Taxpayers' Responses to Tax and Administrative *Notches* across the Revenue Distribution*

Santiago Garriga† Jorge Puig‡ Darío Tortarolo§

Abstract

We estimate the response of firms and self-employed workers to two revenue taxes—Monotributo and Gross Receipts Tax—across the turnover distribution using rich administrative data from Argentina. We exploit several revenue-dependent discontinuities that provide incentives to underreport taxable income combined with a bunching design to estimate sales elasticities. We also explore heterogeneities by firm size, type of activities, and type of taxpayers. In the case of small firms, we find sizeable bunching below the thresholds that is stronger for higher tax incentives. The response is larger in sectors that have more space for manipulation, such as service-based activities. In the case of medium and large firms, bunching is more muted but even large firms are able to underreport their gross sales to avoid facing higher costs; in particular, these firms seem to find more costly the indirect administrative cost of becoming a collection agent than the direct fiscal cost of the Gross Receipts Tax. We cannot rule out, however, that large firms adjust other margins to compensate for the higher tax pressure.

JEL CODES: I38, H22, J46

KEYWORDS: elasticity of taxable income, notches, bunching, self-employed

*We would like to thank Alan Auerbach, Antoine Bozio, Julien Grenet and Emmanuel Saez for research guidance and very fruitful discussions. Darío gratefully acknowledges financial support from the Burch Center for Tax Policy and Public Finance.

†Paris School of Economics, France. Email: garrigasantiago@gmail.com

‡Department of Economics, UNLP. Email: jorge.puig@econo.unlp.edu.ar

§Univ. of Nottingham & Institute for Fiscal Studies. Email: dario.tortarolo@nottingham.ac.uk

1 Introduction

A central question in public and labor economics is how workers and businesses respond to governments' action within the economy e.g., to taxes, subsidies, welfare programs, etc. Pioneer studies showed simple correlations between some of these policy options and different outcomes, ignoring most of the endogeneity problems. The marginal revolution and the availability of large administrative datasets, have led to a boom of research that tried to understand what are the behavioral responses to government policies (Angrist & Pischke, 2010). Nevertheless, the evidence is still scarce and inconclusive, particularly in developing countries where governments are more reluctant to share this type of information.

The vast majority of the empirical literature has focused on labor supply responses to taxes considering the standard margins; the extensive i.e., decision to work or not, and the intensive i.e., how much to work. However, individuals may react in many dimensions including job effort, career choice, form and timing of compensation, and tax avoidance and evasion, among other dimensions. Therefore, a proper measure of the costs induced by taxation should consider all these behavioral responses; and, in principle, these margins of response to a certain tax could be captured by the reported income. Indeed, Feldstein (1999) stated that observed responses in taxable income provide a broader concept of behavior as compared to the traditional labor supply analysis, where the elasticity of taxable income (ETI) summarize all possible responses. Therefore, if the ETI is greater than the labor supply elasticity, the resulting deadweight loss induced by taxation will also be greater.[1]

In this paper we study the response of self-employed workers and firms to two different revenue taxes, the *Monotributo* and the Gross Receipts Tax.[2] Our laboratory is Argentina, which offers very rich administrative data and quasi experimental variation provided by discontinuities in both tax schedules. The *Monotributo* is characterized by eleven income categories with different tax liabilities. This federal tax is a simplified regime for small entities that replaces the income and the value added tax. The Gross Receipts Tax is characterized by a three-bracket schedule with tax rates and thresholds that vary by sector and are increasing in income. This is a sub-national tax levied on gross receipts in every stage of the supply chain. Importantly, the discontinuities of *Monotributo* affect relatively small businesses and self-employed workers, while those embedded in the Gross Receipts Tax affect relatively large firms. The number of discontinuities in the tax liability, jointly with the existence of two taxes that share the same tax base, provide an ideal setting to estimate revenue responses for taxpayers located at very different parts of the gross income distribution. We look at small and large taxpayers, as well as self-employed workers and firms that vary in size.

[1] The study of the elasticity of taxable income is relatively new as compared to the standard labor supply elasticity (Saez *et al.*, 2012).

[2] The Gross Receipts Tax is also known as Turnover Tax (in Spanish, *Impuesto sobre los Ingresos Brutos* or IIBB). We use these terms interchangeably.

Besides the experiment, an interesting feature of our setting is that it takes place in a context of relatively high evasion, informality, inflation, and somehow, weak enforcement.

Our research design exploits the presence of *notches* along the revenue distribution to estimate behavioral responses to taxes using the bunching approach developed by Kleven & Waseem (2013).[3] This technique approximates individuals' responses to a given policy by comparing the observed distribution of taxpayers with a counterfactual one. Under some conditions, it can be shown that the excess of mass of the observed distribution around a given *notch*, can be understood as the individual response to the tax, and therefore be interpreted as the elasticity of taxable income.[4]

Our findings show substantial bunching of small taxpayers below the thresholds of the first eight categories within the *Monotributo* regime. The absence of *missing mass* above the discontinuities suggests that some entities face substantial adjustment frictions or inattention. The intensity of bunching is stronger for higher discrete jumps in average tax rates. In the remaining three notches, however, bunching is less evident, which can be justified by a higher composition of taxpayers with less flexibility to adjust their reported revenue. We also document that taxpayers in services react the most, followed by agriculture and manufacturing, and finally wholesale and retail. We interpret these behavioral responses as suggestive evidence that a significant number of taxpayers attempt to avoid higher tax liabilities by keeping their reported revenue below the thresholds.

In the case of medium and large taxpayers (approximately above the percentile 75 of the revenue distribution), bunching to turnover tax notches is not as striking as in the first eight notches of *Monotributo* and our results show that large firms are not able to manipulate their reported sales to avoid facing higher tax rates. However, we find some bunching for an administrative discontinuity that forces firms to work as collection agents if they cross a revenue threshold. This could imply that firms find more costly this indirect administrative cost than the direct fiscal cost of the turnover tax. When we split the analysis by sector we find that the manipulation is mainly driven by service and manufacturing firms. As a caveat, our analysis cannot identify, however, general equilibrium effects e.g., cascading, or rule out that taxpayers react in other dimensions e.g., fiscal externalities to other tax bases. We provide, nevertheless, some suggestive evidence on shifting behavior of tax bases across different sectors.

The main contribution of our paper is that we are using a wider distribution, as compared with previous papers, to estimate the elasticity of taxable income. We combine the presence of a large number of number of *notches* with administrative data. A second contribution, although minor, is that we are estimating

[3] Saez (2010) initiated this literature by exploiting *kinks* at which *marginal* tax rates jump to estimate taxable income elasticities. Kleven & Waseem (2013) extended this methodology to consider settings with *notches* i.e., where *average* tax rates change discretely. The survey by Kleven (2016) summarizes the growing literature using bunching estimation techniques.

[4] Indeed, Kleven & Waseem (2013) show that the compensated elasticity is a function of the change in the tax liability and of the change in reported taxable income.

these revenue elasticities for a middle-income country where, for instance, access to micro-data is less likely as compared to developed countries. Previous is not trivial, as the elasticities that result from developed countries studies may no be easily mapped into less advanced contexts as evasion opportunities and enforcement strategies are rather different. Our results are in line with those recent estimations is similar country contexts as in Pakistan, Costa Rica and Honduras.

Our paper contributes to a recent but growing literature on firm responses to tax and administrative regulations in developing countries. A strand of literature exploited *kink* points in firms' budget sets provided by minimum tax schemes where firms pay the largest tax liability between a tax on profits and a tax on turnover e.g., Best *et al.* (2015) in Pakistan and Alejos (2018) in Guatemala. This policy is motivated by the idea that the broader turnover base is harder to evade. In Argentina, however, firms face both national corporate taxes and subnational turnover taxes. In the same vein, Bachas & Soto (2018) use *notches* in the tax schedule to estimate profits' responses to corporate taxes in Costa Rica. They estimate a very high elasticity of reported profits, which is entirely driven by evasion and cost-deductibility, providing support in favor of taxing turnover.

More generally, our article is related to recent research on firm responses to revenue-dependent regulations that provide incentives to underreport revenue, such as administrative *notches* and the value added tax (VAT) registration thresholds where firms are required to register and charge VAT on all sales e.g., Onji (2009) in Japan, Boonzaaier *et al.* (2017) in South Africa, Rauhanen *et al.* (2016) in Finland, Liu *et al.* (2017) in the U.K., and Asatryan & Peichl (2017) in Armenia. Almunia & Lopez-Rodriguez (2018) also show that in a more developed economy, such as Spain, firms underreport revenue to avoid stricter monitoring from the tax authority. Overall, this literature provides suggestive evidence that income under-reporting drives the bunching response of firms. Note, however, that a common feature of all these papers is that the regulations allow to study either small or large firms. Our setting instead applies to small, medium, and very large firms, and thus lets us measure the responses across the revenue distribution in a comparable way.

The magnitude of firm responses to features of the tax code is of utmost importance in the formulation of tax policy and the assessment of efficiency costs.[5] From a policy perspective, this article pins down policy-relevant elasticities that enable a better understanding of two taxes for which there is very limited causal evidence. Evaluating tax-filers' behavioral responses across the turnover distribution is important for several reasons. First, there are still inconclusive estimates of the magnitude of this elasticity, especially for developing countries where the institutional setting differs to the one present in developed economies. For instance, the possibility to shift income due to lower enforcement capacities is higher. Second, unlike previous studies, our setting and data allow us to study the response of large firms, small firms, and self-employed workers (with presumably different elasticities) in a comparable way. Namely, by exploiting

[5]Indeed, Besley & Persson (2013) argue that "... tax design in a developing country context has to take into account the information about behavioral responses needed by governments ...".

discontinuities in average tax rates and using the same source of administrative data which contains the reported tax base that applies to both taxes. The self-employed typically have more flexibility, and therefore the reaction to the policy could be greater. Third, the results from this study will be of great value for national and sub-national governments in developing countries with similar tax systems. Moreover, our paper could also be of interest to policymakers in more developed federal countries, such as the United States, where some states currently have a gross receipts tax e.g., Washington State, and some others are considering its implementation e.g., Oregon. As such, the results from this research exhibit great promise of informing policy debates.

This paper is structured as follows. Section 2 describes the institutional details of the two taxes that are analyzed in the empirical section. It also briefly describes the administrative data. Section 3 presents the analytical framework and behavioral responses at the lower end of the gross income distribution. Section 4 focuses instead on the responses at the upper end of the distribution. Finally, Section 5 ends with some concluding remarks.

2 Institutional setting and data

Argentina is a federal country where businesses' activity is taxed both at the national and the sub-national level. At the national level, medium and large firms face the value added tax (VAT) and the corporate income tax (CIT), and self-employed and small businesses face a simplified tax regime called *Monotributo*. In addition, at the sub-national level firms and self-employed workers are taxed by the Gross Receipts Tax (GRT). Taxes at the national level are collected by the Federal Administration of Public Revenue (AFIP, in Spanish), while taxes at the sub-national level are collected by each provincial i.e., state level, tax administration.

The *Monotributo* and GRT are both levied on gross income before costs and taxes are deducted, which make them more distortive than other standard taxes. In addition, their differential tax schemes provide quasi-experimental variation at different segments of the revenue distribution and different incentives to respond to the tax. In this paper, we will exploit such variation to analyze the response of different taxpayers to these two taxes. For data availability, we focus our attention to the Province of Buenos Aires (PBA), which is the largest province in Argentina in terms of population (16 million), area (11 percent of the country), and gross domestic product (it has the largest gross state product, almost 40 percent of the national total). This implies that if PBA were a country, it would be the sixth economy in Latin America, very close to Chile. In the rest of this section, we describe the specific features of each tax and how they can affect firm's incentives.

2.1 Monotributo

Monotributo is a simplified tax regime for small taxpayers that was created in 1998 with the goal of inducing self-employed and small firms to enroll into the formal system by the simplification of tax duties as well as social security contributions i.e., pension and health coverage.[6]

This regime has three main components, one that refers to taxes and the other two to social security contributions (SSC). First, the regime simplifies the tax duties of taxpayers by the unification of two pre-existing taxes, personal income tax and value added tax. The result of this is an integrated tax, *impuesto integrado*, that gave also the name to those registered into this system i.e., *monotributistas*, and also to the tax i.e., *Monotributo*. Secondly, it contains a component linked to contributions for retirement pension. These are then accounted in the Argentinean pension system (SIPA, *Sistema Integrado Previsional Argentino* of its acronym in Spanish). Finally, the last component comprises health contributions, generally known as *obra social*, and provides access to health care.

Taxpayers are subject to a combined monthly fee that varies by gross income level accrued in the last twelve months based on eleven categories. Table 2 presents these categories and its corresponding upper thresholds since January 2010. The monthly fee comprises a tax that varies by bracket (see Table 3) and a flat SSC component for retirement and health that do not vary by income level (see Table 4). Taken together, this means that if a taxpayer makes 1 cent more than the upper threshold, then the tax fee increases discretely while the SSC component remains constant. Hence, the structure of the regime is characterized by eleven discontinuities or *notches*, as depicted in panel (a) of Figure 1. Two points are worth noting. First, the average tax rate is overall increasing across brackets and decreasing within brackets (this is due to the structure of the tax i.e., a fixed fee for those in a given bracket). Second, the tax schedule is such that the average tax rate increases from 1 to 6 percent with eleven *notches* that create space for behavioral responses. These discontinuities provide strong incentives to locate to the left of the thresholds in order to pay lower taxes.

2.2 Gross Receipts Tax

Each of the 24 provinces into which Argentina is divided imposes a tax on gross revenues from the sale of goods and services, the so called *Impuesto sobre los Ingresos Brutos*. This Gross Receipts Tax (GRT) is the main source of own revenue in all of these jurisdictions. In the Province of Buenos Aires (PBA), the laboratory of this article, it represents about 75 percent of tax revenue and currently raises 4

[6]See law (Law 24.977). As of September 2017, there were about 1.5 million registered self-employed and small firms, which represents 12 percent of the mass of registered workers. Monteiro & Assunção (2012) and Fajnzylber *et al.* (2011) analyzed the effects of a similar program in Brazil called SIMPLES. Both papers found that the program leads to an increase in the probability of registering a business and thus becoming formal.

percent of the national GDP (Figure 21).[7]

The GRT is determined on the basis of gross income accrued by any firm or independent worker selling goods or providing services inside a particular province. Tax rates vary according to the taxpayer activity, annual turnover from the previous year, and the place where the transaction takes place (inside or outside the province). Taxpayers are classified as *contribuyentes locales* when performing the activity inside the province or as *contribuyentes de convenio* when they operate both inside and outside the province (this set of firms is part of the so-called Multilateral Agreement). For firms operating in multiple provinces, there exists an apportionment regime to distribute the tax base. The formula is based on a sales factor and an expenditure factor determined at year t_{-1} (*coeficiente unificado*). In particular, 50 percent of total revenue is distributed according to the ratio of a taxpayer's sales in the taxing province to its overall sales; and the other 50 percent is distributed according to the ratio of a taxpayer's expenditures in the taxing province to its overall expenditure. Surprisingly, firms selling in foreign provinces might be taxed at differential tax rates as if there were internal customs; these are known as "Foreign Jurisdiction" firms.[8]

In the province of Buenos Aires, economic activities are classified into 791 sectors that can be taxed either under a general or a differential regime. Within the *general regime*, there are 647 activities that face a three-bracket progressive tax schedule based on annual turnover from the previous year (or the first two months when the firm is new). That is, a firm's tax rate at year t depends on whether annual turnover at year t_{-1} is above or below a threshold. If a firm crosses the threshold, then the tax rate jumps and is applied to all the tax base.[9] Hence, this revenue threshold represents a *notch* because the average tax rate changes discretely.[10]

The provincial Tax Act from Buenos Aires defines three salient features of the GRT that are critical for the empirical strategy. Without loss of generality, in Figure 1 panels (b)-(d), we summarize these features for the year 2015 (also in Figure 30 we present an extract of the tax code). The law establishes progressive tax rates based on annual gross revenue from the previous year, y_{t-1}, for three broad groups of activities classified in the law as follows:[11]

- **Type A (wholesale and retail)**: the main article sets a tax rate of 5 percent; another article reduces the rate from 5 to 3.5 percent when $y_{t-1} \leq 40$ million pesos; and another article reduces the rate from 3.5 to 3 percent when $y_{t-1} \leq$

[7] The other taxes levied at the provincial level are the following: stamp, car and property, and inheritance tax.

[8] "Foreign Jurisdiction" are mostly firms that are based in a different jurisdiction and that do not have an establishment within the province e.g., they sell in the province of Buenos Aires but they are based in other province and do not have any establishment inside PBA.

[9] Note that the tax base coincides with monthly turnover for *contribuyentes locales* but not for *contribuyentes de convenio* where the apportionment formula must be applied first.

[10] In the *differential regime*, the remaining 144 activities are taxed at differential flat tax rates.

[11] The tax was initially designed such that tax rates increase as the product or service approach the final consumer to alleviate inefficiencies in the supply chain.

1 million pesos. "Foreign Jurisdiction" firms are always taxed at 5 percent.

- **Type B (services)**: the main article sets a tax rate of 3.5 percent; another article increases the rate from 3.5 to 4 percent when $y_{t-1} > 500$ thousand pesos; and another article increases the rate from 4 to 5 percent when $y_{t-1} > 30$ million pesos.

- **Type C (agriculture and manufacturing)**: the main article sets a tax rate of 4 percent; another article sets a rate of 1.75 percent for firms based in PBA with $y_{t-1} > 60$ million pesos; another article sets the rate to 0.5 percent for firms based in PBA with $y_{t-1} > 40$ million pesos; the remaining firms based in PBA with $y_{t-1} \leq 40$ million pesos face a 0 percent tax rate. Hence, in practice "Foreign Jurisdiction" firms are the only ones with a statutory tax of 4 percent.

Formally the tax rate in a given sector is calculated as follows $tax\ rate_{i,j,t} \rightarrow f(turnover_{t-1})$ where *turnover* refers to total sales during previous year (t_{-1}), and i indexes sectors, j months and t years. While the tax base in each sector is naturally defined as total sales in that particular sector in a given month and, therefore, the total amount of tax paid in that sector would be: $tax\ paid_{ijt} = turnover_{ijt} \cdot tax\ rate_{i,j,t}$.

Intuitively, each firm has to discriminate its taxable base by the type of activities they carry out and apply the corresponding tax rate in each case. For example, a large manufacturing firm with gross sales of 65 million in 2014 that also sells to final consumers would be taxed at 1.75 percent for manufacturing sales and 5 percent for retail sales. Importantly, every province passes a new law every year where they either rectify and/or revise tax rates and revenue thresholds. In fact, these parameters have been modified several times during the period of analysis, as shown in Table 5.

The filing timeline for the GRT can be summarized as follows (see Figure 29). Around November each year, the tax code for the following fiscal year is approved.[12] For instance, the tax law that was enforced during the fiscal year 2015 was approved by the end of 2014, and the rates that were paid during 2015, were a function of total turnover in 2014 (t_{-1}). In addition, the deadline to file the tax of previous year e.g., 2014, go, on average, up to three months within the calendar of the following year. This means that when taxpayers have to report the tax base i.e., total turnover, they are aware of the potential rates they will face depending on how much they declare. Therefore, they have room to manipulate the current tax base in order to end up on the left side (low liability) of the cut-off.

There is also a regulatory provision that forces firms with y_{t-1} greater than a certain cutoff to work as *collection agents* for the tax authority. These firms are required to withhold the GRT at source from suppliers and non-final customers as payments on account. While this regulation improves the enforcement and revenue potential of the tax, it also increases significantly the accounting costs of

[12] Note that in Argentina, the fiscal and the calendar year coincide.

the affected companies. Between 2004 and 2009, the threshold that determines whether a firm must act as a collection agent remained fixed at 5 million pesos. In 2010, it was updated to 10 million pesos and remained fixed until 2016 when it was updated to 20 million pesos. The lack of update of the amounts that determine this duty during 2004-2009 and 2010-2015, and the context of high inflation led many SMEs to join the system, causing additional costs due to the lack of administrative structure to meet this fiscal requirement. It is also said that the duplication of the threshold in 2016 affected 30 percent of the collection agents, most of them SMEs, which benefited from a significant cost reduction in their administrative structure.

All these features and the variation (or lack of variation) of some parameters over time provide an ideal setting for estimating the response of firms to tax and administrative *notches*, as explained in the following section. In this paper we will focus on the activities affected by the tax *notches* of the general regime and the administrative *notch* to become a collection agent. We leave the analysis of specific activities taxed at differential tax rates for future research.[13]

From a theoretical point of view, these taxes on gross income are usually criticized due to the introduction of multiple distortions. The main point is that the tax is levied at every stage of the production process, including intermediate business-to-business purchases of supplies, raw materials, and machinery. Besides the distortion of input prices, cascading effect represents a distinctive feature, since the sale price of a good includes the tax accrued in all the previous stages. As a result, cascading encourages vertical integration since firms may seek to avoid taxes on their inputs by producing them in-house. It also affects the competitiveness of goods that are locally produced and traded, either because in the export they cannot fully recover the locally imposed taxes or because in the import they face goods that come from countries with neutral tax schemes, free from indirect tax burden to exportable products. Finally, cascading could increase the cost of capital when capital inputs are taxed, affecting productive efficiency (Libonatti (1998), Moskovits & Susmel (2006) and Keen (2014)).

2.3 Administrative Data

To analyze the behavioral response of taxpayers to the different discontinuities we use administrative tax records from the Province of Buenos Aires, Argentina, for the period 2011-2016. The data cover the universe of firms and self-employed workers who must file their GRT returns electronically on a monthly basis. It contains about 750,000 taxpayers per month and includes standard information such as masked identifiers, type of activity (around 438 different codes), type of

[13]These tax rates vary widely across activities and over time. The number of tax rates has been increasing over time. While in 2007 there were 7 tax rates that vary between 0.1 and 6 percent, in 2016 there were 18 tax rates that vary between 0 and 12 percent. For instance, between 2012 and 2013 tax rates went from 8 to 12 percent for bingo rooms and slot machines but remained constant for the other recreational activities. This represents an interesting setting to perform a difference-in-difference analysis.

business (LLC, Inc, Partnerships, etc.), monthly revenue, monthly tax liability, tax rates, an indicator for exempt activities, tax withholdings, etc. Importantly, all the records were de-identified and all the computations were performed at the Ministry of Economy of the Province of Buenos Aires.[14]

A key feature of this dataset is that it contains the tax base of the two taxes that serve as a quasi-experiment. Gross revenue, is the variable that determines in which category taxpayers fall within the *Monotributo* scheme, and the same variable is used to determine the tax rate and tax liability of the GRT. To the best of our knowledge, this is the first study that exploits this rich administrative database for academic purposes.

3 Behavioral responses at the lower end

In this section we analyze the behavioral responses to revenue taxation of self-employed workers and small businesses located at the lower end of the income distribution. To this end, we focus on the *Monotributo* simplified tax regime and use quasi-experimental variation provided by the discrete increases in average tax rates.

3.1 Conceptual framework

This regime is structured into eleven brackets, including eleven *notches*, as shown in Figure 1 panel (a) and Table 2. The tax schedule can be formalized in the following way:

$$T_b(z) = F_b + [\Delta F_b] \cdot \mathbb{1}(z > z_b^*) \tag{1}$$

where F_b is a fixed tax liability that should be paid in bracket b when revenue is below the *notch* z_b^*, ΔF_b represents a discrete increase in the tax liability above the threshold, and $\mathbb{1}(.)$ is a indicator function for being above this cutoff. This type of *notch* is generally known as a *pure notch*.[15]

Tax *notches*, as opposed to *kinks*, where the marginal tax rates change discretely, create a strictly dominated area under which it is possible to increase both, consumption and leisure at the same time.[16] [17] As a consequence, it generates strong incentives to *move* from a region above the threshold (z^*) to a point below. This happens due to the existence of an implicit marginal tax rate of more than 100 percent over a window of earnings imposed by the *notch*. As long as ΔF_b is

[14] The Ministry of Economy of Buenos Aires is in charge of designing the provincial tax policy and passes a new tax law every year.

[15] The marginal tax rate on either side of z_b^* remains the same, but individuals with before-tax revenue greater than z_b^* incur a discrete increase in their tax liability.

[16] This holds only for *downward notches*, that is to say either in situations where there is an increase in the tax liability or a decrease in the transfer's amount.

[17] Technically, the dominated area is defined such that net earnings in z^* (with tax liability F_b) are equal to net earnings in z^d (with tax liability $F_b + \Delta F_b$).

greater than one, then the taxpayer that crosses the threshold pays a tax liability that is larger than the marginal peso. Therefore, in a frictionless world, the empirical distribution of turnover should present a big mass below the threshold, where the tax liability is lower, and a missing mass above it. This gives rise to *excess bunching* and a *hole* in the earnings distribution.

Under the *Monotributo* scheme, individuals with practically the same gross income can face very different tax liabilities (and therefore their after-tax income will be different). Analytically, individuals face a budget constraint with *notches* at which they have incentives to *bunch*. Figure 2 presents graphically the analytical framework for the marginal "buncher". Without the tax, this agent would reach the indifference curve U_0. With the tax, the new budget constraint will present a *notch* at z^* and the taxpayer reaches the indifference curve U_1. In this case, she will be indifferent between the interior and the corner solution. Two things are worth noting; first, the figure clearly shows that two individuals with practically the same before-tax income, end up with different after-tax income. Second, the tax incentivizes individuals to the right of z^* to reduce their reported income, or work less, in order to pay a lower tax. Hence, bunching is created by individuals coming from above the *notch*.

Counterfactuals and mass. The identification of bunching and missing mass require comparing the empirical distribution with an estimated counterfactual as shown by Kleven & Waseem (2013). The latter is estimated by fitting a flexible polynomial to the empirical density in the following form:

$$c_j = \sum_{i=0}^{p} \beta_i \cdot (z_j)^i + \sum_{i=z_L}^{z_U} \gamma_i \cdot [z_j = i] + v_j \tag{2}$$

where c_j is the number of taxpayers and z_j the income level in a given bin j, p is the order of the polynomial, Z_L and Z_U are the lower and upper bounds of the *excluded range* area, which is generally the area affected by bunching responses. The counterfactual distribution is the prediction that results from the previous regression, with the only exception that those in the excluded area are fitted without including the second term of the regression i.e., dummy variables for each bin in the excluded range. Excess bunching is estimated as the difference between the observed distribution of taxpayers and the counterfactual one within the range between the lower bound (Z_L) and the *notch* (z^*). Likewise, the missing mass is estimated in the same way but focusing on the window between the *notch* (z^*) and the upper bound (Z_U).[18] Kleven & Waseem (2013) also define the share of unresponsive individuals, those that suffer from some optimization frictions and thus, are unable to locate below the *notch*, as the share of individuals located in the dominated region over the number predicted by the counterfactual distribution.

The natural question one may impose to this methodology is how to determine the excluded range. We are agnostic on this and follow Kleven & Waseem

[18] Formally, excess bunching is defined as $\sum_{j=z_L}^{z^*}(c_j - \hat{c})$ while missing mass as $\sum_{j>z^*}^{z_U}(\hat{c} - c_j)$.

(2013), and we then try to do several sensitivity analysis to check the robustness of our results. As they claim in their pioneer work, the lower bound is visually determined as the point where the excess bunching starts to be notorious. The upper bound is somehow more ambiguous to determine. They propose that this threshold should be determined such that the excess bunching (below the *notch*) equals the missing mass (above the *notch*).

3.2 Empirical evidence: sharp bunching

To test the predictions from the previous conceptual framework, we start by simply plotting the empirical distribution of earnings pooling taxpayers in fiscal years 2014, 2015, and 2016 (Figure 3). For clarity, we split the figure into two panels: panel (a) presents the first four *notches* and panel (b) the following seven *notches*. Taxpayers are grouped in bins of $1,000 and, the red dotted vertical lines, identify the location of the *Monotributo* tax *notches*. There are various interesting facts to highlight from this figure. First, we observe sharp bunching in every *notch*, except the three highest located at the right tail of the distribution. Second, we do not observe, at least in this figure, a hole above each *notch* suggesting that some taxpayers' may be inattentive or unable to move below the cutoff. Third, there is also some bunching, in other parts of the distribution, that is unrelated to the presence of *notches* e.g., at 60, 120 and 180 thousands.

Relatedly to the last point, in Figure 4 we present, once again, the empirical distribution of earnings for a non-rounder sample where we remove taxpayers whose reported taxable income is a multiple of $1,000$. We observe the same bunching as in the previous figure although somehow attenuated and, at the same time, a smoother distribution and no bunching at non-*notches* points. Hence, this suggests that the spurious bunching in Figure 3 is mostly driven by rounding behavior of taxpayers when reporting their taxable income.

To compute bunching, missing mass, and counterfactuals, we zoom in on a specific discontinuity, the *notch* at $400k$ pesos (Figure 5). We first group and count taxpayers by mutually exclusive bins and plot this as a function of the value of the bin (blue line with crosses). The solid green line is the result of the predicted values of regression (2) with the caveat within the excluded range that we already mentioned. The red vertical line indicates the *notch* and the dashed black line locates the income value that makes a taxpayer indifferent between this point and the *notch*. The area between these two lines represents the dominated range. The remaining dashed gray lines correspond to the lower and upper bounds, respectively.

The figure confirms various interesting facts. First, there is a significant excess of bunching below the *notch* of around 2.6 times the height of the counterfactual distribution. This means that there is 2.6 times the density that should be expected. In the absence of the *notch*, the marginal "buncher" would have an income of 421,000 pesos, 5 percent higher than the threshold. The bunching estimate of $b = 2.6$ implies that on average "bunchers" reduce taxable income by

2.6 bins or 2,600 pesos. Second, although there is some evidence of missing mass above the *notch*, a substantial amount of taxpayers suffer from some sort of friction or inattention (approximately 80 percent). Third, the start of the bunching behavior is somehow clear and identified by z_L. Fourth, although not perfect, the upper bound is such that the excess bunching and the missing mass are very similar. For completeness, in Figure 6, we repeat the analysis for the first eight *notches*. The main message of this set of figures is roughly the same as above. There is significant bunching right below of each *notch* and relatively little missing mass above.

Robustness. The previous conclusions remain unchanged when exploring additional robustness checks. We explore the following alternative exercises: [a] consider the full sample including multiples of $\$1,000$ (Figure 22); [b] focus on a balanced panel of taxpayers that have filed taxes in years 2014, 2015, and 2016 (Figure 23); [c] break down the bunching analysis by year (Figure 24 and 25); and finally [d] group taxpayers into bins with different sizes (see Figure 26 for three alternative bin sizes). These various exercises are, somehow reassuring, and confirm that the behavioral response that we are documenting is neither spurious nor driven by any ad-hoc decision.[19]

Moreover, as a placebo test or alternative counterfactual, we can exploit the fact that the four notches located at $192k$, $240k$, $288k$ and $400k$ did not exist before 2014. In Figure 7 we plot the pre (green) and post (blue) earnings distribution using the first one as control. We verify that there was no bunching at those thresholds during the pre-2014 fiscal years when, indeed, these notches didn't exist.

Discussion. In the preceding analysis, we documented substantial bunching right below the *Monotributo* thresholds. A natural question is why some notches exhibit more bunching than others. This result can be justified by the first stage reported in panel (a) of Figure 1. In this figure we can see that the average tax rate increases the most at $288k$ and $400k$, which are precisely the two notches that display the sharpest bunching in Figure 4. Hence, reassuringly the differential bunching intensity is consistent with tax incentives.

Another puzzling result is given by the absence of bunching at the top three notches of the regime at $470k$, $540k$, and $600k$ (Figure 27). One explanation that could rationalize this result is the type of activity that each category of *Monotributo* encompasses namely a composition effect. While the first eight notches span both, services and the sale of goods, the last three notches only consider the sale of goods (see Table 3). Therefore, by definition tax incentives in the last three notches are not binding for individuals in the service sector, a group that is typically more responsive to taxes and transfers (Saez, 2010). Hence, the absence of

[19] In those cases where we include round numbers i.e., taxpayers with turnovers that are multiple of $\$1,000$, we consider an alternative regression specification that allows us controlling for bunching at round numbers. In concrete, we run the following specification $c_j = \sum_{i=0}^{p} \beta_i \cdot (z_j)^i + \sum_{i=z_L}^{z_U} \gamma_i \cdot [z_j = i] + \sum_{r \in R} \delta_r \cdot [\frac{z_j}{r} \in \mathbb{N}] + v_j$, where the last term controls for round-number fixed effects.

bunching in the top three notches could be explained by a higher composition of taxpayers selling goods, who presumably face more indivisibilities and frictions making bunching at the thresholds more difficult. Note also that taxpayers providing services and making more than $400k$ are excluded from the *Monotributo* regime and must register in the VAT and pay the income tax. This implies that the burden for taxpayers changes discontinuously when their revenue hits this eligibility threshold, providing further incentives to bunch at $400k$.

The previous two observations lead to the question of whether the response to this simplified regime varies across sectors. To this end, we focus on the discontinuity at $400k$ where there is a mix of taxpayers across sectors. In Figure 8 we break the sample into services, wholesale and retail, and others (including agriculture and manufacturing). Interestingly, we find that taxpayers in the services sector react the most ($b = 4.33$), followed by agriculture and manufacturing ($b = 1.90$), and finally wholesale and retail ($b = 1.15$).

There is some consensus in the field that any evidence of sharp bunching in earnings is likely due to tax evasion or tax avoidance rather than real responses (Kleven, 2016). We adhere to this view and interpret the behavioral responses from this section as convincing evidence that a significant number of taxpayers attempt to avoid higher tax liabilities by keeping their reported revenue below the thresholds.[20] The next logical step consists on translating the behavioral responses into elasticities which is what we present in Table 1. Intuitively, what we do is to scale down the bunching response by the first stage; that is to say, the discrete change in the average tax rate. We provide two alternative reduced form elasticities following Kleven & Waseem (2013). Overall, we found fairly low elasticities in the range of $[0.02; 0.60]$.[21]

Finally, since our data are remitted on a monthly basis, we can also analyze the timing of the bunching and the probability of filing after the due date. Figure 28 presents some preliminary analysis for the notch at $400k$. In these figures we look at: [a] the evolution of total earnings for those just above and just below the notch across the different months of the year (to see whether they accumulate earnings differently throughout the year), [b] date of latest declaration (to check whether they have diverse timing strategies) and [c] number of declarations within a year (to see whether the margin of adjustment is via the rectification of an already submitted declaration).[22] Overall, it seems that they do not have very different filing strategies, as we do not observe distinct patterns in the dimensions that we have just listed.[23]

[20] Unfortunately, in this article we cannot study the benefit side (and perhaps the main purpose) of this simplified regime i.e., whether it effectively serves as a formalization instrument for small taxpayers. This is an extensive-margin question that cannot be addressed with administrative data because one needs to observe both registered and non-registered individuals.

[21] These results are in line with those found in similar papers e.g., Kleven & Waseem (2013), Bachas & Soto (2018) and Devereux *et al.* (2014). Figure 9 puts these estimations into perspective with existing studies.

[22] Taxpayers can rectify a submitted declaration if they realize that there was a mistake or want to introduce a modification.

[23] Other interesting characteristics to further explore the bunching behavior, including age and

4 Behavioral responses at the upper end

In this section we turn the attention to the gross receipts tax (GRT) whose schedule allows us to study the middle and upper part of the revenue distribution.

4.1 Conceptual framework

The GRT notched tax schedule can be formally described as:

$$T(z) = t \cdot z + [\Delta t \cdot z] \cdot \mathbb{1}(z > z^*) \tag{3}$$

where t is the tax rate that should be paid when sales are below the threshold z^*, Δt corresponds to the increase in the tax rate above the notch and $\mathbb{1}(.)$ is a indicator function for being above the cutoff. Unlike *Monotributo*, the notch considered here takes the form of a discontinuity in a proportional tax rate, and thus the threshold represents a discontinuity in both the average and the marginal tax rate. This is shown in Figure 10, where the budget constraint not only shifts down at z^* but also becomes flatter. This type of notch is generally known as *proportional tax notch*. Nonetheless, incentives operate in the same way as in *Monotributo*, since firms have incentives to bunch right below the discontinuities to avoid a discrete jump in the tax burden.[24]

It is also worth noting that the bunching approach are related to the regression discontinuity (RD) design. The latter essentially exploits notched incentives, but in situations where the assignment variable that determines whether the firm is above or below the relevant threshold is not subject to manipulation. Given the large size of firms around these notches, one could expect frictions and indivisibilities to operate more strongly and, thus, bunching to be weaker or nonexistent i.e., gross revenue is harder to manipulate. In such cases, one could exploit an RD design to measure the revenue elasticity to the GRT.

4.2 Empirical evidence: little bunching

Recall from panels (b)-(d) of Figure 1 that the tax scheme of the general regime varies by three broad groups: retail and wholesale, services, agriculture and manufacturing. Hence, in this subsection we analyze each of these groups separately. We start the empirical analysis by plotting the revenue distribution as we did in the previous section for the simplified regime. We do so by pooling data from

gender, are not available in the dataset that we have accessed to.

[24]The setting and the figure that we have just described, although correct, is probably more accurate to represent the situation of self-employed or individual taxpayers rather than large firms which are the ones present at the top of the sales distribution. Therefore, the proper way to frame this tax and related incentives, is to associate it to the literature of the elasticity of corporate taxable income (ECTI) that measures the response of corporate taxable income to changes in the corporate income tax rate. Refer to Devereux *et al.* (2014) for more on the analytical framework regarding large firms, and also to Bachas & Soto (2018).

years 2013 to 2015, a period in which the tax rates and revenue thresholds remained unchanged (see Table 5). Since the determination of tax rates is based on annual income from year t_{-1}, the densities consider revenue from years 2012-2014.[25]

Figure 11 presents the results for the two notches in the retail and wholesale sector. In the first notch at 1 million pesos, there is no visible bunching. This is consistent with the absence of bunching in the top three notches of *Monotributo*, a result that perhaps can be explained by the relatively small first-stage change in average tax rates at these thresholds. In the second notch at 40 million pesos, however, there is a small spike to the left of the threshold and a hole to the right. Although bunching is not as striking as in the first eight notches of *Monotributo*, this suggests that some large firms are able to manipulate their gross revenue to avoid a higher tax burden in the following year. Unlike the first notch, note that at 40 million pesos the average tax rate goes from 3.5 to 5 percent. This stronger first stage applied to a broad gross revenue tax base can indeed be translated into very high rates on profits. For example, assuming a profit margin of 10 percent (which is in the ballpark of empirical estimates), a turnover tax of 5 percent corresponds to a 50 percent tax on profits i.e., the tax on profits is increasing by 15 percentage points.[26]

Figure 12 presents the results for the two notches in the service-based sector. Note that in this sector revenue is closer to profits than the other two sectors because it is less intensive in intermediate inputs and, thus, the first stage tax variation and bunching incentives operate less strongly. In the first notch at 500k pesos, there seems to be a small spike to the left of the threshold, albeit not as sharp as in the 400k notch of *Monotributo* which is very close to it. Recall that self-employed workers providing services are excluded *de facto* from the simplified regime if their annual income is higher than 400k pesos and must register in the VAT. Hence, when we analyze the notch at 500k we are mostly dealing with registered firms instead of independent workers which could have less flexibility to manipulate reported sales, or they could compensate the additional tax burden by adjusting other margins e.g., inflating costs to pay less income tax. In the second notch at 30 million pesos, there is no visible bunching either.

Finally, in Figure 13, we present the results for the two notches in manufacturing. Note that this is the sector that faces the lowest tax rates. In fact, from Table 5 on can see that before 2008 this sector was fully exempted from the turnover tax to alleviate inefficiencies from the cascading effect. In this case, the two notches operate up in the revenue distribution and therefore we plot them together. From

[25] Figure 29 presents a diagram with an example of the standard timing of fiscal year, tax code approval, and deadline to file.

[26] Profit margin is defined as reported profits over revenue. In practice, margins vary greatly by sector and industry. For example, average profit margin ranges from 5 to 15 percent in Costa Rica (Bachas & Soto, 2018). Similarly, Almunia & Lopez-Rodriguez (2018) report stable profit margins between 5 and 6 percent for Spanish firms. An example of low-margin product-based sectors are grocery stores and supermarkets with net profit margins ranging from 1 to 3 percent. An example of a high-margin sector is pharmaceutical with net profit margins of about 18 percent. In service-based sectors, bookkeeping and payroll services firms can reach margins of 20 percent.

the figure we can see that there is very little manipulation of annual sales right below the taxable thresholds. Again, although we do not observe sharp bunching, this result is striking as this figure includes "super firms" in the top 0.1 percent of the sales distribution.

Administrative notch. Recall from the institutional setting, that the GRT also has a regulatory provision that obliges firms with annual sales above a threshold to work as *collection agents* for the tax authority. Since this regulation increases significantly the accounting costs of companies, they have incentives to bunch strictly below the threshold. Between 2010 and 2016, the threshold that determines whether a firm must act as a collection agent remained fixed at 10 million pesos.

In Figure 14 we plot the distribution of firms around 10 million pesos. From the figure we can see a rather clear mass to the left and a missing mass to the right of the threshold.[27] Note that in this case, the degree of bunching, albeit small, is stronger than in the tax notches. This result strikes us as remarkable given that these are very large firms and it is more difficult to manipulate broad tax bases such as income from sales. Hence, it is very likely that this is capturing underreporting behavior rather than real production effects. We interpret this result as suggestive evidence that firms find more costly the indirect administrative cost of working as a collection agent than the direct fiscal cost of the GRT tax. When we split the analysis by sector we find that the manipulation is mainly driven by service and manufacturing firms (Figure 15).

First stage and compliance. Unlike *Monotributo*, note that for the GRT we can also explore the empirical first stage because the data we use is generated from turnover tax returns. The goal here is to confirm whether tax rates jump (and to what extent) around the different cutoffs for the different sectors. Note also that, if the distribution of revenue is smooth at the discontinuity, in a second stage, one could test whether the tax variation from the first stage affects the production/sales of firms in the following year.[28]

In Figures 16 through 18, we plot the fraction of firms paying the statutory tax rate mandated to the right of the threshold. The vertical line indicates the *notch* at which the statutory tax rate changes discretely. Under perfect compliance, the line should be flat at zero and jump to one at the threshold. The sample, bins, and x-axis are the same as in the density figures.

These figures show that, although firms comply with the tax law, compliance is far from perfect making our research design fuzzy. Moreover, one can see that firms make two type of mistakes: to the left of the notch we have some firms paying a higher tax rate than mandated (overrate) and to the right of the notch we have firms paying a lower tax rate than mandated by law (underrate). The intensity of these behavioral mistakes also vary by the way the law is written for

[27] Note that in the case of administrative *notches* it is less straightforward to determine the dominated area given the difficulty to assign a specific value, or cost, to administrative tasks.

[28] See Figure 31 for the regression discontinuity results.

each sector.[29] Hence, we conclude that the way tax laws are written also matters for tax compliance. This is a novel result in public finance that has important implications for revenue collection and vertical equity considerations.

Related to this point, we also show that non-compliance falls over time. In Figure 19, we split the previous analysis by year for the first notch in services, retail and wholesale. The figure shows that while "overrate" mistakes to the left are fairly constant over time, "underrate" mistakes to the right are reduced across years. This is possibly explained by asymmetric enforcement actions of the tax authority.

Finally, we replicate the first stage figures but distinguishing between *local* firms, those that operate within PBA, and firms with *convenio* i.e., operate both inside and outside PBA. We present these results for the retail and wholesale, and service sector in Figure 20; the difference between the two types of firms is clear. *Local* firms face present a higher compliance to the first stage implied by the thresholds, as there is a clear, and larger, discontinuity in the tax rates they face as soon as they pass the *notch* as compared to their counterparts with *convenio* (the discontinuity for these firms is way more fuzzy). This difference in first stage, could be due either to better monitoring by the tax administration office, or, naturally, by less evasion possibilities compared to those that operate outside the province.[30] In addition, the discontinuity in the tax liability as companies pass the cutoff, is more sharp in those thresholds located to the right of the turnover distribution i.e., where large firm are supposed to be (compare the two figures on the left with those on the right in Figure 20).

Discussion. In this section we document behavioral responses for medium and large firms to the *notched* schedule of the turnover tax.[31] Although less sharp that in *Monotributo*, the importance of this group of firms in terms of GDP and tax revenue raises concerns about the efficiency (direct) cost of taxes on gross sales. The differential response to the *notches* from the two taxes we study could be explained by firm-size characteristics, such as the number of employees or the level of fixed assets, that might prevent large firms from engaging in evasion actions such as the underreporting of their gross revenue. Kleven *et al.* (2016) make this point theoretically, arguing that larger and more complex firms are less likely to reach a colluding agreement to evade taxes, as there is a higher chance that one of the employees may act as a "whistleblower".[32]

[29] In section 2, where we described the GRT, we listed the tax rates for the different sectors following the exact order as they appear written in the official tax code. See also Figure 30 where we provide extracts, in Spanish, of how the tax code is written in practice.

[30] Besides these differences in first stages, both types of firms show no bunching at the cutoff (figures available upon request).

[31] We also look at other discontinuity; there is a sort of an entry *notch* in the GRT schedule. New firms whose turnover during the first two months of activity are below certain threshold face a lower tax rate as compared to those that are above. We look at whether firms bunch at this entry *notch* but found no evidence of such behavior.

[32] Another important feature of the turnover tax that was not studied in this paper is the differential and high tax rates imposed on "foreign jurisdiction" firms, which works as an internal custom. In future work we would like to analyze cross-border effects of the turnover tax on "for-

Note also that the low degree (in-existence) of bunching by large firms does not necessarily imply that the efficiency costs is low and firms are absorbing the additional burden of the tax. In fact, lowering revenue is only one of the possible responses to a higher tax rate. For instance, we cannot rule out responses in other margins or even fiscal externalities to other tax bases. For example, this could lead firms to inflate costs in order to pay a lower profit tax or value added tax. There could be various behavioral adjustments, for instance a change in location, adjustments in investments, sales shifting to different activities or localities, among others. We discuss some preliminary explorations below.

First, we look at whether firms split into different subsidiaries in order not to face the discrete jump in the average tax rate. We can look at this in the microdata by checking whether there is a differential survival rate between those below and above the *notch*. In Figure 32 we plot the survival probability for those taxpayers on both sides of the cutoff in 2012 (we repeat the exercise but considering 2013 and 2014 as baseline years, respectively).[33] As it is possible to infer from the chart, we find no evidence of such a behavior. Second, there may also exist another behavioral response, for instance shift tax bases across different sectors. In our setting, the tax schedule of the service sector dominates the tax schedule of the other two. This means that for any given annual sales, the tax rate that should be paid under the service sector is larger than the one in retail and agro-related activities, respectively. This suggests that as soon as any firm crosses one of the two *notches* associated with the service sector i.e., $500k$ and $30m$, there are incentives to diminish the sales in that sector relative to others. In Figure 33 we analyze this point.[34] For the *notch* with the largest discontinuity in the average tax rate, we observe a decrease in the share of service-related sales after the firm passes the thresholds which is somehow suggestive evidence of shifting behavior to avoid taxes.

5 Conclusions

In this paper we analyze the response of self-employed workers and firms to two different revenue taxes, *Monotributo* and the Gross Receipts Tax. We use administrative tax return data from Argentina and quasi-experimental variation provided

eign jurisdiction" firms.

[33] To do this figure we focus on taxpayers that were just below and above the GRT notches in 2012 and calculate the probability of filing a tax in the subsequent years. We replicate the analysis for those around the notch in 2013 and 2014 respectively. In order to calculate the standard errors we follow a bootstrap procedure.

[34] For this exercise we defined treated firms as those that in 2013 (2014) were above the *notch* but were below in 2011 and 2012 (2012 and 2013 respectively). The treatment year corresponds to the moment were firms cross the threshold. Then, we run a regular event study specification of the form $y_{i,t} = \sum_{j=-3}^{2} \gamma_j \cdot d_{i,t}^{j} + \mu_i + \mu_t + \epsilon_{i,t}$ with firm and time fixed effects, taking the ratio of sales in the service sector relative to total sales as dependent variable. Importantly for the validity of this specification, is the fact that firms cannot manipulate where to locate on either side of the threshold.

by several *notches* in the tax schedules.

In the case of small taxpayers, our findings show substantial bunching right below the thresholds of the first eight categories in the *Monotributo* regime. The intensity of this clustering is stronger for higher discrete jumps in average tax rates. In the remaining three notches, however, bunching is non existent, which can be justified by a higher composition of taxpayers selling goods instead of services i.e., that face higher indivisibilities. We also document that taxpayers in services react the most, followed by agriculture and manufacturing, and finally wholesale and retail. We interpret these behavioral responses as compelling evidence that a significant number of taxpayers attempt to avoid higher tax liabilities by keeping their reported revenue below the thresholds.

In the case of medium and large taxpayers, we find very little or almost inexistent bunching of firms to manipulate their gross sales to avoid facing higher tax rates. However, we find that bunching is stronger for the administrative notch that obliges firms to work as collection agents. This could imply that firms find more costly this indirect administrative cost than the direct fiscal cost of the turnover tax. When we split the analysis by sector we find that the manipulation is mainly driven by service and manufacturing firms.

Overall, these results suggest that taxes imposed on broader tax bases are prone to avoidance for small firms, but are harder to avoid for larger firms since they are in the public eye of tax authorities. Although less sharp that in *Monotributo*, the importance of large firms in terms of GDP and tax revenue raises concerns about the efficiency (direct) cost of taxes on gross sales. The differential response to the notches from the two taxes we study could be explained by firm-size characteristics, such as the number of employees or the level of fixed assets, that might prevent large firms from engaging in evasion actions, such as the underreporting of their gross sales.

Importantly, despite a relatively low response of medium and large firms, we cannot rule out that companies react on other margins. Furthermore, our identification strategy does not let us identify or measure the cascading effect of the turnover tax, which is presumably the most relevant distortion introduced by this tax. This type of question could be eventually addressed if researchers were given access to other administrative databases that allow to follow business-to-business transactions. One example of this type of data is the one generated from the withholding regime of collection agents or from the value added tax.

References

Alejos, Luis. 2018. *Firms' (mis)reporting under a minimum tax: Evidence from Guatemalan corporate tax returns*. Working Papers. University of Michigan.

Almunia, Miguel, & Lopez-Rodriguez, David. 2018. Under the Radar: The Effects of Monitoring Firms on Tax Compliance. *American Economic Journal: Economic Policy*, **10**(1), 1–38.

Angrist, Joshua D., & Pischke, Jörn-Steffen. 2010. The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, **24**(2), 3–30.

Asatryan, Zareh, & Peichl, Andreas. 2017. *Responses of Firms to Tax, Administrative and Accounting Rules: Evidence from Armenia*. Tech. rept.

Bachas, Pierre Jean, & Soto, Mauricio. 2018 (July). *Not(ch) your average tax system : corporate taxation under weak enforcement*. Policy Research Working Paper Series 8524. The World Bank.

Besley, Timothy, & Persson, Torsten. 2013. Chapter 2 - Taxation and Development. *Pages 51–110 of:* Auerbach, Alan J., Chetty, Raj, Feldstein, Martin, & Saez, Emmanuel (eds), *Handbook of Public Economics*. handbook of public economics, vol. 5, vol. 5. Elsevier.

Best, Michael Carlos, Brockmeyer, Anne, Kleven, Henrik Jacobsen, Spinnewijn, Johannes, & Waseem, Mazhar. 2015. Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan. *Journal of Political Economy*, **123**(6), 1311–1355.

Boonzaaier, Wian, Harju, Jarkko, Matikka, Tuomas, & Pirttilä, Jukka. 2017. *How do small firms respond to tax schedule discontinuities? Evidence from South African tax registers*. Tech. rept.

Devereux, Michael P., Liu, Li, & Loretz, Simon. 2014. The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records. *American Economic Journal: Economic Policy*, **6**(2), 19–53.

Fajnzylber, Pablo, Maloney, William F., & Montes-Rojas, Gabriel V. 2011. Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program. *Journal of Development Economics*, **94**(2), 262–276.

Feldstein, Martin. 1999. Tax Avoidance and the Deadweight Loss of the Income Tax. *The Review of Economics and Statistics*, **81**(4), 674–680.

Keen, Michael. 2014. Targeting, cascading and indirect tax design. *Indian Growth and Development Review*, **7**(2), 181–201.

Kleven, Henrik J., & Waseem, Mazhar. 2013. Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan *. *The Quarterly Journal of Economics*, **128**(2), 669–723.

Kleven, Henrik Jacobsen. 2016. Bunching. *Annual Review of Economics*, **8**(1), 435–464.

Kleven, Henrik Jacobsen, Kreiner, Claus Thustrup, & Saez, Emmanuel. 2016. Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries. *Economica*, **83**(330), 219–246.

Libonatti, Oscar. 1998. La imposición sobre los ingresos brutos. *Pages pp 301–326 of: La Reforma Tributaria en Argentina*. editorial FIEL.

Liu, Li, Lockwood, Ben, & Almunia, Miguel. 2017. *VAT Notches, Voluntary Registration, and Bunching: Theory and UK Evidence*. Working Papers.

Lobel, Felipe, Scot, Thiago, & Zúniga, Pedro. 2020. *Corporate Taxation and Evasion Responses: Evidence from a Minimum Tax in Honduras*. University of California, Berkeley.

Monteiro, Joana C. M., & Assunção, Juliano J. 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**(1), 105–115.

Moskovits, Cynthia, & Susmel, Nuria. 2006. El diseño del sistema tributario argentino. *Pages pp. 111–178 of: La presión tributaria sobre el sector formal de la economía*. FIEL.

Onji, Kazuki. 2009. The response of firms to eligibility thresholds: Evidence from the Japanese value-added tax. *Journal of Public Economics*, **93**(5), 766 – 775.

Rauhanen, Timo, Harju, Jarkko, & Matikka, Tuomas. 2016. *The effects of size-based regulation on small firms: evidence from VAT threshold*. Tech. rept.

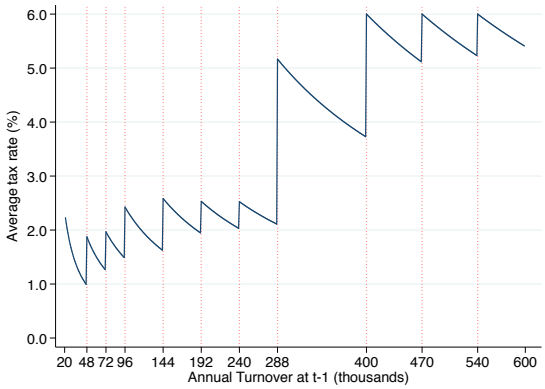
Saez, Emmanuel. 2010. Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, **2**(3), 180–212.

Saez, Emmanuel, Slemrod, Joel, & Giertz, Seth H. 2012. The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature*, **50**(1), 3–50.

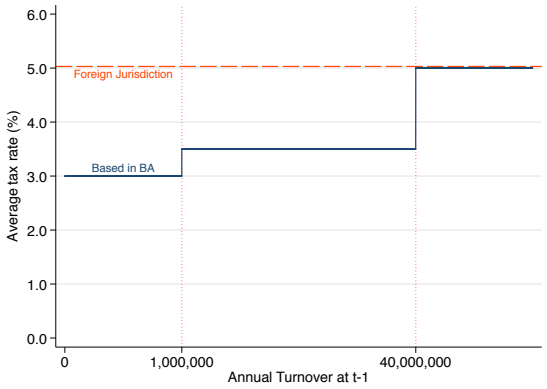
Figures and Tables

Figure 1: The tax schemes: *Monotributo* and *Turnover Tax*

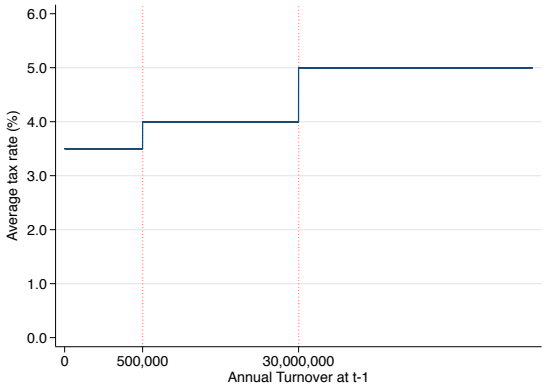
(a) *Monotributo*



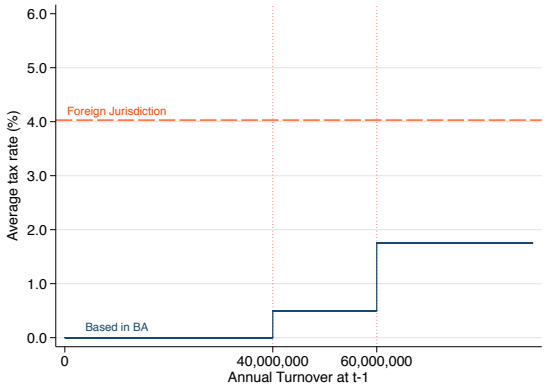
(b) GRT - Wholesale and retail



(c) GRT - Services



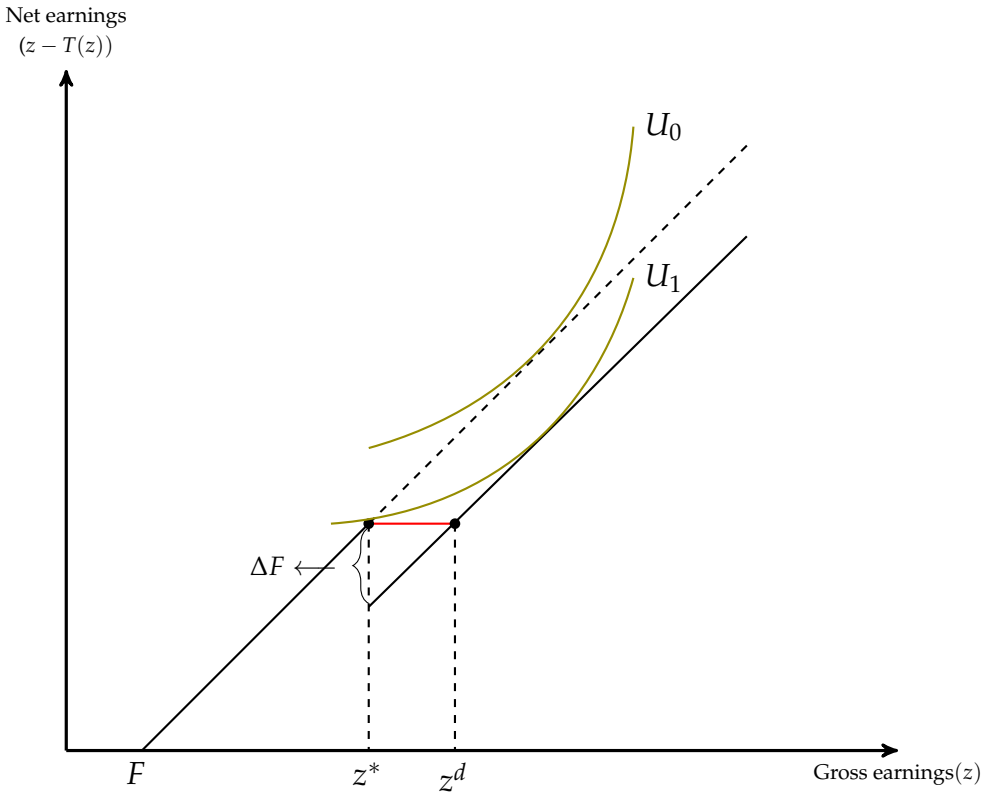
(d) GRT - Agriculture and manufacture



Source: Own elaboration based on official documents and tax laws.
Notes: Panel (a) presents the average tax rate i.e., tax liability as a proportion of turnover, for the *monotributo* tax schedule for retail and wholesale. Panels (b), (c) and (d) present the average tax rate for the GRT for different sectors for the period 2013-2015. These three panels are not drawn to scale. Each dotted red vertical line in the different panels, refers to a different tax *notch*. It is worth highlighting that while the horizontal axis in panel (a) is measured in thousands of pesos, in the other panels it is so in millions. Then, the discontinuities in each of these two taxes occurs at very different parts of the turnover distribution.

Figure 2: *Monotributo* scheme:

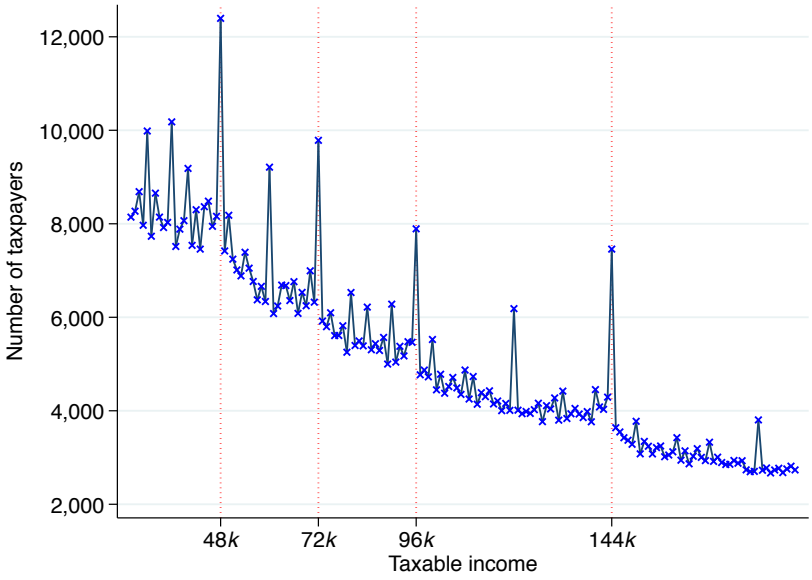
$$T_b(z) = F_b + [\Delta F_b] \cdot \mathbb{1}(z > z_b^*)$$



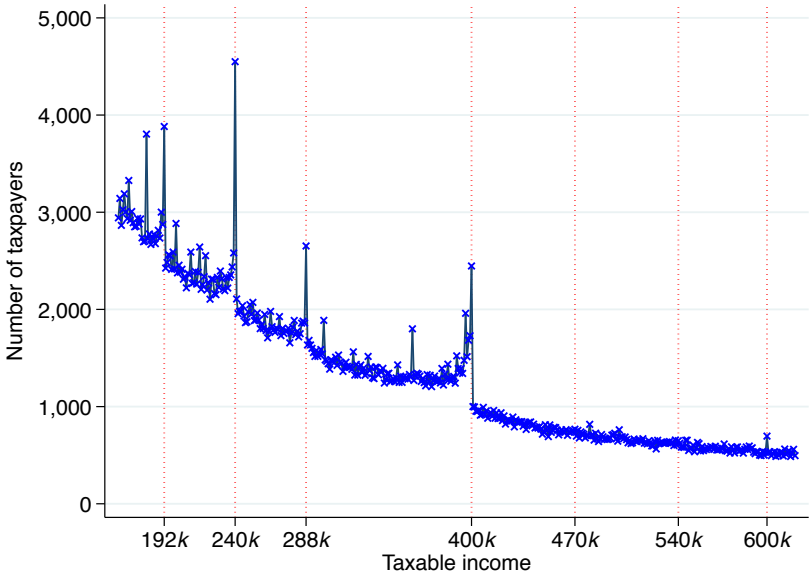
Notes: In this figure we present the standard framework to show the incentives that a rational agent should have to bunch. On the vertical axis we measure after-tax earnings while on the horizontal before-tax earnings. Note that the *notch* that is introduced in this example refers to *pure notch* i.e., there is a downward shift in the budget choice set, but the *slope* remains the same before and after the threshold (choice sets are parallel). Before the introduction of the *notch* the *marginal buncher* enjoys utility U_0. After the tax change she will be in U_1, where she will be indifferent between enjoying consumption at z^* or at the point where the new utility curve is tangent to the new choice set. Taxpayers to the left of the *marginal buncher* will have incentives to bunch at the *notch*. Importantly, no rational taxpayers should be located in the segment given by the red line. This segment corresponds to the *dominated area*. At the *notch*, taxpayers enjoy more consumption and leisure than in the dominated area.

Figure 3: Full sample

(a) First four *notches*



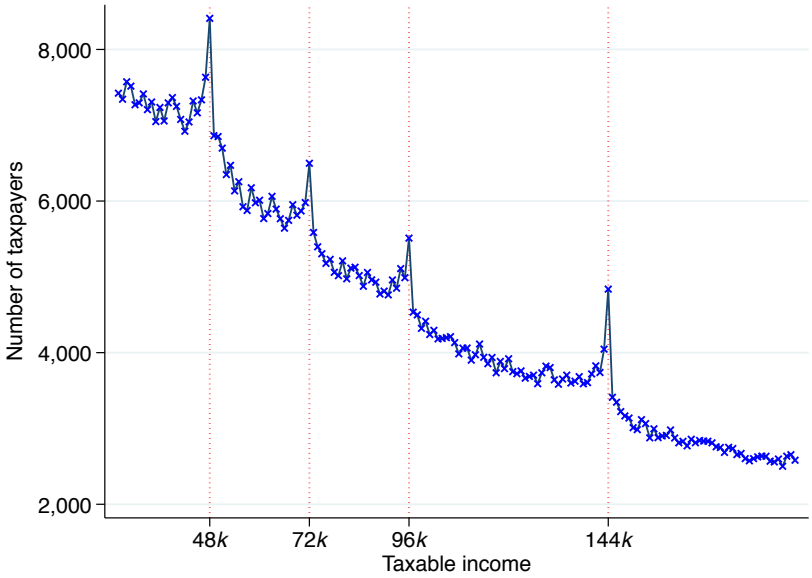
(b) Next seven *notches*



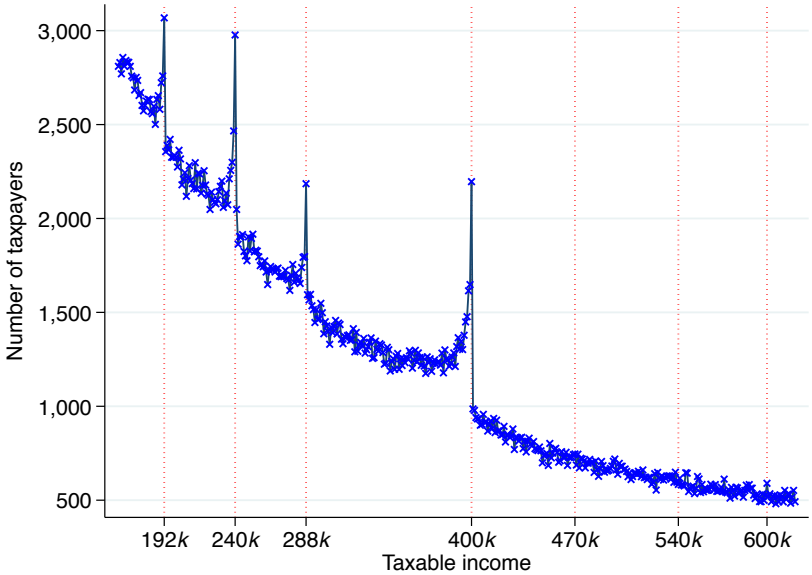
Notes: Taxpayers of the years 2014, 2015 and 2016 are grouped in bins of \$1,000 (full sample). Each dotted red vertical line refers to a different tax *notch*.The horizontal axis measures taxable income which is equal to annual turnover, while the vertical axis counts the number of taxpayers.

Figure 4: Non-rounder sample

(a) First four *notches*

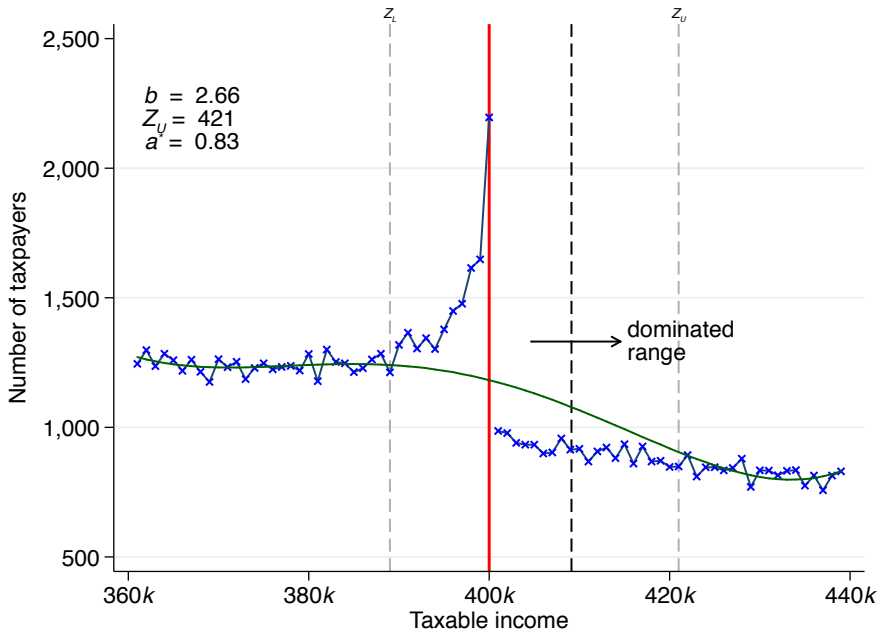


(b) Next seven *notches*



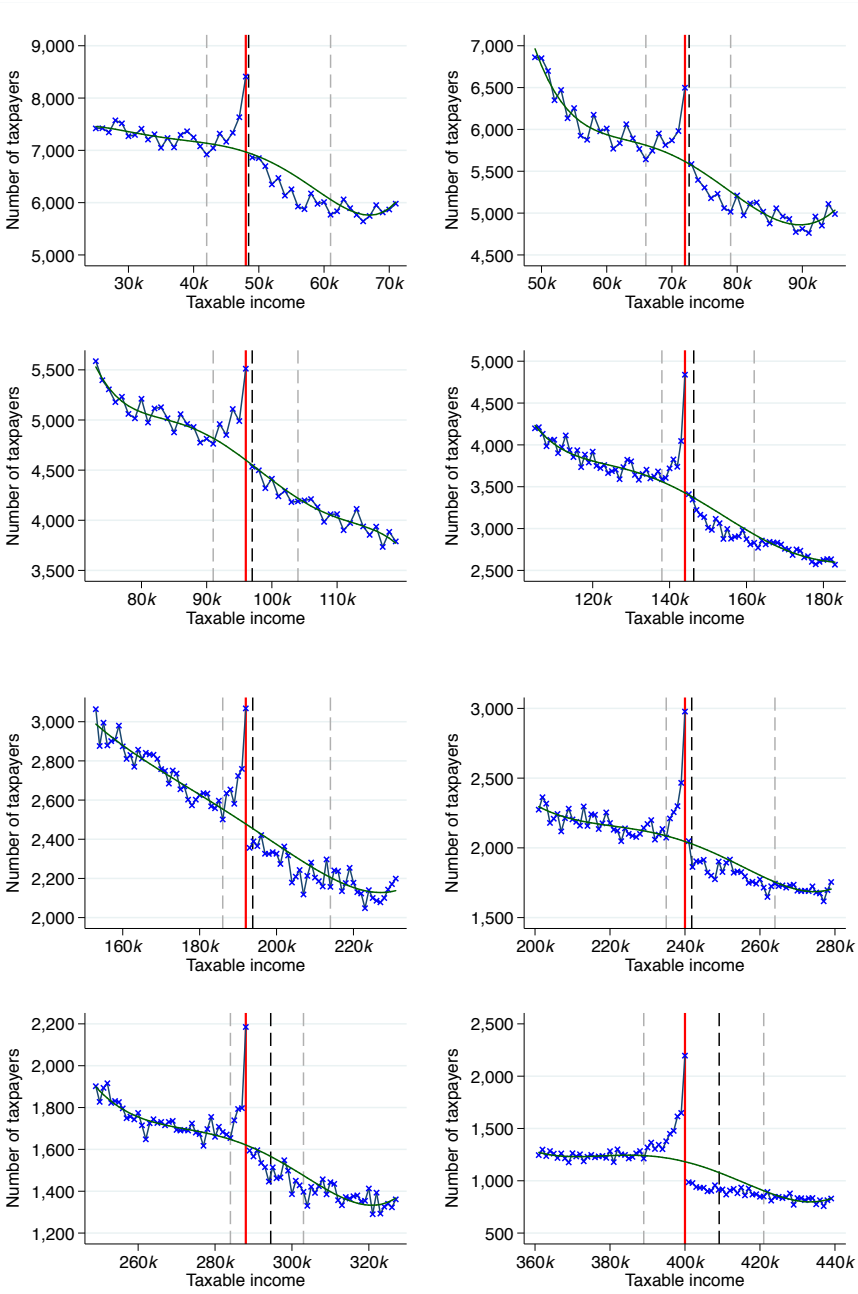
Notes: Taxpayers of the years 2014, 2015 and 2016 are grouped in bins of \$1,000. Each dotted red vertical line refers to a different tax *notch*. The horizontal axis measures taxable income which is equal to annual turnover, while the vertical axis counts the number of taxpayers. We remove those taxpayers whose reported taxable income is a multiple of 1,000.

Figure 5: Bunching and missing mass at *monotributo notch* (400k)



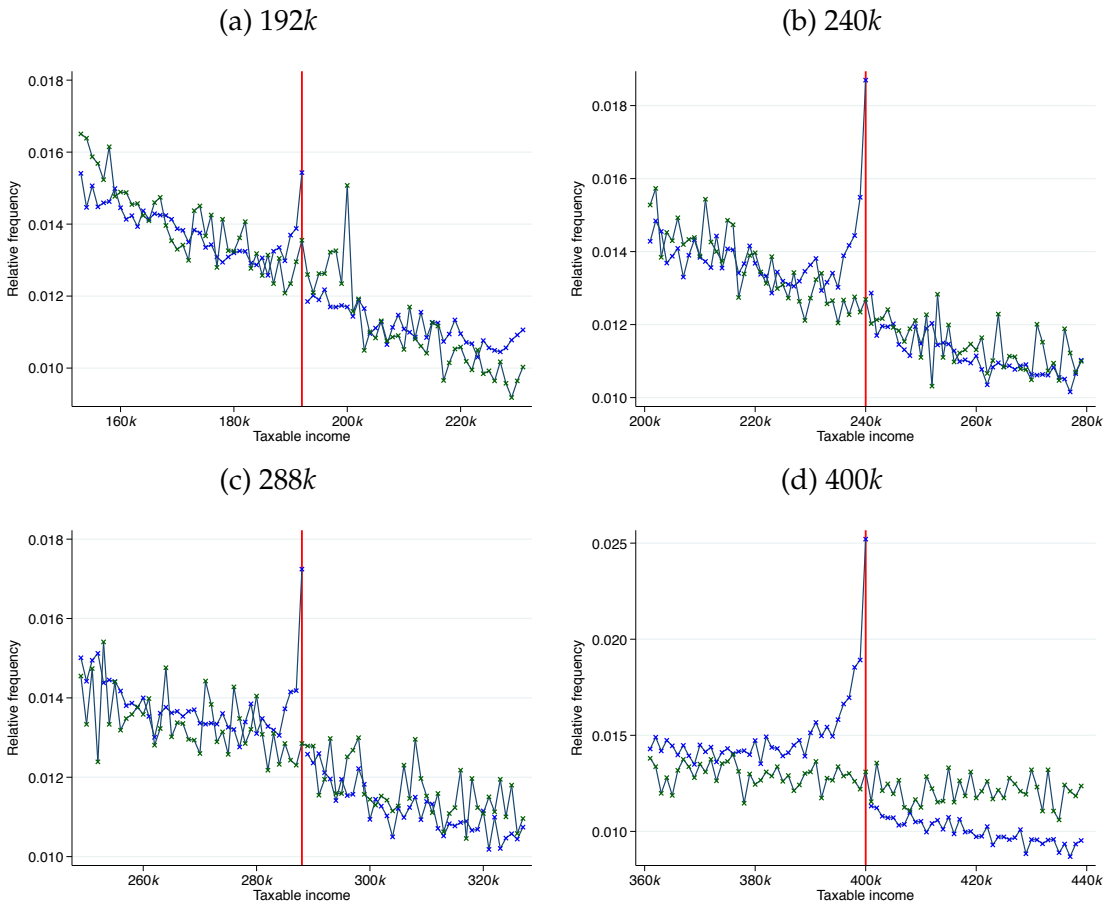
Notes: Taxpayers of the years 2014, 2015 and 2016 are grouped in bins of \$1,000. The solid red vertical line refers to the existing *notch* at 400k. The green solid line is the prediction of regressing equation (2) with the caveat of the dummies in the excluded range. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts the number of taxpayers. We remove those taxpayers whose reported taxable income is a multiple of 1,000. In the left box, b refers to the bunching mass, Z_U to the upper limit of the excluded area and a to the percentage of taxpayers in the dominated range who are unresponsive. The grey dashed lines delimit the excluded area (lower and upper bounds respectively), while the black dashed one identifies the income value that makes the marginal buncher indifferent with the net-income she gets at the *notch*.

Figure 6: Bunching at *monotributo notches*



Notes: Taxpayers of the years 2014, 2015 and 2016 are grouped in bins of $\$1,000$. The solid red vertical lines refer to the different *notches*. The green solid line is the prediction of regressing equation (2) with the caveat of the dummies in the excluded range. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts the number of taxpayers. We remove those taxpayers whose reported taxable income is a multiple of $1,000$. The grey dashed lines delimit the excluded area (lower and upper bounds respectively), while the black dashed one identifies the income value that makes the marginal buncher indifferent with the net-income she gets at the *notch*.

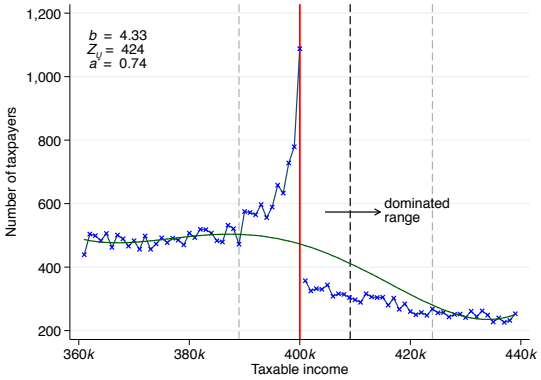
Figure 7: Bunching at *monotributo notches* and *control years*



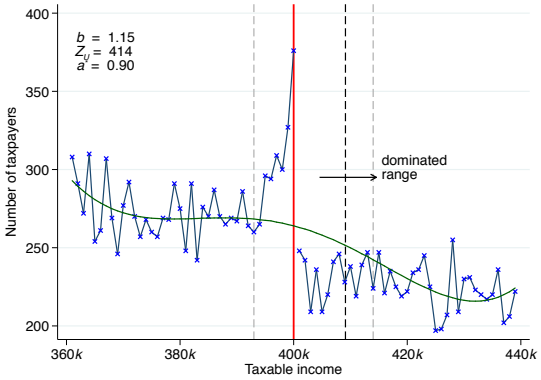
Notes: These figures show that there is no bunching when we use pre-2014 data and the *notches* located at 192k, 240k, 288k, and 400k do not exist yet (green line). The blue lines correspond to taxpayers in fiscal years 2014 to 2016. The vertical axis measures relative frequency while the horizontal axis measures taxable income which is equal to the annual gross income. We remove taxpayers whose reported taxable income is a multiple of 1,000. Note that the bunching at 200k that we observe in panel (a) is due to the presence of an old *notch* existing in the tax scheme (see Table 2).

Figure 8: Bunching at 400k by sector

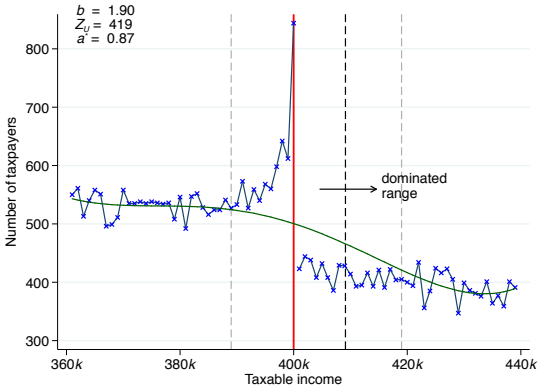
(a) Services



(b) Wholesale and Retail

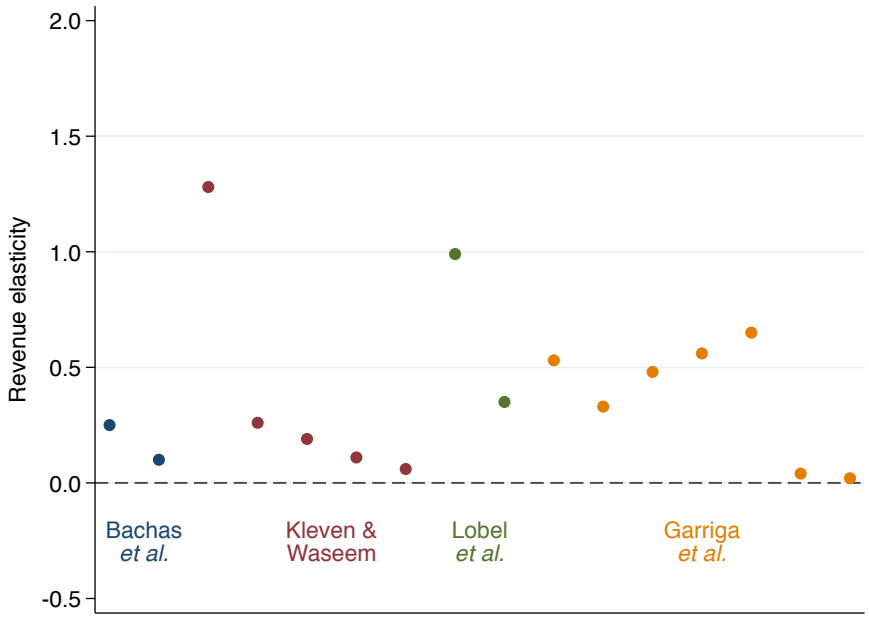


(c) Other (incl. Agriculture and Manufacturing)



Notes: Taxpayers of the years 2014, 2015 and 2016 are grouped in bins of \$1,000. These figures split the sample into services, wholesale and retail, and agriculture and manufacturing. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts the number of taxpayers. We remove taxpayers whose reported taxable income is a multiple of 1,000. In the left box, b refers to the bunching mass, Z_U to the upper limit of the excluded area and a to the percentage of taxpayers in the dominated range who are unresponsive. The grey dashed lines delimit the excluded area (lower and upper bounds respectively), while the black dashed one identifies the income value that makes the marginal buncher indifferent with the net-income she gets at the *notch*.

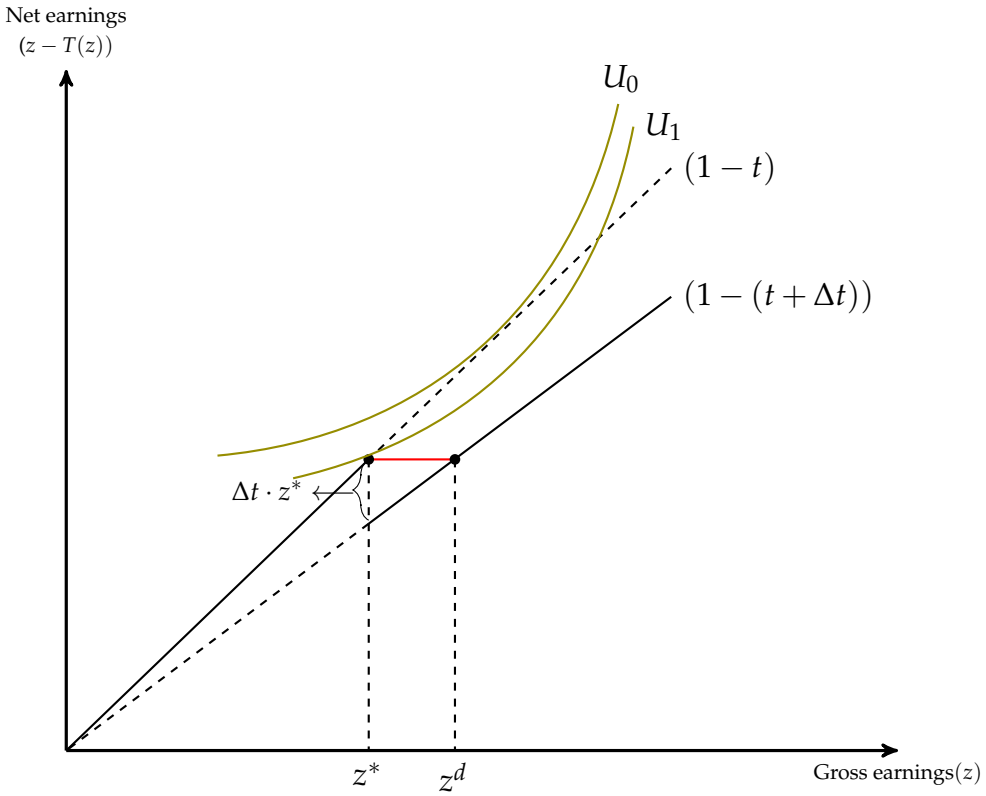
Figure 9: Literature review: revenue elasticity



Notes: This figure plots some of the most recently documented revenue elasticities. From left to right we show the following estimations: the blue dots present the results by Bachas & Soto (2018) for Costa Rica (Table 1), the red ones those by Kleven & Waseem (2013) for Pakistan (Table 2), the green ones recent values for Honduras by Lobel *et al.* (2020) (Table 3), and finally the yellow ones refer to our estimations (Table 6).

Figure 10: GRT scheme:

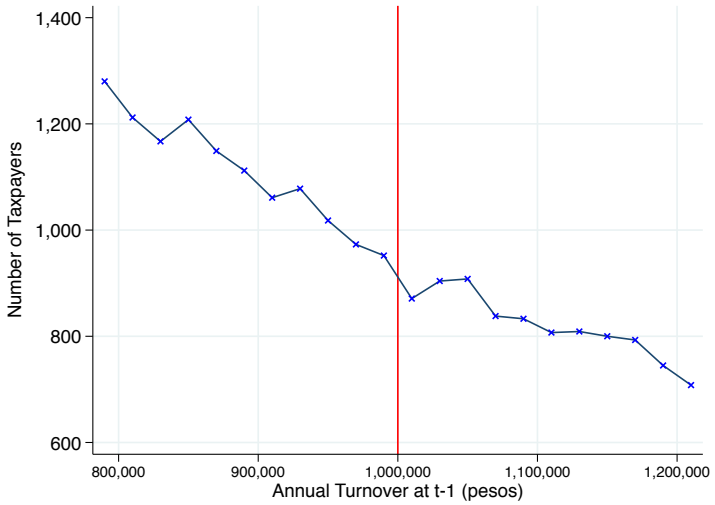
$$T(z) = t \cdot z + [\Delta t \cdot z] \cdot \mathbb{1}(z > z^*)$$



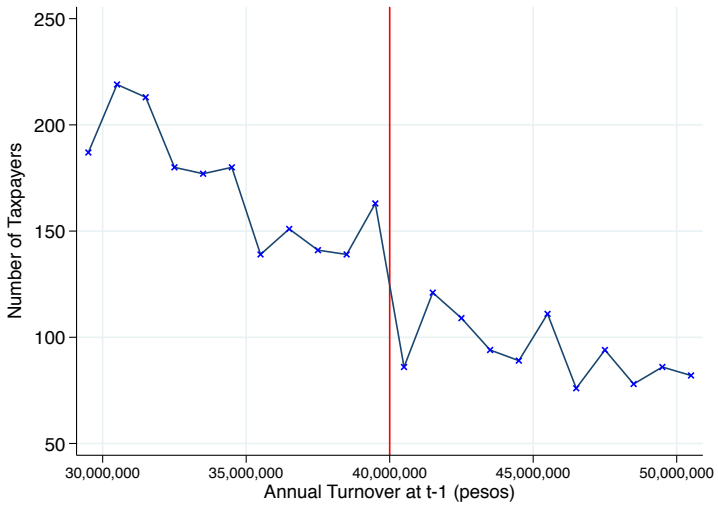
Notes: In this figure we present the standard framework to show the incentives that a rational agent should have to bunch. On the vertical axis we measure after-tax earnings while on the horizontal before-tax earnings. Note that the *notch* that is introduced in this example refers to *proportional notch* i.e., there is a downward shift in the budget choice set but also a change in the *slope*; therefore a change in the marginal tax rate either (marginal tax rate was t and then is $t + \Delta t$). Before the introduction of the *notch* the *marginal buncher* enjoys utility U_0; after the tax change she will be in U_1. Importantly, no rational taxpayers should be located in the segment given by the red line. This segment corresponds to the *dominated area*. At the *notch*, taxpayers enjoy more consumption and leisure than in the dominated area.

Figure 11: Bunching at *GRT notches*: Retail and wholesale

(a) Notch 1: tax rate goes from 3 to 3.5 percent



(b) Notch 2: tax rate goes from 3.5 to 5 percent

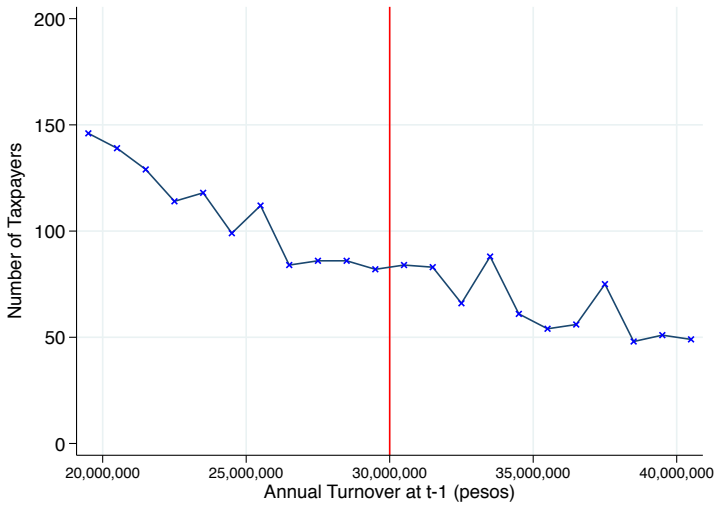


Notes: This graph shows the distribution of reported annual revenue from year t_{-1} using pooled data for the years 2012-2014. The vertical line indicates the *notch* at which the average tax rate changes discretely. The bins are $20k$ pesos wide in panel (a) ($N = 21,226$) and $1m$ pesos in panel (b) ($N = 2,915$), delimited such that no bin contains data both to the left and to the right of the threshold. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts the number of taxpayers. These figures do not consider *foreign jurisdiction* firms, and for those multi-sector firms, we identify their main sector.

Figure 12: Bunching at *GRT notches*: Services

(a) Notch 1: tax rate goes from 3.5 to 4 percent

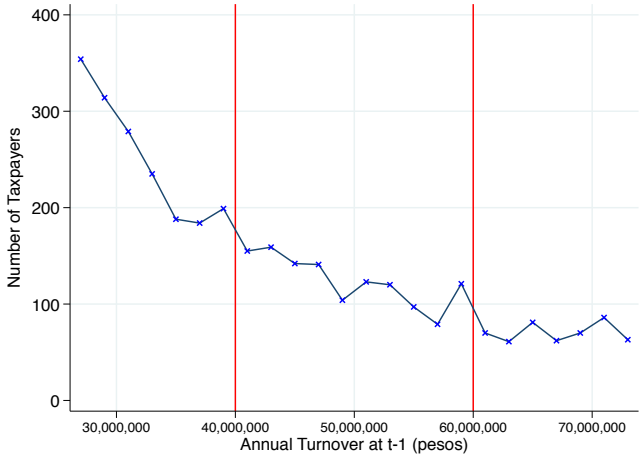
(b) Notch 2: tax rate goes from 4 to 5 percent



Notes: This graph shows the distribution of reported annual revenue from year t_{-1} using pooled data for the years 2012-2014. The vertical line indicates the *notch* at which the average tax rate changes discretely. The bins are $10k$ pesos wide in panel (a) ($N = 20,104$) and $1m$ pesos in panel (b) ($N = 1,910$), delimited such that no bin contains data both to the left and to the right of the threshold. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts the number of taxpayers. These figures do not consider *foreign jurisdiction* firms, and for those multi-sector firms, we identify their main sector.

Figure 13: Bunching at *GRT notches*: Manufacturing

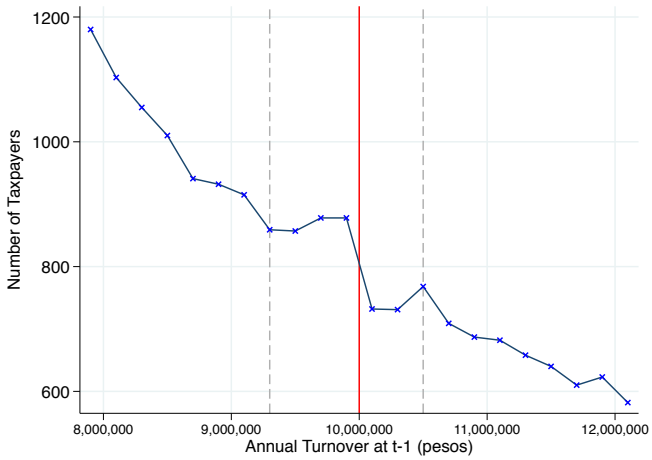
(a) Notch 1 and 2: tax rate goes from 0 to 0.5 and 0.5 to 1.75 percent



Notes: This graph shows the distribution of reported annual revenue from year t_{-1} using pooled data for the years 2012-2014. The vertical line indicates the *notch* at which the average tax rate changes discretely. The bins are $1m$ pesos wide in panel (a) ($N = 3,487$) and $1m$ pesos in panel (b) ($N = 18,030$), delimited such that no bin contains data both to the left and to the right of the thresholds. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts the number of taxpayers. These figures do not consider *foreign jurisdiction* firms, and for those multi-sector firms, we identify their main sector.

Figure 14: Bunching at *administrative notches*

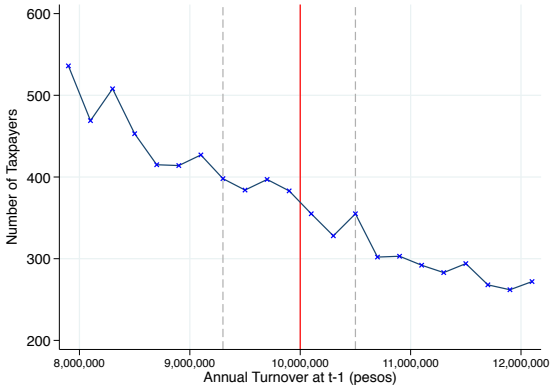
(a) Administrative Notch: firms have to work as collection agents



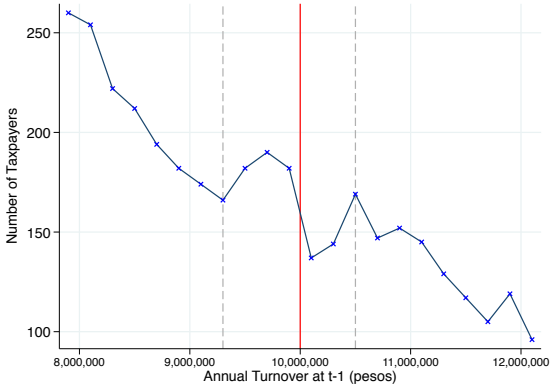
Notes: This graph shows the distribution of reported annual revenue from year t_{-1} using pooled data for the years 2012-2014. We use these years, so as to have the same time window as the one we consider in the main analysis i.e., discontinuities in tax liabilities; nevertheless, we get very similar results if we were to add the year 2015. The red vertical line indicates the administrative *notch* at which firms become collection agents. The bins are $200k$ pesos wide ($N = 18,030$). The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts the number of taxpayers.

Figure 15: Bunching at *administrative notches* by sector

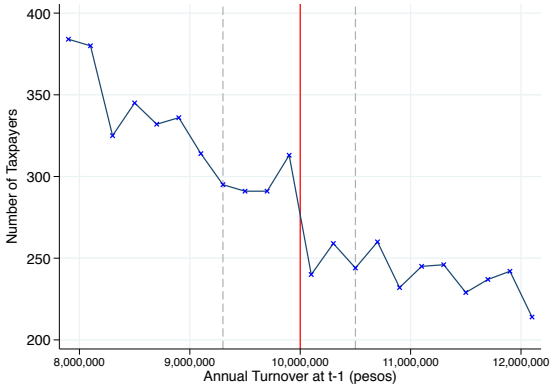
(a) Retail and Wholesale



(b) Services



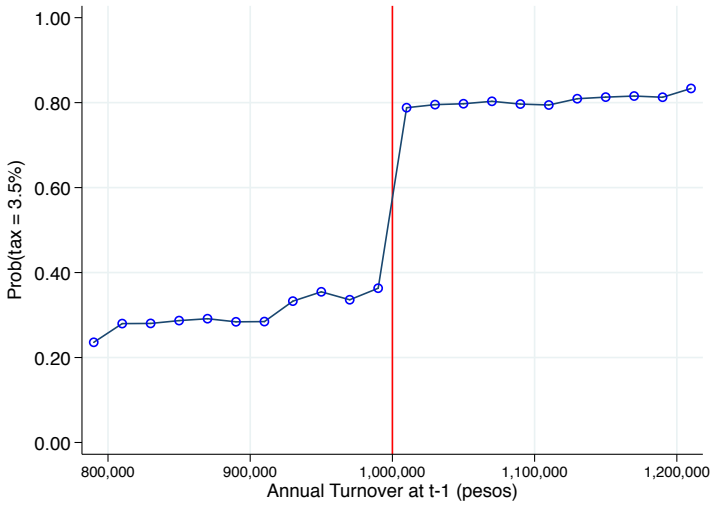
(c) Manufacturing



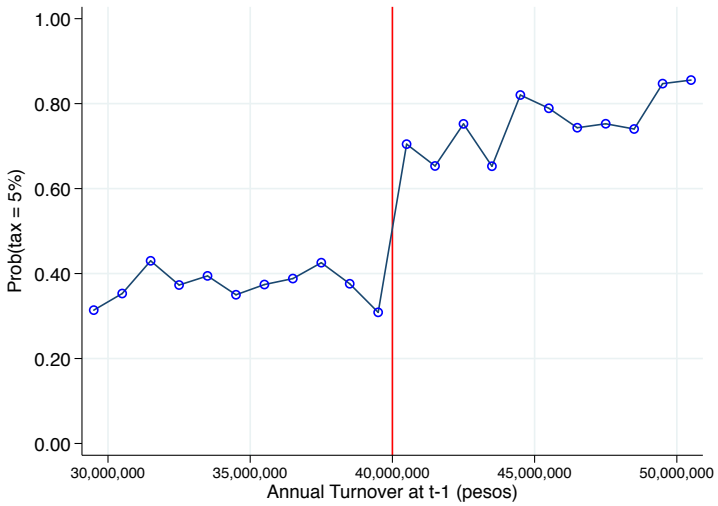
Notes: This graph separates the frequencies from Figure 14 into the main sector of the firm. The red vertical line indicates the administrative *notch* at which firms become collection agents. The bins are 200k pesos wide. $N = 6,254$ in panel (a), $N = 3,678$ in panel (b), and $N = 8,098$ in panel (c). The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts the number of taxpayers.

Figure 16: First stage tax variation and compliance: Retail and wholesale

(a) Notch 1: statutory tax rate goes from 3 to 3.5 percent



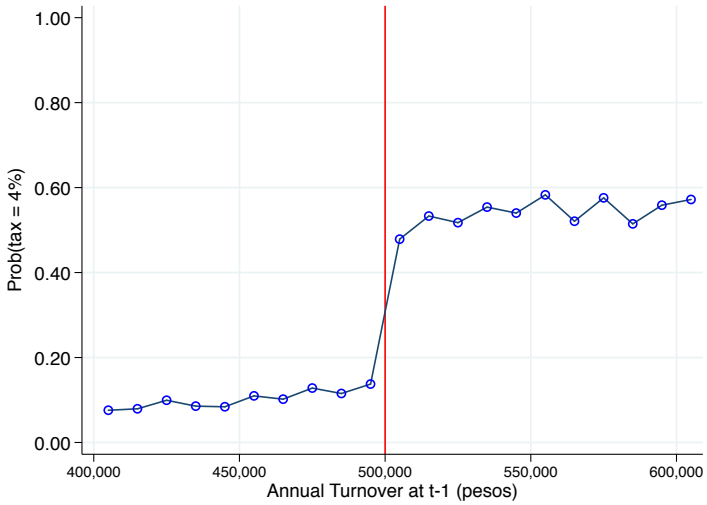
(b) Notch 2: statutory tax rate goes from 3.5 to 5 percent



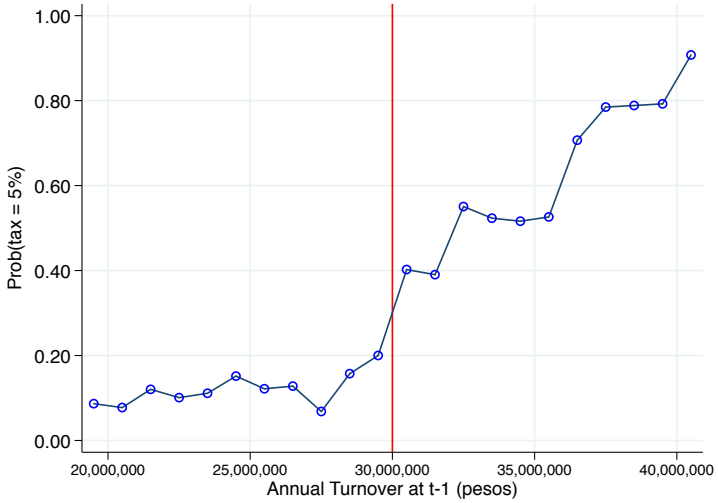
Notes: These graphs show the fraction of firms paying the tax rate that applies to the right of the threshold (vertical axis) as a function of total turnover (horizontal axis). The vertical line indicates the *notch* at which the statutory tax rate changes discretely. Under perfect compliance, we should observe a step function that starts at zero and jumps to one. The sample, bins, and x-axis are the same as in the density figures.

Figure 17: First stage tax variation and compliance: Services

(a) Notch 1: statutory tax rate goes from 3.5 to 4 percent



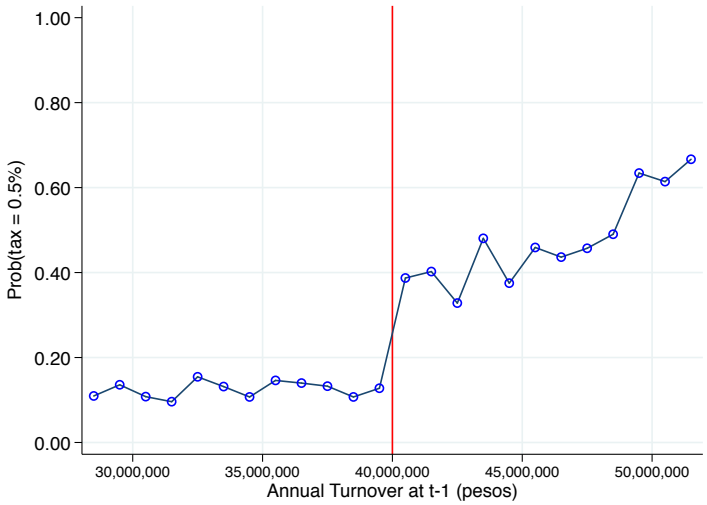
(b) Notch 2: statutory tax rate goes from 4 to 5 percent



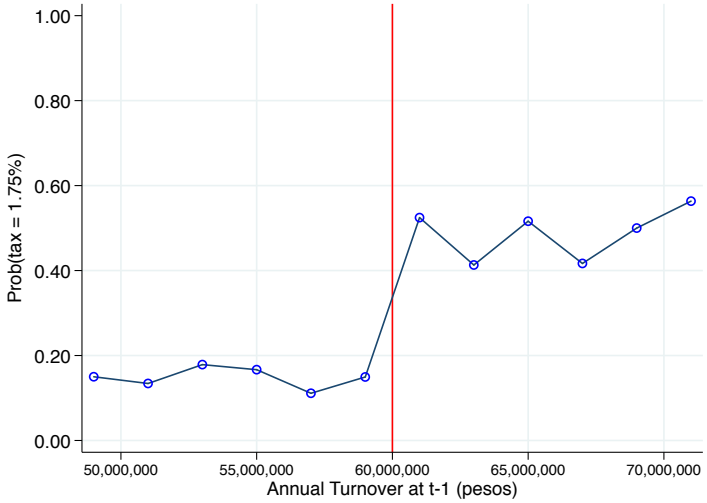
Notes: These graphs show the fraction of firms paying the tax rate that applies to the right of the threshold (vertical axis) as a function of total turnover (horizontal axis). The vertical line indicates the *notch* at which the statutory tax rate changes discretely. Under perfect compliance, we should observe a step function that starts at zero and jumps to one. The sample, bins, and x-axis are the same as in the density figures.

Figure 18: First stage tax variation and compliance: Manufacturing

(a) Notch 1: statutory tax rate goes from 0 to 0.5 percent



(b) Notch 2: statutory tax rate goes from 0.5 to 1.75 percent

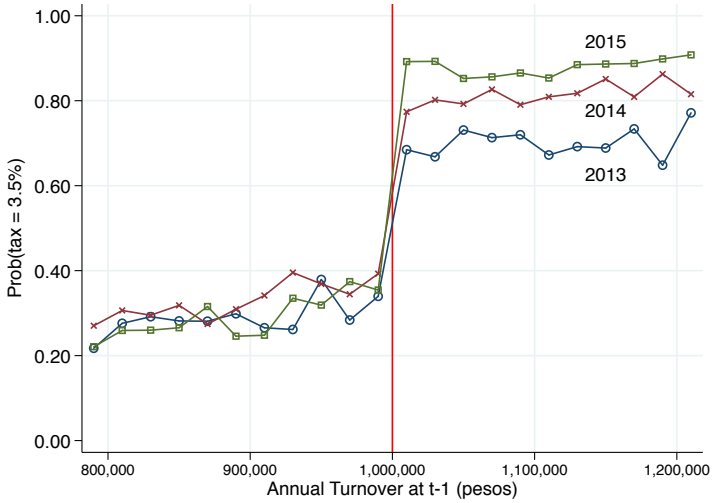


Notes: These graphs show the fraction of firms paying the tax rate that applies to the right of the threshold (vertical axis) as a function of total turnover (horizontal axis). The vertical line indicates the *notch* at which the statutory tax rate changes discretely. Under perfect compliance, we should observe a step function that starts at zero and jumps to one. The sample, bins, and x-axis are the same as in the density figures.

Figure 19: First stage tax variation and compliance, by years

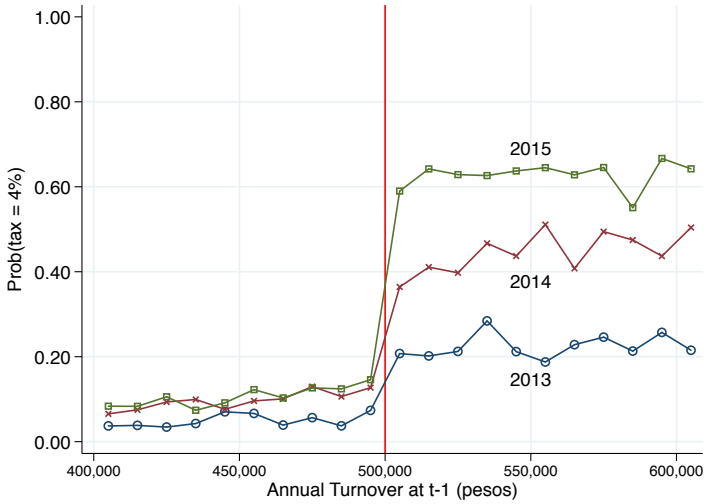
Retail and Wholesale

(a) Notch 1: statutory tax rate goes from 3 to 3.5 percent



Services

(b) Notch 2: statutory tax rate goes from 3.5 to 4 percent

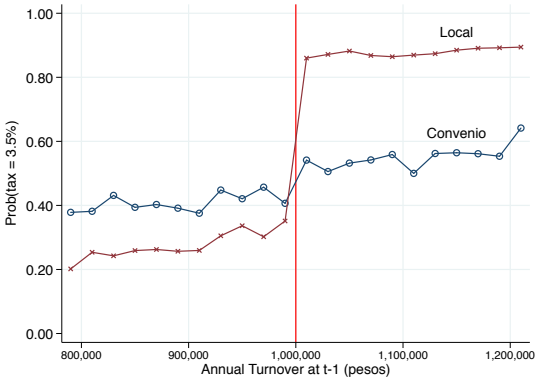


Notes: These graphs show the fraction of firms paying the tax rate that applies to the right of the threshold (vertical axis) as a function of total turnover (horizontal axis). The vertical line indicates the *notch* at which the statutory tax rate changes discretely. Under perfect compliance, we should observe a step function that starts at zero and jumps to one. The sample, bins, and x-axis are the same as in the density figures.

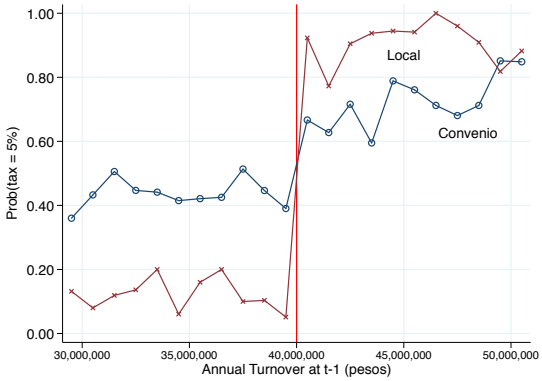
Figure 20: First stage tax variation and compliance, by regimen

Retail and Wholesale

(a) *Notch* at 1*m*

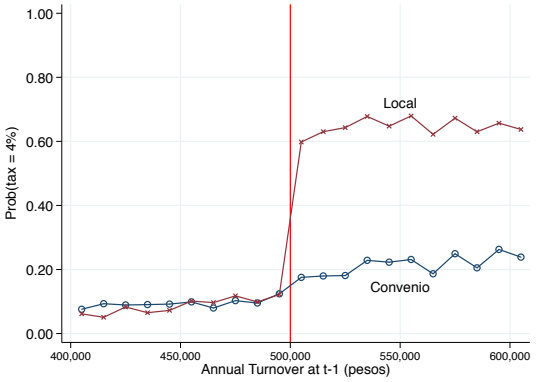


(b) *Notch* at 40*m*

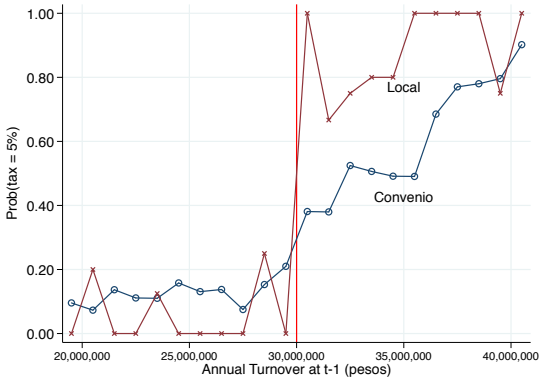


Services

(c) *Notch* at 500*k*



(d) *Notch* at 30*m*



Notes: These graphs show the fraction of firms paying the tax rate that applies to the right of the threshold (vertical axis) as a function of total turnover (horizontal axis). The vertical line indicates the *notch* at which the statutory tax rate changes discretely. Under perfect compliance, we should observe a step function that starts at zero and jumps to one. The sample, bins, and x-axis are the same as in the density figures.

Table 1: Elasticites for the first *notches* in the *Monotributo* schedule

Notch z^*	Tax rate t	Change in tax Δt	*Reduced form* Alt (1)	Alt (2)
(1)	(2)	(3)	(4)	(5)
72,000	1.3%	0.9%	0.528	0.504
96,000	1.6%	1.0%	0.333	0.320
144,000	1.8%	1.6%	0.485	0.456
192,000	2.5%	0.9%	0.683	0.646
240,000	2.8%	0.8%	0.648	0.617
288,000	2.9%	3.8%	0.035	0.034
400,000	4.8%	8.0%	0.016	0.016

Notes: This table presents the reduced-form elasticities for the first 8 discontinuities present in the *Monotributo*. To approximate these elasticities we follow Kleven & Waseem (2013) and, in particular, equations (4) and (5) of the notes that Kleven wrote afterwards to clarify the already published version of the paper (*Calculating Reduced-Form Elasticities Using Notches*, 2018). In column [4] we calculate the elasticity as $\frac{1}{2} \cdot \frac{(\Delta z^*/z^*)^2}{\Delta t/(1-t)}$ while in column [5] using the following formula $\frac{1}{2+\Delta z^*/z^*} \cdot \frac{(\Delta z^*/z^*)^2}{\Delta t/(1-t)}$. As a clarification, we do not present the elasticity for the first *notch* given that when we focus on the sample including rounding numbers the excess of bunching disappears. The change in tax that we consider in the last *notch* is an approximation of the total cost implied by switching to a different regime.

A Appendix Figures and Tables

Table 2: Categories *Monotributo*

Categories *Monotributo*	Upper threshold - Gross Earnings (annual)	
	Jan10-Sep13	Sep13-Jan17
B	24,000	48,000
C	36,000	72,000
D	48,000	96,000
E	72,000	144,000
F	96,000	192,000
G	120,000	240,000
H	144,000	288,000
I	200,000	400,000
J	235,000	470,000
K	270,000	540,000
L	300,000	600,000
Law/Resolution	Law 26.565	Res. General 3.529

Source: Own elaboration based on official documentation i.e., laws and general resolutions.
Notes: The thresholds are expressed in current Argentinean pesos and are not adjusted for inflation.

Table 3: Monthly tax

Period	Categories	*Impuesto integrado*	
		Services	Retail/Wholesale
Jan10-Jan17 Law 26.565	B	39	39
	C	75	75
	D	128	118
	E	210	194
	F	400	310
	G	550	405
	H	700	505
	I	1,600	1,240
	J		2,000
	K		2,350
	L		2,700

Source: Own elaboration based on official documentation i.e., laws and general resolutions.
Notes: The differential rate between the two different columns varies depending on the activity each self-employed works in. These tax duties should be paid on a monthly basis.

Table 4: Social security contributions (subcomponents of the *monotributo*)

Retirement	SIPA
Jan10-Jul12 Law 26.565	110
Jul12-Jan17 Res. General 3.334	157
Health	**Obra Social**
Jan10-Jul12 Law 26.565	70
Jul12-Nov13 Res. General 3.334	100
Nov13-Sep14 Res. General 3.533	146
Sep14-Jul15 Res. General 3.653	233
Jul15-Jun-16 Res. General 3.775	323
Jun16-act Res. General 3.845	419

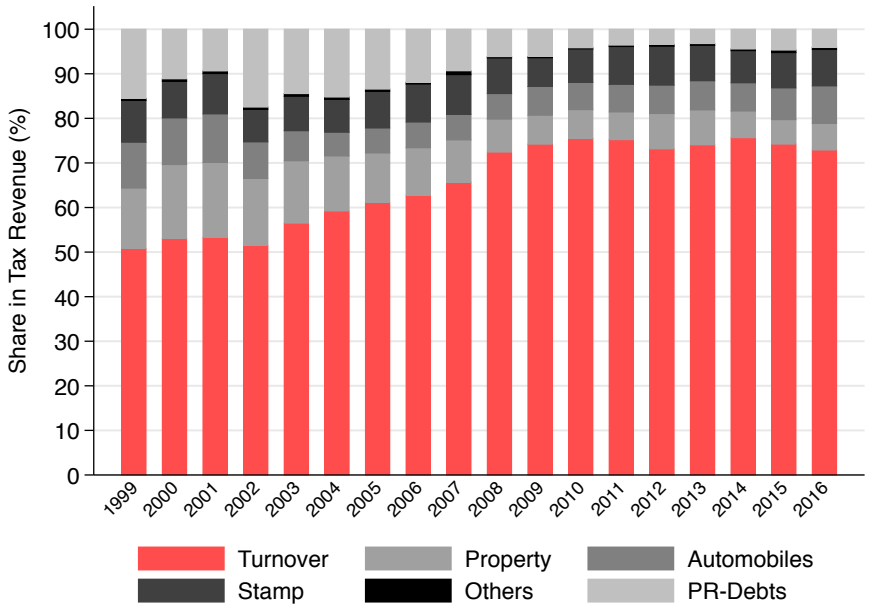
Source: Own elaboration based on official documentation i.e., laws and general resolutions.
Notes: These contribution duties should be paid on a monthly basis.

Table 5: Variation of the Turnover Tax in the Province of Buenos Aires, 2006-2017

Year	Type of Activities									
	Wholesale and Retail			Services			Agriculture (*) and Manufacturing			
	(A)			(B)			(C)			
2017	2.50% <1.3M	3.50% <52M	**5.00%** F.J. or >52M	**3.50%** <650K	4.00% >650K	5.00% >39M	0.00% <52M	0.50% >52M	1.75% >78M	**4.00%** F.J.
2016	3.00% <1.3M	3.50% <52M	**5.00%** F.J. or >52M	**3.50%** <650K	4.00% >650K	5.00% >39M	0.00% <52M	0.50% >52M	1.75% >78M	**4.00%** F.J.
2015	3.00% <1M	3.50% <40M	**5.00%** F.J. or >40M	**3.50%** <500K	4.00% >500K	5.00% >30M	0.00% <40M	0.50% >40M	1.75% >60M	**4.00%** F.J.
2014	3.00% <1M	3.50% <40M	**5.00%** F.J. or >40M	**3.50%** <500K	4.00% >500K	5.00% >30M	0.00% <40M	0.50% >40M	1.75% >60M	**4.00%** F.J.
2013	3.00% <1M	3.50% <40M	**5.00%** F.J. or >40M	**3.50%** <500K	4.00% >500K	5.00% >30M	0.00% <40M	0.50% >40M	1.75% >60M	**4.00%** F.J.
2012 S2	3.00% <1M	3.50% <40M	**5.00%** F.J. or >40M	**3.50%** <30M		5.00% >30M	0.00% <60M		1.00% >60M	**3.00%** F.J.
2012 S1	3.00% <30M		**4.50%** F.J. or >30M	**3.50%** <30M		4.50% >30M	0.00% <60M		1.00% >60M	**3.00%** F.J.
2011	3.00% <30M		**4.50%** F.J. or >30M	**3.50%** <30M		4.50% >30M	0.00% <60M		1.00% >60M	**3.00%** F.J.
2010	3.00% <30M		**4.50%** F.J. or >30M	**3.50%** <30M		4.50% >30M	0.00% <60M		1.00% >60M	**3.00%** F.J.
2009	3.00% <30M		**4.50%** F.J. or >30M	**3.50%**			0.00% <60M		1.00% >60M	**3.00%** F.J.
2008	3.00% <30M		**4.50%** F.J. or >30M	**3.50%**			0.00% <60M		1.00% >60M	**3.00%** F.J.
2007	**3.00%**			**3.50%**			0.00% based in BA		**1.00%** F.J. (Agro)	**1.50%** F.J. (Manuf)
2006	**3.00%**			**3.50%**			0.00% based in BA		**1.00%** F.J. (Agro)	**1.50%** F.J. (Manuf)
2005	**3.00%**			**3.50%**			0.00% based in BA		**1.00%** F.J. (Agro)	**1.50%** F.J. (Manuf)

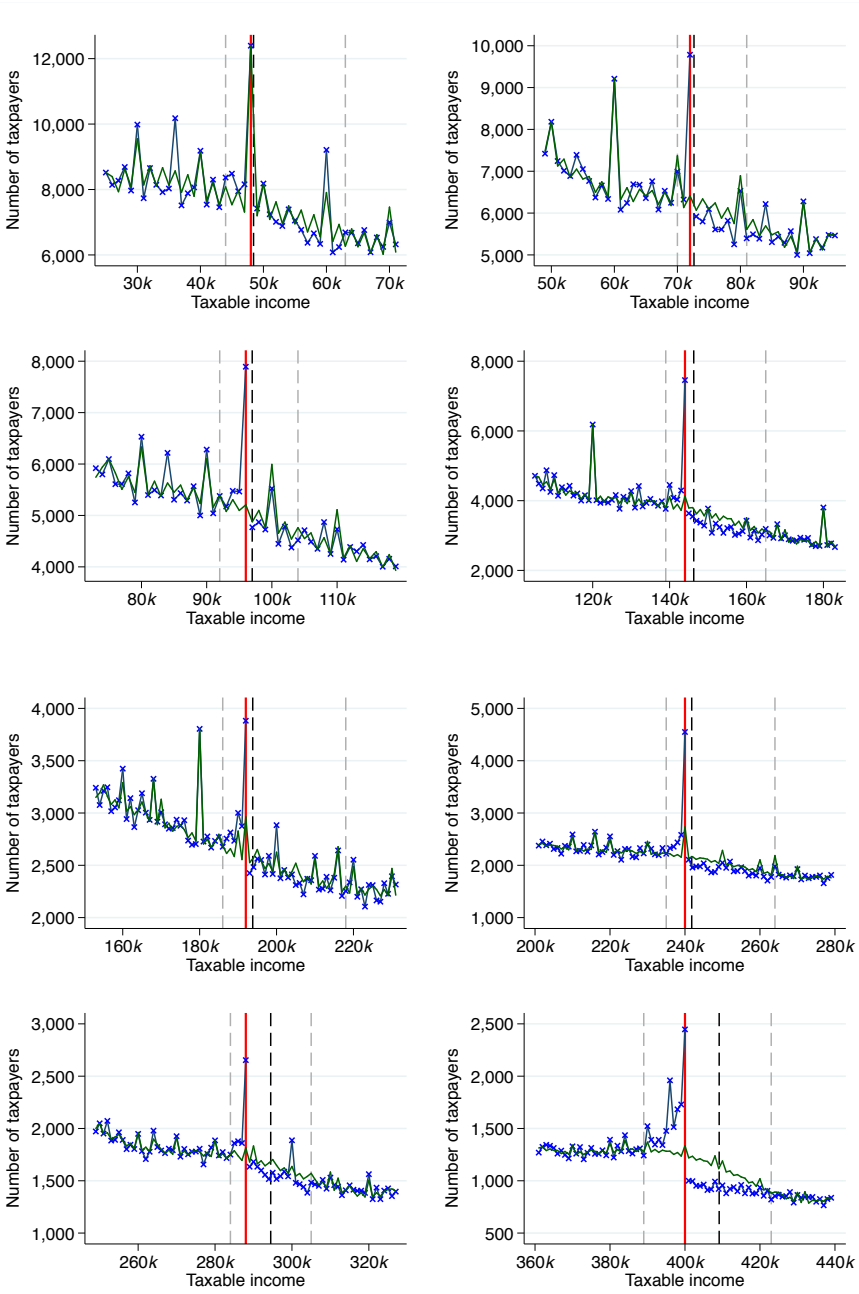
Source: Own elaboration based on the Tax Law of the Province of Buenos Aires 2006-2017.
Note: this table shows the variation of the turnover general tax rates for (A) wholesale and retail, (B) services, and (C) agriculture and manufacturing. In the top row of each year we report the tax rates and in the bottom row we report the condition the determines each firm's tax rate. Gross income thresholds denote annual sales at t_{-1}. F.J. denotes foreign jurisdiction (firms based outside the province and selling inside the province). (*) In agriculture, a tax rate of one percent applies to "Cultivation of cereals, oilseeds, and forage crops" and "Livestock breeding" (two percent for sowing pools with annual turnover at year t_{-1} greater than 10m).

Figure 21: Composition of own revenue (Buenos Aires province, 1999-2016)



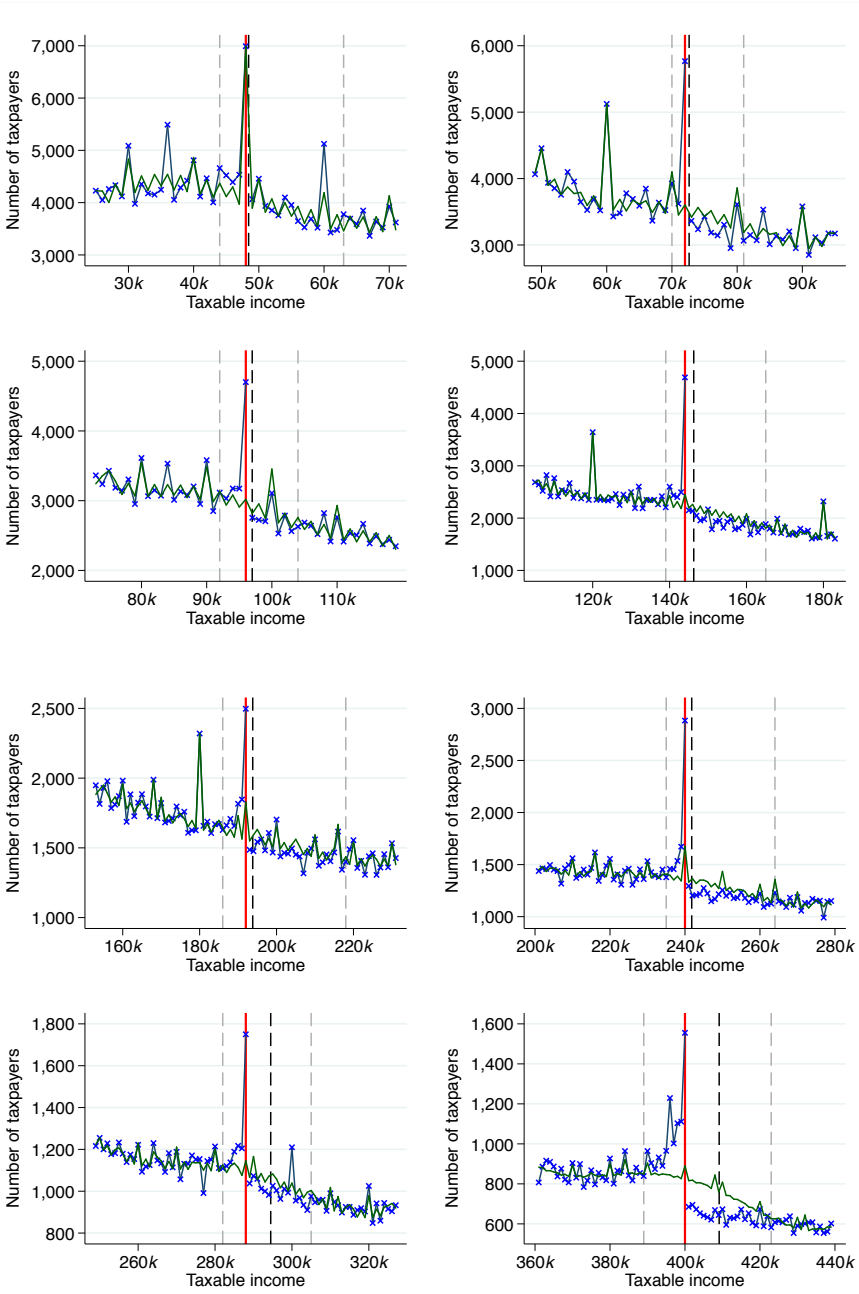
Source: The tax revenue is taken from the *Dirección Provincial de Política Tributaria* while the provincial gdp from the *Dirección Provincial de Estadística.*

Figure 22: Bunching at *monotributo notch* (full sample)



Notes: Taxpayers of the years 2014, 2015 and 2016 are grouped in bins of \$1,000. The solid red vertical lines refer to the different *notches*. The green solid line is the prediction of regressing equation (2) plus the round-number fixed effects i.e., $c_j = \sum_{i=0}^{p} \beta_i \cdot (z_j)^i + \sum_{i=z_L}^{z_U} \gamma_i \cdot [z_j = i] + \sum_{r \in R} \delta_r \cdot [\frac{z_j}{r} \in \mathbb{N}] + v_j$. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts the number of taxpayers. The grey dashed lines delimit the excluded area (lower and upper bounds respectively), while the black dashed one identifies the income value that makes the marginal buncher indifferent with the net-income she gets at the *notch*. These figures consider the full sample i.e., includes multiple of \$1,000.

Figure 23: Bunching at *monotributo notch* (panel sample)



Notes: Taxpayers of the years 2014, 2015 and 2016 are grouped in bins of \$1,000. The solid red vertical lines refer to the different *notches*. The green solid line is the prediction of regressing equation (2) plus the round-number fixed effects i.e., $c_j = \sum_{i=0}^{p} \beta_i \cdot (z_j)^i + \sum_{i=z_L}^{z_U} \gamma_i \cdot [z_j = i] + \sum_{r \in R} \delta_r \cdot [\frac{z_j}{r} \in \mathbb{N}] + v_j$. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts taxpayers. The grey dashed lines delimit the excluded area (lower and upper bounds respectively), while the black dashed one identifies the income value that makes the marginal buncher indifferent with the net-income she gets at the *notch*. These figures consider taxpayers that have filed taxes during the three consecutive years mentioned before and includes multiple of \$1,000.

Table 6: Elasticites for the first *notches* in the *Monotributo* schedule

(full sample)

Notch z^* (1)	Tax rate t (2)	Change in tax Δt (3)	*Reduced form* Alt (1) (4)	Alt (2) (5)
72,000	1.3%	0.9%	0.873	0.822
96,000	1.6%	1.0%	0.333	0.320
144,000	1.8%	1.6%	0.660	0.615
192,000	2.5%	0.9%	0.954	0.893
240,000	2.8%	0.8%	0.648	0.617
288,000	2.9%	3.8%	0.045	0.044
400,000	4.8%	8.0%	0.020	0.019

Notes: This table presents the reduced-form elasticities for the first 8 discontinuities present in the *Monotributo*. To approximate these elasticities we follow Kleven & Waseem (2013) and, in particular, equations (4) and (5) of the notes that Kleven wrote afterwards to clarify the already published version of the paper (*Calculating Reduced-Form Elasticities Using Notches*, 2018). In column [4] we calculate the elasticity as $\frac{1}{2} \cdot \frac{(\Delta z^*/z^*)^2}{\Delta t/(1-t)}$ while in column [5] using the following formula $\frac{1}{2+\Delta z^*/z^*} \cdot \frac{(\Delta z^*/z^*)^2}{\Delta t/(1-t)}$. These elasticities are estimated based on the bunching behavior presented in Figure 22. As a clarification, we do not present the elasticity for the first *notch* the excess of bunching is inexistent. The change in tax that we consider in the last *notch* is an approximation of the total cost implied by switching to a different regime.

Table 7: Elasticites for the first *notches* in the *Monotributo* schedule

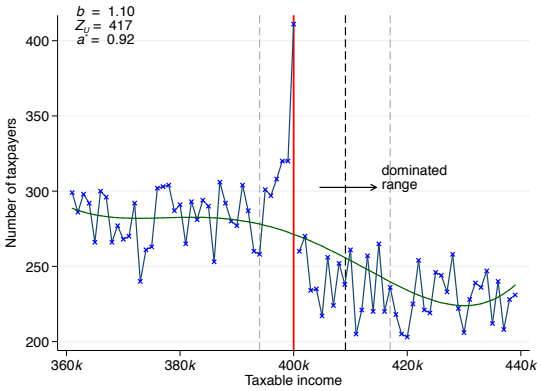
(panel sample)

Notch z^* (1)	Tax rate t (2)	Change in tax Δt (3)	*Reduced form* Alt (1) (4)	Alt (2) (5)
72,000	1.3%	0.9%	1.553	1.433
96,000	1.6%	1.0%	0.333	0.320
144,000	1.8%	1.6%	0.599	0.560
192,000	2.5%	0.9%	0.954	0.893
240,000	2.8%	0.8%	0.648	0.617
288,000	2.9%	3.8%	0.076	0.073
400,000	4.8%	8.0%	0.016	0.016

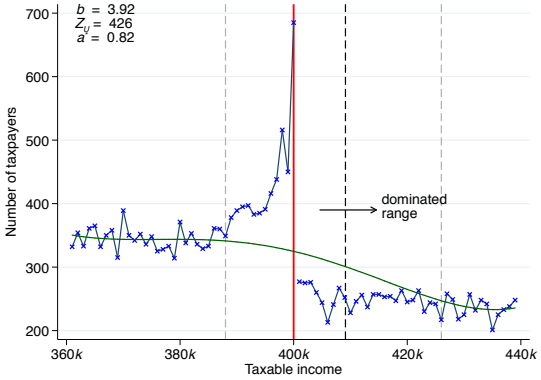
Notes: This table presents the reduced-form elasticities for the first 8 discontinuities present in the *Monotributo*. To approximate these elasticities we follow Kleven & Waseem (2013) and, in particular, equations (4) and (5) of the notes that Kleven wrote afterwards to clarify the already published version of the paper (*Calculating Reduced-Form Elasticities Using Notches*, 2018). In column [4] we calculate the elasticity as $\frac{1}{2} \cdot \frac{(\Delta z^*/z^*)^2}{\Delta t/(1-t)}$ while in column [5] using the following formula $\frac{1}{2+\Delta z^*/z^*} \cdot \frac{(\Delta z^*/z^*)^2}{\Delta t/(1-t)}$. These elasticities are estimated based on the bunching behavior presented in Figure 23. As a clarification, we do not present the elasticity for the first *notch* the excess of bunching is inexistent. The change in tax that we consider in the last *notch* is an approximation of the total cost implied by switching to a different regime.

Figure 24: Bunching at 400k by year

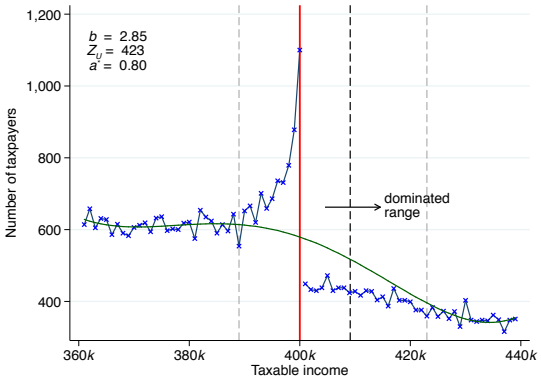
(a) 2014



(b) 2015

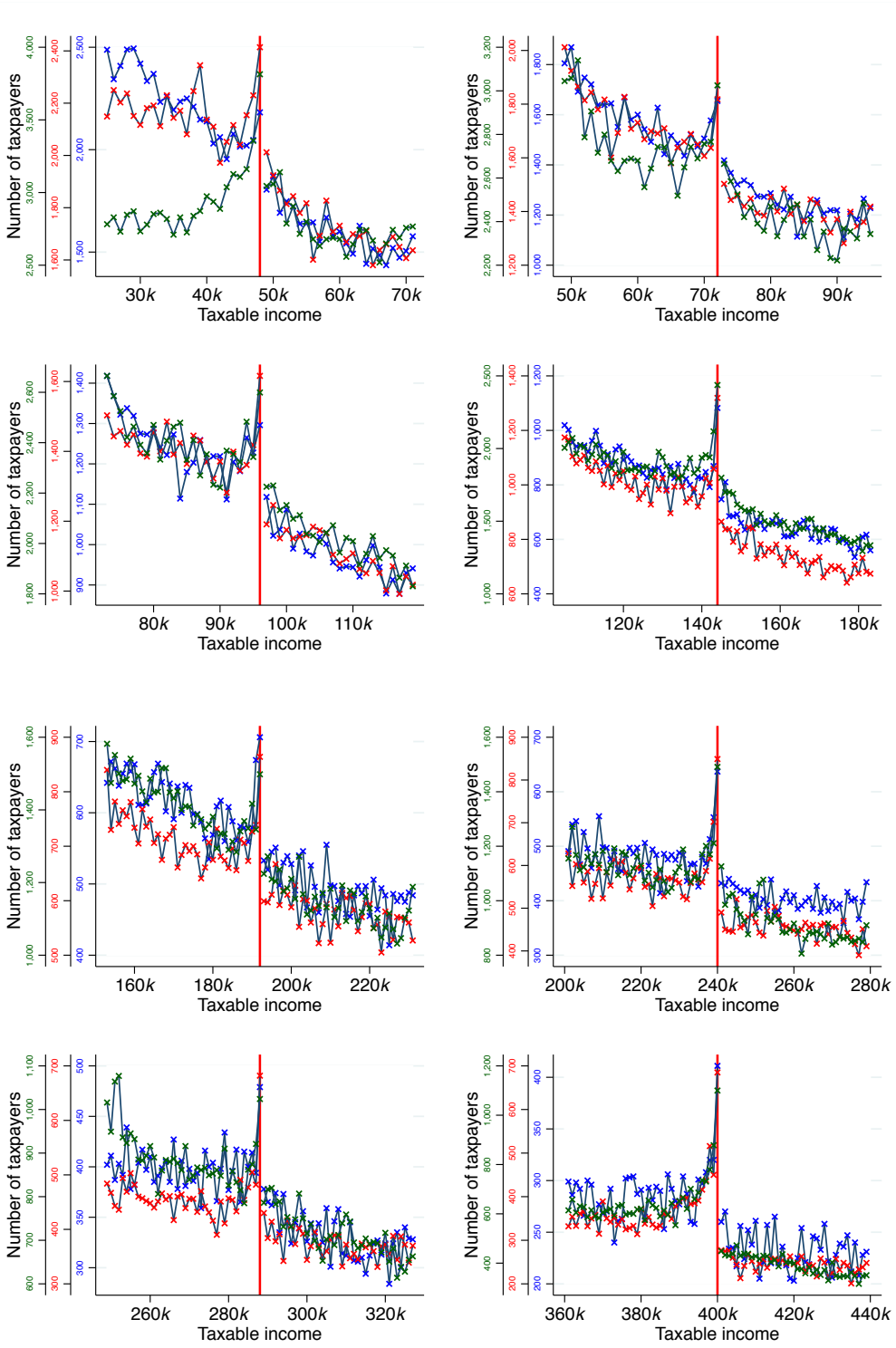


(c) 2016



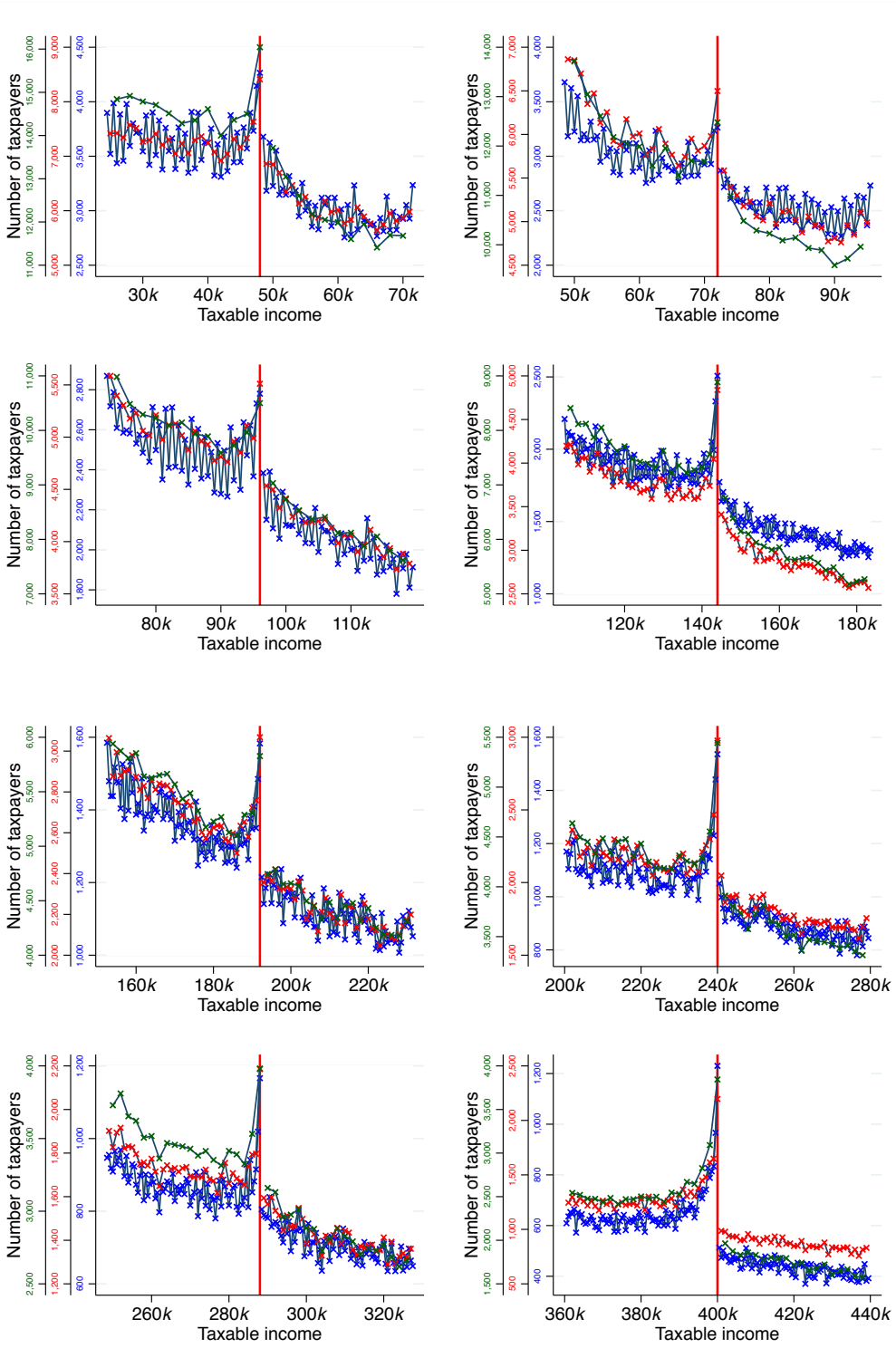
Notes: In this set of figures we focus on taxpayers that have filed taxes in the three years. The solid red vertical lines refer to the different *notches*. The green solid line is the prediction of regressing equation (2) with the caveat of the dummies in the excluded range. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts taxpayers. In the left box, b refers to the bunching mass, Z_U to the upper limit of the excluded area and a to the percentage of taxpayers in the dominated range who are unresponsive. The grey dashed lines delimit the excluded area (lower and upper bounds respectively), while the black dashed one identifies the income value that makes the marginal buncher indifferent with the net-income she gets at the *notch*.

Figure 25: Bunching at *monotributo notches* across years



Notes: Bunching remains when using different bin sizes. In each series, we group taxpayers of the years 2014, 2015 and 2016 using three different bins of \$500 (blue), \$1,000 (red) and \$2,000 (green) respectively. Note that the vertical axis, that counts the number of taxpayers, has a different scale for each series, where the color of the series corresponds to the color of the axis label. The horizontal axis measures taxable income which is equal to the annual turnover. We remove taxpayers whose reported taxable income is a multiple of 1,000 pesos.

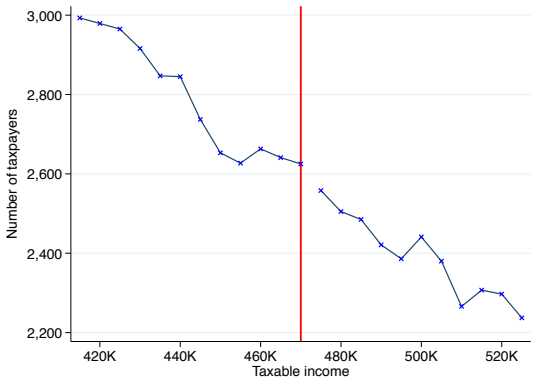
Figure 26: Bunching at *monotributo notches* and bin sizes



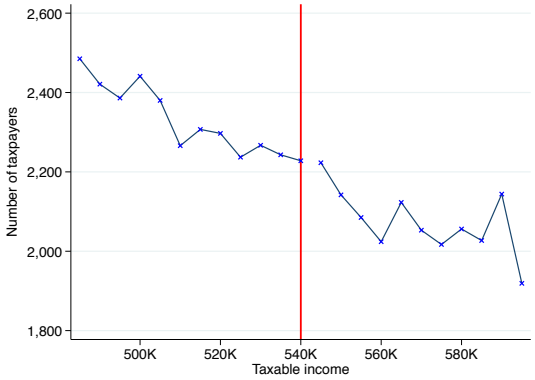
Notes: Bunching remains when using different bin sizes. In each series, we group taxpayers of the years 2014, 2015 and 2016 using three different bins of \$500 (blue), \$1,000 (red) and \$2,000 (green) respectively. Note that the vertical axis, that counts the number of taxpayers, has a different scale for each series, where the color of the series corresponds to the color of the axis label. The horizontal axis measures taxable income which is equal to the annual turnover. We remove taxpayers whose reported taxable income is a multiple of 1,000 pesos.

Figure 27: Bunching at the top *monotributo notches*

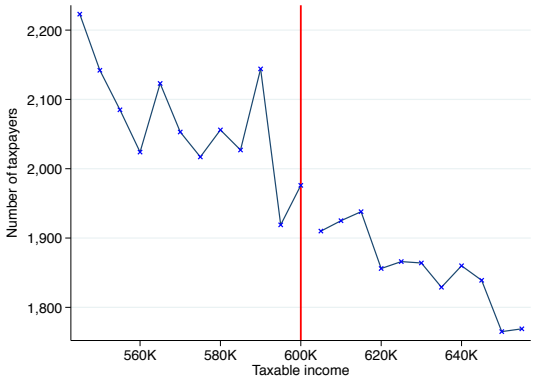
(a) 470k



(b) 540k



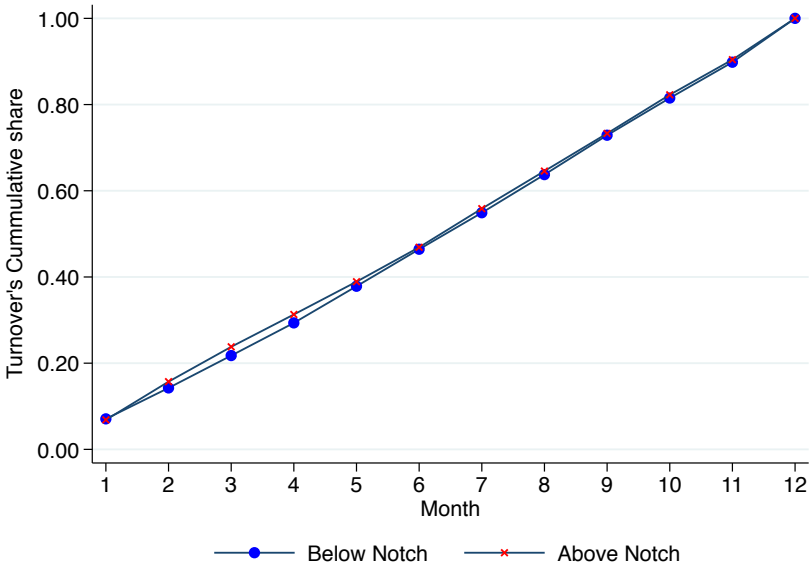
(c) 600k



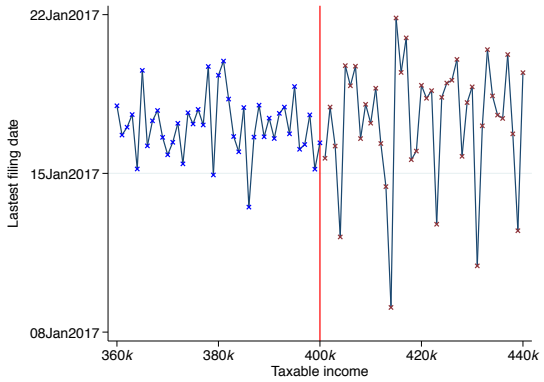
Notes: taxpayers of the years 2014, 2015 and 2016 are grouped in bins of \$1,000. The solid red vertical lines refer to the different *notches*. The horizontal axis measures taxable income which is equal to the annual turnover, while the vertical axis counts taxpayers. We remove taxpayers whose reported taxable income is a multiple of 1,000.

Figure 28: Bunchers' characteristics (at 400k)

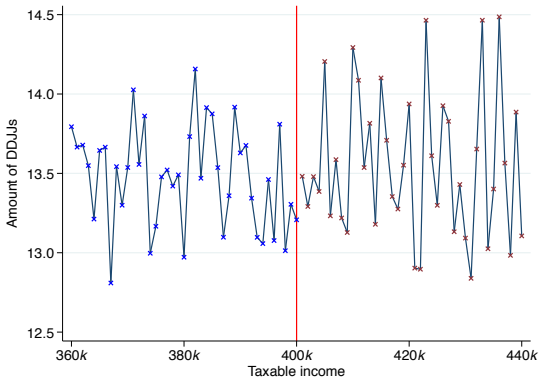
(a) Cumulative share



(b) Latest filing date



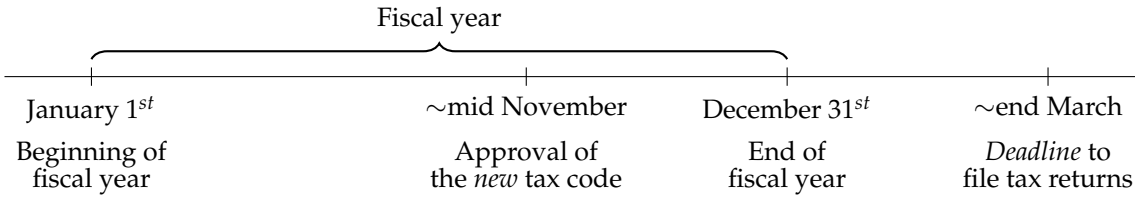
(c) Number of tax returns



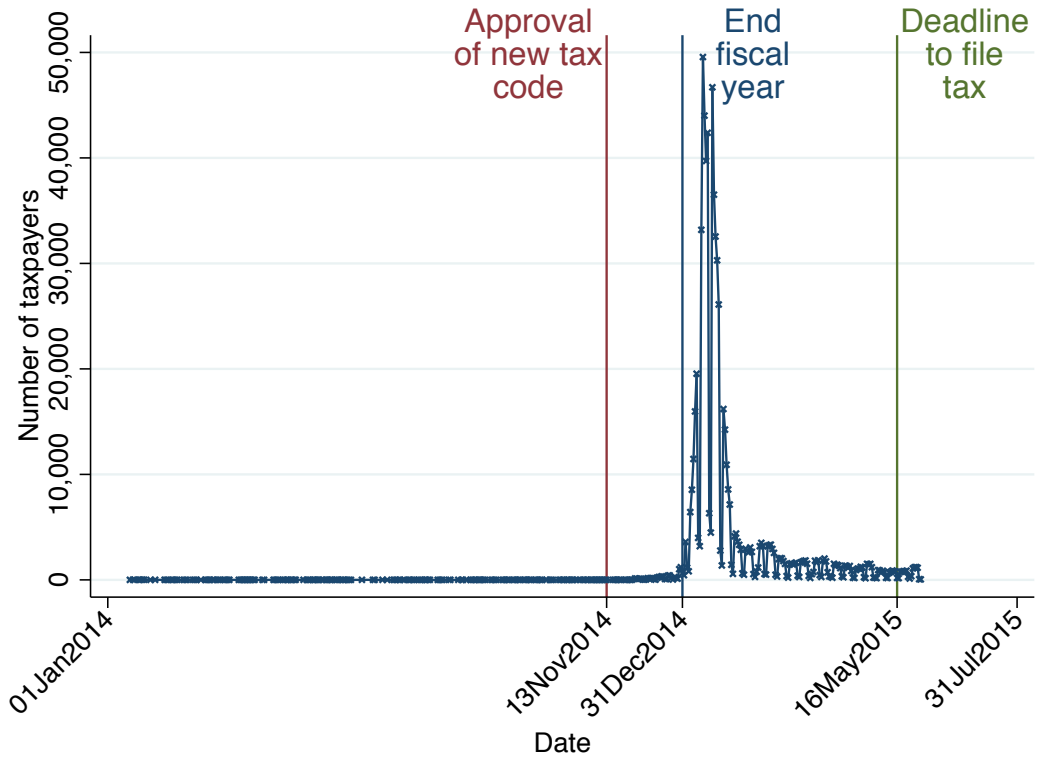
Notes: Panel (a) presents the evolution of the turnover's cumulative share across the different months (year 2016) for those taxpayers just right below the *notch* (bins 399k and 400k) and those just right above (bins 401k and 402k). Panel (b) contains the last filing date for each bin (year 2016). Panel (c) presents the number of declarations each taxpayer files. Note that taxpayers can adjust or correct filing from previous months.

Figure 29: GRT filing diagram

(a) Filing timeline



(b) Last filing date (2014)



Notes: In panel (a) we present the filing timeline of the gross receipts tax. The fiscal year goes from January 1st to December 31st each year. Normally, the tax code for the following fiscal year is approved by mid/end of November. Finally, the last data to file is generally three months after the fiscal year has finished. Simply, this means that a given firm will know the tax code that will be valid starting next year sometime in November and will have time to file the latest declaration three months after the fiscal year ends. In November the firm becomes aware of the *notches* that will be in place, and thus has time to *manipulate* the reported revenue to lower its tax liability. In panel (b) we present an example using data from 2014.

Figure 30: Tax code

Wholesale and retail sector

(a) Article 21

ARTÍCULO 21. De acuerdo a lo establecido en el artículo 223 del Título II del Código Fiscal -Ley Nº 10397 (Texto ordenado 2011) y modificatorias-, fíjanse las siguientes alícuotas generales del impuesto sobre los Ingresos Brutos:
A) Establécese la alícuota del cinco por ciento (5%) para las siguientes actividades, en tanto no tengan previsto otro tratamiento en esta Ley o se encuentren comprendidas en beneficios de exención establecidos en el Código Fiscal o Leyes especiales:

(b) Article 23

ARTÍCULO 23. Establécese en tres con cinco por ciento (3,5%) la alícuota del impuesto sobre los Ingresos Brutos aplicable exclusivamente a las actividades detalladas en el inciso A) del artículo 21, cuando las mismas se desarrollen en establecimiento ubicado en la provincia de Buenos Aires y el total de ingresos gravados, no gravados y exentos, obtenidos por el contribuyente en el período fiscal anterior, por el desarrollo de cualquier actividad dentro o fuera de la Provincia, no supere la suma de pesos cuarenta millones ($40.000.000).

(c) Article 24

ARTÍCULO 24. Establécese en tres por ciento (3%) la alícuota del impuesto sobre los Ingresos Brutos aplicable exclusivamente a las actividades detalladas en el inciso A) del artículo 21, cuando las mismas se desarrollen en establecimiento ubicado en la provincia de Buenos Aires y el total de ingresos gravados, no gravados y exentos, obtenidos por el contribuyente en el período fiscal anterior, por el desarrollo de cualquier actividad dentro o fuera de la Provincia, no supere la suma de pesos un millón ($1.000.000)

Service sector

(d) Article 21

ARTÍCULO 21. De acuerdo a lo establecido en el artículo 223 del Título II del Código Fiscal -Ley Nº 10397 (Texto ordenado 2011) y modificatorias-, fíjanse las siguientes alícuotas generales del impuesto sobre los Ingresos Brutos:

B) Establécese la alícuota del tres con cinco por ciento (3,5%) para las siguientes actividades, en tanto no tengan previsto otro tratamiento en esta Ley o se encuentren comprendidas en beneficios de exención establecidos en el Código Fiscal o Leyes especiales:

(e) Article 25

ARTÍCULO 25. Establécese en cuatro por ciento (4%) la alícuota del impuesto sobre los Ingresos Brutos aplicable a las actividades detalladas en el inciso B) del artículo 21 de la presente Ley, cuando el total de ingresos gravados, no gravados y exentos obtenidos por el contribuyente en el período fiscal anterior, por el desarrollo de cualquier actividad dentro o fuera de la Provincia supere la suma de pesos quinientos mil ($500.000).

(f) Article 26

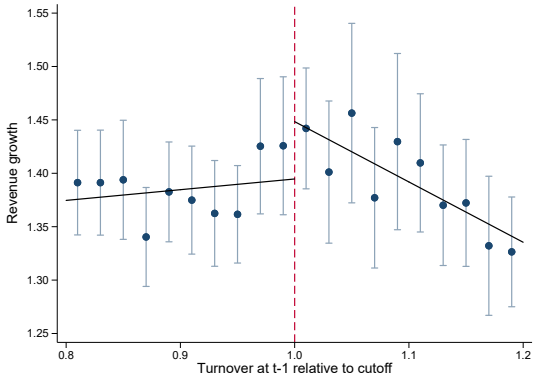
ARTÍCULO 26. Establécese en cinco por ciento (5%) la alícuota del impuesto sobre los Ingresos Brutos aplicable a las actividades detalladas en el inciso B) del artículo 21 de la presente Ley, cuando el total de ingresos gravados, no gravados y exentos obtenidos por el contribuyente en el período fiscal anterior, por el desarrollo de cualquier actividad dentro o fuera de la Provincia supere la suma de pesos treinta millones ($30.000.000).

Notes: In the six panels listed above we present the different articles that contain the tax rates that should be applied to those taxpayers in wholesale and retail, and service sector respectively. Importantly, we keep the order in which they appear in the tax code.

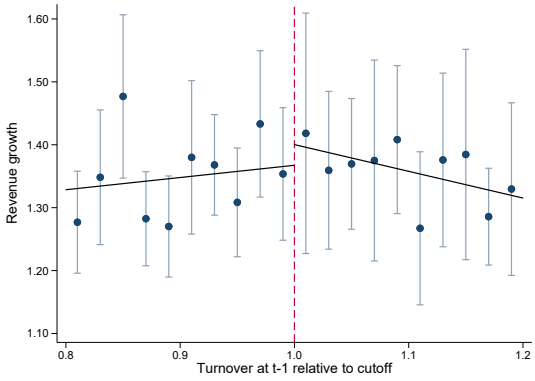
Figure 31: Regression discontinuity estimates and turnover growth

Wholesale and retail

(a) *Notch* at 1m

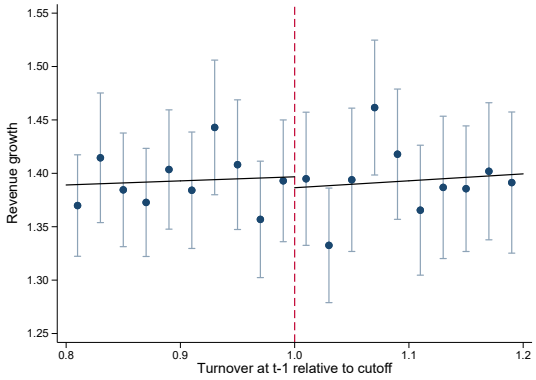


(b) *Notch* at 40m

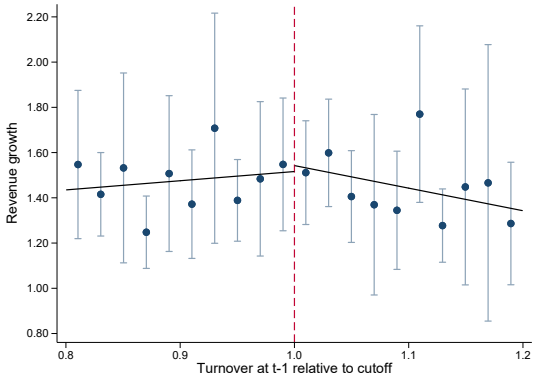


Services

(c) *Notch* at 500k

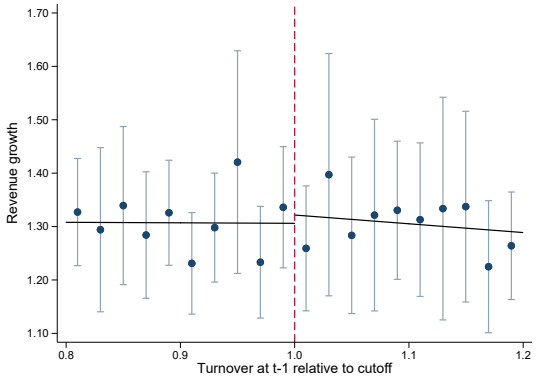


(d) *Notch* at 30m

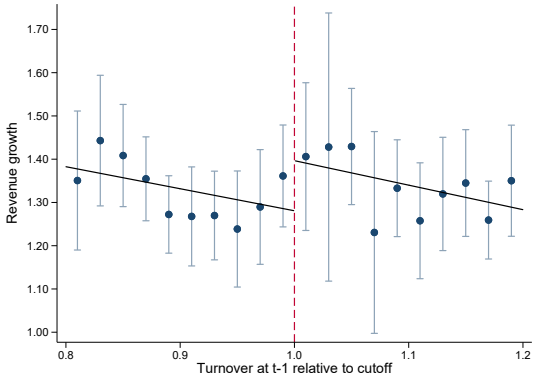


Agro and manufacture

(e) *Notch* at 40m

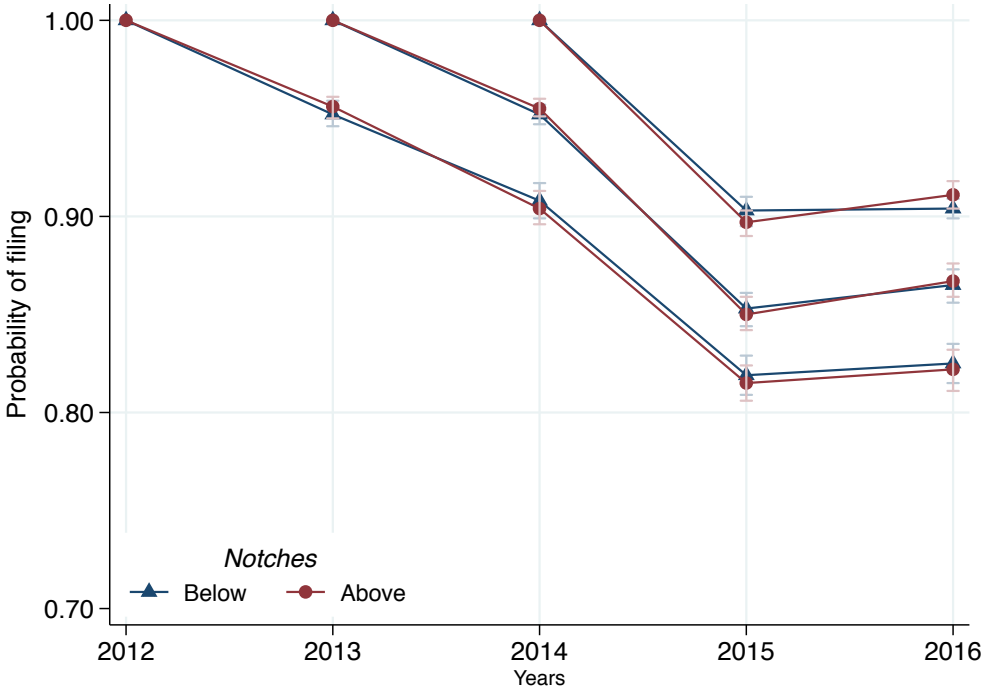


(f) *Notch* at 60m



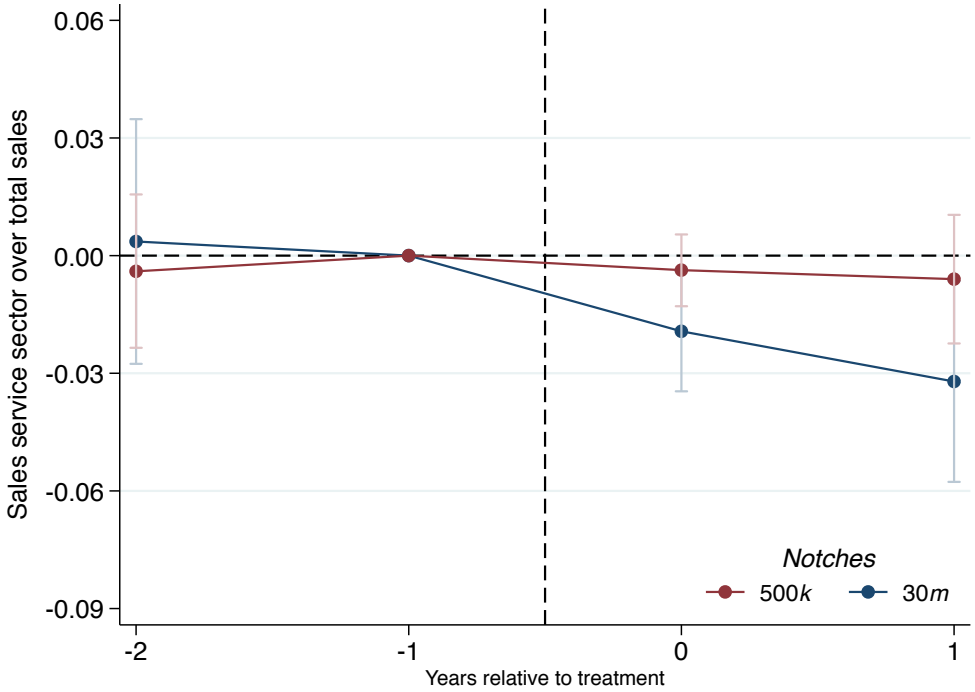
Notes: These graphs shows the regression discontinuity estimates (RD) using as dependent variable the growth rate in turnover between t_{-1} and t using pooled data for the years 2012-2014. The horizontal axis measure the distance to the *notch* while the vertical axis the growth rate in turnover

Figure 32: Other margins: survival probability



Notes: this figure shows the survival probability of taxpayers that were just below at just above the GRT notches in 2012 and calculate the probability of filing a tax in the subsequent years (same for 2013 and 2014). Note that we are pooling tax payers just below/above for the different *notches* and we consider windows of $\$20,000$ $(500,000)$, $\$40,000$ $(1,000,000)$ and $\$1,000,000$ (for higher *notches*) so as to have a minimum amount of observations on each side of all of them. In order to get the standard errors we follow a bootstrap procedure.

Figure 33: Other margins: shift bases across sectors



Notes: this figure plots the gammas coefficients of the following event study specification $y_{i,t} = \sum_{j=-3}^{2} \gamma_j \cdot d_{i,t}^{j} + \mu_i + \mu_t + \epsilon_{i,t}$ with firm and time fixed effects, taking the ratio of sales in the service sector relative to total sales as dependent variable. We identify treated firms as those that in 2013 (2014) were above the notch but were below in 2011 and 2012 (2012 and 2013 respectively). The treatment year corresponds to the first period where the company passes the threshold.