,775 upstream firms. Given our classification of firms as either an upstream or a downstream firm, we can investigate whether the *State-SIMPLES* tax reform affected these business differently, i.e., we can look for heterogeneous responses along the production chain.

The descriptive statistics of upstream and downstream firms in our sample are presented in Table I. First, notice that 88% (94%) of firms classified as upstream (downstream) are in sectors eligible to the *State-SIMPLES* reform.[7] Moreover, 40% downstream firms are in States that participated in the tax reform while only 33% of the upstream firms are located in states that implemented tax reform in 2003. While downstream firms stay longer *in business*, on average

[6]For example, an eligible firm in the state of Pará (Tax reform - *SIMPLES* - January of 2003), with annual revenue of R$90,000 and total expenses of R$50,000 would have to pay R$7,200 in VAT if it decided not to join the *SIMPLES*. On the other hand, if this same firm opted in, it would have to pay a quarter of this amount (R$1,800); a tax burden of 4.5%, significantly lower than 18% under the old scheme.

[7]The number of eligible firms are determined by following Law 9,317/1996 and Complementary Law 127, 1997 available at https://legislacao.presidencia.gov.br/atos/?tipo=LEI&numero=9317&ano=1996&ato=687Izaq1UMJpWTa6e. Our definition is similar to Monteiro and Assuncao (2012) definition of eligible sectors for the *Federal-SIMPLES*, with one slight modification, we consider the sector of food and beverages eligible for the federal tax reform. The results are qualitatively similar using their definition of eligible sectors and are available upon request.

112 months compared to 106 months of upstream firms, the later report a slight higher hours per week - 44.5 versus 43.3 hours/week by downstream firms. Regarding the firms' reported margins, i.e., reported sales, reported total costs and reported non-labor expenses, our data reveals that upstream firms report sales and total costs two times larger than their downstream counterparts.

Regarding the upstream and downstream firms extensive margin decisions, almost all firms reported that their decision to do not register their business is either because they do not want to pay a fee/taxes to the government or they lack the resources to hire an accountant. Among those firms that registered their business, 97% enroll it from the very beginning - bureaucracy and paperwork are pointed out by 70% as the main barrier to tax registry. A larger fraction of upstream entrepreneurs reported that they have an accountant to help them with financial and taxes services and they are registered in the national tax system (*Cadastro Nacional de Pessoas Jurídicas; CNPJ*). Note that, there is no statistical difference related to the decision of having a tax registration number in either States that participated in the reform, which suggest a weak extensive margin response to firm's formalization (see A.1). However, we must control for the hiring of an accountant in our empirical analysis. Also, we do not find evidence of significant difference between upstream and downstream firms time *in business*.

Upstream and downstream entrepreneurs (business owners) are, on average, 41-year old and they manage their business for approximately two years (twenty-seven months). Owners of upstream firms tend to be more educated and concentrate their activities in economic sectors such as agriculture, construction, housing and transportation. Downstream firms, on the other hand, are mainly concentrated in agriculture, manufacturing and construction (Table I).

2.4 Empirical Analysis

2.4.1 Identification Strategy

Our empirical strategy aims to document the relationship between the firms' reported sales and costs and the tax changes. We consider firms located in the five Brazilian States (Pará, Rio Grande do Norte, Paraná, Ceará, and Paraíba) that joined the *Federal-SIMPLES* tax reform and changed their tax code at the State level. Therefore, we explore the state-level variation of the tax reform, while controlling for the firm's eligible economic sector. Ideally, we would like to compare

Table I: Descriptive Statistics - Downstream and Upstream Firms

	Downstream Firms (y_d)			Upstream Firms (y_u)			
Variable	Obs	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Diff
Firms Characteristics							
Eligible Firms	19,304	0.94	0.24	2,672	0.88	0.33	0.06
State-SIMPLES reform	20,312	0.40	0.49	2,775	0.33	0.47	0.07
In business$^{(1)}$	20,288	112.79	113.06	2,773	99.99	103.31	12.80
Hours/week	20,285	43.42	22.26	2,774	44.56	18.34	-1.14
# workers	20,312	1.50	1.07	2,775	1.73	1.42	-0.23
Share middle school	20,312	0.13	0.32	2,775	0.15	0.33	-0.02
Share high school	20,312	0.20	0.36	2,775	0.27	0.40	-0.07
Share college degree	20,312	0.06	0.21	2,775	0.11	0.29	-0.05
Tax Registry$^{(2)}$	20,312	0.15	0.36	2,775	0.30	0.46	-0.15
Market Power$^{(3)}$	19,293	0.62	0.48	2,645	0.49	0.50	0.14
Outside Household$^{(4)}$	20,312	0.65	0.48	2,775	0.71	0.45	-0.05
Accountant$^{(5)}$	20,312	0.15	0.35	2,775	0.25	0.43	-0.10
Reported Margins							
Sales ($\overline{y}$)	20,160	1920.62	5517.09	2,750	4110.13	9187.28	-2189.51
Total Costs	18,749	1434.36	5739.08	2,548	2888.54	7430.96	-1454.18
Non-labor Costs ($\widehat{y}$)	20,312	1037.51	5118.63	2,775	1872.40	6525.49	-834.89
Individual (Entrepreneur) Characteristics							
Age	20,312	41.39	12.52	2,775	41.02	11.93	0.37
Tenure$^{(6)}$	19,752	27.23	13.92	2,725	26.76	12.88	0.47
Education							
Less than Middle School	20,312	0.49	0.50	2,775	0.32	0.47	0.17
Middle School	20,312	0.13	0.34	2,775	0.14	0.35	-0.01
High School (incomplete)	20,312	0.09	0.28	2,775	0.08	0.27	0.01
High School	20,312	0.19	0.39	2,775	0.27	0.45	-0.08
College (incomplete)	20,312	0.03	0.17	2,775	0.06	0.25	-0.04
College	20,312	0.07	0.25	2,775	0.12	0.33	-0.06
Economic Sectors							
Agriculture	20,312	0.12	0.33	2,775	0.15	0.36	-0.03
Manufacturing	20,312	0.20	0.40	2,775	0.08	0.27	0.12
Construction	20,312	0.37	0.48	2,775	0.34	0.47	0.03
Retail	20,312	0.06	0.24	2,775	0.03	0.18	0.03
Housing	20,312	0.07	0.26	2,775	0.16	0.37	-0.09
Transportation	20,312	0.04	0.19	2,775	0.13	0.33	-0.09
Real State	20,312	0.04	0.19	2,775	0.02	0.13	0.02
Education and Health	20,312	0.08	0.27	2,775	0.06	0.23	0.02

Note: (1) Time in business (months); (2) Dummy if firms has tax registry (CNPJ); (3) Dummy if the firm can negotiate its sales price; (4) Dummy if the firm is located outside household; (5) Dummy if the firm hires an accountant; (6) Time that entrepreneur works in the same field (months).

the reported sales and costs of eligible firms after the *State-SIMPLES* reform was introduced in their State with their otherwise reported sales and costs in the absence of the tax reform. However, we only observe firms already operating subject to the *State-SIMPLES* tax system; otherwise they are not in our sample.

Hence, an important aspect in our identification strategy relies on the choice of the control group. We proceed as follows. For each Brazilian State that had implemented its version of the *State-SIMPLES* tax reform, we compare their eligible and non-eligible firms with eligible and non-eligible firms located in States that had not (as of 2003) implemented any alternative tax scheme change following the *Federal-SIMPLES*. In addition, we also explore firms' heterogeneous responses to the tax reform due to their age (i.e., time *in business*) as well as when they were created - similar to Monteiro and Assuncao (2012) and Piza (2018), that is whether a firm was created before or after the date a specific State implemented its version of the *State-SIMPLES* tax reform.7 For specific dates, see Table A.2, Appendix A.1.

We implement a difference-in-differences (DID) strategy, where we separate firms into two categories: eligible (treatment group) and non-eligible (control group) to the new tax regime and out of twelve states, *State-SIMPLES* assumes denotes a tax-reformed state. More formally, we estimate the following regression:

$$\begin{aligned} RM_i &= \alpha_r \text{ Eligible Firms}_i + \alpha_t \textit{ State-SIMPLES Reform}_i \\ &+ \alpha_{r,t} \text{ Eligible} * \text{Reform}_i + \beta' X_i + \epsilon_i \end{aligned} \quad (1)$$

where RM_i represents a reported margin - either reported sales ($\overline{y_i}$) or reported total and non-labor costs ($\widetilde{y_i}$) - of a given firm i, the variable Eligible Firms$_i$ is equal to one (zero) if the firm (does not) belongs to an economic sector eligible to the tax reform, the variable *State-SIMPLES Reform*$_i$ is a dummy variable that equals to one if the a firm i is located in a State that implemented the *State-SIMPLES* reform in the year 2003, and zero otherwise. The coefficient we are interested is $\alpha_{r,t}$ that captures the interaction between being in a eligible sector and in a State that adopted a tax reform in 2003. In other words, this coefficient represents the difference in the reported margin for firms in eligible sectors versus non-eligible and firms located in States with a tax reform versus those located in States without a tax reform. With the variable X_i, we also control for the *ECINF* economic sector fixed effects, firm's individual characteristics such as location (whether it operates outside the owner's household), workers education level, time *in business*, production chain position (whether an upstream or a downstream firm) and average hours per week, whether firm has an accountant, and if it has market power to decide sales prices. We also include the owner's age and education as controls.

We are mainly interested in investigating whether same eligible sector firms along the production chain respond differently to the *State-SIMPLES* tax reform. We rewrite equation (1) making it explicit our dummy variable Upstream(U)_i that assumes a value equals to one if the firm is classified as an upstream firm, and zero otherwise (downstream), as follows

$$\begin{aligned} RM_i &= \alpha_r \text{ Eligible Firms}_i + \alpha_t \textit{ State-SIMPLES Reform}_i + \alpha_u \text{ Upstream(U)}_i \\ &+ \alpha_{r,t} \text{ Eligible} * \text{Reform}_i + \alpha_{r,t,u}\ U\text{*Eligible} * \text{Reform}_i + \beta'\ X_i + \epsilon_i. \end{aligned} \tag{2}$$

For this particular equation we are interested in the coefficient $\alpha_{r,t,u}$ a threefold difference that accommodates the heterogeneous effects of a firm being in the eligible sector, located in a State that changed its tax system in 2003 and being an upstream firm compared to downstream firms in non-eligible sectors and located in states without tax reform.

Regarding heterogeneous responses due to differences in the time firms are *in business*, we estimate two models. First, in addition to controlling for firm's age, we consider a simple interaction of the variable Eligible Firms_i with the variable *In Business* $(B)_i$. Hence, we rewrite equation (1) to explicit the variable *In Business* $(B)_i$ and its interaction with $\text{Eligible} * \text{Reform}_i$, i.e.,

$$\begin{aligned} RM_i &= \alpha_r \text{ Eligible Firms}_i + \alpha_t \textit{ State-SIMPLES Reform}_i + \alpha_b \textit{ In Business(B)}_i \\ &+ \alpha_{r,t} \text{ Eligible} * \text{Reform}_i + \alpha_{r,t,b}\ B\text{*Eligible} * \text{Reform}_i + \beta'\ X_i + \epsilon_i. \end{aligned} \tag{3}$$

Second, within each of the five treated States, we consider only firms created after the tax reform on their respective States. For example, the State of Pará opted in the *SIMPLES* tax reform in January of 2003. We consider firms that are created, for instance, up to nine months after January of 2003, i.e., firms created between January and October of 2003. For comparison, we select as our control group young firms in States that did not adhere to the *SIMPLES* tax reform. By "young firms" we mean firms that were *in business* for less than nine month (the oldest age of a treated firm). Hence, we also estimate the following regression

$$\begin{aligned} RM_i &= \alpha_r \text{ Eligible* Firms}_i + \alpha_t \textit{ State-SIMPLES Reform}_i \\ &+ \alpha_{r,t} \text{ Eligible*} * \text{(after)Reform}_i + \beta' X_i + \epsilon_i \end{aligned} \quad (4)$$

where the variable Eligible* Firms_i is equal to one for firms in eligible economic sectors and *in business* for up to nine months and zero if the firm is either in a non-eligible sector or it is older than nine months; and the variable Eligible* $*$ (after)Reform_i captures the impact of the tax reform on the reported margins (costs, sales) for eligible young firms in states with the tax reform vis-à-vis older eligible firms, firms in non-eligible firms and young non-eligible firms.[8] The main goal with this approach is to control for any difference between firms in the treatment and control groups that is constant with respect to the time in business similar to Monteiro and Assuncao (2012).

2.4.2 Empirical Results

In this section we present our main empirical results focusing on the most complete empirical specification. Additional results are presented in Appendix A.1. Table II presents our estimates for reported sales ($\overline{y}$) and reported non-labor costs ($\widehat{y}$) responses for all firms in our sample. We focus on these two reported margins for two reasons. First, from the point of view of a business owner, these are the easiest margins to manipulate. Second, the *ECINF* reports that many entrepreneurs that responded the survey had difficulties to identify their total costs but showed detailed responses to itemized expenses which is confirmed in missing observations former but not the latter. We believe this might be an indication that not only entrepreneurs do not fully control their business finances but also that these itemized expenses provide better information about these business and, hence, should be used in our analysis (instead of total aggregate costs). Since labor market

[8]Firms located in states without a reforms are considered young if created less than 9 months. For the treated group, to be considered young a firm must have been created less than PA= 9, PR= 9, RN = 9 CE= 8, PB=6.

regulations may impose different incentives for the firm to misreport labor costs (Jales, 2018), we focus on non-labor expenses as these are harder to be detected by the tax authorities.

Estimations based on equation (1) are presented on columns (I) and (IV) of Table II. Columns (II) and (V) and columns (III) and (VI) report the estimated results according to the specifications in equations (3) and (4), respectively. First, we observe an almost symmetrical effect for reported sales and non-labor expenses - columns (I)-(II) and (IV)-(V) for our complete sample. Eligible firms report larger values for both variables. Being in a State that carried out a tax reform is also associated with larger values for reported sales and non-labor costs. The difference between the results in columns (I)-(II) versus those in columns (IV)-(V) is that the last ones are precisely estimated.

Consider, for instance, column (V) of Table II. It shows that an eligible sector firm, operating in one of the twelve States we consider, increased its reported non-labor costs by about R$1,187, conditional on all observable variables. As non-eligible firms report non-labor expenses of R$549 versus R$1,213 for eligible firms in the unconditional mean, this represents almost a 100% increase for eligible firms. Additionally, firms located in States that implemented a *State-SIMPLES* tax reform also reported larger costs (R$542) when compared to firms in States without a tax reform. More important to our analysis, the interaction between being eligible and being in a State that changed the tax code shows a reduction in reported non-labor costs of R$516 by eligible firms in treated States compared to non-eligible firms and/or in States without a tax reform (column (IV), Table II). We also find that allowing for interactions with the variable *In Business* $(B)_i$ does not significantly affect our main results (equation 3). And, no effect of the tax reform is found for newly created firms regarding their reported non-labor costs either.

Columns (III) and (VI), Table II, refer to estimations when we consider young firms only. These results reveal that these firms respond almost symmetrically regarding both margins - reported sales and reported non-labor cost - however with two remarkable distinctions. First, the estimated coefficients are of opposite signs when compared to the full sample model results, i.e., columns (I-II) and (IV-V). Second, the only model precisely estimated refers to the reported sales margin. That is, we find a significant impact of the *State-SIMPLES* tax reform on reported sales ($\overline{y}$). More specifically, we find a smaller reported sales for firms in the treated States (R$3,103; column III, Table II). A significant impact of the tax reform is identified on reported sales of newly created firms in eligible sectors on treated States versus newly created firms in either non-eligible sectors or in States without a tax reform. We estimate an increase of R$3,377 on the reported sales for our treated group.

Table II: Reduced-form responses in reported sales and non-labor costs

	Reported Sales ($\overline{y}$)			Reported Non-Labor Costs ($\widehat{y}$)		
	(I)	(II)	(III)	(IV)	(V)	(VI)
Eligible Firms	734.5	734.4	-1383.2	1187.3*	1187.3*	398.4
	(1.74	(1.73	(1.02)	(3.95	(3.95	(1.66
State-SIMPLES reform	430.8	429.9	-3103.2***	542.3***	542.2***	-1167
	(1.15	(1.14)	(1.93)	(1.85)	(1.84)	(1.02)
Eligible*Reform	-467.8	-561.6	3377.5***	-516.9*	-536.9**	1185.8
	(1.43)	(1.47)	(2.03)	(3.76)	(2.88)	(1.01)
B*Eligible*Reform		0.855		0.181		
		(1.13)		(0.25)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
# observations	20155	20155	1660	20313	20313	1675

Note: Standard errors clustered by census tract. Controls include time in business, Market Power, Outside Household, Accountant, Hours/week, owner's education, owner's age and age2, owner's tenure, firm's worker education and ECINF sector dummies *** p < 0.01, ** p < 0.05, * p < 0.10.

Last, we investigate whether upstream firms respond differently than downstream firms to the tax reform with respect to their reported margins. The results presented in Table III are based on equation (2). Columns (I) and (IV) present our results for reported sales and reported non-labor costs, respectively. And, the results with respect to interactions with the variable *in business* and for a sub-sample of newly created firms are presented in columns (II) and (V) and (III) and (VI), respectively.

We first notice that the results presented in Table III reinforce those presented in Table II. Two important remarks. First, upstream firms reported sales are significantly larger than those reported by downstream firms (R$1,020, column (I), Table II) and they are associated to a significant positive effect of this reported margin for eligible firms in our sample (R$732, column (I), Table II). Second, by acknowledging a firm's position in the production chain we are able to provide better information regarding the level of reported sales. We observe a much smaller reaction by the upstream firms (-R$1,876, column (III), Table III), but its not statistically different from the responses of reported margins of downstream firms of eligible sectors in States that participated in the tax reform in 2003 *vis-à-vis* non-eligible firms in States that did not reformed their tax codes. Similar to our previous estimates (Table II), we find that reported non-labor costs responses ranges from R$528 to R$543 (columns (IV)-(V), Table III), being lower for our treated group when compared to our control group.

We find evidence that the effect of the *State-SIMPLES* tax reform on newly created firms is positive and statistically significant. That is, the tax reform is associated with an increase of reported sales of R$3,469 (column (III), Table III) compared to the responses of newly created firms in non-eligible sectors or firms located in States without the tax reform. In response to the tax reform, newly created upstream firms in eligible economic sectors would increase their reported sales by R$1593 (i.e., R$3469 - R$1876 column (III), Table II) in States that changed their tax code, compared to newly created upstream firm in non-eligible sectors in States that did not alter their tax systems. Notice that the reduction of upstream firms reported non-labor costs is about R$116 (i.e., R$1240 - R$1154, column (VI), Table II) smaller than those by downstream firms after the *State-SIMPLES* tax reform. Upstream responses to tax changes in reported sales and non-labor costs are, however, not statistically different from by the downstream firms.[9]

Finally, we conduct two robustness checks (Table A.5, Appendix A.1). In the first exercise, we consider as a control group only the States that had already implemented their versions of the *State-SIMPLES* tax reform. This sample consists of fifteen States that were not used in our main exercise. As expected, we do not find any effect on reported sales or non-labor costs. Then, we randomize treated and non-treated states, making sure to keep at least 2 (out of five) treated states in one group. Again, we do not find any effect on reported sales and non-labor costs.[10]

[9]We also investigate whether the *State-SIMPLES* tax reform led to a higher formalization rate for the firms in our sample. As in de Paula and Scheinkman (2010), we use the variable *Tax Registry* but using only 12 states, being 5 of them considered to be treated, we find that the tax reform had no effect on firms' formalization, a result in line with Piza (2018).

[10]We also checked for balancing tests using our treated versus control sample with respect to the observable variables, see table A.4.

Table III: Reduced-form responses in reported sales and non-labor costs:
Upstream X Downstream Firms - Limited Tax Capacity

	Reported Sales ($\overline{y}$)			Reported Non-Labor Costs ($\widehat{y}$)		
	(I)	(II)	(III)	(IV)	(V)	(VI)
Eligible Firms	731.9***	730	-1259.2	1182.5*	1182.0*	475.1
	(1.8)	(1.78)	(0.94)	(4.19)	(4.16)	(1.77)
State-SIMPLES reform	430.4	429.3	-2993.5***	541.6***	541.4***	-1097.3
	(1.15)	(1.15)	(1.90)	(1.86)	(1.86)	(0.97)
Upstream (U)$^{(2)}$	1020.1*	1015.6*	2559.0**	340.1	339.2	1351
	(3.71)	(3.67)	(2.69)	(1.14)	(1.12)	(1.32)
Elilgible*Reform	-474.3	-550.7	3469.3***	-528.7*	-543.8***	1239.9
	(1.35)	(1.24)	(2.05)	(3.21)	(2.11)	(1.04)
B*Eligible*Reform		0.689			0.135	
		(0.64)			(0.13)	
U*Eligible*Reform	63.87	-109.1	-1876.7	116.4	56.12	-1154.6
	(0.12)	(0.13)	(1.19)	(0.23)	(0.07)	(1.33)
U*B*Eligible*VAT reform		1.859			0.633	
		(0.5)		(0.2)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
# observations	20155	20155	1660	20313	20313	1675

Note: Standard errors clustered by census tract. Controls include time in business, Market Power, Outside Household, Accountant, Hours/week, owner's education, owner's age and age2, owner's tenure, firm's worker education and and ECINF sector dummies *** p < 0.01, ** p < 0.05, * p < 0.10. *** p < 0.01, ** p < 0.05, * p < 0.10.

2.5 Discussion

Our empirical analysis have shown that there is suggestive evidence that eligible firms in States that had implemented a *State-SIMPLES* tax reform in 2003 seem to have reduced reported non-labor expenses more strongly than non-eligible firms in States that did not change their tax system, consistent with the literature (Carrillo et al., 2017; Pomeranz, 2015). Although we find evidence that upstream firms respond in both reported sales and costs with lesser magnitude than their downstream counterparts, our results suggest that these responses are not statistically different across production chain.

Consider the following illustrative example. A typical downstream firm in our sample with annual revenue of R$23,048 (R$1,920 monthly sales according to Table A.4) and input costs of R$17,212 (R$1,434/month, Table A.4) would owe R$1,050 in taxes (VAT) before the *State-SIMPLES*. Following the tax reform and under the new tax scheme, this same firm would have to remit approximately R$484 to the Brazilian tax authority. This means a reduction in the share of tax payments with respect to the reported revenue of 0.024 (i.e., Δ = - 0.024, from 0.045% to 0.021%). Using the estimated reduction in reported non-labor cost of R$520, Table II (mean equals to R$1,037 in Table A.4) this result would lead to an estimated elasticity response of 0.92, i.e., (-R$520/-0.024)*(0.045/R$1,037). Similarly, for a typical young firm with reported sales and costs of R$1,750 and R$1,412, respectively, we can compare the effective tax rate (0.038%) with the one after the reform, 0.021%. Using the estimated increase in reported revenue of R$3377 (Table II) we find a large elasticity of roughly -4.3, i.e., (R$3377/-0.017)*(0.038/R$1,750). Therefore, our results suggest that following a value-added tax decrease, we should observe (i) a decrease in the reported non-labor cost (an elasticity equal to 0.92) of small firms, probably as an attempt to reduce their tax base (value-added) and (ii) a more significant increase in the reported sales of newly created firms (an estimated elasticity of -4.3).

In the next section, we develop a stylized model to study the optimal decision of firms in a limited tax capacity environment, the implication of optimal alternative tax systems for reported margins - sales and production costs - of upstream and downstream firms and the relevant (mechanical and behavioral) mechanisms associated with the optimal tax policies.

3 A Limited Tax Capacity Economy

Consider an economy populated by a large number of identical consumers. There is an infinite number of identical downstream firms (subscript d), an upstream firm (subscript u), and the government. Production of the final good occurs in two stages. An upstream firm produces an intermediate good using only labor as input and sells it to downstream firms - we denote these transactions as business-to-business (B2B) transactions. Combining the intermediate good and labor, downstream firms produce a final good that is sold to consumers - business-to-consumers (B2C) transactions. We assume that upstream and downstream firms operate constant returns to scale technologies (Diamond and Mirrlees, 1971a,b). Consumers supply labor to both downstream and upstream firms. They consume the final good but not the intermediate good. Although one might prefer to think about a multi-sector production process, we find it closer to reality of small firms, however, to start with a one-sector formulation, focusing on homogeneous firms within an industry (sector) and homogeneous households. Our modeling approach can be extended to study n sectors and heterogeneous consumers.

The government taxes upstream and downstream production in order to finance an exogenous revenue requirement. We abstract from labor taxation. We study a multistage tax system (MST) in which the government's set of tax instruments is denoted as $\mathcal{T} = \left(\tau^u, \tau^d, \delta\right)$. The three policy instruments τ^u, τ^d, and δ are meant to tax B2B and B2C transactions. The upstream firm sales of the intermediate good to downstream firms are taxed at an *ad valorem* rate τ^u and τ^d is the *ad valorem* tax on the downstream sales of the final good to consumers. Downstream firms can also claim a fraction $\delta \in [0, 1]$ of the amount of taxes remitted by the upstream firm as their production cost. Misreporting taxes is costly and directly associated with the government's limited tax capacity.

Our modeling approach is flexible enough to allow us to analyze the effects of different tax instruments on firms' revenues and production costs as well as the responses of firms in different stages of production to alternative tax systems. In line with the *SIMPLES* tax reform, we analyze a value-added tax (VAT) and turnover tax (TOT) system, as summarized in Table IV. In a VAT system, both the intermediate and the final good sales are taxed at the same rate, i.e., $\tau^u = \tau^d = \tau_{VAT}$, and the downstream firm can claim the full amount of taxes remitted by the upstream firm as part of its production costs ($\delta = 1$). Hence, the VAT system is represented as $\mathcal{T}_{\mathcal{VAT}} = (\tau_{VAT}, \tau_{VAT}, 1)$. On the other hand, a TOT system, $\mathcal{T}_{\mathcal{TOT}} = \left(\tau^u_{TOT}, \tau^d_{TOT}, 0\right)$, is such that upstream and downstream firms are taxed at different tax rates but taxes remitted by the upstream firm can not be claimed as production costs by the downstream firm (i.e., $\delta = 0$).

3.1 Consumers, Downstream and Upstream Firms

Consumers are endowed with one unit of time, which they can supply to the upstream firm (l_u), supply to a downstream firm (l_d), or spend on leisure $z = 1 - l_u - l_d$. Workers are paid the same wage rate w regardless whether they work at the upstream or a downstream firm.

A representative consumer values consumption and leisure according to the utility function $U(c, z)$, where c represents the consumption of the final good y_d. We assume that $U(\cdot)$ is a strictly

Table IV: VAT, TOT and MST Systems: $\mathcal{T} = \left(\tau^u, \tau^d, \delta\right)$

Tax System	Description	$\mathcal{T} = \left(\tau^u, \tau^d, \delta\right)$
Multistage Tax (MST)	Taxation of intermediate and final goods sales at different tax rates. Zero to full refund.	$\mathcal{T}_{\mathcal{MST}} = \left(\tau^u, \tau^d, \delta\right)$
Value-Added Tax (VAT)	Taxation of intermediate and final goods sales at the same tax rate. Full refund.	$\mathcal{T}_{\mathcal{VAT}} = (\tau_{VAT}, \tau_{VAT}, 1)$
Turnover Tax (TOT)	Taxation of intermediate and final goods sales at different tax rates. No refund.	$\mathcal{T}_{\mathcal{TOT}} = \left(\tau^u_{TOT}, \tau^d_{TOT}, 0\right)$

concave, twice continuously differentiable function, separable in consumption and leisure, that satisfies the Inada conditions. Hence, the consumer solves the following maximization problem

$$\max_{c,l_u,l_d} \quad U\left(c, 1 - l_u - l_d\right)$$
$$\text{subject to } p_d c = w\left(l_u + l_d\right),$$

which implies the following equilibrium condition

$$w \quad = \quad \frac{U_z\left(c, 1 - l_u - l_d\right)}{U_c\left(c, 1 - l_u - l_d\right)} p_d, \tag{5}$$

where $U_c = \partial U/\partial C$, $U_z = \partial U/\partial z$. We normalize the wage rate to 1 (i.e., $w = 1$). Labor is our *numeraire* (Keen, 2008) and it is not taxed.

Downstream firms produce and sell the final good to consumers in order to maximize profits subject to input (labor and intermediate goods) and tax misreporting costs. They purchase the intermediate good y_u from the upstream firm at a price p_u, and combine it with labor l_d in order to produce the final good y_d. A constant returns to scale production function $F\left(y_u, l_d\right)$ is increasing in both arguments y_u and l_d and it is strictly concave. Downstream firms operate in a perfect competitive market (Diamond and Mirrlees, 1971a). These are common assumptions in the efficiency analysis literature and, by assuming them, we can restrict our attention to the role of a limited tax capacity government on the optimal design of tax instruments and tax systems.

Taxation affects downstream firms profit maximization at the following two margins in a multistage tax system $\mathcal{T}_{\mathcal{MST}} = \left(\tau^u, \tau^d, \delta\right)$. A downstream firm has to remit τ^d on the amount of final goods sold to consumers, but it can deduct from its production cost a fraction (δ) of taxes τ^u that the upstream firm had remitted due to the intermediate good sales. However, the government's limited tax capacity allows a profit-maximizing downstream firm to misreport (i) the amount of intermediate goods purchased and (ii) the amount of final goods sold.

A downstream firm reports sales of the final good in the amount of $\overline{y_d}$ and intermediate goods' purchases equal to $\widehat{y_u}$ to the tax authority. Notice that, in limited tax capacity economy, the amounts reported may not be the equal to the their true amounts, i.e., $\overline{y_d} \neq y_d$ and $\widehat{y_u} \neq y_u$. Our modelling approach accommodates a downstream firm that either misreport at both margins or misreport at just one of them (either regarding its purchase of inputs or sales of the final good). With respect to the amount of input (intermediate good) purchased from the upstream firm, a downstream firm has an incentive to report an amount higher than the actual amount purchased (over-report inputs purchases, i.e., $\widehat{y_u} > y_u$). In doing so, it would be entitled to a larger refund of taxes claimed to be paid by the upstream firm. On the other hand, the downstream firm can, at a cost, under-report its final good sales ($\overline{y_d} < y_d$) in order to remit less taxes to the tax authority.

It is important to reiterate that misreporting is costly and margin-specific. This means that when deciding to misreport taxes, a downstream firm has to consider the costs associated with the misreporting of the total amount of final good's sales as well as the costs associated with misreporting intermediate good's purchases. Due to, for instance, auditing procedures, the government limited tax capacity might affect these two margins differently. We assume that misreporting costs are strictly increasing, convex and differentiable functions of the difference between the actual amounts sold and purchased and the ones reported to the tax authority, rather than their respective tax liabilities (Best et al., 2015; Slemrod, 2001).[11] For instance, let $J\left(\widehat{y_u} - y_u\right)$ be the misreporting cost associated with the purchase of intermediate goods in the (actual) amount of y_u while reporting an amount of $\widehat{y_u}$ to the government. Similarly, misreporting final good sales entails a cost of $H\left(y_d - \overline{y_d}\right)$. We assume that $J(0) = 0$, $H(0) = 0$.

[11] We are assuming that the misreporting resource cost depends only on output evaded rather than the amount of taxes evaded (Yitzhaki, 1974). Our analysis goes through if one assumes that the misreporting cost takes a nonlinear (but positive) relation with tax rates. Also, we are focusing on firms with constant returns to scale technology and our analysis could easily be extended to many firms with an aggregated unreported misreport cost function defined above.

Thus, a downstream firm chooses labor demand (l_d), the amount of the intermediate good (y_u) and how much to report to the tax authority ($\widehat{y_u}$, $\overline{y_d}$) in order to maximize profits, as follows

$$\begin{aligned}\Pi^d\left(l_d, \overline{y_d}, \widehat{y_u}\right) = p_d F\left(y_u, l_d\right) - \tau^d p_d \overline{y_d} - \left[l_d + p_d H\left(F\left(y_u, l_d\right) - \overline{y_d}\right)\right] \\ - \left[p_u y_u + p_u J\left(\widehat{y_u} - y_u\right) - \delta\left(\tau^u p_u \widehat{y_u}\right)\right].\end{aligned} \tag{6}$$

Equation (6) reports first-order conditions for the maximization problem with respect to $l_d, y_u, \overline{y_d}, \widehat{y_u}$ respectively:

$$p_d F_{l_d}(\cdot) - 1 - p_d H_{\Delta\overline{y_d}}(\cdot) F_{l_d}(\cdot) = 0, \tag{7}$$

$$p_d F_{y_u}(\cdot) - p_u + p_u J_{\Delta\widehat{y_u}}(\cdot) - p_d H_{\Delta\overline{y_d}}(\cdot) F_{y_u}(\cdot) = 0, \tag{8}$$

$$-\tau^d p_d + p_d H_{\Delta\overline{y_d}}(\cdot) = 0, \tag{9}$$

$$\delta\tau^u p_u - p_u J_{\Delta\widehat{y_u}}(\cdot) = 0, \tag{10}$$

where $\Delta\overline{y_d} = y_d - \overline{y_d}$ and $\Delta\widehat{y_u} = y_u - \widehat{y_u}$ denote the misreported downstream (final good) sales gap and the downstream (intermediate good) purchase gap, respectively; $H_{\Delta\overline{y_d}}(\cdot) = \partial H\left(\Delta\overline{y_d}\right)/\partial\Delta\overline{y_d}$, $J_{\Delta\widehat{y_u}}(\cdot) = \partial J\left(\Delta\widehat{y_u}\right)/\partial\Delta\widehat{y_u}$, $F_{y_u}(\cdot) = \partial F\left(y_u, l_d\right)/\partial y_u$, and $F_{l_d}(\cdot) = \partial F\left(y_u, l_d\right)/\partial l_d$. Equation (7) equalizes the marginal cost and the marginal benefit of using labor to produce the final good, while equation (8) reflects the marginal cost and benefit of using the intermediate good to increase the final good production. The trade-offs regarding misreporting the final goods sales and the intermediate good purchases are captured by equations (9) and (10), respectively.

The (perfectly competitive) upstream firm produces an intermediate good y_u using labor l_u. It has a linear technology that converts labor into output, i.e., $y_u = l_u$. In an economy with limited tax capacity, the upstream firm can misreport its output or, equivalently, the amount of goods it sells to downstream firms. Let $\overline{y_u}$ be the amount reported by the upstream firm to the tax authority, which may not be necessarily equal to the actual amount of intermediate goods produced (y_u). We abstract from cross-validation and, hence, the amount $\overline{y_u}$ can still be different than the amount claimed to be purchased by the downstream firms ($\widehat{y_u}$).[12] For simplicity, we assume that

[12]If cross-validation is possible, downstream and upstream firms can either tell the truth (y^u) or collude (both

the upstream firm misreporting cost function is the same as the downstream firm, i.e., $J\left(y_u - \overline{y_u}\right)$. Notice, however, that while downstream firms have an incentive to over-report the purchase of the intermediate good, the upstream firm would under-report the sales of the intermediate good.

Hence, a upstream firm chooses $l_u, \overline{y_u}$ in order to maximize after-tax profits

$$\Pi^u\left(l_u, \overline{y_u}\right) = p_u l_u - \tau^u p_u \overline{y_u} - l_u - p_u J\left(l_u - \overline{y_u}\right). \tag{11}$$

At the firm's optimum, the following equilibrium conditions must hold

$$p_u - 1 - p_u J_{\Delta \overline{y_u}}(\cdot) = 0, \tag{12}$$

$$-\tau^u p_u + p_u J_{\Delta \overline{y_u}}(\cdot) = 0, \tag{13}$$

where $\Delta \overline{y_u} = y_u - \overline{y_u}$ is the upstream (intermediate good) sales gap. Equations (12) and (13) imply that the price of the intermediate good is simply $p_u = 1/\left(1 - \tau^u\right)$.

3.2 Unlimited vs. Limited Tax Capacity Economy

Before we move on to the government's optimal taxation problem, it is important to discuss the implications of the tax authority limited tax capacity on the firms' decisions to misreport the amount of sales and purchases. We model the government limited tax capacity via (upstream/downstream) tax misreporting costs as a sheltering with a resource cost model, similar to Chetty (2009)'s approach. In other words, in a limited tax capacity economy any reported amount that is lower than the actual (true) amount would imply in a benefit for the firms, i.e., less taxes are remitted at the margin. This marginal benefit can be, in equilibrium, larger or equal to the cost to misreport sales and purchases. On the hand, an unlimited tax capacity environment is characterize by a marginal cost of misreporting one additional unit of revenue (cost) larger than the amount of taxes saved.[13] The following definition summarizes the conditions that characterize

reporting in the range $[\widehat{y_u, y_u}]$) and splitting the net gains from evasion. In this last case, one of these firms would necessarily remit larger taxes than it really should. Our framework can accommodate this situation.

[13] Notice that we are only considering the jointly cases of either perfect or imperfect tax enforcement of both upstream and downstream taxation. Of course, it is possible to envision a situation in which only one of these production stages are subject to tax enforcement. For instance, the government could perfectly enforce taxes levy on upstream firms while only imperfectly taxing downstream firms and vice-versa. This can be captured in our model by restricting the margins of evasion of either the upstream or the downstream firms.

either an unlimited or a limited tax capacity economy.

Definition 1. *The government tax capacity is defined as either unlimited or limited according to the following condition*

	Tax Capacity	
	Unlimited	Limited
Upstream	$p_u J_{\overline{y_u}}(0) > \tau^u p_u$	$p_u J_{\overline{y_u}}(0) \leq \tau^u p_u$
Downstream	$p_d H_{\overline{y_d}}(0) > \tau^d p_d$	$p_d H_{\overline{y_d}}(0) \leq \tau^d p_d$
	$p_u J_{\widehat{y_u}}(0) > \delta\tau^u p_u$	$p_u J_{\widehat{y_u}}(0) \leq \delta\tau^u p_u$

In the case of a limited tax capacity economy, the upstream firm reports sales in the amount of $\overline{y_u}$ and a downstream firm reports final good sales of $\overline{y_d}$ and the purchase of intermediate (upstream) goods of $\widehat{y_u}$ such that the following equilibrium conditions are satisfied

$$J_{\overline{y_u}}\left(y_u - \overline{y_u}\right) = \tau^u p_u, \tag{14}$$

$$H_{\overline{y_d}}\left(y_d - \overline{y_d}\right) = \tau^d p_d, \tag{15}$$

$$J_{\Delta\widehat{y_u}}\left(\widehat{y_u} - y_u\right) = \delta\tau^u p_u. \tag{16}$$

Assume a given level of production by the upstream firm and a downstream firm y_u, y_d, respectively. The revenue efficiency argument implies that a reduction of either the refund rate ($\downarrow \delta$) or the tax rate on the upstream output ($\downarrow \tau^u$) discourages, at the margin, a downstream firm to misreport the amount of the intermediate good it purchases, equation (16). That is, with a lower refund rate or a lower tax on the upstream production, downstream firms have less incentive to overreport their purchases of intermediate goods from the upstream firm (i.e., they over report less, $\downarrow \widehat{y_u}$). We can see from equations (14) and (15) that reductions in the upstream and downstream tax rates, i.e., $\downarrow \tau^u$ and $\downarrow \tau^d$, respectively, encourage firms to underreport less (i.e., $\uparrow \overline{y_u}$ and $\uparrow \overline{y_d}$ and report amounts closer to the actual amount sold y_u and y_d, respectively). In summary, a reduction in any of the policies (i.e., $\downarrow \tau^u$, $\downarrow \tau^d$, or $\downarrow \delta$), discourages sales or purchase misreporting.

Moreover, the equilibrium conditions of the upstream and downstream firms allow us to analyze how the government tax instruments may distort the scale of (upstream and downstream) production, as well the optimal combination of inputs in the downstream production stage. Equations

(7) - (10) imply that the optimal decisions of a downstream firm regarding its demand for inputs must satisfy the following conditions:

$$1 = F_{l_d}(\cdot)\left(1-\tau^{d}\right)p_d, \tag{17}$$

$$\frac{(1-\delta\tau^{u})}{(1-\tau^{u})} = F_{y_u}(\cdot)\left(1-\tau^{d}\right)p_d. \tag{18}$$

For a downstream firm, the latter condition, equation (18) equalizes the marginal cost and the marginal benefit of using the intermediate good (instead of labor), paying a price $p_u = (1-\tau^u)^{-1}$, to increase the production of the final good production. Notice that this decision takes into account that the upstream firm costly misreports its sales and a downstream firm costly misreports the amount purchased of the intermediate good. Taxes on upstream and downstream production distort not only the scale of production but also the input composition of the final good produced by downstream firms. In other words, for a given refund rate δ, taxation of the intermediate good sector distorts downstream firm's choice of inputs away from the intermediate good. Combining equations (17) and (18), these partial equilibrium effects are clearly expressed by the marginal rate of technical substitution between (downstream) labor (l^d) and the intermediate good (y^u). That is, $MRTS^{MST}_{l^d,y^u} = F_{l_d}(\cdot)/F_{y_u}(\cdot) = (1-\tau^u)\,/\,(1-\delta\tau^u)$.

At this point, it is illustrative to compare a MST system with two other well-known tax systems, a VAT and a TOT. In a VAT system $\mathcal{T}_{\mathcal{VAT}} = (\tau_{VAT}, \tau_{VAT}, 1)$, taxation does not distort the downstream firm's choice between labor and the intermediate good inputs, i.e., $F_{l_d}(\cdot) = F_{y_u}(\cdot)$. The marginal rate of technical substitution in this tax system is, hence, equal to one. By design, the VAT system corrects for any distortions that the taxation of the intermediate (upstream) good might cause in the downstream firm optimal choice of production inputs by allowing a downstream firm to claim the full amount of taxes remitted by the upstream firm as part of their production costs. On the other hand, in a TOT system $\mathcal{T}_{\mathcal{TOT}} = \left(\tau^u_{TOT}, \tau^d_{TOT}, 0\right)$, a downstream firm can not deduct the upstream firm tax payments as part of their costs. As a result, the marginal rate of technical substitution in this tax system $MRTS^{TOT}_{l^d,y^u} = (1-\tau^u_{TOT})$, i.e., $F_{l_d}(\cdot) < F_{y_u}(\cdot)$. The main consequence of a zero refund rate (i.e., $\delta = 0$) as in a TOT system is that the use of intermediate goods as inputs of final good production is taxed twice - first, directly at the upstream stage of production and again, albeit indirectly, at the final production (downstream) stage by not allowing the refund of taxes already collected in the production chain. Notice that while this cascading effect of the intermediate good taxation illustrates a source of production inefficiency in a TOT system that arises in general equilibrium (taxation cascades through the production chain), such production distortions can be avoided in, for instance, in a VAT system.

3.3 The Government

In an economy with limited tax capacity and a multistage tax system $\mathcal{T}_{\mathcal{MST}} = \left(\tau^u, \tau^d, \delta\right)$, the government finances an exogenous revenue requirement R with tax revenue which depends on the amount of sales reported by the upstream firm ($\overline{y_u}$) and downstream firms ($\overline{y_d}$), as well as the amount of the intermediate good downstream firms report they purchased from the upstream firm ($\widehat{y_u}$). Thus, the government budget constraint in a multistage tax system $\mathcal{T}_{\mathcal{MST}} = \left(\tau^u, \tau^d, \delta\right)$ is as follows

$$R = T_{MST}\left(\overline{y_d}, \overline{y_u}, \widehat{y_u}\right) = \tau^d p_d \overline{y_d} + \tau^u p_u \overline{y_u} - \delta \tau^u p_u \widehat{y_u}. \tag{19}$$

The upstream firm and the downstream firms decisions to misreport their sales and purchases affect the amount of taxes collected by the government. This, in turn, constrains the optimal choice of policy instruments $\left(\tau^u, \tau^d, \delta\right)$. In a multistage tax system $\mathcal{T}_{\mathcal{MST}} = \left(\tau^u, \tau^d, \delta\right)$, equation (19) illustrates that the government can raise revenue by increasing tax rates on either the upstream output ($\uparrow \tau^u$) or the downstream output ($\uparrow \tau^d$), or on both ($\uparrow \tau^u, \tau^d$). Revenue could also be increased if the government reduces the refund rate ($\downarrow \delta$).

Changes in these policies have important effects on firms along the production chain. First, changes in the upstream *ad valorem* tax rate τ^u affect only the upstream firm decisions. For instance, a increase in τ^u might either decrease the firm's real output level (higher taxation discourage production) or increase the level of misreporting (the firm reports a lower level of output produced and sold to a downstream firm). Second, decrease in the downstream tax rate τ^u and increase in the refund rate δ reduces the (effective) tax burden on downstream firms and thus creating an incentive for them to increase their output level. While the effect on production is clearer, the changes in these policy instruments might have different (asymmetric) impacts on the misreporting behavior of a downstream firm. On one hand, a lower *ad valorem* tax rate τ^d reduces (at the margin) the level of output sales a downstream firm misreport to the government. On the other hand, if a higher refund rate δ is offered, a downstream firm has an incentive to overreport the amount of (intermediate) goods it purchases from the upstream firm.

3.4 Welfare Maximization Problem

The government's goal is to maximize the economy's welfare by choosing tax rates and refund rate $\mathcal{T} = \left(\tau^u, \tau^d, \delta\right)$ subject to an exogenous revenue requirement R as follows:

$$W = U\left(c, 1 - l_u - l_d\right) + \lambda\left[T\left(\overline{y_d}, \overline{y_u}, \widehat{y_u}\right) - R\right] \tag{20}$$

where λ is the multiplier on the government revenue requirement constraint, i.e., the (endogenous) marginal cost of public funds, equation (19). In the welfare maximization problem, the government takes into account the equilibrium reactions of consumers and firms to a distortionary tax system, i.e., equations (7) - (10), (17), and (18). In this stylized framework, the government's problem amounts to maximize the consumer's indirect utility function $v\left(p_d\left(\mathcal{T}\right)\right)$ subject to the revenue requirement.

Consider the optimality conditions of the consumer utility maximization problem and the profit maximization problems of a representative downstream firm and the upstream firm, equations (7) - (9), (12) - (13) and $p_u = \left(1 - \tau^u\right)^{-1}$, respectively. We write the (mis)reported amounts $\overline{y_d}$, $\overline{y_u}$, and $\widehat{y_u}$, as functions of the government policies, i.e., $\overline{y_d}\left(p_d\left(\mathcal{T}\right), p_u\left(\tau^u\right), \mathcal{T}\right)$, $\overline{y_u}\left(p_d\left(\mathcal{T}\right), p_u(\tau^u), \tau^u\right)$ and $\widehat{y_u}\left(p_d\left(\mathcal{T}\right), p_u\left(\tau^u\right), \tau^u, \delta\right)$, respectively. This means that the reported amounts respond directly and indirectly to changes in the multistage tax system. For instance, consider the reported amount of the intermediate good produced by the upstream firm, i.e., $\overline{y_u}\left(p_d\left(\mathcal{T}\right), p_u(\tau^u), \tau^u\right)$. In equilibrium, $\overline{y_u}$ are directly affected by the price of the final good p_d, the price of the intermediate good p_u and the *ad valorem* tax rate on the upstream production τ^u. Notice, however, $\overline{y_u}$ is indirectly affected by all policies of the multistage tax system via their effect on p_d and by the effect of changes in τ^u that affect p_u. In other words, changes in $\mathcal{T} = \left(\tau^u, \tau^d, \delta\right)$ directly affect p_d, which indirectly leads to changes in $\overline{y_u}$. Similar direct and indirect effects of the policy instruments on the reported amounts can be identified for $\overline{y_d}\left(p_d\left(\mathcal{T}\right), p_u\left(\tau^u\right), \mathcal{T}\right)$ and $\widehat{y_u}\left(p_d\left(\mathcal{T}\right), p_u\left(\tau^u\right), \tau^u, \delta\right)$.

Hence, the government's welfare objective function can be written as follows

$$\begin{aligned} W &= v\left(p_d\left(\mathcal{T}\right)\right) \\ &+ \lambda\left\{\tau^d p_d\left(\mathcal{T}\right)\overline{y_d}\left(p_d\left(\mathcal{T}\right), p_u\left(\tau^u\right), \mathcal{T}\right)\right. \\ &+ \tau^u p_u\left(\tau^u\right)\overline{y_u}\left(p_d\left(\mathcal{T}\right), p_u\left(\tau^u\right), \tau^u\right) \\ &- \left.\delta\tau^u p_u\left(\tau^u\right)\widehat{y_u}\left(p_d\left(\mathcal{T}\right), p_u\left(\tau^u\right), \tau^u, \delta\right) - R\right\}. \end{aligned} \tag{21}$$

In the next section, we solve the government's problem for the optimal choice of policy instruments in a MST system, i.e., $\mathcal{T} = \left(\tau^u, \tau^d, \delta\right)$. The complete derivations of the government's welfare maximization problem are presented in the Appendix A.3 to improve readability. We compare the results for a limited versus an unlimited tax capacity economy, as well as the optimal policy instruments in a VAT and TOT systems.

3.5 Optimal Tax Systems in a Limited Tax Capacity Economy

3.5.1 Limited Tax Capacity: Optimal Multistage Tax System

Recall that in a limited tax capacity economy, it is costly for the upstream and downstream firms to misreport to the tax authority the amounts purchased and sold. This feature of our model economy has important implications for the optimal choice of the policy instruments τ^u, τ^d, and δ, starting with their mechanical effects. The mechanical effects of an increase in the upstream tax rate (M_{τ^u}), the downstream tax rate (M_{τ^d}) and the tax refund rate (M_δ) are, respectively,

$$M_{\tau^u} = -(\alpha/\lambda)y_d p_{d,\tau^u} + \tau^d p_{d,\tau^u}\overline{y_d} + \left(p_u + \tau^u p_{u,\tau^u}\right)\left(\overline{y_u} - \delta\widehat{y_u}\right) \tag{22}$$

$$M_{\tau^d} = -(\alpha/\lambda)y_d p_{d,\tau^d} + \left(p_d + \tau^d p_{d,\tau^d}\right)\overline{y_d} \tag{23}$$

$$M_\delta = -(\alpha/\lambda)y_d p_{d,\delta} + \tau^d p_{d,\delta}\overline{y_d} - \tau^u p_u\widehat{y_u} \tag{24}$$

where α is the marginal utility of income and the interpretation of these effects is straightforward. Consider, for instance, the normalized mechanical effect M_{τ^u}, equation (22), of an increase in the upstream tax rate τ^u. The term $-(\alpha/\lambda)y_d p_{d,\tau^u}$ is the known direct welfare effect of a change in the upstream tax rate. This first effect, also present in an unlimited tax capacity case, refers to the final good price response p_{d,τ^u} to changes in τ^u, weighted by the demand of the final good demand y_d, which is evaluated at the social marginal utility of income (α/λ).[14] The second and third terms of equation (22) represent the mechanical effects of changes in the policy instrument τ^u on the amount a downstream firm reports its sales, i.e., $\tau^d p_{d,\tau^u}\overline{y_d}$, and on the reported upstream sales and the reported downstream purchases of the intermediate good. That is, $(p_u + \tau^u p_{u,\tau^u})(\overline{y_u} - \delta\widehat{y_u})$. More specifically, for a given downstream tax rate and reported sales of the final goods, i.e., τ^d and $\overline{y_d}$, respectively, the second term of equation (22) captures how an increase in the upstream tax rate τ^u affects, via the price of the final good, the reported revenue of a downstream firm. The last term of equation (22) represents the direct and indirect effects (via prices) of a change in the upstream tax rate on the upstream reported output and on the downstream reported purchases. In our notation, the direct mechanical effects are represented by the terms $p_u\overline{y_u}$ and $p_u\delta\widehat{y_u}$, while the terms $\tau^u p_{u,\tau^u}\overline{y_u}$ and $\tau^u p_{u,\tau^u}\delta\widehat{y_u}$ represent the indirect mechanical effects of an increase in the upstream tax rate τ^u. Similarly, equations (23) and (24) define the normalized mechanical effects M_{τ^d} and M_δ of an increase in the downstream tax rate τ^d and in the refund tax rate δ, respectively.

Manipulating the relevant first-order conditions assuming interior solution - i.e., $(\partial W/\partial t)/\lambda = 0$ for $t \in \mathcal{T} = \{\tau^u, \tau^d, \delta\}$ - and applying the envelope theorem, we can write the changes in the consumer's welfare for a given change in the policy instruments τ^u, τ^d, and δ as follows (see Appendix A.3 for details),

$$\begin{bmatrix} \varepsilon_{\overline{y_d},\tau^u} & \varepsilon_{\overline{y_u},\tau^u} & \varepsilon_{\widehat{y_u},\tau^u} \\ \varepsilon_{\overline{y_d},\tau^d} & \varepsilon_{\overline{y_u},\tau^d} & \varepsilon_{\widehat{y_u},\tau^d} \\ \varepsilon_{\overline{y_d},\delta} & \varepsilon_{\overline{y_u},\delta} & \varepsilon_{\widehat{y_u},\delta} \end{bmatrix} \begin{bmatrix} -\tau^d p_d(\mathcal{T})(\overline{y_d}) \\ -\tau^u p_u(\tau^u)(\overline{y_u}) \\ +\delta\tau^u p_u(\tau^u)(\widehat{y_u}) \end{bmatrix} = \begin{bmatrix} -M_{\tau^u}\tau^u \\ -M_{\tau^d}\tau^d \\ -M_\delta\delta \end{bmatrix} \tag{25}$$

where $\varepsilon_{\overline{y},t} = [(\partial\overline{y}/\partial j)(t)/(\overline{y})$. To easy notation, we denote the behavioral effects of policy changes as the elasticities of the misreported sales and purchase gaps to a particular policy instrument $t \in \mathcal{T} = \{\tau^u, \tau^d, \delta\}$, adjusted for the value of their respective gap, namely, $p_d(\mathcal{T})\overline{y_d}$, $p_u(\tau^u)\overline{y_u}$,

[14] As it is standard in the literature, this term becomes $-(\alpha/\lambda)y_d$ with the assumption of unit (specific) taxes and full incidence on consumers.

and $p_u(\tau^u)\widehat{y_u}$, as follows:

$$\mathcal{B}_{\overline{y_d},t} = p_d(\mathcal{T})\overline{y_d}\varepsilon_{\overline{y_d},t} \tag{26}$$

$$\mathcal{B}_{\widehat{y_u},t} = p_u(\tau^u)\widehat{y_u}\varepsilon_{\overline{y_u},t} \tag{27}$$

$$\mathcal{B}_{\overline{y_u},t} = p_u(\tau^u)\overline{y_u}\varepsilon_{\widehat{y_u},t} \tag{28}$$

For instance, consider the behavioral effect on the misreported (final good) sales gap of a change in the upstream tax rate, equation (26). The term $\mathcal{B}_{\overline{y_d},\tau^u}$ represents the elasticity of misreported downstream revenue, weighted by the value of a downstream misreported tax base. This allows us to measure by how much the sales revenue of a downstream firm changes due to changes in the taxation of the upstream sales of the intermediate good. In addition, we can compare the behavioral response of a firm's revenue to the mechanical effect of tax collection after a marginal increase in taxes using more meaningful misreported values. The characterization of the optimal multistage tax system $\mathcal{T} = \{\tau^{u*}, \tau^{d*}, \delta^*\}$ in a limited tax capacity economy is presented in the Proposition 3.

Proposition 1. *In limited tax capacity economy, the optimal multistage tax system* $\mathcal{T} = \{\tau^{u*}, \tau^{d*}, \delta^*\}$ *is the solution of the system of equations represented by equation (25). The optimal tax policies are functions of multistage behavioral effects* $\mathcal{B}_{\overline{y_d},t}, \mathcal{B}_{\widehat{y_u},t}\mathcal{B}_{\overline{y_u},t}$*,* $t \in \mathcal{T} = \{\tau^{u*}, \tau^{d*}, \delta^*\}$*, equations (26) - (28), and the (normalized) mechanical welfare effects of* τ^{u*}*,* τ^{d*} *and* δ^* *(i.e.,* $M_{\tau^{u*}}$ $M_{\tau^{d*}}$ *and* M_{δ^*}*, respectively), equations* (22) - (24).

Proof. See Appendix A.3. ■

Proposition 3 shows that the optimal policy instruments are determined by two main effects, one behavioral and another mechanical. In fact, when choosing the optimal τ^{u*}, τ^{d*} and δ^*, the government faces a trade-off between the mechanical effect and the revenue efficiency effect, where the latter is captured by the (semi-)elasticities of the misreported sales and purchases. On one hand, if the misreporting elasticity is small relative to the relevant policy mechanical effect, it is socially optimal to increase (decrease) tax rates (refund rate). On the other hand, if the misreported gap (semi-)elasticity is large relative to the actual output elasticity, the revenue efficiency argument becomes stronger and it is optimal to lower the upstream tax rate and/or the tax rate on the (final good) downstream output. This occurs such that the optimal policy reduces the incentives to misreport sales and costs, consequently reducing distortion in the economy.

Expressing the solution of the government's welfare maximization problem in terms of the elasticity of the reported sales and costs highlights two key advantages of our approach. First, we show which are the statistics to evaluate welfare after each of theses policies change. Second, this allows us to communicate with our empirical analysis (Section *2)* as we estimate three intensive margin responses ($\varepsilon_{\overline{y^d},\tau}$, $\varepsilon_{\overline{y^u},\tau}$ and $\varepsilon_{\widehat{y_u},\tau}$).

3.5.2 Limited Tax Capacity: Optimal VAT and TOT Systems

Next, we investigate the optimal tax instruments of the two tax systems considered in the *State-SIMPLES* tax reform (Section 2), i.e., a VAT and a TOT systems. Assuming a limited tax capacity economy, first, we rewrite equation (25) for the optimal VAT system $\mathcal{T}^*_{\mathcal{VAT}} = (\tau^*_{VAT}, \tau^*_{VAT}, 1)$, and after some manipulation we have:

$$\tau^*_{VAT} = \frac{1}{\left(\widetilde{\mathcal{B}}_{\overline{y_d},\tau^*_{VAT}} + \widetilde{\mathcal{B}}_{\widehat{y_u},\tau^*_{VAT}} - \widetilde{\mathcal{B}}_{\overline{y_u},\tau^*_{VAT}}\right)} M_{\tau^*_{VAT}} \tag{29}$$

where $\widetilde{\mathcal{B}}_{y,\tau^*_{VAT}} = \left(\mathcal{B}_{y,\tau^*_{VAT}}/\tau^*_{VAT}\right)$, for $y \in \{\overline{y_d}, \overline{y_u}, \widehat{y_u}\}$. That is, in order to implement the optimal VAT system, the government must take into account the (normalized) mechanical welfare effects of τ^*_{VAT}, as well as three additional behavioral responses, $\widetilde{\mathcal{B}}_{\overline{y_d},\tau^*_{VAT}}$, $\widetilde{\mathcal{B}}_{\widehat{y_u},\tau^*_{VAT}}$, and $\widetilde{\mathcal{B}}_{\overline{y_u},\tau^*_{VAT}}$, namely, the effects of τ^*_{VAT} on a downstream reported sales of the final good and purchases of the intermediate good and the behavioral effect on the upstream firm reported sales, respectively. Notice that these behavioral effects are weighted by their respective misreported tax bases; that is, $p_d\left(\mathcal{T}^*_{\mathcal{VAT}}/\tau^*_{VAT}\right)$ and $p_u\left(\tau^*_{VAT}\right)/\tau^*_{VAT}$. On one hand, the higher the elasticities of reported sales and the tax rate (i.e., the higher $\widetilde{\mathcal{B}}_{\overline{y_d},\tau^*_{VAT}}, \widetilde{\mathcal{B}}_{\overline{y_u},\tau^*_{VAT}}$), the smaller is the tax rate τ^*_{VAT}, but at a greater inefficiency. On the other hand, a high elasticity of reported downstream purchases, i.e., $\widetilde{\mathcal{B}}_{\widehat{y_u},\tau^*_{VAT}}$ implies a higher τ^*_{VAT} and reduces tax inefficiency.

In the case of a TOT system $\mathcal{T}^*_{\mathcal{TOT}} = \left(\tau^{u*}_{TOT}, \tau^{d*}_{TOT}, 0\right)$, if we assume that there are no cross-effects such that $\varepsilon_{\Delta\overline{y_u},\tau^d_{TOT}} = \varepsilon_{\Delta\overline{y_d},\tau^u_{TOT}} = 0$, the tax instruments are determined by the solution of the following system of equations:

$$\begin{bmatrix} \tau^{d*}_{TOT} \\ \tau^{u*}_{TOT} \end{bmatrix} = \frac{\begin{bmatrix} \widetilde{\mathcal{B}}_{\overline{y_u},\tau^{u*}_{TOT}} & -\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{u*}_{TOT}} \\ -\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{d*}_{TOT}} & \widetilde{\mathcal{B}}_{\overline{y_d},\tau^{d*}_{TOT}} \end{bmatrix}}{\left(\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{d*}_{TOT}}\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{u*}_{TOT}}\right) - \left(\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{d*}_{TOT}}\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{u*}_{TOT}}\right)} \begin{bmatrix} M_{\tau^{d*}_{TOT}} \\ M_{\tau^{u*}_{TOT}} \end{bmatrix} \tag{30}$$

The optimal TOT system must take into account the revenue gains of the τ^{u*}_{TOT} and τ^{d*}_{TOT} mechanical effects, weighted by the changes in the misreported sales due to changes in the tax instruments behavioral effects.[15] In fact, the optimal TOT system effectively distorts allocations in the economy such that, at the social optimum, the tax wedge must be equal to the ratio between the mechanical and the behavioral responses.

The expression (40) highlights that the optimal upstream tax rate (τ^{u*}_{TOT}) and the optimal downstream tax rate (τ^{d*}_{TOT}) in a TOT system are to be set such that the behavioral responses of both the upstream and downstream firms must be taken into account along with the correspondent mechanical effects of both taxes. For instance, the optimal downstream tax rate must weight the mechanical effect of the (direct) downstream taxation $M_{\tau^{d*}_{TOT}}$, weighted by $\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{u*}_{TOT}} / \left(\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{d*}_{TOT}}\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{u*}}\right) - \left(\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{d*}_{TOT}}\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{u*}_{TOT}}\right)$ and the mechanical effect of the (indirect) upstream taxation $M_{\tau^{u*}_{TOT}}$, weighted by $-\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{u*}_{TOT}} / \left(\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{d*}_{TOT}}\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{u*}_{TOT}}\right) - \left(\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{d*}_{TOT}}\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{u*}_{TOT}}\right)$. It is reasonable to expect that the direct effect of taxes on the misreported margins are larger than the cross-effects, i.e., $\left(\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{d*}_{TOT}}\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{u*}_{TOT}}\right) > \left(\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{d*}}\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{u*}_{TOT}}\right)$. Therefore, $\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{u*}_{TOT}} M_{\tau^{d*}_{TOT}} > -\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{u*}_{TOT}} M_{\tau^{u*}_{TOT}}$, which puts a higher (positive) weight on the tax own mechanical effect in the determination of the optimal tax rate. Finally, notice that the behavioral effect of a given tax rate, either τ^{u*}_{TOT} or τ^{d*}_{TOT}, acts to decrease the optimal tax rates; an effect that could be amplified by the behavioral cross-effects. The following proposition summarizes our findings regarding the VAT and the TOT systems.

Proposition 2. *The optimal value-added (VAT) tax rate τ^{*}_{VAT} is characterized by equation (29). And, the optimal tax instruments of a turnover tax system (TOT), i.e., τ^{u*}_{TOT} and τ^{d*}_{TOT}, are characterized by the equation (30).*

Proof. See Appendix A.3. ■

[15] Our model revenue efficiency effect is similar to Best et al. (2015)'s, equation (5), pp. 1318.

With a general solution for a multistage optimal taxes, we show that any simpler optimal tax systems can also be characterized using the correspondent reported sales and costs responses to tax changes.[16] Next, we conduct quantitative exercises to evaluate numerically the optimal tax instruments exploring the effects of misreporting costs on the optimal tax rates and tax systems.

4 A Numerical Exercise

4.1 Functional Forms and Parameter Values

In this section, we present a numerical analysis of our limited tax capacity economy. Since we cannot analytically find the equilibrium solution, we rely on several numerical exercises to illustrate the main mechanisms of our economy, the optimal multistage tax system and how changes in key parameters of the model affect the sales and purchases of the upstream and a downstream firms and the consumer's welfare. For our numerical exercises, we choose the functional forms and parameter values in line with the standard literature.

We assume that functional forms for the utility function, the production functions of a downstream firm and the upstream firms are as follows $u\left(C, z\right) = \gamma \log(C) + (1-\gamma)\log\left(1 - l_u - l_d\right)$, $y_d = F\left(y_u, l_d\right) = y_u^{\alpha} l_d^{1-\alpha}$ and $y_u = F\left(l_u\right) = l_u$, respectively. The consumer is endowed with one unity of time and leisure $z = 1 - l_u - l_d$. The parameter γ is the relative weight on the utility of consumption, α is the intermediate (upstream) good share of the downstream production function, and $(1-\alpha)$ is the downstream labor share. We set γ such that the equilibrium condition from the consumer utility maximization problem is satisfied, i.e., $l_u + l_d = \gamma$, and the consumer (optimally) spends one third of her time working. We set $\alpha = 0.30$ such that the labor share in the downstream firm is about two-thirds and the upstream input share is one-third of the downstream output.

Misreporting cost functions are assumed to be convex functions of the misreported gaps. In our numerical exercises, we assume the following functional forms: $H\left(y_d - \overline{y_d}\right) = \left(y_u^{\alpha} l_d^{1-\alpha} - \overline{y_d}\right)^{\theta_H}$,

[16] It is possible (i) to show that the VAT $\left(\tau_{VAT}^{u*}, \tau_{VAT}^{d*}, 1\right)$ and the RST $\left(0, \tau_{RST}^{d*}, 0\right)$ systems are equivalent in terms of the economy's welfare in an unlimited tax capacity economy and (ii) to characterize the optimal RST instrument with limited tax capacity. The government's inability to collect taxes and enforce a tax system breaks the VAT-RST equivalence result. In the absence of audit instruments, not only downstream firms can misreport their input purchases but upstream firms can misreport their output sales. To restore the equivalence between these two systems, the government would have, for instance, to subsidize downstream purchases and increase the tax on upstream sales. Results are available upon request.

$J\left(\widehat{y_u}-y_u\right)=\left(\widehat{y_u}-y_u\right)^{\theta_J}$ and $J\left(y_u-\overline{y_u}\right)=\left(l_u-\overline{y_u}\right)^{\theta_J}$. As a benchmark, we assume that these misreporting functions have the same curvature, i.e., we set $\theta_D = \theta_{\widehat{U}} = \theta_U = 1.50$. We will conduct several exercises to explore the differences between the misreporting cost parameters to understand their implications on firm's behavior and the potential heterogeneous responses across the production chain, i.e., the upstream versus a downstream misreporting behavior, as well as within the downstream production stage, i.e., a downstream misreported sale of the final good versus a downstream misreported purchase of the intermediate good.[17] Finally, the exogenous revenue requirement R is set to 0.10, such that it is about $20-25$ percent of the downstream output in an economy with unlimited tax capacity. Our main results are robust to reasonable variations around this benchmark set of parameters.

4.2 VAT and TOT Systems: Optimal Policies and Welfare

Our quantitative analysis starts with the two tax systems considered in the *SIMPLES* tax reform. In a VAT system, both intermediate and final goods sales are taxed at the same rate ($\tau^u = \tau^d = \tau_{VAT}$) and a downstream firm can claim the full amount of taxes remitted by the upstream firm as part of its production costs ($\delta_{VAT} = 1$). On the other hand, a TOT system $\mathcal{T}_{\mathcal{TOT}} = \left(\tau^u_{TOT}, \tau^d_{TOT}, 0\right)$ is such that upstream and downstream firms are taxed at different tax rates, but taxes remitted by the upstream firm can not be claimed as production costs by a downstream firm ($\delta_{TOT} = 0$).

Table V presents our results for both tax systems in a limited limited tax capacity economy For comparison, we also show the optimal policies under the assumption of unlimited tax capacity. Recall that the key distinction between these two cases is the firms' ability to misreport their sales and purchases; or, in other words, the government's ability to collect the taxes on the actual amounts produced and sold. First, notice that in an unlimited tax capacity economy downstream firms are taxed at the same rate either in a VAT or in a TOT system. That is, $\tau^*_{VAT} = \tau^{d*}_{TOT} = 0.23$. The key difference between these two systems comes down to the taxation of the upstream firm and the refund rate. In a VAT system, the upstream firm is taxed at same rate as a downstream

[17] In our benchmark numerical exercises, we ignore the case that upstream and downstream firms cooperate to misreport sales and purchases of the upstream good, splitting the costs to misreport their transactions to the government. That is, we assume they have the same misreport cost functional forms.

firm (i.e., $\tau^*_{VAT} = 0.23$), but the later can claim one-hundred percent of it as production costs in its profit maximization problem (i.e., $\delta *_{VAT} = 1$). Second, by design the refund rate is zero in a TOT system and, hence, it is optimal not to tax the upstream firm (i.e., $\tau^{u*}_{TOT} = 0.0$). And, third, despite the differences in the policy instruments, the optimal VAT and TOT are equivalent in terms of the economy's welfare in an unlimited tax capacity economy.

The government limited tax capacity leads to higher optimal VAT taxes on both the upstream firm and a downstream firm, i.e., $(\tau^*_{VAT}, \tau^*_{VAT}, \delta^*_{VAT}) = (0.27, 0.27, 1.00)$, compared to $(0.23, 0.23, 1.00)$ in an unlimited tax capacity environment (Table V). Prices of the intermediate and the final goods are higher and firms misreport in all margins. That is, we observe misreporting of intermediate good sales and purchases by the upstream and downstream, respectively, and misreporting of sales of the final good (downstream). Because by design $\delta^*_{VAT} = 1$, although the upstream firm is taxed, the VAT system corrects for distortions such taxation may cause in the downstream optimal choice of production inputs - labor versus the intermediate good. The marginal rate of technical substitution imply that $F_{l_d}(\cdot) = F_{y_u}(\cdot)$ and consumers are worse off vis-à-vis the unlimited tax capacity case as they experience a reduction in their leisure time due to more hours worked in the upstream firm.

A TOT system in a limited tax capacity economy is such that it is now optimal to tax the upstream firm (Table V). The optimal tax rate on the intermediate good is about fourteen percent (i.e., $\tau^{u*}_{TOT} = 0.136$). On the other hand, the tax rate on the final (downstream) good is only slightly lower. Hence, the optimal TOT system is expressed as follows $\left(\tau^{u*}_{TOT}, \tau^{d*}_{TOT}, \delta^*_{TOT}\right) = (0.136, 0.228, 0.00)$. Notice that, in this case, the marginal rate of technical substitution $MRTS^{TOT}_{l^d, y^u} = F_{l_d}(\cdot)/F_{y_u}(\cdot) = (1 - \tau^{u*}_{TOT}) = 0.864$ is distorted when compared to a TOT system in an unlimited tax capacity economy or a VAT system ($MRTS^{VAT}_{l^d, y^u} = F_{l_d}(\cdot)/F_{y_u}(\cdot) = 1$). This result illustrates the fact that the use of the intermediate good as an input in the final good production is taxed twice in a TOT system - first directly (higher rate) at the upstream stage of production and again, albeit indirectly, at the final production stage given that a downstream firm is not allowed to claim taxes already collected in the production chain.

Our numerical results reveals that a TOT system implies a higher consumer's welfare than a

Table V: VAT and TOT: Optimal Policies, Output, Prices and Welfare

Policies	Output					Prices		Welfare
(I)	(II)		(III)			(IV)		(V)
	Produced		Reported					
$(\tau^{u*}, \tau^{d*}, \delta^*)$	y_d^*	y_u^*	$\overline{y_d}^*$	$\widehat{y_u}^*$	$\overline{y_u}^*$	p_d^*	p_u^*	U^*
Value-Added Tax System (VAT)								
Unlimited								
$(0.23, 0.23, 1.00)$	0.179	0.099	–	–	–	2.400	1.303	-0.923
Limited								
$(0.27, 0.27, 1.00)$	0.179	0.102	0.147	0.133	0.069	3.105	1.315	-0.934
Turnover Tax System (TOT)								
Unlimited								
$(0.00, 0.23, 0.00)$	0.179	0.099	–	–	–	2.400	1.000	-0.923
Limited								
$(0.136, 0.228, 0.00)$	0.178	0.089	0.156	–	0.081	2.458	1.152	-0.931

Benchmark: $R = 0.10$, $\gamma = 0.33$, $\alpha = 0.30$; $\theta_D = \theta_{\widehat{U}} = \theta_U = 1.50$.

VAT system in a limited tax capacity economy. The optimality of a TOT system compared to a VAT system can be attributed to two main reasons. First, because a downstream firm can not claim the tax paid by the upstream as part of its production cost, the design of a TOT system reduces the incentive a downstream firm has to overreport the purchases of the intermediate good. Second, and more relevant, a TOT system allows for differential tax rates along the production chain. In this particular example, despite both (upstream and downstream) firms having similar misreporting marginal costs, it is still optimal to tax the upstream firm at a lower rate in order to reduce the inefficiency associated with the cascading effect.[18]

Interestingly, our numerical exercise shows that going from an unlimited to a limited tax capacity economy, it is optimal to tax upstream firms at a lower tax rate than downstream firms even when upstream and downstream firms face the same misreporting costs. In our theoretical model, the behavioral effects of policy changes capture this efficiency motive through the elasticities

[18] Firms are taxed at higher rates if we restrict the optimal TOT system to tax upstream and downstream firms at the same rate and the refund rate to be equal to zero, i.e. $(\tau^*_{TOT}, \tau^*_{TOT}, \delta^*_{TOT}) = (0.29, 0.29, 0.00)$. That would entail a smaller aggregated welfare compared to VAT.

of the misreported sales gaps to the policy instruments, i.e., the *ad valorem* taxes on upstream and downstream reported sales. This result can be viewed as a reinterpretation of Diamond and Mirrlees (1971a) when they "conclude that we want efficiency for these private production possibilities taken together." This can be highlighted in our limited tax capacity environment. Going from an unlimited to a limited tax capacity economy, the tax rate on the downstream reported sales remains the same ($\tau^{d*}_{TOT} = 0.23$), while upstream firms are taxed at a substantially larger tax rate - it raises from from $\tau^{u*}_{TOT} = 0$ in the unlimited case to $\tau^{u*}_{TOT} = 0.136$ in the limited tax capacity economy.

Our empirical findings for the informal firms in Brazil (Section 2) suggest that upstream firms are as responsive to tax changes as downstream firms, which can be associated to similar misreporting costs. If the estimated effects of upstream reported sales and non-labor costs were significant, that would imply a smaller (in magnitude) response. We conduct an additional exercise retaking into consideration that it is less costly for the upstream firm than a downstream firm to misreport its good sales, as suggested by our empirical results, for instance, if $\theta_U = 1.50$ and $\theta_D = 1.30$, respectively (instead of our benchmark values $\theta_D = \theta_U = 1.50$), the optimal TOT system is now as follows $\left(\tau^{u*}_{TOT}, \tau^{d*}_{TOT}, \delta^{*}_{TOT}\right) = (0.02, 0.227, 0.00)$. This means that it is optimal to reduce the upstream firm taxation, while keeping the tax rate on the final (downstream) good at the same level. In fact, the total reduction in the upstream tax rate from an unlimited tax capacity environment can be decomposed into (i) a revenue efficiency argument associated with chain effects of upstream firm's taxation with optimal choice of inputs by downstream firms, from $\tau^{u*}_{TOT} = 0.23$ to $\tau^{u*}_{TOT} = 0.136$, and (ii) a relative higher cost for upstream firms to misreport sales and expenses, equation (13) vis-à-vis equation (9), from $\tau^{u*}_{TOT} = 0.136$ to $\tau^{u*}_{TOT} = 0.02$.

4.3 Optimal Multistage Tax Policies and Welfare

In this section, we study an optimal multistage tax (MST) system. The three policy instruments τ^u, τ^d, and δ of a MST system are not constrained by designed as either in a VAT or TOT systems, respectively $\mathcal{T}_{\mathcal{VAT}} = (\tau_{VAT}, \tau_{VAT}, 1)$ and $\mathcal{T}_{\mathcal{TOT}} = \left(\tau^u_{TOT}, \tau^d_{TOT}, 0\right)$. In a MST system $\mathcal{T}_{\mathcal{MST}} = \left(\tau^u, \tau^d, \delta\right)$, the government can optimally choose any (interior) combination of the upstream and downstream *ad valorem* tax rates (i.e., τ^{u*} and τ^{d*}, respectively), and the refund rate δ^*. In an unlimited tax capacity economy, the optimal MST system is equivalent to a particular TOT system, where only the sales of the final good produced by a downstream firm is taxed at a 23 percent tax rate (Table VI). The sales of the intermediate (upstream) good are not taxed and the refund rate is zero. That is, in an unlimited tax capacity economy only one policy instrument should be optimally used by the tax authority. Hence, in an unlimited tax capacity economy the optimal MST system is $\left(\tau^{u*}, \tau^{d*}, \delta^*\right) = (0.00, 0.23, 0.00)$.

In a limited tax capacity economy the government's inability to fully tax firms leads it to optimally use all policy instruments at its disposal. In this case, the optimal multistage tax system includes a 12 percent tax rate on the intermediate good sales, a 24 percent *ad valorem* tax rate on the sales of the final (downstream) good, and a refund rate of 19 percent (i.e., a downstream firm can claim about one-fifth of the taxes remitted by the upstream firm as its production cost). The multistage tax system is, hence, $\mathcal{T} = \left(\tau^{u*}, \tau^{d*}, \delta^*\right) = (0.12, 0.24, 0.19)$, Table VI. In this case, the marginal rate of technical substitution is $MRTS^{MST}_{l^d, y^u} = F_{l_d}(\cdot)/F_{y_u}(\cdot) = (1 - \tau^{u*}_{MST}) \big/ (1 - \delta^*_{MST}\tau^{u*}_{MST}) = 0.90$, which implies that the optimal choice of inputs - labor versus the intermediate good - in the production of the final good is less (more) distorted when compared to an optimal TOT (VAT) system in a limited tax capacity economy.

Table VI: Multistage Tax System: Optimal Policies, Output, Prices and Welfare

Policies	Output					Prices		Welfare
(I)	(II)		(III)			(IV)		(V)
	Produced		Reported					
$\left(\tau^{u*}, \tau^{d*}, \delta^*\right)$	y_d^*	y_u^*	$\overline{y_d}^*$	$\widehat{y_u}^*$	$\overline{y_u}^*$	p_d^*	p_u^*	U^*
Unlimited								
$(0.00, 0.23, 0.00)$	0.179	0.099	–	–	–	2.402	1.000	−0.9231
Limited								
$(0.12, 0.24, 0.19)$	0.1789	0.0910	0.1541	0.0912	0.090	2.486	1.135	−0.9297

Benchmark: $R = 0.10$, $\gamma = 0.33$, $\alpha = 0.30$; $\theta_D = \theta_{\widehat{U}} = \theta_U = 1.50$.

In an unlimited tax capacity economy, the upstream firm and a downstream firm cannot misreport their purchases and sales (i.e., $\overline{y_d}^* = \widehat{y_u}^* = \overline{y_u}^* = 0$). This is no longer the case in a limited tax capacity economy. Under the optimal MST system, the upstream firm underreports its sales to the tax authority and a downstream firm misreports on its two margins. That is, a downstream firm overreports the purchase of the intermediate good and underreports the sales of the final goods. For instance, a downstream firm reports a final good production and sales to consumers equivalent to 86 percent of what it is actually produces and sells. The optimal MST imply higher prices of both the intermediate good and the final good. In particular, the price of the intermediate good (p_u^*) is 13 percent higher when firms can misreport sales and purchases. Consumers are worse off in a limited *vis-à-vis* in an unlimited tax capacity economy - they consume slightly less but experience a significant reduction in their leisure, mainly due to more hours worked in the production of the final (downstream) good. In sum, intermediate and final goods outputs are lower, good prices are higher and the consumer is worse off under a MST system in a limited tax capacity economy.

While the optimal VAT, TOT, and MST systems are equivalent with respect to the consumer's welfare in an unlimited tax capacity economy, a comparison of these three tax systems in a limited tax capacity economy reveals interesting differences. First, as one would expect, the consumer is better off under a MST system - the tax authority is not constrained in its policy instruments choice. Second, the intermediate (upstream) good is taxed at a lower rate under an optimal MST and the *ad valorem* tax rate on the final (downstream) good is the highest in an optimal VAT system. Third, while the final good production is pretty much the same across tax systems, we observe significant differences regarding the production of the intermediate good. In this order, the upstream firm produces more under a VAT system than in a MST or a TOT system. This can possibly be tied to a higher price of the intermediate good in the optimal VAT system, as well as the fact that under this tax system a downstream firm can claim one-hundred percent of taxes paid by the upstream firm. Fourth, the underreporting of the upstream production is the smallest (largest) under an optimal MST (VAT) system, while a downstream firm underreport more under a VAT system than under a MST system. And, finally, an optimal MST system reduces the input (labor vs intermediate good) distortions, expressed in terms the marginal rate of technical substitution ($MRTS$), that a TOT system creates. That is, $MRTS^{TOT}_{l^d,y^u} = 0.86 < MRTS^{MST}_{l^d,y^u} = 0.90 < MRTS^{VAT}_{l^d,y^u} = 1.00$.

4.4 A Multistage Tax System: Additional Exercises

We conduct a series of numerical exercises to better understand the implications of an optimal multistage tax system. In particular, we investigate how the optimal tax system in a limited tax capacity economy changes under different assumptions regarding key parameters of the model (Table A.6). Second, we investigate how an equal share of inputs in the downstream production technology would affect our results. Recall that the upstream production technology is linear and we set $\alpha = 0.50$ - that is, the intermediate (upstream) good and the labor have equal shares in a downstream firm production function. In this case, the optimal multistage tax system is such that the upstream (downstream) firm is taxed at a lower (higher) rate but a downstream firm can claim the full amount of taxes remitted by the upstream firm as part of its production costs.

Next, we analyze how changes in the parameters of the misreport cost functions affect the optimal multistage tax system. In a limited tax capacity economy where it is more costly to misreport sales and purchases, for instance, $\theta_D = \theta_{\widehat{U}} = \theta_U = 1.30$ (*vis-à-vis* the benchmark values $\theta_D = \theta_{\widehat{U}} = \theta_U = 1.50$), the upstream firm and a downstream firm produce and sell more intermediate and final goods, respectively. Prices fall and the firms misreport less in all three margins $\overline{y_d}$, $\overline{y_u}$, and $\widehat{y_u}$ (Table A.7). The optimal tax system is such that the upstream firm is taxed at a 6 percent tax rate, the downstream tax rate is 22 percent and a downstream firm can claim only 3 percent of the taxes remitted by the upstream firm.

To conclude this section, we conduct three additional exercises where we vary only one misreporting cost parameter θ at a time, while keeping the other two at their benchmark levels (Table A.7). Our numerical exercises reveal that higher marginal costs to misreport sales - either sales of the intermediate (upstream) good to a downstream firm or final (downstream) good sales to consumers - lead to welfare gains when optimal policies are implemented since this approaches to unlimited tax capacity environment. In specific term, this occurs mainly due to an increase in leisure that more than compensates a (small) reduction in consumption. On the other hand, consumers are worse off in a economy where it is more costly for a downstream firm to misreport its purchase of the intermediate good. In this case, a profit maximizing downstream firm reduces the amount of sales it misreports to the tax authority while truthfully reporting the amount of the intermediate good it purchases from the upstream firm. While the misreport intermediate good purchase gap is zero at the downstream production stage, the misreport sales gap is larger at the upstream firm - it goes from $(y_u^* - \overline{y_u}^*) = 0.001$ in the benchmark case to $(y_u^* - \overline{y_u}^*) = 0.04$ when it is costly for a downstream firm to misreport its intermediate good purchase. Overall, this result can be attributed to a higher taxation and a higher price of the intermediate (upstream) good.

4.5 Discussion

We interpret our empirical results in light of our benchmark model. Our empirical results of Section 2 suggest that upstream firms' reported margin responses to tax reductions are smaller, albeit not statistically significant, to those responses observed by downstream firms. Assuming the misreporting cost functions used in our numerical exercise and the equilibrium conditions of the upstream and downstream firms, we can illustrate how the reported margins elasticity of response to a tax change can be directly associated with a firm's misreporting cost. In fact, the larger the upstream firm misreporting cost *vis-à-vis* the misreporting cost of a downstream firm, the smaller it must be the elasticity of response of the former relative to the latter.

Consider, for instance, a downstream firm optimal choice of reported sales ($\overline{y_d}$), equation (9). Recall that this equation represents the trade-off regarding misreporting the final goods sales, and assuming $H\left(y_d - \overline{y_d}\right) = \left(y_d - \overline{y_d}\right)^{\theta_H}$, we have $\tau_d = \theta_H(y_d - \overline{y_d})^{\theta_H - 1}$. Total differentiating (the log of) this expression we obtain $\partial log(y_d - \overline{y_d})/\partial log(\tau_d) = 1/(\theta_H - 1)$, where the left-hand side is the model construct equivalent to our estimated elasticity. For downstream firms this elasticity is estimated at 4.3 (Section 2.5) for young firms, which implies a misreported cost parameter of approximately 1.23 (i.e., $\theta_H = 1.23$). Based on our empirical results, one would expect that the corresponding misreporting cost parameter of upstream firms to be larger[19]. In this regard, our numerical results presented in Table A.7, column (III), suggest a reasonable approximation for young firms in our sample, since it shows the opposiste, which is a larger misreporting cost for downstream firm, In that case it is optimal to tax downstream sales at a higher rate (to the benchmark case) and policy instruments related to upstream firm transactions, i.e., upstream *ad valorem* tax rate and tax

[19] Note that our numerical exercises have normalized the labor supply to one. This leads to a unique interpretation of the relation between larger misreporting costs with smaller Θ, different from our empirical exercises that positively associates misreporting cost size with Θ

rebate, should be very small. We should expect the opposite of that interpretation. This result can be, in part, attributed to a revenue efficiency argument as highlighted by the marginal rate of technical substitution between (downstream) labor (l^d) and the intermediate good (y^u). That is, $MRTS^{MST}_{l^d,y^u} = F_{l_d}(\cdot)/F_{y_u}(\cdot) = (1-\tau^u)/(1-\delta\tau^u)$, where we can clearly see the cascading effect of taxing the intermediate (upstream) good sales.

Regarding the estimated reported non-labor cost response, recall that we found an elasticity of 0.92, which following the same approach described above implies a misreporting cost parameter of approximately 2 (i.e., $\theta_H = 2.08$). Notice that the misreporting costs parameter associated with non-labor costs are implied to be bigger than those related to misreporting sales cost parameter. And, hence, the θ parameters used in our numerical exercises reported in Table A.7 is well in the range of those implied by our empirical analysis. For instance, the results presented in Table A.7, column (IV), suggest that it optimal to increase taxation of upstream sales and to reduce the tax rebate to downstream firms.

5 Conclusions

Understanding how firms, in particular small and potentially informal firms, behave and react to tax policies in a limited tax capacity environment is key to improve policy design, inform tax system choices and quantify their welfare implications. In this paper, we approach these issues from a positive and a normative perspective. First, we investigate a tax reform and a survey of informal firms (*ECINF*) in Brazil, a developing country with a large informal sector. For small upstream and downstream firms, the *SIMPLES* tax reform of 2003 represented a change from VAT system, in which both production stages were taxed at the same tax rate and downstream firms could claim the full amount of taxes remitted by upstream firms as part of their production costs, to a turnover tax (TOT) system - upstream and downstream firms can be taxed at different tax rates depending on their sizes and sector eligibility. Our empirical analysis reveals that the *State-SIMPLES* tax reform had a substantial effect on firms in different stages of production. Within the same sector, eligible upstream firms in states that implemented the *SIMPLES* tax reform in 2003 seem to have responded similarly than their downstream counterparts increasing reported costs. Young downstream firms show reduction in the reported sales, again similar to results we obtain for upstream firms.

We develop a theoretical model with upstream and downstream firms, consumers and a limited tax capacity government. The government's goal is to maximize the economy's welfare by choosing *ad valorem* taxes on upstream and downstream sales and refund rate, subject to an exogenous revenue requirement. We characterize the optimal tax instruments under alternative tax systems. In particular, and in line with the SIMPLES tax reform, we characterize the optimal value-added tax and turnover tax system. The main contribution of our model is to highlight the role of each misreporting margin of firms along the production chain - upstream (intermediate goods) sales, downstream (intermediate goods; inputs) purchases and downstream (final goods) sales, on the optimal choice of tax instruments. These effects are expressed in the form of elasticity of the misreported sales and purchase gaps to the policy instruments and they are summarized by behavioral and mechanical effects, a reinterpretation of Diamond and Mirrlees (1971a) findings.

We also present several quantitative exercises in order to further our understanding regarding the optimal tax design under limited tax capacity. Our numerical evaluation reinforces the role of misreporting costs in levelling the sufficient statistics used to derive the optimal taxes on firms under limited tax capacity.

References

Besley, T. and S. Coate (1991): "Public Provision of Private Goods and the Redistribution of Income," *The American Economic Review*, 81, 979–984.

Besley, T. and T. Persson (2013): "Taxation and Development," CEPR Discussion Papers 9307, C.E.P.R. Discussion Papers.

Best, M. C., A. Brockmeyer, H. J. Kleven, J. Spinnewijn, and M. Waseem (2015): "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan," *Journal of Political Economy*, 123, 1311–1355.

Carrillo, P., D. Pomeranz, and M. Singhal (2017): "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," *American Economic Journal: Applied Economics*, 9, 144–64.

Chetty, R. (2009): "Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance," *American Economic Journal: Economic Policy*, 1, 31–52.

Cuff, K., S. Mongrain, and J. Roberts (2017): "Shades of Grey: Business Compliance with Fiscal and Labour Regulations," Discussion papers, Department of Economics, Simon Fraser University.

de Paula, A. and J. A. Scheinkman (2010): "Value-Added Taxes, Chain Effects, and Informality," *American Economic Journal: Macroeconomics*, 2, 195–221.

Dharmapala, D., J. Slemrod, and J. D. Wilson (2011): "Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs," *Journal of Public Economics*, 95, 1036–1047.

Diamond, P. A. and J. A. Mirrlees (1971a): "Optimal Taxation and Public Production I: Production Efficiency," *The American Economic Review*, 61, 8–27.

——— (1971b): "Optimal Taxation and Public Production II: Tax Rules," *The American Economic Review*, 61, 261–278.

Emran, M. S. and J. Stiglitz (2005): "On selective indirect tax reform in developing countries," *Journal of Public Economics*, 89, 599–623.

Fajnzylber, P., W. F. Maloney, and G. V. Montes-Rojas (2011): "Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program," *Journal of Development Economics*, 94, 262 – 276.

Gadenne, L., T. Nandi, and R. Rathelot (2018): "Taxation and Supplier Networks: Evidence from India," Discussion papers, University of Warwick.

Gordon, R. and W. Li (2009): "Tax structures in developing countries: Many puzzles and a possible explanation," *Journal of Public Economics*, 93, 855–866.

Jales, H. (2018): "Estimating the effects of the minimum wage in a developing country: A density discontinuity design approach," *Journal of Applied Econometrics*, 33, 29–51.

Keen, M. (2008): "VAT, tariffs, and withholding: Border taxes and informality in developing countries," *Journal of Public Economics*, 92, 1892–1906.

Keen, M. and J. Mintz (2004): "The optimal threshold for a value-added tax," *Journal of Public Economics*, 88, 559–576.

Kopczuk, W. and J. Slemrod (2006): "Putting Firms into Optimal Tax Theory," *The American Economic Review*, 96, 130–134.

Monteiro, J. C. and J. J. Assuncao (2012): "Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises," *Journal of Development Economics*, 99, 105–115.

Piza, C. (2018): "Out of the Shadows? Revisiting the impact of the Brazilian SIMPLES program on firms' formalization rates," *Journal of Development Economics*, 134, 125 – 132.

Pomeranz, D. (2015): "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, 105, 2539–69.

Slemrod, J. (2001): "A General Model of the Behavioral Response to Taxation," *International Tax and Public Finance*, 8, 119–128.

Ulyssea, G. (2018): "Firms, Informality, and Development: Theory and Evidence from Brazil," *American Economic Review*, 108, 2015–47.

Yitzhaki, S. (1974): "Income tax evasion: A theoretical analysis," *Journal of Public Economics*, 3, 201–202.

A.1 Appendix

A.1 *SIMPLES* Tax Reform: Additional Details

Table A.1: The *SIMPLES* Tax Reform: Eligible Economics Activities

	Economic Activities
Retail and Trade	Vegetable products, Beverage, Meat and food, Garment and accessories, Decoration articles. Books and magazines, Construction material, home appliance, machines and electric supplies, transport equipment, pharmaceutics and chemistry, oil goods, supermarkets, leisure articles.
Construction	Industry and work related
Manufacturing	ware, construction material and ceramics, metallic instruments, wood objects bamboo, wicker and agave manufacturing , furniture, paper goods, rubber goods manufacturing , leather goods manufacturing, plastic goods manufacturing , textile goods manufacturing, garment manufacturing, shoes manufacturing, food manufacturing, printing and editing, medical material and hygiene products manufacturing.
Transportation	cargo and people transportation, charter freight, maritime freight, air transportation.
Other Services	loading, food, furniture restoration, plumber and electrics, sewing, clothes rental, lm laboratory, laundering, pressing and dying, gardening, entertainment TV installation, articles renting, tourism.

Source: Monteiro and Assuncao (2012)

Table A.2: The *SIMPLES* Tax Reform: Participating States and Dates

States	Data	States	Data
Maranhão	03/24/1997	Santa Catarina	05/08/2000
Espírito Santo	04/24/1997	Acre	07/19/2000
Pernambuco	12/29/1997	Alagoas	10/03/2001
Goiás	06/04/1998		
Amapá	06/17/1998	Pará	01/07/2003
Mato Grosso do Sul	07/08/1998	Rio Grande do Norte	01/28/2003
Bahia	11/04/1998	Paraná	01/29/2003
São Paulo	11/19/1998	Ceará	04/02/2003
Sergipe	12/22/1999	Paraíba	04/28/2003
Rio de Janeiro	12/29/1999		
Distrito Federal	12/29/1999		
Rondônia	12/30/1999		

Source: de Paula and Scheinkman (2010).

Table A.3: Descriptive Statistics.

	All Firms - ECINF 2003(1)			2003 and No Tax Reform(2)			
Variable	# Obs.	Mean	Std. Dev.	# Obs.	Mean	Std. Dev.	Diff.(3)
Firms Characteristics							
Eligible Firms	52,293	0.90	0.31	21979	0.93	0.25	-0.04
State-SIMPLES reform	23,104	0.45	0.50	23090	0.39	0.49	0.06
In business(4)	54,841	105.21	106.23	23064	111.22	111.99	-6.01
Hours/week	54,814	44.36	22.04	23,062	43.55	21.83	0.81
# workers	54,905	1.71	1.52	23,087	1.52	1.12	0.19
Share middle school	54,905	0.13	0.30	23,087	0.14	0.32	-0.01
Share high school	54,905	0.23	0.38	23,087	0.20	0.37	0.02
Share college degree	54,905	0.08	0.25	23,087	0.06	0.22	0.02
Tax Registry(5)	54,871	0.17	0.38	23,087	0.17	0.37	0.00
Market Power(6)	52,164	0.59	0.49	21,941	0.61	0.49	-0.01
Outside Household(7)	54,905	0.65	0.48	23,087	0.66	0.47	-0.01
Accountant(8)	54,905	0.16	0.36	23,087	0.16	0.37	0.00
Reported Margins							
Sales ($\overline{y}$)	53,974	2617.63	7874.93	22,913	2186.64	6123.94	430.99
Total Costs	51,422	1739.40	6884.85	21,300	1612.45	5990.63	126.95
Non-labor Costs ($\widehat{y}$)	54,905	1254.799	6042.686	23,087	1138.87	5315.94	115.93
Individual Characteristics							
Age	54,905	40.72	12.34	23,087	41.34	12.45	-0.62
Tenure(9)	53,608	25.90	13.70	22,480	27.17	13.79	-1.27
Education							
Less than Middle School	54,905	0.42	0.49	23,087	0.47	0.50	-0.05
Middle School	54,905	0.12	0.33	23,087	0.13	0.34	-0.01
High School (incomplete)	54,905	0.09	0.28	23,087	0.09	0.28	0.00
High School	54,905	0.23	0.42	23,087	0.20	0.40	0.02
College (incomplete)	54,905	0.04	0.20	23,087	0.03	0.18	0.01
College	54,905	0.10	0.30	23,087	0.07	0.26	0.03
Economic Sectors							
Agriculture	54,905	0.11	0.31	23,087	0.12	0.33	-0.02
Manufacturing	54,905	0.14	0.34	23,087	0.19	0.39	-0.05
Construction	54,905	0.30	0.46	23,087	0.37	0.48	-0.07
Retail	54,905	0.09	0.28	23,087	0.06	0.23	0.03
Housing	54,905	0.09	0.28	23,087	0.08	0.28	0.01
Transportation	54,905	0.07	0.25	23,087	0.05	0.21	0.02
Real State	54,905	0.06	0.25	23,087	0.03	0.18	0.03
Education, Health	54,905	0.10	0.30	23,087	0.08	0.27	0.02

Note: (1) All firms with non-missing data in at least one of the reported margins; (2) Firms located in either states that adopted *State-SIMPLES* in 2003 or in states without a state tax reform; (3) Reports the difference of the mean of each variable for the two group of firms;(4) Time in business (months); (5) Dummy if firms has tax registry (CNPJ); (6) Dummy if the firm can negotiate its sales price; (7) Dummy if the firm is located outside household; (8) Dummy if the firm hires an accountant; (9) Time that entrepreneur works in the same field (months).

A.2 Limited Tax Capacity Economy - Theoretical Results

A.2.1 Government's Welfare Maximization Problem

The government's goal is to maximize the economy's welfare by choosing tax rates and refund rate. Hence, the government's welfare objective function can be written as

$$\begin{aligned} W &= v\left(p_d\left(\mathcal{T}\right)\right) \\ &+ \lambda\left\{\tau^d p_d\left(\mathcal{T}\right)\overline{y_d}\left(p_d\left(\mathcal{T}\right), p_u\left(\tau^u\right), \mathcal{T}\right)\right. \end{aligned} \quad (31)$$

Table A.4: Balancing Tests for eligible firms

Variables	*State-SIMPLES* Reform					
	Yes$^{(1)}$	No$^{(2)}$	Std. Dev.	p-value$^{(3)}$	Yes$^{(1)}$ # obs.	No$^{(2)}$ # obs.
Firms Characteristics						
In business$^{(4)}$	103.28	106.58	1.54	0.03	10793	8998
Hours/week	44.43	44.17	0.32	0.40	10785	8994
# of workers	1.70	1.73	0.03	0.21	10802	9014
Share middle school	0.14	0.12	0.00	0.000***	10802	9014
Share high school	0.22	0.23	0.01	0.33	10802	9014
Share college degree	0.04	0.04	0.00	0.18	10802	9014
Tax Registry$^{(5)}$	0.18	0.15	0.01	0.000***	10797	9006
Market Power$^{(6)}$	0.62	0.61	0.01	0.15	10355	8587
Outside household$^{(7)}$	0.63	0.61	0.01	0.000***	10802	9014
Accountant$^{(8)}$	0.17	0.14	0.01	0.000***	10802	9014
Reported Margins						
Sales ($\overline{y}$)	2317.81	2456.40	118.76	0.24	10710	8957
Costs	1784.98	1786.02	110.94	0.99	10099	8176
Non-labor Costs ($\widehat{y}$)						
Individual (Entrepreneur) Characteristics						
Age	40.61	40.61	0.18	0.99	10802	9014
Tenure$^{(9)}$	26.11	26.34	0.20	0.25	10538	8726
Education						
Less than Middle School	0.46	0.47	0.01	0.063*	10802	9014
Middle School	0.14	0.12	0.01	0.001***	10802	9014
High School (incomplete)	0.10	0.09	0.00	0.12	10802	9014
High School	0.22	0.23	0.01	0.17	10802	9014
College (incomplete)	0.04	0.04	0.00	0.22	10802	9014
College	0.05	0.05	0.00	0.14	10802	9014
Economic Sectors						
Agriculture	0.12	0.13	0.01	0.03	10802	9014
Manufacturing	0.17	0.14	0.01	0.000***	10802	9014
Construction	0.32	0.36	0.01	0.000***	10802	9014
Retail	0.10	0.10	0.00	0.50	10802	9014
Housing	0.07	0.06	0.00	0.012**	10802	9014
Transportation	0.04	0.04	0.00	0.041**	10802	9014
Real State	0.04	0.04	0.00	0.39	10802	9014
Education, Health	0.10	0.10	0.00	0.84	10802	9014

Note: (1) Eligible firms located in states with State-SIMPLES; (2) Eligible firms located in states without a state-tax reform; (3) Reports the difference of the mean of each variable for the two group of firms;(4) Time in business (months); (5) Dummy if firms has tax registry (CNPJ); (6) Dummy if the firm can negotiate its sales price; (7) Dummy if the firm is located outside household; (8) Dummy if the firm hires an accountant; (9) Time that entrepreneur works in the same field (years). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

And, to emphasize the effects of the optimal policies on the misreported purchases and sales of the intermediate good and the sales of the final good, as well as on the misreported gaps (i.e., the difference between actual (y_d, y_u) and misreported $(\overline{y_d}, \overline{y_u}, \widehat{y_u})$ amounts), we rewrite the welfare

Table A.5: Robustness Checks

A. Control group: States previously treated

		Reported		
	Tax registry	Sales	Costs	Non-labor Costs
	(I)	(II)	(III)	(IV)
Eligible Firms	0.0106	302.4	1099.4**	950.1**
	(0.0242)	(193.0)	(389.8)	(331.2)
State-SIMPLES	-0.0630***	-365.7	-54.78	88.92
	(0.0291)	(261.6)	(244.9)	(206.5)
Elilgible*Reform	0.0483	132.9	-68.45	-95.68
	(0.0302)	(254.7)	(237.8)	(171.1)
Controls	Yes	Yes	Yes	Yes
# observations	34214	33605	31909	34215

B. Control group: States randomly treated assigned

		Reported		
	Tax registry	Sales	Costs	Non-labor Costs
	(V)	(VI)	(VII)	(VIII)
Eligible Firms	0.0617**	558.6	1237.5*	1110.2*
	(0.0243)	(378.2)	(270.8)	(258.8)
State-SIMPLES reform	0.025	-65.89	-34.56	161.6
	(0.0540)	(351.3)	(232.0)	(129.9)
Elilgible*Reform	-0.0298	44.78	-33.88	-198.4
	(0.0581)	(351.9)	(227.8)	(128.2)
Controls	Yes	Yes	Yes	Yes
# observations	20313	20155	18748	20313

Note: (1) Firms located in states with State-SIMPLES; (2) Firms located in states without a state-tax reform; (3)...... *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

function, equation (31), as follows

$$\begin{aligned} W(\mathcal{T}) &= v(p_d(\mathcal{T})) \\ &+ \lambda\left\{\tau^d p_d(\mathcal{T})\left[y_d - \Delta\overline{y_d}\left(p_d(\mathcal{T}), p_u(\tau^u), \mathcal{T}\right)\right]\right. \\ &+ \tau^u p_u(\tau^u)\left[y_u - \Delta\overline{y_u}\left(p_d(\mathcal{T}), p_u(\tau^u), \tau^u\right)\right] \\ &- \left.\delta\tau^u p_u(\tau^u)\left[y_u - \Delta\widehat{y_u}\left(p_d(\mathcal{T}), p_u(\tau^u), \tau^u, \delta\right)\right] - R\right\} \end{aligned} \tag{32}$$

where $\Delta\overline{y_d}\left(p_d(\mathcal{T}), p_u(\tau^u), \mathcal{T}\right) = y_d - \overline{y_d}\left(p_d(\mathcal{T}), p_u(\tau^u), \mathcal{T}\right)$, $\Delta\overline{y_u}\left(p_d(\mathcal{T}), p_u(\tau^u), \tau^u\right) = y_u - \overline{y_u}\left(p_d(\mathcal{T}), p_u(\tau^u), \tau^u\right)$ and $\Delta\widehat{y_u}\left(p_d(\mathcal{T}), p_u(\tau^u), \tau^u, \delta\right) = y_u - \widehat{y_u}\left(p_d(\mathcal{T}), p_u(\tau^u), \tau^u, \delta\right)$.

The total welfare effects of changing τ^u, τ^d and δ can be written as, respectively

$$\begin{aligned}
\left(\frac{\partial W/\partial\delta}{\lambda}\right)\tau^u &= M_{\tau^u} - \tau^d p_d\left(\mathcal{T}\right)\Delta\overline{y_d}\left(\varepsilon_{\Delta\overline{y_d},\tau^u}\right)\\
&- \tau^u p_u\left(\tau^u\right)\left(\Delta\overline{y_u}\right)\left(\varepsilon_{\Delta\overline{y_u},\tau^u}\right) + \delta\tau^u p_u\left(\tau^u\right)\left(\Delta\widehat{y_u}\right)\left(\varepsilon_{\Delta\widehat{y_u},\tau^u}\right)\\
\left(\frac{\partial W/\partial\delta}{\lambda}\right)\tau^d &= M_{\tau^d} - \tau^d p_d\left(\mathcal{T}\right)\Delta\overline{y_d}\left(\varepsilon_{\Delta\overline{y_d},\tau^d}\right)\\
&- \tau^u p_u\left(\tau^u\right)\left(\Delta\overline{y_u}\right)\left(\varepsilon_{\Delta\overline{y_u},\tau^d}\right) + \delta\tau^u p_u\left(\tau^u\right)\left(\Delta\widehat{y_u}\right)\left(\varepsilon_{\Delta\widehat{y_u},\tau^d}\right)\\
\left(\frac{\partial W/\partial\delta}{\lambda}\right)\delta &= M_{\delta} - \tau^d p_d\left(\mathcal{T}\right)\Delta\overline{y_d}\left(\varepsilon_{\Delta\overline{y_d},\delta}\right)\\
&- \tau^u p_u\left(\tau^u\right)\left(\Delta\overline{y_u}\right)\left(\varepsilon_{\Delta\overline{y_u},\delta}\right) + \delta\tau^u p_u\left(\tau^u\right)\left(\Delta\widehat{y_u}\right)\left(\varepsilon_{\Delta\widehat{y_u},\delta}\right) \qquad (33)
\end{aligned}$$

where, for instance, the elasticity of downstream sales reported gap ($\Delta\overline{y_d}$), downstream purchases reported gap ($\Delta\widehat{y_u}$) and upstream sales reported gap ($\Delta\overline{y_u}$) with respect to the tax policies τ^u, τ^d and δ can be written as, respectively:

$$\begin{aligned}
\varepsilon_{\Delta\overline{y_d},\tau^u} &= \left(\tfrac{\partial\Delta\overline{y_d}}{\partial\tau^u}\right)\left(\tfrac{\tau^u}{\Delta\overline{y_d}}\right) & \varepsilon_{\Delta\widehat{y_u},\tau^u} &= \left(\tfrac{\partial\Delta\widehat{y_u}}{\partial\tau^u}\right)\left(\tfrac{\tau^u}{\Delta\widehat{y_u}}\right) & \varepsilon_{\Delta\overline{y_u},\tau^u} &= \left(\tfrac{\partial\Delta\overline{y_u}}{\partial\tau^u}\right)\left(\tfrac{\tau^u}{\Delta\overline{y_u}}\right)\\
\varepsilon_{\Delta\overline{y_d},\tau^d} &= \left(\tfrac{\partial\Delta\overline{y_d}}{\partial\tau^d}\right)\left(\tfrac{\tau^d}{\Delta\overline{y_d}}\right) & \varepsilon_{\Delta\widehat{y_u},\tau^d} &= \left(\tfrac{\partial\Delta\widehat{y_u}}{\partial\tau^d}\right)\left(\tfrac{\tau^d}{\Delta\widehat{y_u}}\right) & \varepsilon_{\Delta\overline{y_u},\tau^d} &= \left(\tfrac{\partial\Delta\overline{y_u}}{\partial\tau^d}\right)\left(\tfrac{\tau^d}{\Delta\overline{y_u}}\right)\\
\varepsilon_{\Delta\overline{y_d},\delta} &= \left(\tfrac{\partial\Delta\overline{y_d}}{\partial\delta}\right)\left(\tfrac{\delta}{\Delta\overline{y_d}}\right) & \varepsilon_{\Delta\widehat{y_u},\delta} &= \left(\tfrac{\partial\Delta\widehat{y_u}}{\partial\delta}\right)\left(\tfrac{\delta}{\Delta\widehat{y_u}}\right) & \varepsilon_{\Delta\overline{y_u},\delta} &= \left(\tfrac{\partial\Delta\overline{y_u}}{\partial\delta}\right)\left(\tfrac{\delta}{\Delta\overline{y_u}}\right)
\end{aligned}$$

A.2.2 Unlimited Tax Capacity: The Optimal Multistage Tax System

To gain intuition about the government's optimal tax policy, consider first an unlimited tax capacity economy. In this case, the government is able to prevent that the upstream and the downstream firms misreport their sales and purchases. The mechanisms and measures though which the government can achieve this result (e.g., auditing, tax administration, among others(is beyond the scope of this paper. What is important for the discussion that follows is the fact that both the upstream firm and a downstream firm report to the tax authority the actual (true) amount produced and sold.

In order to satisfy the revenue requirement R, the government has to take into account the effect of changes in the taxation (τ^u) of the upstream firm output, which is an input of production of a downstream firm, *vis-à-vis* the effects of changes of the policy instruments that directly affect the sales of the final good and the cost of production of a downstream firm, i.e, τ^d and δ, respectively. This be achieved by changes in the policy instruments such that the price of the final goods p_d (and, hence the consumer's welfare) and the government's tax revenue (net of input responses) both remain constant. In other words, the optimal policy trade-off is such that the marginal rate of substitution between τ^u and τ^d, keeping prices constant (i.e., $\left(\left(\partial p_d/\partial\tau^u\right)/\left(\partial p_d/\partial\tau^d\right)\right)p_d y_d$, must be equal to $(1-\delta)\left[\tau^u y_u\left(\partial p_u/\partial\tau^u\right) + p_u y_u + \tau^u p_u\left(\left(\partial y_u/\partial\tau^u\right) + \left(\partial y_u/\partial p_u\right)\left(\partial p_u/\partial\tau^u\right)\right)\right]$, which is the marginal rate of substitution between τ^u and τ^d, while keeping the government revenue constant. Notice that this condition illustrates the trade-off the government faces when choosing policies that affect the two stages of production. That is, the government must balance the distortion and benefits its policies might cause both in the upstream and in the downstream stage.

While these terms capture the trade-off the tax authority faces when choosing the optimal policy at different stages of production, the government must also consider the trade-off within the downstream stage. The trade-off here is between the taxation (τ^d) of the final product y_d, which reduces firm's revenue, and the refund (δ) of taxes remitted in the upstream stage of production, which reduces the firm's cost of production and, hence, increases the profits of a downstream firm. In the notation of our model, these two effects are represented by $\left(\left(\partial p_d/\partial \tau^d\right)/\left(\partial p_d/\partial \delta\right)\right) p_d y_d$ and $-\tau^u p_u y_u + (1-\delta)\tau^u p_u \left(\partial y_u/\partial \delta\right)$, respectively.

As we noted before, the refund rate δ affects direct or indirectly the price the (downstream) final good prices and the misreported purchases and sales of the upstream and downstream firms. Before moving on to the limited tax capacity case, we derive a sufficient condition for $\delta = 1$, i.e., a one-hundred percent refund rate. Consider, for instance, a reduction of δ, starting from one. This change would only imply a smaller net gain in the revenue collected by the government, i.e., $\tau^u p_u y_u$ and $\tau^d \left(\partial p_d/\partial \delta\right) y_d$, than the net welfare gain experienced by consumers in the economy, i.e., $\left(-\left(\alpha/\lambda\right) y_d + \tau^d p_d \left(\partial y_d/\partial p_d\right)\right)\left(\partial p_d/\partial \delta\right)$, where α comes from the Roy identity $\partial W/\partial p_d = -\alpha y_d$. In words, this means that by reducing the refund rate the government would gain less in revenue than what the consumer would lose in welfare. In this case, the only reason to reduce the refund rate from a one-hundred percent rate is to collect more tax revenue at the expenses of final consumers. As expected, the price-elasticity of the demand for the final (downstream) good $\left(1+\epsilon_{p_d}^{y_d}\right)$ is greater than $\left(\left(1/\tau^d\right)\left[\left(\alpha/\lambda\right) + \left(\tau^u p_u y_u\right)/\left(p_d y_d\right)\right]\right)$. Thus, in an unlimited tax capacity economy, it is optimal for the government to set $\delta = 1$ and to allow a downstream firm to fully deduct taxes paid by the upstream firm as part of its production cost. In fact, if this sufficient condition is satisfied, then $\delta = 1$ and the optimal tax system is characterized by a tax on the final good only; a result similar to Diamond and Mirrlees (1971a). Notice that, from the point of view of the government's tax revenue, a one-hundred percent tax refund is equivalent to a zero tax on the upstream output (i.e., $\tau^u = 0$). On the other hand, if the abovementioned condition is not satisfied, the solution of the planner's problem calls for $\delta < 1$ and it is optimal to distortion the production of the intermediate good and to tax the upstream firm. In this case, a cascading effect of the upstream taxation, equation (**??**), can be observed in a multistage tax system $\mathcal{T} = \left(\tau^u, \tau^d, \delta\right)$.

A.2.3 Limited Tax Capacity: The Optimal Multistage Tax System

Manipulating the relevant first-order conditions and applying the envelope theorem, we can write the changes in the consumer's welfare for given a change in the policy instruments τ^u, τ^d, and δ as follows, respectively

$$\left(\frac{\partial W\left(\mathcal{T}\right)}{\partial \tau^u}/\lambda\right)\tau^u = M_{\tau^u}\tau^u - \tau^d p_d\left(\mathcal{T}\right)\Delta\overline{y_d}\varepsilon_{\Delta\overline{y_d},\tau^u} - \tau^u p_u\left(\tau^u\right)\Delta\overline{y_u}\varepsilon_{\Delta\overline{y_u},\tau^u} + \delta\tau^u p_u\left(\tau^u\right)\Delta\widehat{y_u}\varepsilon_{\Delta\widehat{y_u},\tau^u} \quad (34)$$

$$\left(\frac{\partial W\left(\mathcal{T}\right)}{\partial \tau^d}/\lambda\right)\tau^d = M_{\tau^d}\tau^d - \tau^d p_d\left(\mathcal{T}\right)\Delta\overline{y_d}\varepsilon_{\Delta\overline{y_d},\tau^d} - \tau^u p_u\left(\tau^u\right)\Delta\overline{y_u}\varepsilon_{\Delta\overline{y_u},\tau^d} + \delta\tau^u p_u\left(\tau^u\right)\Delta\widehat{y_u}\varepsilon_{\Delta\widehat{y_u},\tau^d} \quad (35)$$

$$\left(\frac{\partial W\left(\mathcal{T}\right)}{\partial \delta}/\lambda\right)\delta = M_{\delta}\delta - \tau^d p_d\left(\mathcal{T}\right)\Delta\overline{y_d}\varepsilon_{\Delta\overline{y_d},\delta} - \tau^u p_u\left(\tau^u\right)\Delta\overline{y_u}\varepsilon_{\Delta\overline{y_u},\delta} + \delta\tau^u p_u\left(\tau^u\right)\Delta\widehat{y_u}\varepsilon_{\Delta\widehat{y_u},\delta} \quad (36)$$

where the terms $\varepsilon_{\Delta y,t}$, $\Delta y \in \left\{\Delta\overline{y_d}, \Delta\overline{y_u}, \Delta\widehat{y_u}\right\}$ and $t \in \mathcal{T} = \left\{\tau^u, \tau^d, \delta\right\}$, represent the downstream misreported sales ($\Delta\overline{y_d}$) and purchases ($\Delta\overline{y_u}$), gaps (semi-)elasticities, and the upstream misreported sales misreported gap ($\Delta\widehat{y_u}$) (semi-)elasticity with respect to each tax policy instrument in the multistage tax system $\mathcal{T} = \left\{\tau^u, \tau^d, \delta\right\}$. For instance, $\varepsilon_{\Delta\overline{y_d},\tau^u} = \left(\partial\Delta\overline{y_d}/\partial\tau^u\right)/\left(\tau^u/\Delta\overline{y_d}\right)$ is the (semi-)elasticity of the downstream misreported sales gasp ($\Delta\overline{y_d}$) with respect to the upstream tax rate τ^u. The other elasticities follow along the same lines.[20]

Multistage tax system - Interior refund. Our first result is to show that in a limited tax capacity economy the optimal refund rate neither zero nor one-hundred percent. In other words, the optimal refund rate is interior, i.e., $\delta \in (0,1)$.[21] The solution of the government's welfare maximization problem must satisfy equation (36), which represents the changes in the consumer's welfare for a given a change in the refund rate δ. Notice that such policy change entail not only mechanical effects, equation (24), but also behavioral effects - the latter determined by the (semi-)elasticities of the downstream misreported (final goods) sales ($\Delta\overline{y_d}$) and (intermediate goods) purchases ($\Delta\widehat{y_u}$) gaps with respect to the refund rate δ, i.e., $\varepsilon_{\Delta\overline{y_d},\delta} = \left(\frac{\partial\Delta\overline{y_d}}{\partial\delta}\right)\left(\frac{\delta}{\Delta\overline{y_d}}\right)$ and $\varepsilon_{\Delta\widehat{y_u},\delta} = \left(\frac{\partial\Delta\widehat{y_u}}{\partial\delta}\right)\left(\frac{\delta}{\Delta\widehat{y_u}}\right)$, respectively, and the (semi-)elasticity of $\varepsilon_{\Delta\overline{y_u},\delta} = \left(\frac{\partial\Delta\overline{y_u}}{\partial\delta}\right)\left(\frac{\delta}{\Delta\overline{y_u}}\right)$ of the upstream misreported (intermediate goods) sales ($\Delta\overline{y_u}$) gap with respect to δ.

Regarding the signs of these elasticities, first notice that it is reasonable to expect that the elasticity $\varepsilon_{\Delta\overline{y_d},\delta}$ is negative. A lower refund rate increases the production cost of a downstream firm and, hence, *ceteris paribus*, leads to an increase of the final good price. Consumers react to this change by reducing their demand for the final good, which leads downstream firms to decrease their own demand for labor and the intermediate good produced by the upstream firm. In equilibrium, the firm's optimal condition with respect to $\overline{y_d}$, equation (9), must still be satisfied for a higher final good price. This implies that the marginal cost of misreporting has to increase, too. This occurs via an increase of the downstream misreported sales gap $\Delta\overline{y_d}$. And, hence, $\varepsilon_{\Delta\overline{y_d},\delta} < 0$. Next, the first-order condition of a downstream profit maximization problem with respect to $\widehat{y_u}$, equation (10), suggests that a decrease in the refund rate δ leads a downstream firm to increase the (intermediate good) misreported purchase gap $\Delta\widehat{y_u}$. That is, the misreported purchase gap $J_{\Delta\widehat{y_u}}\left(\widehat{y_u} - y_u\right)$ has to decrease once the refund rate falls, which occurs by increasing the (misreported) purchase gap. Thus, this (semi-)elasticity with respect to the refund rate is positive, i.e., $\varepsilon_{\Delta\widehat{y_u},\delta} > 0$. Finally, we argue that the (semi-)elasticity of upstream misreported sales gap $\Delta\overline{y_u}$ with respect to the tax refund is zero, i.e., $\varepsilon_{\Delta\overline{y_u},\delta} = 0$. To see this result, consider the upstream firm equilibrium condition, equation (13), which implies that $\tau^u = J_{\Delta\overline{y_u}}(\cdot)$. For a given reduction in δ, the left-hand side of this equation (13) remains constant as τ^u is fixed and p_u does not depend on the refund rate. Hence, in equilibrium, the right-hand-side has to remain constant. This implies that the upstream misreported sales gap $\Delta\overline{y_u} = y_u - \overline{y_u}$ is not affected by changes in the rebate rate and, hence, $\varepsilon_{\Delta\overline{y_u},\delta} = 0$.

Consider now an increase of the refund rate, starting from $\delta = 0$. This policy change has a first-order negative impact on how much a downstream firm misreports final good sales and a second-order positive effect on how much it misreports intermediate good purchases. It is reasonable to argue that the first-order effect dominates the second-order effect. A downstream firm misreports less and it is optimal to increase, from zero, the refund rate δ. On the other hand, reducing the refund rate from its upper bound of one-hundred percent has two first-order effects on a downstream misreporting behavior. First, a lower refund rate (relative to $\delta = 1$) reduces the amount a downstream firm misreports its final good sales, which has a positive impact on the consumer's welfare and it is measured by $\left(\left(\partial W/\partial\delta\right)/\lambda\right)\delta = M_\delta - \tau^d p_d \Delta\overline{y_d}\left(\varepsilon_{\Delta\overline{y_d},\delta}\right) > 0$. Also, the same policy

[20]Expressing the solution of the government welfare maximization problem in terms of the elasticity of the misreported gaps allows us to capture a feature revealed by our empirical analysis (Section 2).In particular, larger and significant responses on the part of upstream firms suggest differences in concealment costs, unobserved by the researcher, which could be captured by our model.

[21]Our result is more general than Best et al. (2015)'s as we allow for downstream firms to misreport on both revenue and production costs margins.

change affects the negatively the consumer's welfare via a downstream change in its misreported purchase gap, measure by $\left(\left(\partial W/\partial\delta\right)/\lambda\right)\delta = M_\delta - \tau^d p_d \Delta\overline{y_d}\left(\varepsilon_{\Delta\overline{y_d},\delta}\right) + \tau^u p_u\left(\tau^u\right)\left(\Delta\widehat{y_u}\right)\left(\varepsilon_{\Delta\widehat{y_u},\delta}\right) < 0$. We argue that the effect on welfare of a reduction of the refund rate via its effect on a downstream firm misreported sales gap is stronger than its effect on the firm misreported intermediate good purchase gap. That is, the net effect of lowering the refund rate from a one-hundred percent rate is positive. This result reasonably requires that a sufficiently small (normalized) mechanical welfare effect of δ (M_δ) and that the elasticity of the upstream firm misreported sales gap $(\Delta\overline{y_u})$ with respect to the tax refund $(\varepsilon_{\Delta\widehat{y_u},\delta})$ to be sufficiently large. Hence, the optimal refund rate is interior, i.e., $\delta \in (0,1)$. In a limited tax capacity economy, it is optimal for the government to allow a downstream firm to claim a fraction of the taxes remitted by the upstream firm as part of its production costs.

Proposition 3. *In limited tax capacity economy, the optimal multistage tax system $\mathcal{T} = \left\{\tau^{u*}, \tau^{d*}, \delta^*\right\}$ is the solution of the system of equations represented by equation (25). The optimal tax policies are functions of multistage behavioral effects $\mathcal{B}_{\overline{y_d},t}, \mathcal{B}_{\widehat{y_u},t}\mathcal{B}_{\overline{y_u},t}$, $t \in \mathcal{T} = \left\{\tau^{u*}, \tau^{d*}, \delta^*\right\}$, equations (26) - (28), and the (normalized) mechanical welfare effects of τ^{u*}, τ^{d*} and δ^* (i.e., $M_{\tau^{u*}}$ $M_{\tau^{d*}}$ and M_{δ^*}, respectively), equations* (22) - (24).

The optimal policies can be expressed as follows:

$$\tau^{u*} = \frac{\begin{array}{l} M_\delta\left[\mathcal{B}_{\overline{y_d},\tau^u}\mathcal{B}_{\widehat{y_u},\tau^d} - \mathcal{B}_{\overline{y_d},\tau^d}\left(\mathcal{B}_{\widehat{y_u},\tau^u} - M_{\tau^d}\right)\right. \\ \left. + \left(\mathcal{B}_{\widehat{y_u},\tau^u} - M_{\tau^d}\right)M_{\tau^u}\right] \end{array}}{\begin{array}{l} \mathcal{B}_{\overline{y_d},\tau^u}\left(-\mathcal{B}_{\widehat{y_u},\tau^d}\mathcal{B}_{\overline{y_u},\delta} + \mathcal{B}_{\widehat{y_u},\delta}\mathcal{B}_{\overline{y_u},\tau^d}\right) \\ +\mathcal{B}_{\overline{y_d},\tau^d}\left(\mathcal{B}_{\overline{y_u},\delta}\left(\mathcal{B}_{\widehat{y_u},\tau^u} - M_{\tau^d}\right) - \mathcal{B}_{\widehat{y_u},\delta}\mathcal{B}_{\overline{y_u},\tau^u}\right) \\ +\mathcal{B}_{\overline{y_d},\delta}\left(\mathcal{B}_{\overline{y_u},\tau^d}\left(M_{\tau^d} - \mathcal{B}_{\widehat{y_u},\tau^u}\right) + \mathcal{B}_{\widehat{y_u},\tau^d}\mathcal{B}_{\overline{y_u},\tau^u}\right) \\ +\left(\mathcal{B}_{\widehat{y_u},\delta}\mathcal{B}_{\overline{y_u},\tau^u} + \mathcal{B}_{\overline{y_u},\delta}\left(M_{\tau^d} - \mathcal{B}_{\widehat{y_u},\tau^u}\right)\right)M_{\tau^u} \end{array}}$$

$$\tau^{d*} = \frac{\begin{array}{l} M_\delta\left[\mathcal{B}_{\overline{y_d},\delta}\mathcal{B}_{\widehat{y_u},\tau^d} + \mathcal{B}_{\widehat{y_u},\delta}\left(M_{\tau^u} - \mathcal{B}_{\overline{y_d},\tau^d}\right)\right] \\ \times\left[\mathcal{B}_{\overline{y_d},\tau^u}\mathcal{B}_{\widehat{y_u},\tau^d} - \mathcal{B}_{\overline{y_d},\tau^d}\left(\mathcal{B}_{\widehat{y_u},\tau^u} - M_{\tau^d}\right)\right. \\ \left. +M_{\tau^u}\left(\mathcal{B}_{\widehat{y_u},\tau^u} - M_{\tau^d}\right)\right] \end{array}}{\begin{array}{l} \left[\mathcal{B}_{\overline{y_d},\tau^u}\mathcal{B}_{\widehat{y_u},\tau^u} - \mathcal{B}_{\overline{y_d},\delta}\left(\mathcal{B}_{\widehat{y_u},\tau^d} - M_{\tau^d}\right)\right] \\ \times\left[\mathcal{B}_{\overline{y_d},\tau^u}\left(-\mathcal{B}_{\widehat{y_u},\tau^d}\mathcal{B}_{\overline{y_u},\delta} + \mathcal{B}_{\widehat{y_u},\delta}\mathcal{B}_{\overline{y_u},\tau^d}\right)\right. \\ +\mathcal{B}_{\overline{y_d},\tau^d}\left(\mathcal{B}_{\overline{y_u},\delta}\left(\mathcal{B}_{\widehat{y_u},\tau^d} - M_{\tau^d}\right) - \mathcal{B}_{\widehat{y_u},\delta}\mathcal{B}_{\overline{y_u},\tau^u}\right) \\ +\mathcal{B}_{\overline{y_d},\delta}\left(\mathcal{B}_{\widehat{y_u},\tau^d}\mathcal{B}_{\overline{y_u},\tau^u} - \mathcal{B}_{\overline{y_u},\tau^d}\left(\mathcal{B}_{\widehat{y_u},\tau^u} - M_{\tau^d}\right)\right) \\ \left. -\left(\mathcal{B}_{\overline{y_u},\delta}\left(\mathcal{B}_{\widehat{y_u},\tau^d} - M_{\tau^d}\right) + \mathcal{B}_{\widehat{y_u},\delta}\mathcal{B}_{\overline{y_u},\tau^u}\right)M_{\tau^u}\right] \end{array}}$$

$$\delta^* = \frac{\mathcal{B}_{\overline{y_d},\tau^u}\mathcal{B}_{\widehat{y_u},\tau^d} - \mathcal{B}_{\overline{y_d},\tau^d}\mathcal{B}_{\widehat{y_u},\tau^u} + \mathcal{B}_{\overline{y_d},\tau^d}M_{\tau^d} + \mathcal{B}_{\widehat{y_u},\tau^u}M_{\tau^u} - M_{\tau^u}M_{\tau^d}}{\mathcal{B}_{\overline{y_d},\tau^u}\mathcal{B}_{\widehat{y_u},\delta} - \mathcal{B}_{\overline{y_d},\delta}\mathcal{B}_{\widehat{y_u},\tau^u} + \mathcal{B}_{\overline{y_d},\delta}M_{\tau^d}}$$

where we denote the behavioral effects of policy changes as the elasticities of the misreported sales and purchase gaps to the policy instruments $t \in \mathcal{T} = \left\{\tau^u, \tau^d, \delta\right\}$, adjusted for the value of their respective gap $\left(p_d\left(\mathcal{T}\right)\Delta\overline{y_d}, p_u\left(\tau^u\right)\Delta\overline{y_u}, p_u\left(\tau^u\right)\Delta\widehat{y_u}\right)$, as in equations (26) - (28).

A.2.4 Limited Tax Capacity: Optimal VAT and TOT Systems

The optimal VAT system $\mathcal{T}^*_{\mathcal{VAT}} = (\tau^*_{VAT}, \tau^*_{VAT}, 1)$ is such that

$$\begin{bmatrix}\varepsilon_{\Delta\overline{y_d},\tau_{VAT}} & \varepsilon_{\Delta\overline{y_u},\tau_{VAT}} & \varepsilon_{\Delta\widehat{y_u},\tau_{VAT}}\end{bmatrix}\begin{bmatrix} p_d\left(\mathcal{T}_{\mathcal{VAT}}\right)\left(\Delta\overline{y_d}\right) \\ p_u\left(\tau_{VAT}\right)\left(\Delta\overline{y_u}\right) \\ -p_u\left(\tau_{VAT}\right)\left(\Delta\widehat{y_u}\right)\end{bmatrix}\tau_{VAT} = M_{\tau_{VAT}}\tau_{VAT} \tag{37}$$

which, after some manipulation, we obtain that the optimal value-added tax rate τ^*_{VAT} must satisfy the following condition:

$$\left(\widetilde{\mathcal{B}}_{\overline{y_d},\tau^*_{VAT}} + \widetilde{\mathcal{B}}_{\widehat{y_u},\tau^*_{VAT}} - \widetilde{\mathcal{B}}_{\overline{y_u},\tau^*_{VAT}}\right)\tau^*_{VAT} = M_{\tau^*_{VAT}} \tag{38}$$

where $\widetilde{\mathcal{B}}_{y,\tau^*_{VAT}} = \left(\mathcal{B}_{y,\tau^*_{VAT}}/\tau^*_{VAT}\right)$, for $y \in \{\overline{y_d}, \overline{y_u}, \widehat{y_u}\}$.

In the case of a TOT system $\mathcal{T}^*_{\mathcal{TOT}} = \left(\tau^{u*}_{TOT}, \tau^{d*}_{TOT}, 0\right)$, the tax instruments are determined by the solution of the following system of equations:

$$\begin{bmatrix}\varepsilon_{\Delta\overline{y_d},\tau^u_{TOT}} & \varepsilon_{\Delta\overline{y_u},\tau^u_{TOT}} \\ \varepsilon_{\Delta\overline{y_d},\tau^d_{TOT}} & \varepsilon_{\Delta\overline{y_u},\tau^d_{TOT}}\end{bmatrix}\begin{bmatrix}\tau^d_{TOT}p_d\left(\mathcal{T}_{\mathcal{TOT}}\right)\left(\Delta\overline{y_d}\right) \\ \tau^u_{TOT}p_u\left(\tau^u_{TOT}\right)\left(\Delta\overline{y_u}\right)\end{bmatrix} = \begin{bmatrix}M_{\tau^u_{TOT}}\tau^u_{TOT} \\ M_{\tau^d_{TOT}}\tau^d_{TOT}\end{bmatrix} \tag{39}$$

which implies that the optimal upstream tax rate (τ^{u*}_{TOT}) and the optimal downstream tax rate (τ^{d*}_{TOT}) in a TOT system are to be set such that the behavioral responses of both the upstream and downstream firms must be taken into account along with the correspondent mechanical effects of both taxes, i.e.,

$$\begin{bmatrix}\tau^{d*}_{TOT} \\ \tau^{u*}_{TOT}\end{bmatrix} = \frac{\begin{bmatrix}\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{u*}_{TOT}} & -\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{u*}_{TOT}} \\ -\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{d*}_{TOT}} & \widetilde{\mathcal{B}}_{\overline{y_d},\tau^{d*}_{TOT}}\end{bmatrix}}{\left(\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{d*}_{TOT}}\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{u*}_{TOT}}\right) - \left(\widetilde{\mathcal{B}}_{\overline{y_u},\tau^{d*}_{TOT}}\widetilde{\mathcal{B}}_{\overline{y_d},\tau^{u*}_{TOT}}\right)}\begin{bmatrix}M_{\tau^{d*}_{TOT}} \\ M_{\tau^{u*}_{TOT}}\end{bmatrix} \tag{40}$$

Proposition 4. *The optimal value-added (VAT) tax rate τ^*_{VAT} is characterized by equation (38). And, the optimal tax instruments of a turnover tax system (TOT), i.e., τ^{u*}_{TOT} and τ^{d*}_{TOT}, are characterized by the equations (39) and (40).*

A.3 Limited Tax Capacity Economy - Numerical Results

We also conduct a series of numerical exercises to better understand the implications of an optimal multistage tax system. In particular, we investigate how the optimal tax system in a limited tax capacity economy changes under different assumptions regarding key parameters of the model. First, Table A.6 presents the results for different values of γ , α and R, the utility weight parameter, the input share in the downstream production function and the revenue requirement, respectively. If agents value the utility from consumption more, i.e., a higher relative weight on the utility derived from consumption (γ), the production of the intermediate and the final goods increase and their prices fall. Going from the benchmark value $\gamma = 0.33$ to, for instance, $\gamma = 0.70$, we observe that tax rates and refund rate are smaller, the optimal multistage tax system changing from $\left(\tau^{u*}, \tau^{d*}, \delta^*\right) = (0.12, 0.24, 0.19)$ to $(0.06, 0.13, 0.03)$. In this latter case, despite more output is being produced and consumed (at lower prices), the consumer's welfare goes down. Because agents value consumption more, in order to consume more they have to produce more and, hence, work more. The increase in welfare due to the increase in consumption is, however, more than compensated by a reduction in their welfare due to the fact that they are now enjoying less leisure.

Second, we investigate how an equal share of inputs in the downstream production technology would affect our results. Recall that the upstream production technology is linear and we set $\alpha = 0.50$ - that is, the intermediate (upstream) good and the labor have equal shares in a downstream firm production function. In this case, the optimal multistage tax system is such that the upstream (downstream) firm is taxed at a lower (higher) rate but a downstream firm can claim the full amount of taxes remitted by the upstream firm as part of its production costs. That is, $\mathcal{T} = \left(\tau^{u*}, \tau^{d*}, \delta^*\right) = (0.025, 0.276, 1.00)$. The upstream firm underreports about eighty percent of its intermediate good production, while a downstream firm misreports three percent of its final goods sales (compared to two and fourteen percent in the benchmark case, respectively). As expected, a lower exogenous revenue requirement R leads to lower taxes on both stages of production as well as a lower refund rate. The overall production level increases, prices fall and the consumer experiences higher welfare.

Table A.6: Multistage Tax System, Relative Consumption Utility (γ), Upstream Input Share (α) and Revenue Requirement (R): Optimal Policies, Output, Prices and Welfare

			$\alpha^* = 0.30,\ R^* = 0.10$		$\gamma^* = 0.33,\ R^* = 0.10$		$\gamma^* = 0.33,\ \alpha^* = 0.30$	
		Benchmark	$\gamma = 0.50$	$\gamma = 0.70$	$\alpha = 0.50$	$\alpha = 0.70$	$R = 0.05$	$R = 0.07$
		(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Policies	τ^{u*}	0.122	0.077	0.061	0.025	0.134	0.046	0.079
	τ^{d*}	0.236	0.153	0.130	0.276	0.167	0.125	0.167
	δ^*	0.189	0.050	0.032	1.000	0.025	0.066	0.098
Output	y_d^*	0.1790	0.2713	0.3256	0.1650	0.1788	0.1791	0.1791
	y_u^*	0.0920	0.1420	0.1720	0.1650	0.2210	0.0960	0.0939
Reported	$\overline{y_d}^*$	0.1542	0.2609	0.3181	0.1311	0.1664	0.1722	0.1667
	$\widehat{y_u}^*$	0.0922	0.1420	0.1720	0.1653	0.2210	0.0960	0.0940
	$\overline{y_u}^*$	0.0910	0.1420	0.0365	0.1647	0.2207	0.0959	0.0937
Prices	p_d^*	2.4869	2.2248	2.1563	2.7613	2.4369	2.1330	2.2609
	p_u^*	1.1352	1.0829	1.0647	1.0256	1.1526	1.0481	1.0849
Welfare	U^*	-0.9296	-1.0918	-1.1890	-0.9611	-0.9258	-0.8833	-0.9017

Benchmark: $R = 0.10$, $\gamma = 0.33$, $\alpha = 0.30$; $\theta_D = \theta_{\widehat{U}} = \theta_U = 1.50$.

Next, we analyze how changes in the parameters of the misreport cost functions affect the optimal multistage tax system. In a limited tax capacity economy where it is less costly to misreport sales and purchases, for instance, $\theta_D = \theta_{\widehat{U}} = \theta_U = 1.30$ (*vis-à-vis* the benchmark values $\theta_D = \theta_{\widehat{U}} = \theta_U = 1.50$), the upstream firm and a downstream firm produce and sell more intermediate and final goods, respectively. Prices fall and the firms misreport less in all three margins $\overline{y_d}$, $\overline{y_u}$, and $\widehat{y_u}$ (Table A.7). The optimal tax system is such that the upstream firm is taxed at a 6 percent tax rate, the downstream tax rate is 22 percent and a downstream firm can claim only 3 percent of the taxes remitted by the upstream firm. Compared to the benchmark case $\left(\tau^{u*}, \tau^{d*}, \delta^*\right) = (0.12, 0.24, 0.19)$, both the intermediate and the final good are taxed at lower rates, but the refund rate is also smaller, i.e., $\left(\tau^{u*}, \tau^{d*}, \delta^*\right) = (0.06, 0.22, 0.03)$.[22]. As the economy produces more of the intermediate and final goods, agents consume more and work relatively less, which improves consumer's welfare.

[22]Since the worker's labor supply is bounded, i.e., $(l_u + l_d) \in [0, 1]$, it implies that upstream and downstream firms' output is also bounded in the same range.

Finally, we conduct three additional exercises where we vary only one misreporting cost parameter θ at a time, while keeping the other two at their benchmark levels (Table A.7). When it is less costly for a downstream firm to misreport its final goods sales, i.e., $\theta_D = 1.30$ (and $\theta_{\widehat{U}} = \theta_U = 1.50$), the tax rate on the intermediate (upstream) good and the refund rate fall dramatically to about one-third of their benchmark values. The misreport gap also falls significantly - $(y_d^* - \overline{y_d}^*) = 0.0029$, compared to 0.0248 in the benchmark. If, on the other hand, misreporting sales is more costly for the upstream firm $\theta_U = 1.30$, keeping the other parameters at the benchmark level $(\theta_{\widehat{U}} = \theta_D = 1.50)$, we observe a drop in the production of the intermediate good along with an increase in its taxation $(\tau^{u*} = 0.19)$. Notice that, although the production of the final (downstream) good falls and agents consume less, consumers are better off as they work less hours in the upstream firm and, hence, enjoy more leisure.

Table A.7: Multistage Tax System and Misreporting Cost Parameters $(\theta_D, \theta_{\widehat{U}}, \theta_U)$
Optimal Policies, Output, Prices and Welfare

		Benchmark	$\theta_D, \theta_{\widehat{U}}, \theta_U = 1.30$	$\theta_{\widehat{U}}, \theta_U = 1.50$ $\theta_D = 1.30$	$\theta_D, \theta_U = 1.50$ $\theta_{\widehat{U}} = 1.30$	$\theta_D, \theta_{\widehat{U}} = 1.50$ $\theta_U = 1.30$
		(I)	(II)	(III)	(IV)	(V)
Policies	τ^{u*}	0.122	0.058	0.045	0.135	0.191
	τ^{d*}	0.236	0.222	0.225	0.231	0.207
	δ^*	0.189	0.029	0.041	0.161	0.080
Output	y_d^*	0.1790	0.1791	0.1791	0.1789	0.1785
	y_u^*	0.0920	0.0950	0.0960	0.0910	0.0860
	$\overline{y_d}^*$	0.1542	0.1763	0.1762	0.1552	0.1594
Reported	$\widehat{y_u}^*$	0.0922	0.0950	0.0960	0.0910	0.0861
	$\overline{y_u}^*$	0.0910	0.0950	0.0960	0.0510	0.0860
Prices	p_d^*	2.4869	2.4090	2.4085	2.4821	2.4632
	p_u^*	1.1352	1.0612	1.0470	1.1507	1.2348
Welfare	U^*	−0.9296	−0.9240	−0.9241	−0.9385	−0.9286

Benchmark: $R = 0.10$, $\gamma = 0.33$, $\alpha = 0.30$; $\theta_D = \theta_{\widehat{U}} = \theta_U = 1.50$.

Our numerical exercises reveal that smaller marginal costs to misreport sales - either sales of the intermediate (upstream) good to a downstream firm or final (downstream) good sales to consumers - lead to welfare gains. This occurs mainly due to an increase in leisure than more than compensate a (small) reduction in consumption. On the other hand, consumers are worse off in a economy where it is less costly for a downstream firm to misreport its purchase of the intermediate good ($\theta_{\widehat{U}} = 1.30$ and $\theta_D = \theta_U = 1.50$). In this case, a profit maximizing downstream firm reduces the amount of sales it misreports to the tax authority (i.e., $(y_d^* - \overline{y_d}^*) = 0.0237$, compared to an amount of 0.0248 in the benchmark), while truthfully reporting the amount of the intermediate good it purchases from the upstream firm (i.e., $(y_u^* - \widehat{y_u}^*) = 0$). While the misreport intermediate good purchase gap is zero at the downstream production stage, the misreport sales gap is bigger at the upstream firm - it goes from $(y_u^* - \overline{y_u}^*) = 0.001$ in the benchmark case to $(y_u^* - \overline{y_u}^*) = 0.04$ when it is costly for a downstream firm to misreport its intermediate good purchase. Overall, this result can be attributed to a higher taxation and a higher price of the intermediate (upstream) good.