



Title:

**Evasion vs. real production responses among firms to taxation:
Bunching evidence from Argentina**

Thesis by:
Ana Gamarra Rondinel

Promotor:
Jean Hindriks

Reader:
Romain Houssa

Academic year:
2016 - 2017

Master thesis nominated to obtain the academic degree of:
Master 120 en Sciences Économiques, Finalité Approfondie

ACKNOWLEDGEMENT

I am grateful to Professor Jean Hindriks for his advices and suggestions which made this thesis possible. His interest and support on this subject determine my research topic. I am also thankful to Romain Houssa for reading my thesis. Finally, I dedicate this study to my mother who died when I was working this thesis.

TABLE OF CONTENT

I.	Introduction	5
II.	Theoretical framework	7
	2.1 The trade-off in Best <i>et al.</i> (2014) framework	7
	2.2 Extension	10
	2.2.1 Turnover evasion	11
	2.2.2 Single tax	11
III.	Institutional background and data	12
	3.1 Corporate taxation in Argentina	12
	3.2 Data	13
	3.3 Limitations	14
IV.	Empirical methodology	15
V.	Results	18
VI.	Evaluation	21
VII.	Conclusion	24
	References	25
	Appendix	27

LIST OF TABLES:

Appendix: Table AI	27
Appendix: Table AII	28
Appendix: Table AIII	29
Appendix: Table AIV	31
Appendix: Table AV	32

LIST OF FIGURES:

Appendix: Figures I	33
Appendix: Figures II	35
Appendix: Figures III	40
Appendix: Figures IV	45
Appendix: Figures V	50

EVASION vs. REAL PRODUCTION RESPONSES AMONG FIRMS TO TAXATION: BUNCHING EVIDENCE FROM ARGENTINA

Abstract:

The centerpiece in public economics is that optimal tax policies and tax instruments should promote production efficiency even in second-best environments. This theoretical prediction has been widely accepted in developed and developing countries, although it has been derived from models that ignore tax evasion. Once enforcement constraints are acknowledged, we show that contrary to the theoretical prediction, production efficiency is not anymore the centerpiece of the model while instead revenue efficiency becomes more relevant. This paper analyzes empirically such trade-off between revenue and production efficiency in the choice of tax instruments in Argentina. We use a production inefficient tax policy, the simplified tax regime, which affects firms' behavior on compliance and real output. Using the bunching approach and administrative tax data covering all corporate income tax returns for the years 1997-2011, we find that medium firms cluster around a profit rate of 0.09 which might suggest evidence of a behavioral response to the policy.

Keywords: Informal firms, tax evasion, SMEs, bunching approach, simplified tax regime, corporate taxation.

I. INTRODUCTION

In the last decades the introduction of globalization and market liberalization has exacerbated poverty and inequality in developing countries. Hence, additional domestic revenue to finance poverty reduction has become a need; as Besley and Persson (2013, p.2) claim "tax lies at the heart of state development". Latin American countries for instance rely more on indirect taxation and resource revenue, the Value Added Tax (henceforth, VAT) and the non-tax revenue are the primary sources of revenue collection¹. Moreover, the presence of large informal economies is one of the causes of this dependence to indirect taxation; as a result, the income tax is insignificant and highly concentrated in few large firms². Furthermore, the economic growth of more than one decade in the region has promoted an entrepreneurship spirit which in turn has created numerous small and medium enterprises (henceforth, SMEs); in the literature of tax design, those enterprises belong to the so called 'hard-to-tax' sectors³. Although this sector provides a great dynamic to the economy, the tax raised by it is minimal due to the large informality. For this reason, we believe that the relation between informality, tax evasion and domestic revenue mobilization in developing countries is highly connected with the taxation of SMEs: is it worth in terms of costs-benefit to tax this sector?, what should be the optimal tax structure of a country with numerous SMEs and limited tax capacity? Indeed, such relation challenges academic economists and policy practitioners of developing countries.

¹ Along 1990 and 2009, the revenue collected by the VAT was around half of the tax revenue in Argentina; whereas the non-tax revenue was 12%. Based on the ICTD Government Revenue Dataset (ICTD GRD, 2014).

² According to Schneider *et al.* (2010), the shadow economy in Argentina during 1997-2007 represented 25.3%GDP. Moreover, only 0.1% of firms remit 49% of tax revenue in Argentina (International Tax Dialogue, 2007).

³ Over our period of study (1997-2011), large firms account for around 0.2% of total registered taxpayers in the corporate income tax structure; while, medium firms accounts for 67.6% and small firms accounts for 32.2%. Based on the Federal Administration of Public Revenue of Argentina (AFIP).

The broad literature of optimal taxation relies on the production efficient theorem of Diamond and Mirrlees (1971) which suggests that tax systems should maintain full production efficiency even in second-best environments; for this reason, authors recommend to avoid taxes on turnover, trade and intermediate inputs that distort production efficiency. However, any tax policy for developing countries based on that premise ignores the issues faced by those countries as imperfect enforcement, limited tax capacity, informality and tax evasion. Therefore, Kleven *et al.* (2016) insist that in limited tax capacity environments third-best policies are more suitable although it implies revenue creation at the expense of production efficiency. The objective of this study is to analyze empirically such trade-off between revenue and production efficiency in the choice of tax instruments in a developing country. For that, we exploit a production inefficient policy called Simplified Tax Regimes (henceforth, SR) according to which firms are taxed either on their profits or turnover depending on which tax liability is larger⁴. This policy was implemented in 1998 in Argentina with two purposes: fight informality and reduce evasion; in other words, transform “ghosts” into taxpayers and “icebergs” into fully taxpayers⁵, based on the idea that a larger tax base is more difficult to evade.

We begin our analysis by presenting a simple theoretical framework based on the model of Best *et al.* (2014) which we extend to introduce turnover evasion to account for the Argentinean setting. The optimality conditions of this model suggest that in countries with limited tax capacity may be desirable to deviate from full production efficiency in order to increase compliance. To evaluate this theoretical prediction we use administrative data from the Federal Administration of Public Revenue (AFIP, in Spanish) covering the tax returns of all firms filling the corporate income tax between 1997 and 2011. However, this dataset has two drawbacks, it is not in a micro level and do not represent firms in the simplified regime. To overcome these limitations we adjust the estimation strategy of Best *et al.* (2014) but we maintain the main idea that the simplified regime gives rise to non-standard kink points due to the joint and discontinuous change of the tax rate and the tax base at a cutoff. Such kinks -as authors suggest- influences firms’ behavior on compliance and real production differently, and give rise to the excess mass around the kink. Three are our main findings. First, we observe bunching for all firms and in all years, especially in 1997 (the last year before the introduction of the policy) and in 1998 among medium firms. This result confirms our hypothesis that the introduction of the policy provides medium firms of additional incentives to reduce their turnover (‘legally’ or ‘illegally’) and to comply in costs. Also, the speed of the response supports the hypothesis that bunching is due to evasion responses rather than real output responses, as previous studies allude (Mosberger 2016; Lediga *et al.* 2016). Second, we observe that firms cluster around a profit rate of 0.09 which might be evidence of a behavioral response to the policy. We make use of the reference point argument to explain this result as in earlier studies (Kleven 2016;

⁴ They were first introduced in the 1970s in Latin America to deal with the difficulty to apply the VAT to small taxpayers. Later, in the 1990s (1997, in Brazil and 1998, in Argentina) were expanded to the rest of the continent, except in Venezuela, Panama and El Salvador. Nowadays, there are three types of simplified tax regimes: ones that replace the CIT, others that replace the VAT and others that replace both plus social security contributions (applied only in Argentina, Brazil and Uruguay). The criterion of qualification is mainly the turnover; although some regimes have additional objective parameters as physical area, electricity bill and number of employees. Some countries have more than one regime as Bolivia (3), Chile (5), Mexico (3), Uruguay (2) and Peru (2). The universe of taxpayers covered is: self-employed, and micro and small unincorporated enterprises mainly in the commercial and service sectors. Finally, the tax calculation commonly used is a monthly fixed quota.

⁵ Kanbur and Keen (2014) divide the population of taxpayers by different forms of compliance and non-compliance. Among the later we can find “ghosts” and “icebergs”. The former is the invisible taxpayer who should be registered for tax purposes but do not, hence is outside the tax net. While, the later is a registered taxpayer who illegally misreport their costs and/or output in order to reduce her tax liability.

Mosberger 2016; Carrillo *et al.* 2014; Castro and Scartascini 2015). Third, in line with previous work, we present evidence that firms bunch around zero which suggest that deductions might have an important role in the analysis.

The paper is placed into the broad literature of firms' behavioral responses to taxes (Kopczuk and Slemrod 2006; Munk 1978), the recent bunching literature (Saez 2010; Chetty *et al.* 2011; Kleven 2016; Kleven and Waseem 2013) and the studies of optimal taxation of firms with limited tax capacity (Emran and Stiglitz 2005; Gordon and Li 2009; Kanbur and Keen 2014; Dharmapala *et al.* (2011); Keen 2007; Keen 2012; Boadway and Sato 2009; Abramovsky *et al.* 2014). Our contributions to the literature are threefold. First, few studies (Best *et al.* (2014) and Waseem (2013) for Pakistan; Carrillo *et al.* (2014) for Ecuador; Castro and Scartascini (2015) for Argentina; Mosberger (2016) for Hungary and Lediga *et al.* (2016) for South-Africa) analyze the relation between informal firms, evasion and taxation in developing countries, our paper fills this gap and provides direct empirical evidence on firms' margin responses to a widespread and questionable policy in Latin America with scarce quantitative evidence. Second, we contribute to the nascent literature that uses bunching approach to estimate corporate responses to tax changes that is object of a controversial debate in the public economic literature. As Saez (1999) points out, the majority of studies focus on the effect of marginal tax rates in the personal income tax (henceforth, PIT); whereas little attention has been paid to the CIT. Third, we take distance from Best *et al.* (2014) in three aspects, we do not assume *a priori* that bunching represents evasion responses, also we give a crucial role to turnover evasion and finally, our results are more pessimistic than Best *et al.* (2014) results because the simplified regime in Argentina increases the number of evaders by creating "icebergs" in both regimes. All in all, the novelty of the paper relies on the topicality of the subject, the analytical approach and the social interest of the topic.

The set up of the rest of this paper is as follows. In the next section we present a theoretical framework for analyzing the trade-off between revenue and production efficiency. In section III we present the relevant institutional background and describe the data we use in the empirical analysis. In section IV we explain the empirical strategy used in estimating firms' behavioral responses. In section V we present our results from the bunching evidence and bunching estimation. In section VI we present some experiences of the impact of the policy. Finally, section VII briefly concludes.

II. THEORETICAL FRAMEWORK

The purpose of this section is to develop a conceptual framework to examine the trade-off between production and revenue efficiency in the presence of tax evasion. We begin with a brief review of Best *et al.* (2014) who model this trade-off for Pakistan, we take this model as our baseline model and propose two extensions to make it compatible with the Argentinean setting.

2.1 The trade-off in Best *et al.* (2014) framework

We assume a static model of optimal taxation of firms, with no uncertainty. In a partial equilibrium framework with no inputs and only one final good, firms choose how much output y to produce and how much costs to report to the tax administration \hat{c} . We assume firms *only* can over-report their costs $\hat{c} > c(y)$ in order to reduce their reported profits (and therefore tax

liability). The tax liability depends on firms' output and reported costs: $T(y, \hat{c}) = \tau[y - \mu\hat{c}]$, where μ is the tax base parameter which determines the tax regime the firm belongs to, either to the simplified tax regime with a turnover tax base ($\mu = 0$) or to the general regime with a pure profit tax base ($\mu = 1$)⁶. When firms misreport they run into some costs $g(\hat{c} - c(y))$, those can be for instance: the risk of being audited, the fine paid when the firm is caught by the tax administration, the productivity losses from operating in cash, the costs of not keeping accurate accounting books or the costs from changing the production process to eliminate verifiable evidence (Kopczuk and Slemrod 2006; Best *et al.* 2014). Those costs are convex with the level of tax evasion⁷. Also, profits depend positively on after-tax income and negatively on reported costs. Moreover, we assume a small open economy where firms are price-takers; hence we normalize the price of the final good so that turnover and output are identical⁸. As a result, we have the true after-tax profit: $\pi(y, \hat{c}) = y - c(y) - \tau[y - \mu\hat{c}] - g(\hat{c} - c(y))$ and the reported after-tax profit: $\hat{\pi}(y, \hat{c}) = y - \hat{c} - \tau[y - \mu\hat{c}]$. Therefore, the optimization problem of a representative firm is:

$$\max_{y, \hat{c}} \pi(y, \hat{c}) = y - c(y) - \tau[y - \mu\hat{c}] - g(\hat{c} - c(y))$$

FOCs:

$$\frac{d\pi}{dy} = 0 \rightarrow c'(y) = 1 - t \left(\frac{1-\mu}{1-t\mu} \right) = 1 - t_E \quad \dots (1)$$

$$\frac{d\pi}{d\hat{c}} = 0 \rightarrow g'(\hat{c} - c(y)) = t\mu \quad \dots (2)$$

Where, $t \left(\frac{1-\mu}{1-t\mu} \right)$ is the effective marginal tax rate (t_E).

The first of these expressions determines the real output level. Note that the effective tax rate reduces the marginal return to real output, it is a wedge created by a distortionary tax which depends on the statutory tax rate and the tax base; $\frac{dt_E}{dt} \geq 0, \frac{dt_E}{d\mu} \leq 0$. Therefore, an increase of the tax rate ($\uparrow \tau$) or a large tax base ($\downarrow \mu$) rises the effective tax rate, which in turn reduces the marginal return of real output and, hence decreases firms' real output level. The second expression determines the level of evasion and it is increasing in the statutory tax rate and the tax base. Thus, an increase of the tax rate ($\uparrow \tau$) or a narrow tax base ($\uparrow \mu$) rises the marginal return of misreporting ($t\mu \uparrow$) and, as a result firms have more incentives to evade in costs⁹.

When firms belong to the general regime, they are taxed on their profits ($\mu = 1$) and the statutory tax rate is the profit tax, hence equations (1) and (2) become: $c'(y) = 1, g'(\hat{c} - c(y)) = \tau_\pi$. We observe in the first expression that the wedge disappears, so firms' production

⁶ If we do not consider the tax base parameter in the tax liability, then we are in a case where all firms are only taxed by a non-distortionary and time-invariant profit tax. But, with the introduction of the simplified tax regime, firms can voluntarily choose to which regime belong to.

⁷ As a result, we have three possible cases: no misreporting ($\hat{c} - c(y) = 0$), over-reporting ($\hat{c} - c(y) > 0$) and under-reporting ($\hat{c} - c(y) < 0$). Is not unreasonable to think that firms under-report costs; in fact Carrillo *et al.* (2014) show that when the audit probability is a decreasing function of the profit rate, firms under-report revenues and costs in order to 'look small' and, hence avoid being audited.

⁸ Therefore, net income can be referred to profits and gross income to turnover or output. So, the profit tax is over net income while the turnover tax is over gross income.

⁹ For example, if the government improves its tax capacity by increasing the number of inspections or the amount of fines, then is more costly for the firm to evade. For example, in the reform of 2004 the tax administration introduced a fine (100 -3 000 pesos) plus the precautionary closing of the business. Five years later, in the reform of 2009, the fine was modified to 50% of the single tax.

decision is undistorted. The second expression implies that firms have incentives to evade in the general regime because the marginal return of misreporting is the profit tax. Likewise, when firms belong to the simplified tax regime, they are taxed on their turnover ($\mu = 0$) and the statutory tax rate is the turnover tax, hence equations (1) and (2) become: $c'(y) = 1 - t_y$, $g'(\hat{c} - c(y)) = 0$. In this case, a wedge distorts the production decision of firms by reducing the marginal return to real output; however, in this regime firms get no benefits from misreporting costs.

The model sets a government who can only raise revenue by taxing firms; this is not an unreasonable assumption in developing countries where due to high administrative costs income taxes are concentrated on the CIT¹⁰. Moreover, the government sets the tax base μ and the tax rate τ in the presence of tax evasion in order to maximize welfare. This simple stylized model does not include the behavior of households, we assume that the only individuals in the economy are the firms' owners, hence their consumption depend on the after-tax profits obtained. Therefore, the government maximizes welfare that is the aggregate consumption or after-tax profit subject to an exogenous revenue requirement R . The optimization problem of the government is:

$$\begin{aligned} \max_{t, \mu} W &= \pi(y, \hat{c}) \\ \text{st: } R &\leq T(y, \hat{c}) \end{aligned}$$

Where, $T(y, \hat{c}) = \tau[y - \mu\hat{c}]$

$$\mathcal{L}: \pi(y, \hat{c}) + \lambda(T(y, \hat{c}) - R)$$

FOCs:

$$\begin{aligned} \frac{d\mathcal{L}}{dt} = 0 &\rightarrow \lambda \geq 1 \\ \frac{d\mathcal{L}}{d\mu} = 0 &\rightarrow \lambda \geq 1 \end{aligned}$$

λ denotes the (endogenous) marginal costs of public funds; intuitively, is the 'price' the government has to run through to collect revenue in order to finance public expenditure.

In this setting, the government has two instruments for tax policy, the tax rate and the tax base. Henceforth, from this model Best *et al.* (2014) state the following conditions of optimality:

- Lemma 1: When there is no evasion or government's concern is production efficiency, then the optimal tax base is $\mu = 1$.

Proof: Perfect enforcement implies no evasion, so firms report their true costs $\hat{c} = c(y)$ hence $g(0) = 0$. Therefore, firms' optimization is stated as: $\max_y \pi(y) = y - c(y) - \tau[y - c(y)]$ where the result is an efficient output decision: $c'(y) = 1$. Consequently, with perfect tax enforcement the theorem of Diamond and Mirrless (1971) holds, full production efficiency is maintained with $\mu = 1$ and a pure profit tax.

¹⁰ The empirical evidence suggests that low fiscal capacity countries rely heavily on firms' taxation; for instance, in Argentina the CIT revenue represents 65% of the tax income revenue, while the PIT represents only 31% (mean over 1990-2010). Based on the ICTD Government Revenue Dataset (ICTD GRD, 2014).

- Proposition 1: When there is evasion or government's concern is revenue efficiency, then the optimal tax base is $\mu = 0$.

Proof: With imperfect tax enforcement, firms have an incentive to misreport $\hat{c} > c(y)$ although it has some costs $g(\hat{c} - c(y)) > 0$. In such context, firms optimization is: $\max_{y, \hat{c}} \pi(y, \hat{c}) = y - c(y) - \tau[y - \mu\hat{c}] - g(\hat{c} - c(y))$ which results in an inefficient output decision $c'(y) = 1 - t_y$. Consequently, firms deviate from optimality but have no incentives to misreport.

Proposition 1 is a generalization of the optimal tax rule (Lemma 1) in the presence of tax evasion, it captures the trade-off between production and revenue efficiency in the choice of the tax base and reflects the notion that “a broader base is harder to evade” (Best *et al.* 2014, p.2). For policy matter, if the revenue efficiency concern is stronger than the production efficiency concern, then it will be socially optimal to move towards the turnover tax by broadening the tax base $\mu = 0$; therefore the production wedge increases producing a second-order welfare loss; while the evasion rate decreases leading to a first-order welfare gain. Conversely, if the production efficiency concern is stronger than the revenue efficiency concern it will be socially optimal to move towards a profit tax setting by $\mu = 1$; therefore, firms' real output decision is undistorted at the expense of evasion. The former describes the case of developing countries with imperfect tax enforcement and weak tax capacity, while the later describes the reality of a developed country.

2.2 Extension

2.2.1 Turnover evasion

In the baseline model we assume firms can only over-report costs¹¹, here we extend the model to allow for output evasion; that is, firms can over-report costs $\hat{c} > c(y)$ and under-report output $\hat{y} < y$, and by doing so they run into some costs $g(\hat{c} - c(y), y - \hat{y})$. In this setting, firms choose how much output \hat{y} and costs \hat{c} report to the tax administration, hence the optimization problem of the representative firm is:

$$\max_{\hat{y}, \hat{c}} \pi(\hat{y}, \hat{c}) = y - c(y) - \tau[\hat{y} - \mu\hat{c}] - g(\hat{c} - c(y), y - \hat{y})$$

FOCs:

$$\frac{d\pi}{d\hat{y}} = 0 \rightarrow g'_{\hat{y}} = \tau \quad \dots (1)$$

$$\frac{d\pi}{d\hat{c}} = 0 \rightarrow g'_{\hat{c}} = \tau\mu \quad \dots (2)$$

Similar to the baseline model, equation (1) determines the level of turnover evasion and equation (2), the level of costs evasion. Both are increasing with the statutory tax rate but only equation (2) is affected positively by the tax base. Moreover, firms in the general regime ($\mu = 1$) have the same incentives to evade in output and in costs because the marginal return of misreporting is the profit rate: $g'_{\hat{y}} = g'_{\hat{c}} = \tau\pi$. This result is reasonable in the general regime

¹¹ We did so based on Best *et al.* (2014, p.4) idea that “it may be easier to fabricate costs than to conceal revenues”. However, we think Argentinean firms have strong incentives to under-report turnover because is the main criterion of qualification and for categorization in the simplified regime.

where evade in costs and/or in turnover reduces reported profits and, hence tax liability. While, firms in the simplified regime ($\mu = 0$) have only incentives to under-report turnover: $g'_{\hat{y}} = \tau_y$, $g'_{\hat{c}} = 0$. The result from the government's optimization problem remains the same: $\lambda \geq 1$.

Therefore, the introduction of turnover evasion to the baseline model does not contradict our previous results; in fact, it could further our understanding. With this simple extension we capture the idea that firms under-report turnover despite the regime they belong to; however, they have more incentives to evade in the general regime than in the simplified regime because $0 < \tau_y < \tau_\pi$.

2.2.2 Single tax

In this second extension we assume a fixed fee rather than a turnover tax. We do so because Argentinean firms in the simplified regime pay a fixed fee called single tax which is determined by the tax administration and depends on the category in which the firm is classified (see Table AVIII); therefore, the single tax depends positively on the reported turnover $\theta(y)$. The optimization problem for a representative firm with *only* costs evasion is:

$$\max_{y, \hat{c}} \pi(y, \hat{c}) = y - c(y) - \mu_1 \theta(y) - \mu_2 \tau [y - \hat{c}] - g(\hat{c} - c(y))$$

FOCs:

$$\frac{d\pi}{dy} = 0 \rightarrow c'(y) = \frac{1 - \mu_1 \theta'_y - \mu_2 \tau}{1 - \mu_2 \tau} \quad \dots (1)$$

$$\frac{d\pi}{d\hat{c}} = 0 \rightarrow g'_{\hat{c}} = \mu_2 \tau \quad \dots (2)$$

Now, we introduce turnover evasion to see if there is any significant change in the results. Note that in this case the single tax is decreasing in the under-reported turnover $\theta(\hat{y})$. The optimization problem for a representative firm with turnover and costs evasion is:

$$\max_{\hat{y}, \hat{c}} \pi(\hat{y}, \hat{c}) = y - c(y) - \mu_1 \theta(\hat{y}) - \mu_2 \tau [\hat{y} - \hat{c}] - g(\hat{c} - c(y), y - \hat{y})$$

FOCs:

$$\frac{d\pi}{d\hat{y}} = 0 \rightarrow g'_{\hat{y}} = \mu_1 \theta'_{\hat{y}} + \mu_2 \tau \quad \dots (3)$$

$$\frac{d\pi}{d\hat{c}} = 0 \rightarrow g'_{\hat{c}} = \mu_2 \tau \quad \dots (4)$$

When there is only costs evasion (equation 1 and 2), firms in the general regime ($\mu_1 = 0, \mu_2 = 1$) maintain full production efficiency and over-report costs: $c'(y) = 1$, $g'_{\hat{c}} = \tau_\pi$; while, firms in the simplified regime ($\mu_1 = 1, \mu_2 = 0$) deviate from production efficiency and comply in costs: $c'(y) = 1 - \theta'_y$, $g'_{\hat{c}} = 0$. When there is turnover and costs evasion (equation 3 and 4), firms in the general regime ($\mu_1 = 0, \mu_2 = 1$) evade in turnover and costs: $g'_{\hat{y}} = g'_{\hat{c}} = \tau_\pi$; while firms in the simplified regime ($\mu_1 = 1, \mu_2 = 0$) only under-report turnover: $g'_{\hat{c}} = 0$, $g'_{\hat{y}} = \theta'_{\hat{y}}$. In both cases, the government chooses the tax rate τ and the tax base μ_1, μ_2 to maximize welfare subject to the revenue requirement: $R \leq \mu_1 \theta(y) + \mu_2 T(y, \hat{c})$ when there is only costs evasion, or $R \leq \mu_1 \theta(\hat{y}) + \mu_2 T(\hat{y}, \hat{c})$ when there is turnover and costs evasion. The result from the government's optimization problem in both cases remains the same: ≥ 1 . Therefore, this extension leads to the same predictions of the baseline model; consequently use a single tax or a turnover tax should not alter the main insight of the model.

The theoretical work gives qualitative predictions on how should be the optimal taxation of firms under the presence of tax evasion. In the empirical work we will find the quantitative implications of such predictions by exploiting a production inefficient tax policy in Argentina, the simplified tax regime.

III. INSTITUTIONAL BACKGROUND AND DATA

3.1 Corporate taxation in Argentina

The CIT is a crucial source of revenue in Argentina and raises 4.4%GDP which represents about 29% of all central tax revenues¹². The tax system is residence-based and the current CIT is 35%. In our period of study we find about 132 298 firms in the general regime filling tax returns each year and around 1 561 711 active firms registered in the simplified regime each year (Table AI). However, the revenue collected in the simplified regime is insignificant in Argentina's tax system, it accounts for 3% of all CIT revenue and 1% of total tax revenue (mean over 1998 to 2011)¹³.

The introduction of the simplified regime in Nov. 1998 had two purposes: fight informality (first order motivation) and reduce tax evasion (second order motivation). This scheme, is one of the few in Latin America that links informality, taxation and social protection¹⁴, it consists of a single tax composed by two components: a monthly tax which replaces the income tax (PIT and CIT) and the VAT, and a social security component that includes retirement benefits and health coverage. To keep the model simple, we ignore the social security contributions and focus only on the monthly tax and the CIT¹⁵. Furthermore, the SR classifies firms among eight categories (A-H) depending on their reported turnover, and each category has a different monthly tax (see Table AII). Importantly, in order to follow Best *et al.* (2014) methodology, we transform the single tax to a turnover tax; so that, the turnover tax we get is the minimum amount that a firm allocated in the lowest category (category A) can pay¹⁶. However, when we compare between categories, the turnover tax of category A is in fact the highest turnover tax in proportional terms, suggesting (as mentioned in the reports of the tax administration) certain regressivity in the SR.

The idea of the SR is to give incentives to small taxpayers (ie. self-employed, unincorporated small business as cooperatives and irregular societies of up to three members) to voluntarily comply with their tax obligations by offering access to a retirement plan and health insurance, and by minimizing the costs of compliance. Note that once adhered to it, a minimum period of permanence (one calendar year) is required; as well, when the firm renounces to continue in the SR it has to fulfill the tax and social security obligations by the general regime the first day of

¹² Mean over 1997-2011, based on the ICTD Government Revenue Dataset (ICTD GRD, 2014).

¹³ Nevertheless, the SR of Argentina is the second country in Latin America (after Brazil's simplified tax regime, SIMPLES), collecting the highest revenue (ILO, 2014).

¹⁴ Since 2004, the SR contains complementary special regimes for specific 'hard-to-tax' sectors, those are: the Regime of Social Inclusion and Promotion of Independent Work, the Simplified Regime for Effectors of Local Development and Social Economy, the Special Regime for Workers Associated to Labor Cooperatives and the Simplified Regime for Domestic Service Workers.

¹⁵ So we use the terms "single tax" and "monthly tax" interchangeably throughout the paper.

¹⁶ The turnover tax is the single tax (annual) divided by the lowest and highest annual turnover base of each category; thus, we have a minimum and maximum turnover tax in each category. To follow as close as possible the methodology of Best *et al.* (2014), we choose the minimum turnover tax from the lowest category (category A).

the month following the renounce. To be an eligible firm the highest turnover obtained from the principal activity in the previous year should not exceed the limit established by the tax administration (see Table AII); in addition, the firm has to meet some objective parameters (see Table AII), it cannot be an importer, it cannot reclaim credits on inputs and the firm is obliged to report its turnover and objective parameters every tax year in order to confirm they meet the requirements¹⁷. However, firms in the SR do not need to fill any tax return because are excluded from the VAT and CIT and are not obliged to keep accounting books.

During our period of analysis we identify three reforms in the CIT structure. The first reform (Dec. 1998) increased the profit tax rate from 33% to 35%, the second reform (July 2004) raised the single tax of the highest categories and the third reform (Dec. 2009) broadened the tax base (see Table AII). Note that the second and the third reforms introduced additional modifications that we deliberately ignore in order to center our attention on the changes of the tax rate and the tax base. Therefore, the profit tax rate, the turnover tax rate and the tax base undergo changes during the time period we study, Table AI (Panel A) shows these variations for all the period, which as Best *et al* (2014) we exploit in our empirical analysis. For practical reasons we divide our period of study based on the policy and the three reforms identified:

Table I: Tax rates and kink points					
Reforms	1997	1998	1999/2003	2004/2009	2010/2011
τ_y (min.)	-	0.033	0.033	0.033	0.020
τ_y (total,min.)	-	0.066	0.066	0.088	0.11
τ_π	0.33	0.33	0.35	0.35	0.35
τ_y (min.)/ τ_π	-	0.1	0.09	0.09	0.05
τ_y (total,min.)/ τ_π	-	0.2	0.18	0.25	0.31

Note: τ_y (min.) refers to the monthly tax and τ_y (total,min.) refers to the monthly tax plus social security contributions.

3.2 Data

Our analysis is based on administrative data from the Federal Administration of Public Revenue (AFIP, in Spanish), it covers all corporate income tax returns for the period 1997-2011¹⁸. It is worth to highlight that our dataset contains *only* firms filling the corporate income tax returns, hence only firms in the general regime (henceforth, GR)¹⁹. Before the introduction of the policy, there was only one regime in the CIT structure, the general regime, and firms outside any regime were informal firms that we call “ghosts”. So, before the policy, firms can only move from the general regime to informality and conversely. Is important to mention that inside the general regime there were firms misreporting costs and turnover, we call them “icebergs” in the general regime. However, after the policy, there were two regimes co-existing in the CIT structure, the general regime and the simplified regime; as a result, we identify six flows: from

¹⁷ Since 2004, firms in the simplified regime have to report their turnover and objective parameters every four months (May, September and January).

¹⁸ In Argentina, the tax year coincide with the calendar year, it runs from January 1 to December 31. However, the last day for firms to present their corporate income tax return to the tax administration is the 30th June. Note that we do not consider the tax year 2001 because the observations were not sufficient because in 2001 the tax administration modified the format of the tax returns.

¹⁹ Note that the electronic filling (Presentación de DDJJ y Pagos) is optional for all firms; but around 97% of firms used this method from 2006 to 2010 (CIAT, <http://www.ciat.org/index.php/es/productos-y-servicios/ciatdata/anexos-estadisticos.html>) which ensures less measurement error in the data, as Best *et al.* (2014) suggest.

the GR to the SR and conversely; from the GR to informality and conversely, and from the SR to informality and conversely. Note that in the simplified regime we also have firms under-reporting turnover, we call them “icebergs” in the simplified regime. We restrict our analysis to the outflow from the GR to the SR because, as explained above, our data only allow us to observe the behavior of firms in the GR. Flows between formality and informality, although they provide interesting insights, are not the focus of this study; consequently, we center our attention on the second motivation of the policy, evasion reduction.

For practical purposes, we focus on three variables: turnover, gross profits and taxable income, although the information available on the corporate income tax returns includes all kind of firms’ balance sheet items as total costs, total purchases, tax liabilities, total assets, initial and final stocks, etc. Figures I1 and I2 in the Appendix identify the variables needed for the empirical analysis; when we use the calculated profit (reported turnover minus reported costs) or the reported gross profit, the distribution of the profit rate (profits as a share of turnover) is above zero and too on the right of the cutoff even when social security contributions are included. Therefore, in the empirical analysis we take the reported *taxable income*²⁰ to calculate the profit rate.

Finally, for practical purposes in the original data published in the website of the tax administration, firms in the general regime are classified in 17 tax thresholds according to their reported turnover (see Table AIV), although all of them are taxed on their profits by a flat tax rate. We take advantage of this format and divide our data in groups according to their eligibility and size²¹: all firms ([1 – 17] tax thresholds), eligible or small firms ([1 – 2] tax thresholds), non-eligible firms ([3 – 17] tax thresholds), medium firms ([3 – 15] tax thresholds) and large firms ([16 – 17] tax thresholds).

3.3 Limitations

Two things are important to highlight. First, our data is not in micro level, tax returns are already grouped in 17 thresholds based on their turnover, ranged from 0 to more than 500 000 000 pesos (Table AIV). That is to say, we do not have the turnover, gross profit and taxable income of “firm i in tax year t ” instead we have the turnover, gross profit and taxable income of “all firms in the j tax threshold for the tax year t ”; in other words, our original data is already binned. If we assume all firms within each tax threshold are *homogeneous* we can get the mean value of the turnover, gross profit and taxable income reported by a “representative firm in the j threshold for the tax year t ”. Therefore, we only have 34 observations in each tax year from 1997 to 2000, and 51 observations from 2002 to 2011, making a total of 646 observations for the whole period. This is so because our variable of interest is the profit rate (reported taxable income as a fraction of turnover) and to calculate it we only need two variables, the taxable income and the turnover; also, because from 1997 to 2000 firms could only report positive or negative taxable income; while, from 2002 to 2011 (after the modification of the format of the tax returns in 2001) they could report positive, negative or zero taxable income.

In consequence, we identify three significant limitations. First, the fact that our data is not completely disaggregated introduces measurement error in the bunching estimates. Second, as

²⁰ Note that firms report their taxable income which is the gross profit increased or decreased by the tax base modifying items, so that the profit tax rate is levied on it.

²¹ Based on the “Boletín Oficial de la República de Argentina, No. 24/2001”.

we will see in the next section, the empirical density distribution lacks of completely smoothness. Finally, the lack of smoothness and the aggregated data impede us to calculate the standard errors using the bootstrapped method by random re-sampling from the estimated residuals since the number of observations we have in each tax year is too small.

In the empirical bunching literature, Best *et al.* (2014) use the density distribution of the profit rate to analyze the existence of bunching in the CIT structure of Pakistan for different years (2006/09-2008-2010) and for groups of firms (high-rate vs. low-rate firms). Mosberger (2016) reproduce the methodology of Best *et al.* (2014) for Hungary to analyze the introduction of a minimum tax scheme in the CIT for different years (2008/2009 and 2010/2011) and by economic sectors (construction and manufacturing vs. other sectors). On the other side, Chetty *et al.* (2011) study the discontinuous change of the PIT in the U.S. (1994/2001) in two groups of the population (married women vs. single men, and military vs. teachers) using the density distribution of the distance of the taxable income with respect to the top tax threshold. Similarly, Lediga *et al.* (2016) study the CIT structure in South-Africa, using the density distribution of the taxable income relative to the first tax threshold. Finally, Saez (2010) uses the density distribution of the earnings of self-employed and wage earners to study the effect of the marginal tax rates of the PIT in 1960/1963, 1964/1969 and 1970/1972 in the U.S. We decide to build our bunching evidence on Best *et al.* (2014) because of three reasons. First, Saez (2010) as well as Lediga *et al.* (2016) studies the marginal changes in the tax rates within the general tax regime (PIT and CIT, respectively). However, Best *et al.* (2014) study the marginal changes in the tax rate and the tax base of two different regimes. For this reason, as in Best *et al.* (2014) we use an indicator of profitability, the profit rate, which captures both discontinuous changes. Second, we cannot use the distribution of the taxable income among firms because the cutoff in Argentina depends on the turnover rather than the taxable income, since the criterion of qualification to the SR is the turnover. Moreover, when we use the density distribution of the turnover we have no clear bunching, the same happens when we use the distance with respect to the last category of the SR. Finally, following Best *et al.* (2014) approach allow us to exploit three types of changes, on the turnover tax rate, the profit tax rate and the tax base.

IV. EMPIRICAL METHODOLOGY

Our empirical methodology builds on Best *et al.* (2014), but is restricted to the limitations of our data. With the introduction of the simplified regime, firms are either taxed on their profits (general regime) or on their turnover (simplified regime) based on the following equation:

$$T(y, \hat{c}) = \max[t_{\pi}(y - \hat{c}), t_y y]$$

The tax liability is continuous and at the border firms are indifferent between both regimes when:

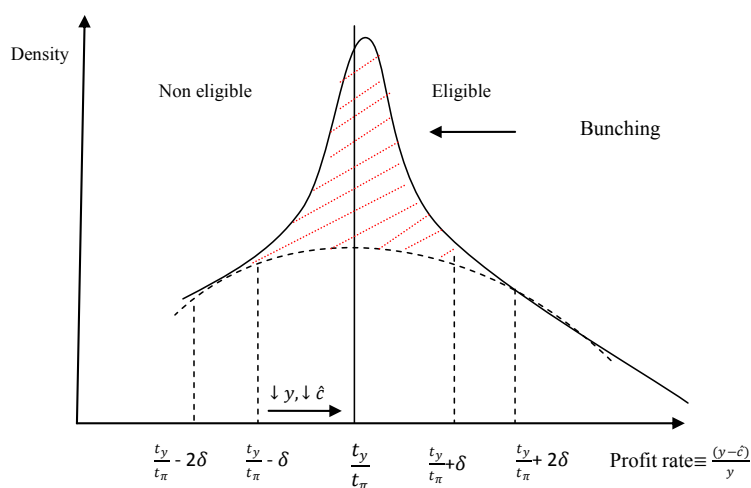
$$\begin{aligned} t_{\pi}(y - \hat{c}) &= t_y y \\ \Leftrightarrow \hat{\pi} &\equiv \frac{(y - \hat{c})}{y} = \frac{t_y}{t_{\pi}} \end{aligned}$$

Hence, the profit rate (the ratio between reported profits and reported turnover) equalizes a fixed cutoff t_y/t_{π} when firms switch between the simplified regime and the general regime. If the

profit rate is above the cutoff then the firm belongs to the general regime; while if it is below, it belongs to the simplified regime. According to Best *et al.* (2014, p.11), the introduction of the simplified regime gives rise to non-standard kink points that affects firms' behavior on real output and compliance decisions differently, authors claim those kink points are non-standard because of the joint change in the tax rate and the tax base.

Under this framework, the introduction of the simplified regime cause an outflow of firms from the general regime to the simplified regime; once in the new regime, firms reduce their real output and decide to comply in costs, both changes (assuming diminishing returns to scale) increases their profit rate and creates bunching at the cutoff. Unfortunately, we cannot observe those behavioral changes because of our data; hence, we take distant from Best *et al.* (2014) and argue that the cutoff separates non-eligible firms (below the cutoff) from eligible firms (above the cutoff), both in the general regime. Note that we call eligible firms to those who meet the turnover criteria of the simplified regime; therefore, they can apply at any moment to the simplified regime and pay a turnover tax. Our hypothesis is that after the introduction of the policy firms in the general regime face two main incentives to shift their profit rate to the right and bunch around the cutoff. That is, medium and small firms have an incentive to reduce turnover 'legally' (a real response) or 'illegally' (an evasion response); first, to reduce reported profits and hence tax liability; second, in the case of medium firms, to become eligible and move to the simplified regime, and in the case of small firms, to be classified in the lowest categories of the SR. In addition, both groups of firms have an incentive to reduce over-reporting costs to be consistent with the information reveal to the tax administration. Note that the behavioral decisions of medium and small firms are observable to us because to apply to the SR firms need to prove they meet the turnover criteria for at least 12 months.

Figure I: Bunching methodology



Source: Based on Best *et al.* (2014, p.30) and Saez (2010, p.187).

In Figure I the dash line represents the distribution before the policy and the solid line the distribution after the policy. Observe that bunching comes from below the cutoff, bunching represents intensive margin responses on real output and evasion²². Based on previous studies

²² Studies as Devereux *et al.* (2012) and Dekker *et al.* (2016) contradict such approach, suggesting that bunching is also the result of extensive responses from firms above the cutoff.

(Saez 2010; Kleven and Waseem 2013; Best *et al.* 2014) we assume a smooth distribution and homogenous responsiveness across firms, that is “there exists a single marginal buncher who reveals the bunching response” (Best *et al.* 2014, p.12). Importantly, in the bunching literature authors as Mosberger (2016, p.6) consider the existence of optimization frictions such as adjustment costs, lack of information and uncertainty which makes the bunching an excess mass around the cutoff rather than a clear spike exactly at the cutoff, we will come back to this idea in the following section.

We now turn to the methodology used in the bunching estimation. We estimate the excess mass at the cutoff, the area between the empirical distribution and the counterfactual distribution for all firms in all years (1997, 1998, 1999/2003, 2004/2009 and 2010/2011)²³. Such excess mass represents both evasion and real responses. Our estimation strategy builds on Chetty *et al.* (2011) and Saez (2010) as well as in the bunching empirical literature (Best *et al.* 2014; Lediga *et al.* 2016, Mosberger 2016, Dekker *et al.* 2016, Devereux *et al.* 2012, Kleven 2016). Three aspects of the methodology deserve our attention.

First, the counterfactual density- what the distribution would have looked like absent the kink-is estimated from the empirical density by fitting a third-order polynomial in almost all cases, excluding data around the kink, as specified in regression (1). The order of the polynomial is chosen to optimize the fit and to keep a reasonable degree of freedom (d.f.=number of observations-order polynomial)²⁴. Second, “the bunching window is chosen as the area around the kink that is visibly affected by the bunching” (Best *et al.* 2014, p.20). For estimation purposes we define a middle band around the kink $[\tau_y/\tau_\pi - \delta, \tau_y/\tau_\pi + \delta]$, and two surroundings bands, $[\tau_y/\tau_\pi - 2\delta, \tau_y/\tau_\pi - \delta]$ and $[\tau_y/\tau_\pi + \delta, \tau_y/\tau_\pi + 2\delta]$ below and above the kink as depicted in Figure I. As in Saez (2010, p.187), the parameter δ measures the width of those bands and its choice matters when estimating excess bunching using regression (1): “if δ is too small, the amount of excess bunching will be underestimated; but if δ is too large it will be overestimated”. As Saez (2010) suggests we select δ graphically to ensure that the full excess mass is included in the band $[\tau_y/\tau_\pi - \delta, \tau_y/\tau_\pi + \delta]$ ²⁵, for this reason we always select the middle band so that the kink points are in the centre. Third, there is an unanimous consensus in the literature that the data must be binned, that is the dataset should be grouped in small intervals (bins); so that the observations are central values (in our case, the mean). Note that our data is already grouped in bins (or tax thresholds) based on firms’ turnover; however, we should mention that those bins are not symmetric.

The estimation methodology has four steps. First, we estimate a regression:

$$d_j = \sum_{i=0}^q \beta_i (\pi_j)^i + \sum_{j=\pi_L}^{\pi_U} \gamma_i \cdot I[\pi_j = i] + v_j \quad (1).$$

²³ In our study the excess mass can be over-estimated because as Dekker *et al.* (2016, p.18) suggest “when pooling data we observe some firms more than once; as a result, we attribute bunching behavior to those firms in every period, although the behavioral decision made is only once”.

²⁴ Note that we use a hypothetical counterfactual rather than the empirical density distribution before the policy because the excess mass of the bunching around the kink is diffuse, also because as we can observe in Figures III, the pre-SR distribution is higher than the post-SR distribution.

²⁵ With a larger dataset and smoother density estimates Saez (2010, p.196) suggests choosing the width parameter with an econometric method. Lediga *et al.* (2016) and Dekker *et al.* (2016), for example, use an asymmetric bunching window determined endogenously by the Schwarz’s Bayesian Information Criterion (BIC). However, the existing literature generally use symmetric bunching windows determined by visual inspection, although the main critique to such method is that it is vulnerable to the researcher’s discretion.

Where, d_j is the number of firms in bin j , $[\pi_L, \pi_U]$ is the bunching window, q is the order of the polynomial, π_j is the profit rate in bin j and j is the index of the bin. Second, we fit a polynomial excluding the data near the kink, the idea is to predict the values from the regression omitting the contribution of the dummies around the kink: $\hat{d}_j = \sum_{i=0}^q \hat{\beta}_i (\pi_j)^i$. Third, we estimate the excess mass that is the number of firms who locate near the kink relative to the counterfactual: $\hat{B} = \sum_{j=\pi_L}^{\pi_U} (d_j - \hat{d}_j)$. Finally, we estimate the empirical bunching, that is the excess mass around the kink relative to the average density of the counterfactual distribution in the bunching window as in Dekker *et al.* (2016, p.7): $\hat{b} = \frac{\hat{B}}{\sum_{j=\pi_L}^{\pi_U} \hat{d}_j / (\pi_U - \pi_L + 1)}$

Finally, our criterion of selection is based on the significance of the estimate ($\hat{b} \geq 2$), on the smoothness of the counterfactual, on the order of the polynomial ($q \leq 3$) and on the size of the band, a smaller band around the kink is always preferred, because according to the bunching approach, the intensive margin distortions created by kinks are very local (Kleven 2016, p.17).

V. RESULTS

First, in Figure II we show bunching evidence for different years (1997, 1998, 1999/2003, 2004/2009 and 2010/2011) and for groups of firms depending on their eligibility (eligible and non-eligible firms) and size (small, medium and large firms). Our goal is to show evidence that firms do indeed bunch and that this bunching is around the tax kink. Second, in Figure III we make a placebo analysis by comparing the empirical density of the profit rate in 1998, 1999/2003, 2004/2009 and 2010/2011 with respect to 1997, the year before the introduction of the SR. Based on the theoretical predictions we expect to see no bunching at 1997. Third, in Figure IV we make identification checks to test if the excess mass at the tax kink is indeed a response to the tax system. In this part of the analysis we exploit the variations in the tax kink by comparing the reforms applied in Dec.1998 ($\tau_\pi \uparrow$), Jul.2004 ($\tau_y \uparrow$) and Dec.2009 ($\downarrow \mu$). Based on the theoretical predictions we expect to see the bunching moving with the different tax kinks. Finally, we use the bunching to estimate the magnitude of the excess mass around the kink point.

We begin our analysis by presenting evidence of bunching, Figure III1 shows the empirical density distribution of the profit rate (reported taxable income as a fraction of turnover) for all firms and different time periods. The figure shows that firms cluster near to the kink point equal to the ratio of the two tax rates $\tau_y(\text{min.})/\tau_\pi$ which are: 1998 (0.1), 1999/2003 (0.09), 2004/2009 (0.09) and 2010/2011 (0.05), (see Table I). This Figure rise one of our three findings, there exist bunching for all firms in all years: 1997 ($\hat{b} = 2.77$), 1998 ($\hat{b} = 2.07$), 1999/2003 ($\hat{b} = 2.72$), 2004/2009 ($\hat{b} = 3.86$) and 2010/2011 ($\hat{b} = 5.82$). Surprisingly, we observe sharp and significant bunching in 1997 and 1998 (the year previous the policy and the year of the policy); moreover, in Figure III1 we observe that bunching in 1997 is sharper than in any year. Two facts can partly explain this unexpected result. According to the tax administration's reports (AFIP, 2006), the introduction of the simplified regime responds to the requests of small taxpayers made to the parliament of Argentina, probably based on the experience of Brazil where the simplified regime was introduced in 1997. So, it is not surprising that Argentineans firms might have anticipated the policy and have modified their behavior in 1997. This

explanation can only rest on the assumption that Argentineans firms have rational expectations. Bunching in 1998 is less astonishing because although the implementation of the policy was in Nov. 1998, it was approved by the parliament in June 1998 and announced on July 1998 through important advertising campaigns; so, firms could have modified their behavior from July 1998 to June 1999 (the last day for firms to present their tax returns to the tax administration). However, bear in mind that by the time the tax changes were announced most of the real decisions corresponding to tax year 1998 had already taken place. Hence, the speed of reaction provides supporting evidence that behavioral changes are driven by evasion responses rather than real responses. A similar intuition is followed by Waseem (2013), Saez (1999) and Mosberger (2016). Therefore, bunching in 1997 and in 1998, confirms the presumption that firms can adapt rapidly their behavior; indeed as Lediga *et al.* (2016, p.7) suggest is more difficult to adjust real output in a short period due to adjustment costs and optimization frictions (i.e. existing long-term contracts).

For a better understanding we divide the full sample by firms' eligibility, the density distributions are plotted in Figure II2 for eligible firms and in Figure II3 for non-eligible firms. At a first glance we observe clear bunching only in 1997 (Panel A) and in 1998 (Panel B) for eligible firms; whereas we observe bunching in all years for non-eligible firms. If we disaggregate non-eligible firms, we see bunching in all years for medium firms (Figure II4); but no bunching for large firms (Figure II5). These findings confirm our initial hypothesis, we expect large firms do not re-optimize their production and evasion behavior because it may be too costly for them to reduce their turnover to the level required by the SR, also those firms are more likely to have international activities and might see a commercial advantage to be registered in the VAT. However, we do expect behavioral changes among medium and small firms. Medium firms have more incentives to decrease their turnover 'legally' or 'illegally' to reduce their tax liability and/or to become eligible and move to the SR, such behavior increases their reported profit rates and move the distribution to the right clustering medium firms around the kink point, as observed in Figure II4 (Panel C and D). Small firms follow a similar pattern, although their goal is to be classified in the lowest categories of the SR and pay a lower turnover tax. Our hypothesis is also confirmed when we compare the number of taxpayers in the years after the policy with the year before the policy, we observe negative growth rates in the immediate years after the policy (1998-2003) in all firms²⁶. More specifically, we observe that the number of eligible and medium firms decreased by 51% and 30% in 1998, and by 68% and 37% in 1999, respectively, and continue decreasing in the subsequent years. As well, the growth rates of the turnover and the taxable income compared to 1997 are negative for small firms in all the period.

Our second finding is consistent with previous studies, Figure III (Panel E), Figure II3 (Panel C and D) and Figure II4 (Panel C) show a slight bunching around the zero profit rate²⁷. Two intuitions might justify this result. First, the high prevalence of losses in taxable income (see Table AI) and zero taxable income may partly be due to generous deductions offered in Argentina which is a common case in developing countries²⁸. For instance, Lediga *et al.* (2016)

²⁶ Note that since 2004 the economy of Argentina improved after more than a decade in crisis (ILO, 2013).

²⁷ As we explained earlier, in 1997 and 1998 we have no bunching around zero because we have no firms reporting zero taxable income since the corporate income tax returns only report positive or negative taxable income. This form of reporting changed in 2001.

²⁸ Firms in Argentina can deduct for: loss carry forward, donations, differences in amortization, salaries to directors, expenditure of representation, contribution to private pension plans, etc. Based on: Law Corporate Income Tax 649/97 (III, art.87).

find bunching at zero income for small firms in South-Africa due to loss carry forward provisions; as well, Dekker *et al.* (2016) find that a large share of bunching in Netherlands was driven by tax deductions. Second, in line with our hypothesis, we think some small firms remain in the GR although they met the turnover criteria but not the objective parameters, reporting a very low taxable income which places the distribution around zero. To address this, we estimate the observed bunching around zero for all firms, non-eligible and medium firms²⁹; however, we got no value for the bunching estimates because the predicted values were zero. The intuitive explanation is that absent the policy, in the hypothetical scenario, there is no firm bunching at the zero profit rate; this means, before the policy there is no firm bunching at zero but after the policy many firms cluster around zero. Therefore, consistent with our hypothesis, there is no other explanation than a response to tax incentives, the policy might have incentive firms to under-report turnover which in turn result in zero taxable income, and hence zero profit rate.

We now turn to our last finding. First, in Figure IV we observe the distribution moves to the right over the period, this move is clear for all firms (Figure IV1), non-eligible firms (Figure IV3) and medium firms (Figure IV4) which strengthens our initial hypothesis. We stress again that the reduction in turnover ‘legally’ or ‘illegally’ and in cost evasion by small and medium firms, increases the reported profit rate and produces a right-shift of the distribution. Surprisingly, we observe the distribution more to the right in 2010/2011 than in any other period, and more to the left in 1997 and in 1998. This confirms our first result, the bunching observed in the immediate years after the policy (1998, 1999/2003) represent mostly evasion responses because the speed of reaction is difficult to conceal with real responses in the presence of optimization frictions and adjustment costs. Second, Table AV provides compelling evidence that firms learn over time. At the beginning, in 1998, firms bunch below the kink point we calculate $\tau_y(\text{min.})/\tau_\pi$ that we call ‘kink point (theory)’; in 1999/2003 the bunch is placed around the kink point, in 2004/2009 exactly at the kink point and in 2010/2011 above it. The idea that bunching is more centered on the kink point over time may indicate that firms adapt slowly. This also explains why bunching estimates in 2010/2011 for non-eligible firms ($\hat{b} = 0.56$) and for medium firms ($\hat{b} = 0.31$) are close to zero and insignificant. Therefore, bunching is not a clear spike at the kink point but instead is a diffuse mass around it; in other words, we have an asymmetric bunching which is sometimes below the kink, other times at the kink and others above it³⁰. However, the ‘kink point (practice)’-where the bunching is placed in reality- is predominantly around 0.09 (see Table AV), this provides strong evidence that firms respond to the tax structure because there is no other reason for firms to cluster around a profit rate of 0.09 other than the presence of the SR. Note that such kink point is calculated using only the taxation component of the single tax ($\tau_y(\text{min})$), without considering the social security contributions, which suggest that firms respond only to the taxation component of the SR.

We found two arguments in the empirical bunching literature that can provide some lights to these findings. First, Kleven (2016, p.23) points out that the creation of a threshold between both regimes makes firms consider the 0.09 threshold as a reference point introduced by the policy. For example, Mosberger (2016) explains the Hungarian asymmetric bunching with this argument, the author argues that in countries where the tax authority is not credible firms bunch

²⁹ Note that we do not estimate for large nor for small firms because of the few observations in both groups.

³⁰ Is important to mention that the bunching estimates are fairly sensible with respect to the choice of the bunching window (lower, middle, upper and all bands) and the order of the polynomial, which confirms that firms bunch diffusely around the kink point.

above the threshold to avoid being audited³¹. Note that in this case firms above the cutoff belong to the GR and firms below the cutoff to the SR. Second, in line with this argumentation, Carrillo *et al.* (2014, p.9) points out that the reported profit rate is one of the key characteristics the tax authority consider when determining whether to audit; for this reason, a very risk averse firm or a firm just below the kink point (in our case, medium firms) might prefer to report the kink point in order to avoid being audited.

Finally, the reforms do not provide too many insights because of three reasons. First, the reform of Dec.1998 ($\tau_\pi \uparrow$) slightly decreased the kink point from 0.1 to 0.09; henceforth in Figure IV we do not observe a substantial move of the distribution, also such change could have been overlapped by the introduction of the SR in Nov. 1998. Second, the reform of Jul.2004 ($\tau_y \uparrow$) increased the turnover tax rate of the highest categories in the SR; therefore, the kink point is not affected because we take the turnover tax rate of the lowest category. Finally, the third reform of Dec.2009 ($\downarrow \mu$) does not affect the ratio of the two tax rates $\tau_y(\text{min.})/\tau_\pi$ because is a change in the tax base.

All in all, three main findings are worth to point out. Our results show evidence that firms do indeed bunch around the kink point; however, contrary to what we expect, bunching in 1997 and in 1998 is sharper than in any year, which can only be explained by rational expectations of firms and evasion responses. Moreover, we found bunching around zero which might be the result of generous deductions and evasion responses. Finally, consistent with our original hypothesis, firms cluster around a profit rate of 0.09 which might suggest behavioral responses to the policy; in particular, to the taxation component of the SR. Due to the lack of data we are not able to go further with the analysis and estimate the responses.

VI. EVALUATION

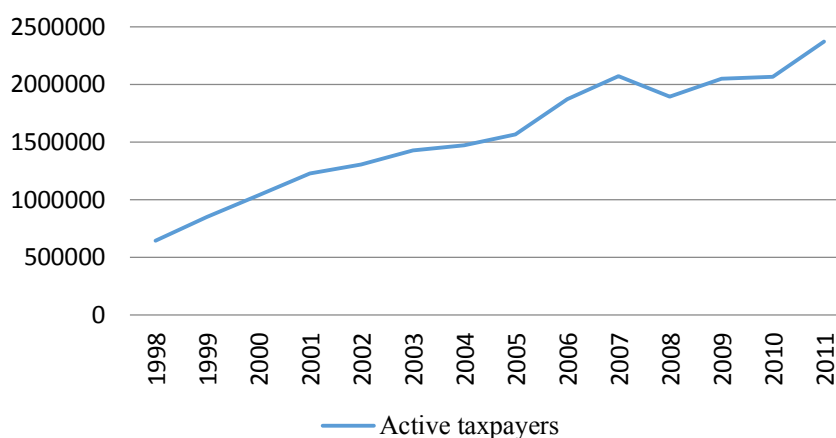
In this last section, we present some general information about the simplified regime that unfortunately we did not have access. As mentioned earlier, the policy had two motivations: fight informality and reduce evasion. Seems that both are the main concerns of the tax administration in Argentina probably because informality and the number of self-employed increased sharply since 1970s. Therefore, small firms are often associated with subsistence entrepreneurship, home-based and low educational levels. Those firms cover a significant part of all salaried employment; for instance, in 2011, 19% of the total employment was in firms of up to 5 employees, and another 19% in firms of 6 to 25 employees (Van Elk and Kok 2014, p.47). In addition, the tax administration reports (AFIP, 2006) refer to the existence of ‘involuntary informality’ as a result of a complex tax system and high administrative costs³². Unsurprisingly, there are few studies about tax evasion in Argentina due to the limited available data, the latest study of Gómez Sabaini and Jiménez (2011, p.29) estimates a tax evasion rate in the VAT of 21.2%, in 2006 and in the income tax of 49.7% in 2005. Since its implementation the simplified regime was successful in attracting taxpayers, as we can observed in Figure II, the

³¹ This line of argumentation supports the idea of Castro and Scartascini (2015) that suggest that the level of compliance also depends on individual’s subjective beliefs about the levels of enforcement and penalties, the behavior of other taxpayers, etc.

³² According to “Doing business 2014” (World Bank), Argentina is ranked on the 126 position over 189 countries in the ease of registering a business.

number of active taxpayers registered in the regime continuously increased and it almost quadrupled in our period of analysis from 642 167 in 1998 to 2 371 469 in 2011.

Figure II: Active taxpayers in the simplified regime



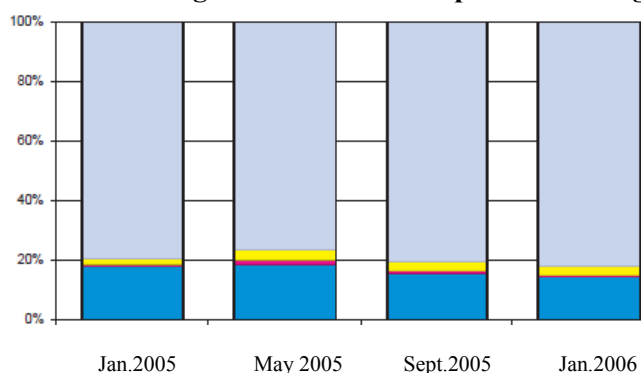
Source: Based on ILO (2013, p.41)

More importantly, the reports (ILO 2013, p.42) show that the majority of active taxpayers in the simplified regime are from the service and commercial sectors; for instance, in 1999, 40.7% of the total number of the registered taxpayers was from the service sector and 34.2%, from the commercial sector. In 2005, the amount increased to 67.4% and 28.7%, and in 2013 it slightly decreased to 61.9% and 18.6%, respectively. On the contrary, the minority are from the industrial and farming sectors probably because those sectors trade with firms registered in the general regime that see as a disincentive to carry out transactions with someone in the simplified regime because of the impossibility of deducting VAT credits from purchases from firms in the simplified regime.

Furthermore, the reports (AFIP, 2006) show that in Feb. 1999 (four months after the implementation of the policy) there were nearly 777 605 adhesions to the regime and by the end of the year the number increased, reaching almost one million of adhesions. According to the reports, 90% of those adhesions were classified in the lowest categories (A, B and C) and around 2% in the highest categories (F, G and H). Moreover, the ILO (2013, p.43) highlights that such numbers continued until 2013, 47% of taxpayers were classified in category A, 46% in 2005 and 54.7% in 2013; whereas only 0.2% of the total taxpayers were classified in the highest category. The agglomeration of taxpayers in the lowest categories of the simplified regime confirm our hypothesis that there exist “icebergs” in the simplified regime, firms who under-report turnover in order to be classified in the lowest categories. As a consequence, there are frequent re-categorizations towards higher categories, as we can observe in Figure III, 79.8% of firms were re-categorized to a higher category in 2005, 82.3% in 2006, 86.7% in 2007, 87.6% in 2008 and 79% in 2009. On the contrary, 17.9% of firms were re-categorized to a lower category in 2005, 14.4% in 2006, 10.8% in 2007, 9.2 % in 2008 and 13.4% in 2009 (ILO 2014:39). As well, the reports confirm that 95.1% of firms registered in 1998 in the simplified regime were from the general regime, henceforth only 4.9% were informal firms. Moreover, from those 95.1% firms, 83.8% of them did not register any tax return in the last year, henceforth were “icebergs” in the GR, and only 16.2% were fully taxpayers. Also, in 2006, 11.4% of taxpayers

registered in the income tax moved to the SR, surprisingly their participation in the income tax revenue was only 0.8%³³ (Van Elk and Kok 2014, p.50).

Figure III: Re-categorization in the Simplified Tax Regime



Source: Report (AFIP June 2006, p.14)

Note: Dark blue corresponds to a re-categorization to a lower category, yellow corresponds to a move from the SR to the VAT, red corresponds to a move from the SR to a sub-regime of the SR for eventual taxpayers and light blue corresponds to a re-categorization to a higher category.

Consequently, despite the high number of active taxpayers in the simplified regime, the tax revenue collected is minimal³⁴ (see Table AI) probably because of the elevated evasion in the regime and because it contains small taxpayers with very low income. The reports (AFIP, 2006) also suggest that the amount of payments in the simplified regime decreased until 2002 and started increasing after 2004. Finally, many features affect the SR as the economic cycle, the inflation and the degree of development. The reports (AFIP, Nov. 2006, p.10) estimate a coefficient of correlation between the amount of payments and the real GDP of 0.84 and between the tax revenue and the real GDP of 0.83 for 1998-2006. As well, inflation affects the simplified regime through prices and wages, for this reason the ILO (2014) suggests that the scarce information available about the simplified regime may be because of its instability due to the continuous updates of the single tax by the tax administration. Moreover, contrary to the general view, the ILO (2013) indicates that cities with high HDI and less informality have a higher number of firms register in the simplified regime, and conversely.

The reports of the ILO (2013, 2014) raise a crucial issue about this policy in Argentina which strengthens our findings. The simplified regime stopped being a “transitory regime” towards formality, a short term policy implemented to motivate a more ‘smooth’ transition of informal firms to formality. This is so, because the weak entry barriers of the SR became strong exit barriers, as ILO (2013, 2014) claim, the SR became a “trap” where small taxpayers are encouraged to remain indefinitely small in order to take advantage of the regime, although most of them have the capacity to register in the GR. As a result, the SR becomes a shelter for “icebergs”, impeding their transition to the GR. Thus, the SR damages revenue and production efficiency because decreases the revenue that can be collected by the general regime and gives incentives to firms in both regimes to under-report turnover. For this reason, the ILO (2013,

³³ For the VAT, Van Elk and Kok (2014, p.50) indicate that only 2.5% of taxpayers registered in the VAT moved to the SR, their payments were only 0.1% of the VAT tax revenue and 61.1% had not made any payment in that year.

³⁴ 70% of the revenue collection corresponding to the SR is for the ANSES (National Administration of Social Security) and 30% is set aside for provincial jurisdictions according to the partnership’s tax regime (Van Elk and Kok 2014, p.47).

2014) argues that the adhesion to the SR should be understood as a short term policy; hence, attention should be paid to the process of "exit" the SR.

VII. CONCLUSION

In this study we analyze the trade-off between production and revenue efficiency faced by governments in developing countries. For that, we select a production inefficient policy, the simplified regime, applied in all Latin America. Under this regime firms are taxed based on their turnover, and have access to a retirement plan and health insurance. We replicate the paper of Best *et al.* (2014) who model the trade-off under the presence of evasion and we introduce turnover evasion to be more consistent with the reality of Argentina. The optimality conditions suggest that in countries with limited tax capacity may be desirable to deviate from full production efficiency in order to reduce evasion. We analyze this empirically using the bunching approach based on the idea that firms are taxed either on their profits or on their turnover depending on which tax liability is larger. Our results are not as robust as we would like to due to the limitations of our data; however, they might give lights to bunching evidence in Argentina, crucial for future research.

Our findings show bunching evidence especially in 1997 and in 1998 among small and medium firms in the general regime. Moreover, bunching around zero corroborates earlier studies that relate it with tax evasion. Furthermore, we found no other explanations for firms to cluster around a profit rate of 0.09 other than a behavioral response to the policy. Overall, our empirical results show that Argentinean firms do indeed bunch around a threshold profit rate; however, further exploration of the anatomy of responses is needed to determine if the bunching is caused by evasion responses. Therefore, contrary to Best *et al.* (2014), we think the policy do not reduce evasion as expected because it increases the number of "icebergs" in both regimes. For this reason, as policy implication, we think the simplified regime should remain as a 'transitory regime' towards the general regime; otherwise, an additional trade-off between informality and evasion should be considered in the choice of tax instruments in developing countries.

REFERENCES

- ABRAMOVSKY, L., KLEMM, A., and PHILLIPS, D., (2014), “Corporate Tax in developing countries: current trends and design issues”, *Institute for Fiscal Studies*, 559-588.
- BESLEY, T and PERSSON, T (2013): “Taxation and Development”, *CEPR Discussion Papers* 9307, C.E.P.R. Discussion Papers.
- BEST, H., SPINNEWIJN, J. and WASEEM, M. (2014), “Production vs Revenue Efficiency with Limited Tax Capacity: Theory and Evidence From Pakistan”, *Journal of Political Economy* 124.
- BOADWAY, R. and SATO, M., (2009), “Optimal Tax Design and Enforcement with an Informal Sector”, *American Economic Journal: Economic Policy*, Vol. 1, N°. 1, 1-27.
- CARRILLO, P., POMERANZ, D. and SINGHAL, M. (2014), “Tax Me if Your Can: Evidence on Firm Misreporting Behavior and Evasion Substitution”, *Working Paper*, Harvard University.
- CASTRO, L. and SCARTASCINI, C. (2015), “Tax compliance and enforcement in the pampas evidence from a field experiment”, *Journal of Economic Behavior & Organization*, 116, 65-82.
- CHETTY, R., FRIEDMAN, J., OLSEN, T. and PISTAFERRI, L. (2011), “Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records”, *Quarterly Journal of Economics*, 126, 749-804.
- DEKKER, V. and STROHMAIER, K. (2016), “A Data-Driven Procedure to Determine the Bunching Window-An Application for the Netherlands”, *mimeo*.
- DEVEREUX, M.P., LIU, L. and LORETZ, S. (2012), “The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records”, *American Economics Journal: Economic Policy*, 6, 19-53.
- DHARMAPALA, D., SLEMROD, J. and DOUGLAS, J. (2011), “Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs”, *Journal of Public Economics* 95, 1036-1047.
- DIAMOND, P. & MIRRLEES, J. (1971), “Optimal taxation and public production I: Production efficiency”, *American Economic Review*, 61(1), 8-27.
- EMRAN, M. and STIGLITZ, J. (2005), “On selective indirect tax reform in developing countries”, *Journal of Public Economics* 89, 599-623.
- Federal Administration of Public Revenue (AFIP), Report June 2006, “Régimen Simplificado para Pequeños Contribuyentes”.
- Federal Administration of Public Revenue (AFIP), Report November 2006, “Evolución de los Regímenes Simplificados para Pequeños Contribuyentes en la República de Argentina”.
- Federal Administration of Public Revenue (AFIP), Statistics-Dataset, <http://www.afip.gob.ar/institucional/estudios/>
- GÓMEZ, J. and JIMÉNEZ, J. (2011), “Estructura tributaria y evasión impositiva en América Latina”, Corporación Andina de Fomento, Documento de Trabajo N° 2011/08.
- GORDON, R. and LI, W. (2009), “Tax structure in developing countries: Many puzzles and a possible explanation”, *NBER Working Paper Series*, N° 11267.
- International centre for tax and development (ICTD): “ICTD Government Revenue Dataset” (ICTD GRD, 2014), <http://www.ictd.ac/en/about-ictd-government-revenue-dataset>
- International Labour Organization (2013), *Desempeño del Monotributo en la formalización del empleo y la ampliación de la protección social*, N°4 Documento de Trabajo.
- International Labour Organization (2014), *Monotributo en América Latina: Los casos de Argentina, Brasil y Uruguay*, FORLAC.
- International Tax Dialogue (2007), “*Small and Medium Enterprises*”-Argentina, October 2007, IMF.

- KANBUR, R. and KEEN, M. (2014), “Thresholds, informality and partitions of compliance”, *International Tax and Public Finance*, Springer, Vol. 21 (4), 536-559.
- KEEN, M. (2007), “VAT, tariffs and withholding: border taxes and informality in developing countries”, *IMF Working Paper*, WP/07/174.
- KEEN, M. (2012), “Taxation and development-Again”, IMF Working Paper, WP/12/220.
- KEEN, M. and MINTZ, J. (2004), “The optimal threshold for a value-added tax”, *Journal of Public Economics* 88, 559-576.
- KLEVEN, H. (2016), “Bunching”, *Annual Review of Economics*, 8.
- KLEVEN, H. and WASEEM, M. (2013), “Using notches to uncover optimization frictions and structural elasticities: theory and evidence from Pakistan”, *The Quarterly Journal of Economics*, 128 (2).
- KLEVEN, H., KHAN, A. and KAUL, U. (2016), “Taxing to develop: When ‘third-best’ is best”, International Growth Centre (IGC), April 2016.
- KLEVEN, H., KNUDSEN, M., THUSTRUP, C., PEDERSEN, S. and SAEZ, E. (2011), “Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark”, *Econometrica*, Vol.79, N°3, pp.651-692.
- KOPZUK, W. and SLEMROD, J. (2006), “Putting firms into Optimal Tax Theory”, *American Economic Review Papers and Proceedings*, 96 (2), 130-134.
- Law No. 24977, Boletín Oficial de la República Argentina, June 1998.
- Law No. 25865, Boletín Oficial de la República Argentina, December 2003.
- Law No. 26565, Boletín Oficial de la República Argentina, November 2009.
- Law Corporate Income Tax, Boletín Oficial de la República Argentina No. 649/97 (III, art.87).
- LEDIGA, C., RIEDEL, N. and STROHMAIER, K. (2016), “Corporate Taxes and Firm Behavior-Evidence from South Africa”, *Working Paper*.
- MOSBERGER, P. (2016), “Essays on responses to taxation”, *Working Paper*.
- MUNK, K. (1978), “Optimal taxation and pure profit”, *The Scandinavian Journal of Economics*, 80, 1-19.
- OECD (2012), “Revenue statistics in Latin America, Special Feature: Taxation and SMEs in Latin America”.
- SAEZ, E. (1999), “Do Taxpayers Bunch at Kink Points?”, *National Bureau of Economic Research*, Working Paper 7366.
- SAEZ, E. (2010), “Do Taxpayers Bunch at Kink Points?”, *American Economic Journal*, 180-212.
- SCHNEIDER, F., BUEHN, A. and MONTENEGRO, C. (2010), “Shadow Economies All over the World: New Estimates for 162 Countries from 1999 to 2007”, *Policy Research Working Paper* 5356, World Bank.
- VAN ELK, K. and KOLK, J. (2014), “Enterprise formalization: fact or fiction (A quest for case studies), ILO, GIZ and BMZ.
- WASEEM, M. (2013), “Taxes, Informality and Income Shifting: Evidence from a Recent Pakistani Tax Reform”, *Working Paper*.
- World Bank, “Doing business 2014”, <http://www.doingbusiness.org/data/exploreeconomies/argentina>