ECONOMIC DEVELOPMENT & INSTITUTIONS

EDI WORKING PAPER SERIES

RED TAPE? THE REVENUE IMPACT OF THE VAT FILING THRESHOLDS*

Jan Luksic†
European Commission

Shekhar Mittal‡
University of California

May 2019

Oxford Policy Management · UKaid from the British people · Université de Namur · Paris School of Economics – École d'économie de Paris · ADE Analysis for economic decisions

Abstract

Value-added tax systems across the world are afflicted with size-dependent regulations.The benefit of such regulations to the tax authority is unclear. In this paper, we use an administrative dataset from the state of Delhi in India to first show that a policy which mandated different frequencies of filing based on self-reported turnover resulted in bunching of firms below the thresholds at all levels. Using the subsequent change in these reporting policies, we provide evidence that such sharp bunching indeed occurs due to the VAT reporting frequency thresholds. We document that such bunching partly occurs due to turnover shifting and underreporting, provide evidence that the observed bunching has no growth consequences for the bunching firms - and find that bunching occurs to similar degree across industries. Second, we calculate the VAT revenue losses due to such bunching. Third, the subsequent withdrawal of the policy allows us to show that in a regime with size-dependent reporting requirements, more frequent reporting is not associated with greater VAT collection. Finally, according to our back of the envelope welfare analysis, the sized-based filing policy is welfare improving if a welfare-maximizing government's objective function assigns important weights to small- and medium-sized enterprises.

JEL codes: H26, H32, O38.

Keywords: Value-added tax, size-based regulation, reporting rules, notches, bunching.

* This study would not have been possible without the generous time, access to data, and answers to questions provided by officials in the Department of Trade and Taxes of the Government of National Capital Territory of Delhi (GoNCTD). We thank Manish Sisodia, Hon'ble Deputy Chief Minister, GoNCTD for his undiminishing support. We thank Jens Dosti and Erik Nettelroth for excellent research assistance. We also thank Pierre Bachas, Youssef Benzarti, Michael Best, Guido Friebel, Paola Giuliano, Adriana Lleras-Muney, Aprajit Mahajan, Juan Matamala, Ricardo PerezTruglia, Anna Rohlfing-Bastian, Divya Singh, Nico Voigtlander, Roman Wacziarg, and participants at the GEM seminar (Anderson School, UCLA), and the seminar at Goethe University. Shekhar Mittal acknowledges financial support from the International Growth Center (India) Grant #89412, from JPAL-GI, and of UK aid from the UK government. The findings, interpretations, and conclusions are those of the authors alone and do not represent the views of the European Commission, GoNCTD, or any other institution.

†European Commission. jan.luksic@ec.europa.eu

‡University of California, Los Angeles. shekhar.mittal.1@anderson.ucla.edu

1 Introduction

Value-added tax (VAT) is considered to be an effective tool to raise revenues by governments across the globe (Keen & Smith, 2006). However, VAT systems across the world are afflicted with size-dependent regulations that form an elementary part of the VAT administration. Not only is there evidence that monitoring effort is being directed on larger firms (Almunia & Lopez-Rodriguez, 2018; Bachas & Jensen, 2017), but also that both the VAT registration thresholds as well as the VAT reporting requirements depend crucially on the reported firm turnover (Ernst & Young, 2015).

However, the benefit of size-dependent filing frequency regulations to the tax authority is unclear. On the one hand, the tax authority may prefer to receive as much information as quickly as possible. The government may also be liquidity constrained and prefer to receive VAT payments from firms at as high a frequency as possible, pending the administrative concerns on the government's end. On the other hand, there are theoretical arguments that the tax authority should economize on administrative and compliance costs by exempting small firms from taxation (Dharmapala *et al.* , 2011). Empirically, the hassle costs associated with filing taxes are certainly substantial, and rising with income (Benzarti, 2017).

Such hassle costs might be particularly crucial in low and middle income countries, where compliance costs are of first order concern. Additionally, compliance costs might be greater for smaller firms - with little benefit to the tax authority, as most of the VAT revenue stems from very large firms (International Tax Dialogue, 2007, 2013).

Our paper is one of the first to carefully investigate how firms respond to multiple VAT filing thresholds on the intensive margin. We do this by taking advantage of the unique setting in the state of Delhi in India. A careful analysis of the VAT filing thresholds, on the intensive margin, fills an important gap in the literature as the VAT filing thresholds are quite ubiquitous, but their impact is only beginning to be studied.In a number of countries, firms close to the VAT registration threshold actively manipulate their reported turnover in order to avoid registering for VAT (Onji, 2009; Gebresilasse & Sow, 2016; Liu *et al.* , 2017; Harju *et al.* , 2016; Boonzaaier *et al.* , 2017). Furthermore, firms may also be responding to VAT filing thresholds at the intensive margin (Asatryan & Peichl, 2017).

Many countries mandate a uniform rate of VAT reporting at either a monthly (e.g., Argentina,

India [GST]), or a bi-monthly (e.g., Barbados), or a quarterly (e.g., Cyprus) frequency. Simultaneously, many countries have policies that mandate different rate of VAT return filing based on firm size. Out of a 2015 sample of 103 countries for which we could determine the reporting frequencies with certainty, 39 countries had size-dependent frequencies of VAT reporting and payments.[1] Size-dependent filing requirements are present in both high-income (e.g., Austria, Germany, Denmark, Finland, France, Ireland, Spain, UK, etc.) as well as low- and middle-income (e.g., Botswana, Colombia, Mauritius, Philippines, South Africa, Swaziland) countries. If firms misreport their turnover in order to avoid filing (and remitting) VAT returns more frequently, it is important to understand the associated costs to the VAT collections and to understand the mechanism through which the firms are able to do it.

But before answering this question empirically, it is important to understand what the "first-best" policy with regards to VAT filing frequencies might be. In an environment without any frictions, both firms and government should be indifferent about the frequency of filing returns. However, a liquidity constrained government, more likely in low and middle income countries, may prefer to receive VAT reports and the associated payments at as high of a frequency as possible. On the other hand, a high frequency of VAT reporting and payments imposes a significant compliance burden on firms, especially on small ones. Therefore, while the government prefers to obtain information and the VAT payments as frequently as possible, it should be amenable to subsidizing smaller firms - those, eligible for VAT - by imposing less stringent filing requirements on them.

Having said that, such policies may lead to firms actively avoiding filing returns more frequently, by either underreporting their true turnover, or by under-producing and thus intentionally limiting their growth. Both these channels are troublesome. The first one creates an environment that fosters a culture of evasion and avoidance, which is especially problematic in the context of the low- and middle-income countries. The second channel results in production inefficiencies, keeping the economy below its potential. In subsequent revisions of our work, we intend to make progress on identifying the dominant channel.

In this paper, we use an administrative dataset from the state of Delhi in India to show that a policy which mandated different frequencies of filing based on self-reported turnover resulted in bunching of firms below each of the filing frequency thresholds. For the five years for which we have the VAT tax returns, firms had to file returns at a specific frequency depending on their

[1] Based on the survey data from Ernst & Young (2015).

declared turnover in the previous financial year. For the first two years, Y1 and Y2 (financial years 2010/11 and 2011/12), there were 3 thresholds of interest: The low threshold at ₹1 million (Indian rupees; ~$15,000), the middle threshold at ₹5 million (~$77,000), and the high threshold at ₹50 million (~$770,000). In the third year, Y3 (financial year 2012/13), the low and the middle thresholds were disbanded but the high threshold still existed. In years Y4 and Y5 (financial years 2013/14 and 2014/15) of our data, all firms had to file returns at a uniform quarterly frequency and there were no filing frequency thresholds.

Our bunching estimates vary from 0.28 for the low threshold to 2.98 for the high threshold. Using the transition to a uniform reporting policy in the subsequent years within our dataset, we show that such sharp bunching in the earlier years indeed occurs due to the VAT filing frequency thresholds. We do this by documenting the immediate cessation of the bunching behavior after the policy change.

We further discuss the three potential bunching channels: underproduction ("production halting"), underreporting (via complete misreporting), and turnover shifting (intial underreporting). We provide indicative evidence that turnover shifting and underreporting are the main channels driving the bunching behavior of the firms below the threshold cut-off.

Second, the bunching behavior results in substantial VAT revenue losses. As per our calculations, for all the thresholds and across all the years, the bunching related losses range from 2% to 8.3% of the VAT contributions of the firms indulging in bunching behavior.

Third, using various regression specifications, we estimate the relationship between the number of returns that firms have to file in a given financial year and the amount of VAT remitted by these firms. We are able to do this due to the rich policy variation that we observe in our period of interest. Our analysis indicates that more frequent reporting does not lead to greater levels of VAT collection.

Finally, we perform a back of the envelope welfare analysis. We calculate the implicit subsidy that the tax authority can offer to all the firms in order to file increased number of returns at each of the thresholds and thereby nullify the revenue losses on account of the bunching behavior. The filing policy is Pareto improving compared to a uniform filing policy if the implicit subsidy is lower than the likely compliance costs incurred by firms for increased frequency of reporting. Our estimates indicate that the minimum implied subsidies are expected to be much lower than the expected compliances costs for filing. Therefore, the sized-based filing policy is welfare improving if the small and medium sized firms are deemed to be important by the welfare-maximizing policy-

maker.

The paper is organized as follows. The following section describes the related literature. Section 3 describes the reforms in the VAT filing system in Delhi and the administrative level data from Delhi that we use for our analysis, and explains the methodology we use for the analysis. Section 4 presents the results of our analysis of the firm response, of the revenue implications, and of welfare. Section 5 concludes.

2 Related Literature

A growing literature has empirically analyzed the effect of the VAT registration threshold. Onji (2009) was the first paper to recognize the reaction of firms to the VAT registration threshold by documenting that large firms masquerade as many small firms in response to the VAT registration threshold in Japan. Liu *et al.* (2017) considered the VAT registration notches and voluntary registration below the VAT registration threshold in the UK. They also describe low cost of inputs relative to sales, high proportion of business-to-customers sales, and high product market competition as determinants of higher bunching. Harju *et al.* (2016) look at the impact of the VAT registration threshold on the behavior of small Finnish firms. They find that the firms actively avoid VAT liability, and that such avoidance is directly caused by the compliance costs of VAT. Furthermore, they find that the bunching behavior persists over the long-term, implying that the VAT registration threshold permanently hinders the growth of small firms. In the context of a developing country, Boonzaaier *et al.* (2017) use tax register data from South Africa to look at the effects of several discontinuities in the tax schedule on the behavior of small firms. They find a moderate level of the bunching response at the VAT registration notch. Along similar lines as the Finnish study, the work on South Africa finds that the bunching firms are less likely to show strong growth dynamics and are more like to be 'stuck' in terms of their profits and sales. Gebresilasse & Sow (2016) look at the firm response to VAT registration threshold in Ethiopia, and estimate substantial bunching. Our study complements the work done on the extensive margin response. We show that significant bunching responses also occur on the intensive margin, at each of the relevant reporting thresholds. We provide evidence of firms actively modifying their reported turnover - either through underreporting, or underproduction - in order to reduce their hassle costs on the intensive margin.

Asatryan & Peichl (2017) look at the effects of both the VAT registration threshold and the so-called administrative thresholds, on the reported turnover by Armenian firms. Their analysis shows a moderate response to the VAT filing frequency threshold. However, this threshold also overlaps with other regulatory thresholds. Furthermore, they use audited tax returns to show that the response is mainly driven by evasion, rather than by underproduction. Finally, Asatryan & Peichl (2017) find no response to the VAT registration threshold notch. Our results are not confounded by other simultaneous regulatory thresholds. Additionally, our work provides a careful analysis of multiple reporting thresholds, at several levels of reported turnover, therefore allowing for the comparison of the revealed compliance costs for firms of different sizes.

We also add to the wide literature discussing VAT policy design theoretically and empirically. Keen & Mintz (2004) look at the optimal VAT registration threshold, and hypothesize that such a threshold will necessarily lead to production inefficiencies via firms bunching below the thresholds. Our analysis empirically confirms that bunching occurs in the case of VAT filing thresholds, analogous to the case of the VAT registration thresholds. Keen & Smith (2006) discuss the problematic aspects of VAT policy in the EU countries, particularly those that result in a greater propensity for tax evasion and fraud. They also look at the possibility of a federal VAT system in the US. Bird & Gendron (2007) discuss the application of VAT in lower income countries, and detail many implementation challenges encountered in such settings: inappropriate thresholds, delayed refunds, and insufficient audits. Mittal & Mahajan (2017) show that strengthened paper trail in the VAT context leads to an increase in tax collection, primarily driven by the behavior of the largest firms. We contribute to the literature on the VAT policy design by showing that reporting thresholds, a crucial VAT policy element in many countries, lead to potentially evasive behavior, particularly in the realm of lower income countries.

Finally, we contribute to the literature on the effects of size-dependent policies on firm behavior more broadly. This literature has shown that size-based regulations can lead to substantial distortions (Gollin, 1995; Guner *et al.* , 2008; Garicano *et al.* , 2016; Almunia & Lopez-Rodriguez, 2018).

3 Methodology

This section describes the methodology we use to derive the results briefly discussed in Section 1. We first describe the methodology used to derive the bunching estimates, and the methodology

to estimate the VAT revenue losses to the tax authority due to the bunching. We then describe the methodology we use to show that there are no apparent revenue benefits to the tax authority by the firms providing more frequent information in the form of the VAT reports. Lastly, we outline our methodology for a back of the envelope social welfare analysis of the VAT reporting thresholds. We begin by describing the policy variation and the data that allows us to do all of the above.

3.1 Delhi: Policy Change

For the five years in Delhi, India for which we have the data, the return filing policy had rich variation. Firms had to file returns at a specific frequency depending their on declared turnover in the previous financial year. For the first two years, Y1 and Y2, there were 3 thresholds of interest: Threshold 1 at ₹1 million (Indian rupees; ~$15,000), threshold 2 at ₹5 million (~$77,000), and threshold 3 at ₹50 million (~$770,000). For the third year, Y3, threshold 1 and threshold 2 were disbanded but the threshold 3 still existed. Y4 onwards, all firms had to file returns at a uniform quarterly frequency and there were no thresholds.

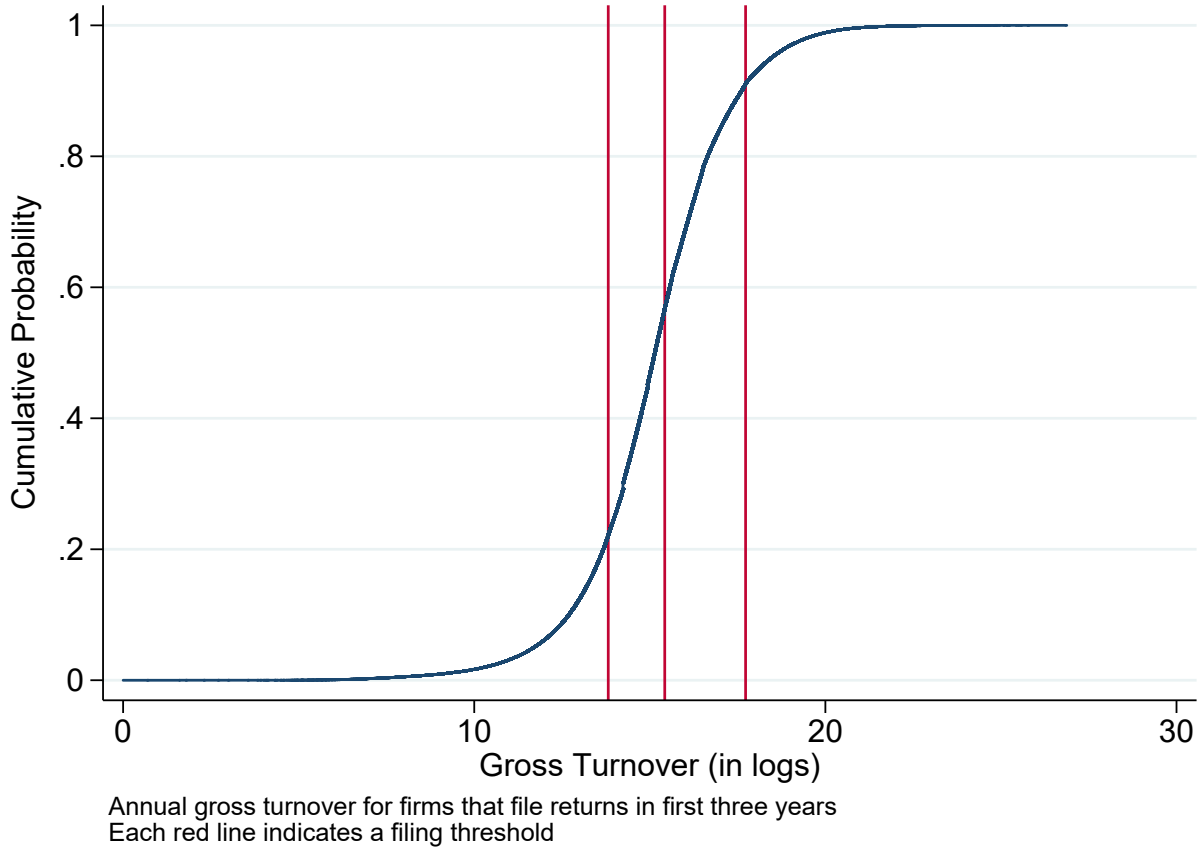
If a firm declared its turnover to be below threshold 1 in the previous year, then it had to file returns at the annual frequency in year 1 and 2. If a firm declared its turnover to be between threshold 1 and threshold 2, then it had to file returns once in every 6 months, twice a year. If a firm declared its turnover to be between threshold 2 and threshold 3, then it had to file returns at a quarterly frequency. Firms with declared turnover above threshold 3 had to file returns at a monthly frequency. Therefore, in years 1 and 2, the tax authority was receiving returns at annual, biannual, quarterly and monthly frequency. In year 3, the tax authority was receiving returns at quarterly and monthly frequency only. In year 4 and 5, the tax authority was receiving returns only at the quarterly frequency. Firms who were initially filing 12 returns every year, by virtue of declaring their turnover to be greater than ₹50 million the previous year, now had to file only 4 returns every year.

This change in policy allows us to pin-down that the bunching behavior that we observe around these thresholds is indeed happening due to the policy of interest and is not a confounded effect of some other policy.

3.2 Data

For our analysis, we use the form 16 data described in the Data section and the subsection on VAT Returns in Mittal & Mahajan (2017). Since all thresholds of interest are defined on annual levels, if

Figure 1: Distribution of Firms by Reported Turnover



Notes: The figure shows the (relatively smooth) cumulative distribution of the firms according to their turnover in our sample, combining the samples from the three years in which the thresholds exist (fiscal years 2010/11-2012/13). The x-axis shows the reported annual firm turnover (in logs). The 3 solid lines indicate the filing frequency thresholds at 1, 5, and 50 million. Therefore, roughly 20% of the firms benefit from the lowest threshold and 90% of the firms benefit from the highest threshold.

a firm is filing multiple returns in a given financial year, we aggregate the values at the financial year level. We have 5 years of tax returns from the state of Delhi, India from the fiscal year 2010-11 until the fiscal year 2014-15. Since the policy thresholds are defined in nominal terms, our analysis is also carried out in nominal values.

3.3 Bunching at Thresholds

Throughout our bunching analysis, we focus on the effects in the vicinity of each of the thresholds. We divide firms into bins of ₹30,000 around the low (annual firm turnover of ₹1 million ~ $15,000) and the middle thresholds (annual turnover of ₹5 million ~ $77,000). Similarly, we divide the firms into bins of ₹300,000 around the high threshold (annual firm turnover of ₹50 million ~ $770,000).

Around each threshold T, we have to determine the excluded area before we can estimate the "true" distribution polynomial. We visually observe the discontinuous increase in the number of firms below the threshold and use such visual observation to determine the starting point, R_1, of the discontinuous increase in the distribution of firms just before the threshold, T. Put differently, the area starting from R_1 and ending at T features an excess mass in the number of firms just before the threshold. Similarly, the upper excluded area, starting at the threshold T and ending at R_2 to the right of a given threshold, features a missing mass in the number of firms immediately after the threshold. We use the convergence method, as described by Kleven & Waseem (2013), to find the values of R_2 for each of the thresholds. Specifically, we choose the value of R_2 such that the area above the counterfactual distribution between R_1 and T, and the area below the counterfactual distribution between T and R_2 are approximately equal.

As a proxy estimate of the counter-factual distribution, we draw a fitted fourth-degree polynomial across all observations after excluding the lower and upper excluded area from R_1 to R_2. We run the following regression to estimate the smooth polynomial:

$$C_j = \sum_{i=1}^{4} \beta_i (B_j)^i + \varepsilon_j, \forall B_j \leq R_1 \& B_j \geq R_2, \tag{1}$$

where C_j denotes the count of firms in a given bin B_j, R_1 denotes the beginning of the lower excluded area range (before the threshold), and R_2 denotes the end of the upper excluded area range (after the threshold). Once we obtain the estimates from Equation (1), we then use the predicted counterfactual in the excluded range as well, and predict the counterfactual number of firms as follows:

$$\hat{C}_j = \sum_{i=1}^{4} \beta_i (B_j)^i \tag{2}$$

We then use the estimated counterfactual to calculate the bunching in the lower excluded area (to the left of the threshold), as follows:

$$b = \frac{\sum_{i \epsilon S} (C_i - \hat{C}_i)}{\hat{C}_{lowerexcluded}}, \tag{3}$$

where S denotes the set of values i, for which the bins B_i are in the lower excluded area, namely: $S = \{i \epsilon \mathbb{N} | B_i \epsilon [T - R_1, T]\}$. The bunching estimate is thus the estimate of the excess mass of firms before the threshold as a share of the counterfactual distribution of firms in the lower excluded

area. In particular the latter value, $\hat{C}_{lowerexcluded}$, is calculated as a weighted average of the counterfactual in the lower excluded area, weighted by the distance of the actual count of firms from the counterfactual distribution, as follows:

$$\hat{C}_{lowerexcluded} = \sum_{i \epsilon S} \mu_i \hat{C}_i, \tag{4}$$

where μ_i is the weight of each bin i, constructed as follows:

$$\mu_i = \frac{C_i - \hat{C}_i}{\sum_{i \epsilon S}(C_i - \hat{C}_i)} \tag{5}$$

We represent the counterfactual distribution of firms with a solid red line in all of the figures showing bunching, while we show the values of R_1 and R_2 using the vertical red dashed lines (refer to Figure 2, Figure 3, and Figure 4).

3.4 VAT Revenue Loss to the Tax Authority

To calculate the tax revenue implications of firm level bunching to avoid increased compliance costs, we need to assume that the firms in the lower excluded area (the excess mass of firms below the threshold), and the firms in the upper excluded area (the missing mass above the threshold) are directly comparable in terms of their unobservable characteristics - apart from their reported turnover and the VAT they remit. We then determine the revenue implication of the threshold regulation as follows. First, we calculate the difference in the average VAT remitted by firms in the upper excluded area and the VAT remitted by firms in the lower excluded area. We then multiply the difference in the average VAT remitted per firm with the extra bunching density, where $B = \sum_{i \epsilon S}(C_i - \hat{C}_i)$. The revenue implication, R, is thus calculated as follows:

$$R = (VAT^{mean}_{above} - VAT^{mean}_{below}) * B \tag{6}$$

3.5 No Benefits to More Information

One question related to the VAT reporting thresholds is why the tax authority might be interested in getting more frequent VAT reports from firms. We test whether the tax authority benefits from the firms providing more frequent information by receiving higher VAT collections when the firms provide more frequent information. We can not simply use VAT remitted as an outcome variable because the VAT remitted by firms above the threshold will be mechanically higher than the VAT

remitted by firms just below the threshold.

We proceed as follows. We group all data together, so that we have a pool of observations covering 2010-2015. We then normalize the VAT collected relative to the size of the firm in terms of the turnover, and run the following regression for firm i in year j:

$$(VAT/Turnover)_{i,j} = \alpha + \beta \cdot NumberReports_{i,j} + \phi_j + \phi_i + \epsilon_{i,j}, \tag{7}$$

where $VAT/Turnover$ is the ratio of a firm- and year-specific VAT remitted and the reported firm turnover, $NumberReports$ is the number of returns filed by the firm given its reported firm size in the given year, ϕ_i and ϕ_j are the firm and the year fixed effects, and ϵ are the heterogeneity robust standard errors.

We additionally test the responsiveness of the VAT collected as a share of turnover to the annual number of submitted reports by regressing the $VAT/turnover$ ratio on each of the reporting categories, while having the annual reporting as the omitted category. The regression specification for a firm i in year j is then:

$$(VAT/Turnover)_{i,j} = \alpha + \sum_{c=semiannual}^{monthly} \beta_c \cdot c + \phi_j + \phi_i + \epsilon_{i,j}, \tag{8}$$

where $VAT/Turnover$ is the ratio of a firm- and year-specific VAT remitted as a share of the reported firm turnover, c can take three values: $semiannual$ is a dummy variable equal to 1 if the firm filed the VAT reports at a semiannual level in a given year, $quarterly$ is a dummy variable equal to 1 if the firm filed the VAT reports at a quarterly level in a given year, $monthly$ is a dummy variable equal to 1 if the firm filed the VAT reports at a monthly level in a given year, ϕ_i and ϕ_j are the firm and the year fixed effects, and $\epsilon_{i,j}$ are the heterogeneity robust standard errors.

Using these regression specifications, we can then estimate the relationship between the number of returns filed annually and the amount of VAT remitted (relative to the firms' reported turnover).

3.6 Social Welfare Analysis

We conduct a simple back of the envelope social welfare analysis. In particular, we recognize that the overall welfare change in our context depends on the sum of the tax authority's revenue losses from the threshold policy, and on the reduction in compliance costs incurred by the firms due to the lower filing frequencies that these firms are now required to abide by. Therefore, we can

calculate the minimum implicit compliance subsidy needed to be given by the tax authority to the firms in a given reporting category in order to at least equalize the revenue losses stemming from the bunching behavior by the firms. If the actual compliance costs for the filing for those firms are higher than the calculated minimum implied subsidies, the overall welfare change stemming from the policy is likely to be positive once considering the implicit compliance subsidies to smaller firms, despite the revenue losses to the tax authority due to the bunching behavior.

We thus calculate the implicit subsidies at each threshold i in the following manner, using the values of revenue losses, R, calculated as explained in the subsection 3.4:

$$ImplicitSubs_i = R_i/n_j, \tag{9}$$

where n_j is the number of firms in the reporting category j below a given threshold i. This implies that for the low threshold, we divide the revenue losses stemming from the threshold by the number of firms reporting and paying VAT at an annual level. For the middle threshold, we divide the revenue losses stemming from the threshold by the number of firms reporting and paying VAT at a bi-annual level. Likewise, for the high threshold, we divide the revenue losses stemming from the threshold by the number of firms reporting and paying VAT at a quarterly level.

We then evaluate whether the implicit subsidies needed to equalize the revenue losses are lower than the likely compliance costs incurred by firms reporting and paying VAT at a certain level of frequency.

4 Results

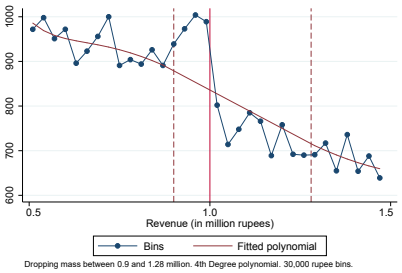
We first discuss the results of estimating the excess bunching at the relevant VAT filing thresholds. In Section 4.3, we discuss the calculation of the VAT revenue loss to the tax authority due to the existence of the arbitrary VAT filing thresholds. Section 4.4 shows that there are no apparent revenue benefits to the tax authority of more frequent information provision by the firms. Section 4.5 concludes by providing the welfare analysis of the size-dependent VAT filing policy.

4.1 Bunching at Thresholds

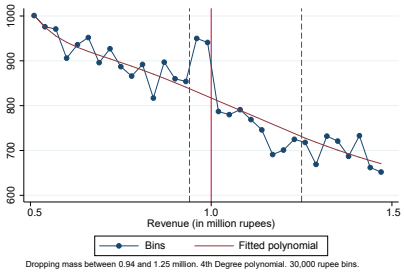
The low threshold, set at the turnover of ₹1 million, was relevant for the size-dependent VAT filing policy in years 1 and 2 (2010-11 and 2011-12) and mandated the change of return filing from an an-

Figure 2: Firm Turnover Distribution at the Low Threshold

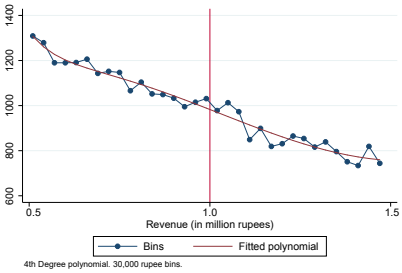
(a) Bunching in Year 1



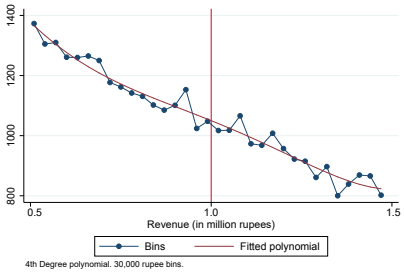
(b) Bunching in Year 2



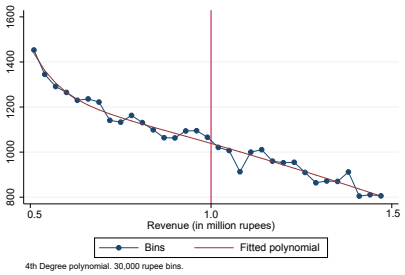
(c) No Bunching in Year 3



(d) No Bunching in Year 4



(e) No Bunching in Year 5



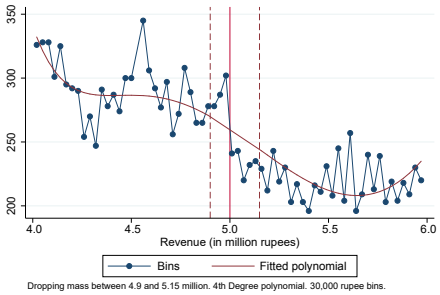
Notes: The figure shows the distribution of the firms around the low threshold (₹1 million), for each of the years in our data (year 1: fiscal year 2010/11 to year 5: fiscal year 2014/15). Panels a) and b) show the bunching behavior by the firms for the years (year 1 and 2) with differential, size-dependent requirements of VAT filing, while panels c), d) and e) document that there is no bunching behavior, with the distribution of firms being smooth around the threshold once the differential reporting requirement is done away with (years 3-5).

nual to a bi-annual frequency. Figures 1a and 1b show that such policy resulted in excess bunching at the threshold, with the bunching estimates of 0.5 for year 1, and 0.28 for year 2, respectively. In years 3, 4, and 5, the policy was discontinued, and all the firms around the threshold were required to file their VAT reports on a quarterly basis. As Figures 1c to 1e illustrate, there was - in contrast to years 1 and 2 - no excess bunching around the low threshold in the later years.

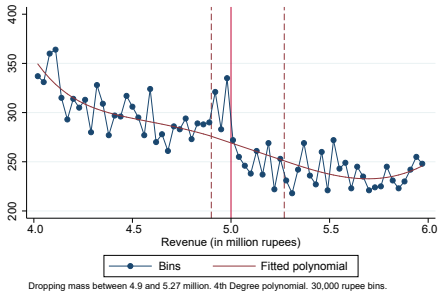
The middle threshold, set at the turnover of ₹5 million, mandated the change of return filing

Figure 3: Firm Turnover Distribution at the Middle Threshold

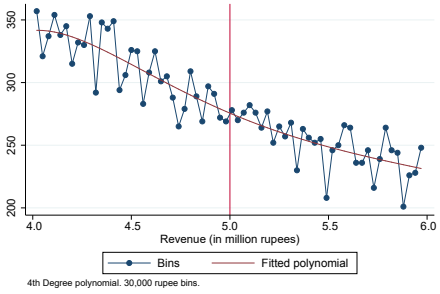
(a) Bunching in Year 1



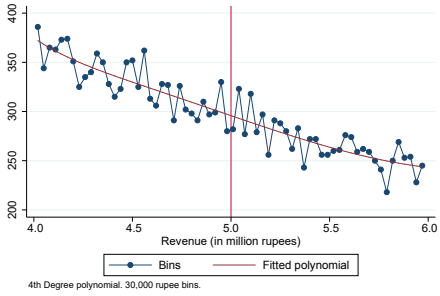
(b) Bunching in Year 2



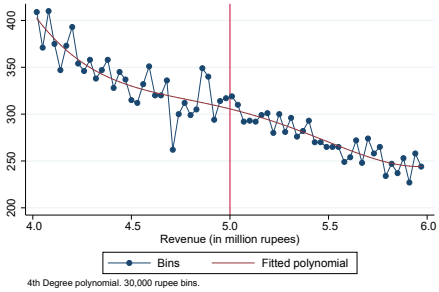
(c) No Bunching in Year 3



(d) No Bunching in Year 4



(e) No Bunching in Year 5

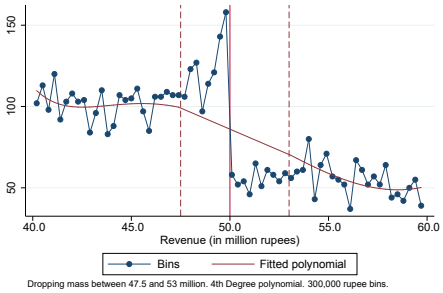


Notes: The figure shows the distribution of the firms around the middle threshold (₹5 million), for each of the years in our data (year 1: fiscal year 2010/11 to year 5: fiscal year 2014/15). Panels a) and b) show the bunching behavior by the firms for the years (year 1 and 2) with differential, size-dependent requirements of VAT filing, while panels c), d) and e) document that there is no bunching behavior, with the distribution of firms being smooth around the threshold once the differential reporting requirement is done away with (years 3-5).

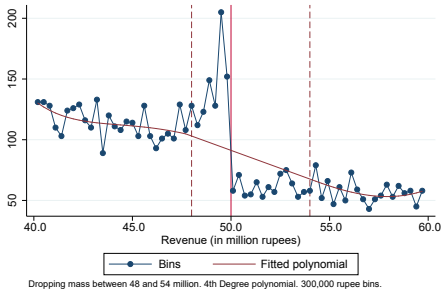
from a bi-annual to a quarterly frequency in years 1 and 2 (2010-11 and 2011-12). Instead of filing the VAT returns twice a year, firms with the reported turnover above ₹5 million in the previous year were required to file their reports four times a year. Figures 2a and 2b show that such a size-dependent policy resulted in significant bunching, with bunching estimates equaling 0.34 and 0.43, for years 1 and 2, respectively. Figures 2c to 2e further show that such bunching disappears, with

Figure 4: Firm Turnover Distribution at the High Threshold

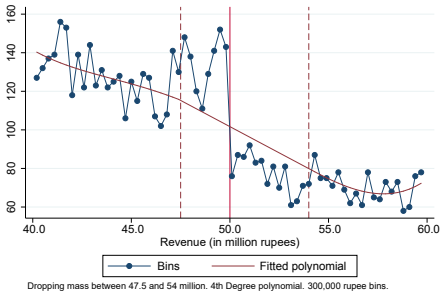
(a) Bunching in Year 1



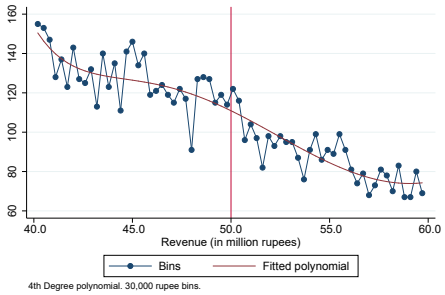
(b) Bunching in Year 2



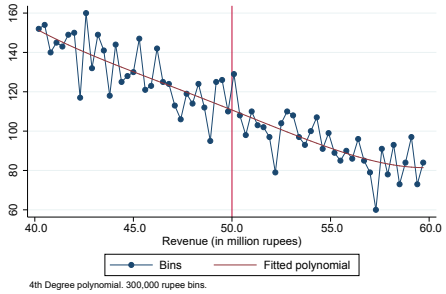
(c) Bunching in Year 3



(d) No Bunching in Year 4



(e) No Bunching in Year 5

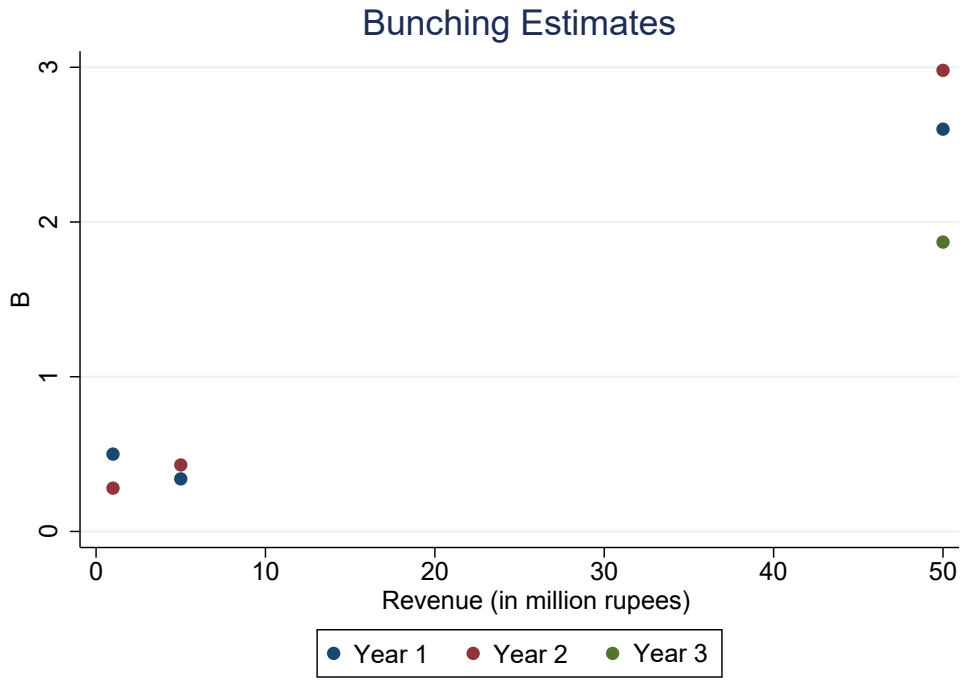


Notes: The figure shows the distribution of the firms around the high threshold (₹50 million), for each of the years in our data (year 1: fiscal year 2010/11 to year 5: fiscal year 2014/15). Panels A-C show the bunching behavior by the firms for the years (years 1, 2, and 3) with differential, size-dependent requirements of VAT filing, while panels D and E document that there is no bunching behavior, with the distribution of firms being smooth around the threshold once the differential reporting requirement is done away with (years 4 and 5).

a much smoother distribution of reported firm turnover, once the threshold policy was done away with in years 3, 4, and 5.

The high threshold, set at the turnover of ₹50 million, mandated the change of return filing from a quarterly frequency to a monthly frequency in years 1, 2, and 3 (2010-11 to 2012-13). In these years, the firms with the reported turnover greater than ₹50 million, in the previous year,

Figure 5: Distribution of Compliance Costs by Turnover Size



Notes: The figure shows the bunching estimates (on the y-axis) against the reported turnover at each of the thresholds, for which the bunching is estimated. The figure provides an indication of increasing compliance costs with the size of the reported turnover thresholds.

had to file VAT returns twelve times a year compared to firms just below the threshold, which had to file VAT returns four times a year. Figures 3a to 3c again indicate substantial bunching due to such filing threshold, with bunching estimates equaling 2.6, 2.98, and 1.87, in years 1, 2, and 3, respectively. Figures 3d and 3e show that after the threshold policy was done away with in years 4 and 5 - with all firms now filing the VAT reports on a quarterly basis - the distribution of reported turnover becomes much smoother, with no bunching at the relevant threshold of ₹50 million. This indicates that the observed bunching indeed occurs due to the filing policy.

Within each of the thresholds, we see that the bunching occurs at approximately the same magnitude. For the low threshold, we see a decrease of bunching in the second year; for the middle threshold, we see a slight increase in the second year bunching; for the high threshold, we see an increase in bunching from year 1 to 2, followed by a decrease in bunching in year 3. Assuming that the level of bunching is a proxy for actual compliance costs incurred by firms, Figure 5 uses these bunching estimates to track the compliance costs across different turnover sizes and plots the bunching estimates on the y-axis versus the turnover sizes for the relevant thresholds on the x-axis. While compliance costs seem to remain more or less at the same level across the low and

middle threshold, we see a sharp increase in the apparent compliance costs at the high threshold. This implies that compliance costs of a differential VAT reporting policy are generally increasing with the reported turnover. One potential factor driving such increasing gradient of compliance costs may be the tax authority's scrutiny, which is increasing with firm size. At the same time, it is important to note that at the high threshold, which exhibits the increase in the bunching behavior, the increase in the filing for the firms at the threshold is 300% (from 4 to 12 VAT filings per year), while at the low and middle thresholds the increase in the VAT filing is 100% (from 1 to 2 VAT filings per year, and 2 to 4 VAT filings per year, respectively). This point is line with the scrutiny gradient, which is increasing with the firms' reported turnover, and thus results in the level of compliance costs, which are in parallel increasing with firm size.

4.2 Bunching Channels

We note two potential channels through which the firms may bunch: underproduction and underreporting. The first channel, underproduction, would imply that the bunching firms halt their production or sales as they get close to the threshold. This channel brings about large welfare losses because of the current reduction in the firm-level profits and due to the shift in the long-run growth trajectory of the bunching firms. If this is the dominant channel, the welfare losses due to the policy (while not considering the implicit subsidies due to the reduced filing frequencies for smaller firms) may be significant, with real and longer-term consequences in the form of stalled firm growth. The second bunching channel encompasses the intentional turnover underreporting by the firms. If such underreporting occurs and is substantial, the welfare losses would occur via the lost tax revenues. One potential sub-channel in the firm turnover underreporting is turnover shifting: in such a case, a portion of the bunching firms illegitimately register a part of their sales for the following filing period in order to avoid passing the relevant reporting threshold. In such a case, the welfare consequences of the policy would be the smallest of the three bunching channel options; the firms would only misreport the current turnover to avoid more frequent filing, but would not affect their actual long term real growth, and would eventually report all of their turnover.

While the available data does not allow us to directly observe the firms' real (rather than reported) production decisions, we use the available information to deduct which of the channels through which the firms may bunch seem to be relatively more important.

Appendix Figure A.I shows the bunching behavior of the firms at the high threshold (₹50 Mil-

lion) by firm type: namely, we classify the firms (based on their self-reported information to the tax authority) as either manufacturers, wholesalers, or retailers. Panels a) to c) show the bunching of the three types of firms in year 1, panels d) to f) focus on the bunching behavior in year 2, while panels g) to i) show the bunching in year 3. Using simple visual inspection, we conclude that there are no significant differences in the firms' bunching behavior both within a sector across years, as well as - more importantly - across different sectors within the same year. We argue that such an observation provides first evidence indicating the most relevant bunching channel: firms in different sectors almost surely face different costs related to production halting. In particular, the nature of the retail sector may mean that it is least costly to "production" halt in it: simply closing "the shop" is enough to halt the "production". This is likely a more difficult task for wholesalers: they are involved in important (and relatively complex) business-to-business relationships with their clients, so that halting their production just in order to avoid reporting more frequently may be more costly for them. Such (implicit or direct) costs of halting production are almost surely the highest in the manufacturing sector: due to a multifaceted nature of the production process, simply halting the production may often not be possible, not the least because of the complex labor relationship and the long-term contracts with the firms' business clients. Given that there are no obvious differences in the bunching behavior across the three sectors - with likely different costs of halting the production - we recognize the production-halting channel as the one which likely plays only a minor, if any, role in the bunching behavior of the firms.

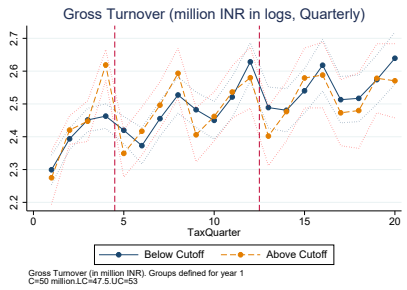
Figure A.II looks at the annual growth rates for the firms: at the low threshold in panels a) and b), at the middle threshold in panels c) and d), and at the high threshold in panels e) to g). We do this to explore whether the firms bunching in a particular year may see an unusually high growth of turnover relative to the comparable non-bunching counterparts. If such sudden growth hikes were to be observed in the post-bunching years, this would provide an indication that the bunching firms artificially reduced their income through underreporting in the bunching year but grew much faster in the post-bunching year as their real activity actually grew even in the bunching year. As all of the panels in the Figure A.II show, such growth hikes in the immediate post-bunching years do not seem to be an actual empirical phenomenon. Instead, we can actually observe no statistically significant differences in the annual growth rates of the bunchers relative to the growth rates of the comparable non-bunching firms. In fact, most of the panels show almost identical annual growth rates of the bunchers and comparable non-bunchers. As such, given the very similar growth rates of the bunchers and the comparable non-bunchers, straightforward underreporting

in the bunching years is unlikely to be a major or the only channel through which the firms bunch.

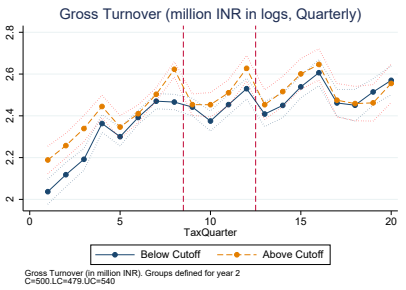
In order to evaluate the possibility of turnover shifting, we look at the highest threshold firms - namely those with annual turnover around ₹50 million. We plot the quarterly gross turnover of the bunchers - those firms right below the threshold, which actually report their turnover on a quarterly basis - with the firms above the threshold cutoff, which report their turnover to the tax authority at a monthly basis. Figure 6 shows the result of such analysis in three panels: each panel identifies the behavior of the firms right below the cutoff (in the lower excluded area in the bunching analysis; the "bunchers") to the firms right above the cutoff (in the upper excluded area in the bunching analysis; the comparable "non-bunchers") for each of the "bunching" years. Panel a) shows that the first year's bunchers report significantly less turnover than the non-bunchers in the fourth quarter - the last quarter relevant for their bunching behavior. Panels b) and c) identify the same pattern for quarters 8 and 12, respectively. Interestingly, in all of the panels, the turnover growth of the bunchers is higher relative to the non-bunchers in the subsequent quarters: quarter 5 in panel a), quarter 9 in panel b), and quarter 13 in panel c). This can be observed by the less negative slope of the blue line (representing the bunchers) relative to the more negative slope of the orange line (representing the non-bunchers).

Figure 6: Log Quarterly Turnover at the High Threshold

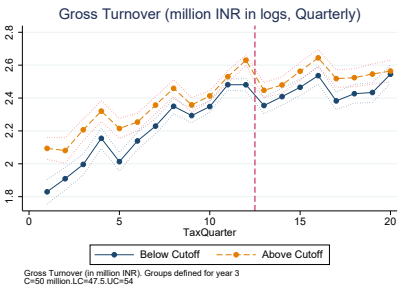
(a) First Year's Bunchers



(b) Second Year's Bunchers



(c) Third Year's Bunchers



Notes: The figure shows quarterly turnover (in logs) for the firms right below and right above the high threshold cutoff in year 1 in panel a), year 2 in panel b), and year 3 in panel c). The blue line represents the firms right below the high threshold cutoff in each of the years: these are the so-called "bunchers", located in the lower excluded area in the bunching estimations. The orange line represents the firms right above the high threshold cutoff in each of the years: these are the so-called comparable "non-bunchers", located in the upper excluded area in the bunching estimations. The red vertical lines represent the quarter of interest and the time when the policy ends, that is: quarters 4 and 12 (in panel a), quarters 8 and 12 (in panel b), and quarter 12 (in panel c), respectively.

Such behavior of the firms is clearly in-line with the hypothesis of initial underreporting and turnover shifting by the bunching firms: Figure 6 shows that the bunching firms underreport their turnover in the last relevant budgetary quarter (compared to the non-bunching firms in the upper

excluded area), but then report more turnover than the non-bunchers in the first quarter of the subsequent financial year. This provides an indication that the bunchers are shifting their turnover in order to avoid more frequent VAT reporting. Notice, however, that the difference between the orange and the blue line right below the end of the budgetary year is larger than the difference between the blue and the orange line at the start of the next budgetary year. This in turn means that while turnover shifting apparently is happening, it cannot account for all of the bunching behavior: it is likely that some of the turnover remain completely unreported. In future work, we will provide the calculations based on Figure 6, which will give us the approximate contribution share of turnover shifting in the bunchers' total underreported revenue amounts.

4.3 VAT Revenue Loss to the Tax Authority

We now compare the VAT contributions of firms in the bunching region below and the region above the threshold. Table 1 shows that the revenue losses are substantial. For the low threshold, the losses amount to nearly ₹900,000 in year 1, and to nearly ₹500,000 in year 2. These amounts are, respectively, 2.3% and 1.8% of the bunchers' (all firms just below the threshold) total VAT contributions, therefore presenting a substantial loss of VAT revenues to the tax authority. For the middle threshold, the losses are generally less substantial, amounting to above ₹700,000 in year 1, and above ₹100,000 in year 2. These amounts respectively equal 1.8% and 0.03% of the bunchers' total VAT contributions in the respective years. The losses are most substantial for the high threshold, amounting to nearly ₹40,000,000 in year 1, and above ₹20,000,000 in years 2 and 3. These amounts are also the highest as a share of the bunchers' total VAT contributions, amounting to 8.3%, 5.2%, and 4.2% in years 1, 2, and 3, respectively. These estimates confirm the findings in Section 4.1 that the bunching is not only the highest, but also the costliest for the highest threshold.

4.4 No Benefits to More Information

This section discusses the regressions looking at the relationship between the VAT-to-turnover ratio and the frequency of VAT reporting, as discussed in the Section 3.5. Table 2 shows the regression results looking at the relationship between the VAT-to-turnover ratios and the yearly number of VAT reports.

Columns 1 and 2 show that the relationship is strikingly negative without including firm fixed effects in the regression specification, implying that the greater the number of VAT reports per year,

Table 1: VAT Revenue Lost and Percentage of the Bunchers' VAT Contribution Lost

	Threshold level		
	Low	Middle	High
Year	(₹1 million)	(₹5 million)	(₹50 million)
2010-11	886,017	730,557	38,373,960
	(2.3%)	(1.8%)	(8.3%)
2011-12	441,600	128,150	21,862,749
	(1.8%)	(0.03%)	(5.2%)
2013-14			22,871,560
			(4.2%)

Notes: Values in Indian rupees. The amounts expressed as a percentage of the bunchers' (i.e. that of firms in the lower excluded area) VAT contributions are in the brackets. The values are calculated as described in the Section 3.4.

the smaller the VAT-to-turnover ratio. When including firm fixed effects in the regression specification, as shown by columns 3 and 4, the relationship turns positive but strongly insignificant.

Table 3 similarly shows the regression results looking at the relationship between the VAT-to-turnover ratios and different reporting categories, with annual reporting being the omitted category. As columns 1 and 2 show, when not including firm fixed effects, all reporting categories are negatively related to the VAT-to-turnover ratios compared to those firm-year observations with annual reporting. Once including firm fixed effects in the regression, the estimates turn positive, but again remain strongly insignificant.

Our interpretation of the results of both of the sets of regressions is that greater than annual frequency of VAT reporting does not lead to more VAT being collected, as measured by the VAT-to-turnover ratios.

4.5 Social Welfare Analysis

In this section we discuss the results of the back of the envelope social welfare analysis that we conducted according to the methodology outlined in Section 3.6. In Table 4 we present the calculated implicit subsidies. The implicit subsidies needed to equalize welfare are very low. At the low threshold for year 1 - an implicit welfare equalizing per firm subsidy of ₹40.77 (less than $1) implies that the welfare change of reducing the frequency would be negative only in the case that compliance costs borne by a given firm reporting and paying VAT at an annual level would be less than $1, which is extremely unlikely. The implicit welfare equalizing per firm subsidy for year

Table 2: VAT Revenue vs. Annual Number of VAT Returns Regressions

VARIABLES	(1) VAT/Revenue	(2) VAT/Revenue	(3) VAT/Revenue	(4) VAT/Revenue
NumberReports	-0.000148***	-0.000314***	0.000474	0.000247
	(3.18e-05)	(0.000100)	(0.000468)	(0.000278)
Firm FE	NO	NO	YES	YES
Time FE	NO	YES	NO	YES
Observations	1,038,331	1,038,331	1,038,331	1,038,331
No. of firms	301,147	301,147	301,147	301,147
R^2	0.000	0.000	0.703	0.703

Notes: Robust standard errors, clustered at the firm level, in parentheses. The regression table presents the results of regressions performed according to the methodology described in the Section 3.5. *** p<0.01, ** p<0.05, * p<0.1.

2, at the low threshold is even smaller. Similarly, the welfare equalizing subsidies for the middle threshold are comparably low.

The implicit welfare equalizing subsidies for the high threshold are an order of magnitude higher, ranging from ₹1327.97 (about $25) in year 3 to ₹2990.02 (about $60) in year 1. The implication of these implicit welfare equalizing per firm subsidies is, however, similar as for the low and the middle thresholds. The actual compliance - that is, filing and paying - costs difference (quarterly level of reporting relative to monthly reporting) associated with VAT of these firms, reporting at a quarterly level is certainly higher than $60. That means that the tax authority subsidizes these firms by reducing the frequency at which they have to file, and yet increases the social welfare. Put another way, this analysis leads us to conclude that even though the tax authority incurs significant losses due to thresholds, its implicit subsidies to small and medium-sized firms are large enough to overcome the revenue losses.

5 Conclusion

In this paper, we identify bunching behavior by firms around the VAT filing thresholds, using an administrative-level dataset from the state of Delhi in India. Our unique dataset - along with rich policy variation - enables us to show that the bunching disappears when the thresholds are done away with. Bunching is the greatest for the high threshold indicating that compliance is most costly for those firms. Using simple graphical analysis, we provide indications that the bunching

Table 3: VAT Revenue vs. VAT Filing Categories Regressions

VARIABLES	(1) VAT/Revenue	(2) VAT/Revenue	(3) VAT/Revenue	(4) VAT/Revenue
SemiAnnualCategory	-0.00685***	-0.00685***	0.00521	0.00516
	(0.000206)	(0.000206)	(0.00782)	(0.00778)
QuarterlyCategory	-0.00312*	-0.00677***	0.00696	0.00606
	(0.00162)	(0.000313)	(0.00823)	(0.00819)
MonthlyCategory	-0.00541***	-0.00713***	0.00670	0.00546
	(0.000286)	(0.000999)	(0.00819)	(0.00717)
Firm FE	NO	NO	YES	YES
Time FE	NO	YES	NO	YES
Observations	1,038,331	1,038,331	1,038,331	1,038,331
No. of firms	301,147	301,147	301,147	301,147
R^2	0.000	0.000	0.703	0.703

Notes: Robust standard errors, clustered at the firm level, in parentheses. Annual reporting is the omitted category. The regression table presents the results of regressions performed according to the methodology described in the Section 3.5. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Implicit Welfare Equalizing Per Firm Subsidies from the Thresholds

	Threshold level		
Year	Low threshold (₹1 million)	Middle threshold (₹5 million)	High threshold (₹50 million)
2010-11	40.77	14.54	2990.02
2011-12	17.31	2.55	1505.49
2013-14			1327.97

Notes: Values in Indian rupees. The implicit minimum welfare equalizing per firm subsidies are computed according to the methodology outlined in the Section 3.6.

behavior likely stems from turnover-shifting and simple underreporting.

We further show that the VAT revenue losses due to the bunching response by firms are substantial, up to 8.3% of the bunchers' total VAT contributions in certain years. We also suggest that there may be no benefits to the tax authority from receiving more frequent information through the VAT filing reports. Finally, our social welfare analysis shows that if the costs of compliance by the firms - and thus the implicit subsidies to those firms by requiring less frequent reporting - are relatively substantial, the welfare impact of the thresholds is positive despite the considerable revenue losses to the tax authority. This is due to relatively large implicit subsidies given to smaller firms as a results of the size-dependent VAT filing policy.

One crucial weakness of our work is our inability to observe the firms' actual production decisions. We only observe the firms' *reported* production (turnover), but not some alternative input- or output-based indicators of the firms' production. This limitation prevents us from examining the channel through which the firms bunch in further detail. Another current limitation of the paper is in the lack of the consideration of the serial correlation of the firms' turnover and thus of their bunching behavior. This may affect the bunching estimatations: a dynamic bunching design a la (Marx, 2019) can, for instance, exploit the panel data structure to better identify and estimate the relevant parameters. In future work, we intend to use such methodology for the bunching estimations.

All in all, our analysis shows that firms' costs of compliance to the VAT policy are quite large, with unclear benefits of more frequent information and payments being made by the VAT registered firms. Tax authorities should thus aim to reduce the compliance costs, which play a substantial role in the behavior of the firms.

More generally, we show that size-dependent VAT reporting can lead to substantial distortions in firm reporting and to potential revenue losses for the tax authority. Given that size-dependent VAT reporting is used in a large number of countries, this finding has potential implications reaching far beyond the Indian context. The overall welfare implications of such size-dependent VAT policy depend on the importance that a welfare-maximizing government gives to the small- and medium-sized enterprises (SMEs). If the social planner deems the SMEs to be important, the revenue losses due to the size-dependent VAT reporting frequency are more than offset by the implicit compliance subsidies given to the SMEs as a result of the size-dependent reporting policy. If the social planner deems the SMEs as relatively unimportant then the size-differentiated policy will not be optimal, and a uniform VAT reporting frequency should be preferred.

Bibliography

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.

ASATRYAN, ZAREH, & PEICHL, ANDREAS. 2017. Responses of Firms to Tax, Administrative and Accounting Rules: Evidence from Armenia. *Working paper*.

BACHAS, PIERRE, & JENSEN, ANDERS. 2017. Size-Dependent Tax Enforcement and Compliance: Global Evidence. *Working paper*.

BENZARTI, YOUSSEF. 2017. How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs. *Working paper*.

BIRD, RICHARD, & GENDRON, PIERRE-PASCAL. 2007. *The VAT in Developing and Transitional Countries*. Cambridge University Press.

BOONZAAIER, WIAN, HARJU, JARKKO, MATIKKA, TUOMAS, & PIRTTILA, JUKKA. 2017. How Do Small Firms Respond to Tax Schedule Discontinuities? Evidence from South African tax registers. *Working paper*.

DHARMAPALA, DHAMMIKA, SLEMROD, JOEL, & WILSON, JOHN DOUGLAS. 2011. Tax Policy and the Missing Middle: Optimal Tax Remittance with Firm-Level Administrative Costs. *Journal of Public Economics*, **95**(9-10), 1036–1047.

ERNST & YOUNG. 2015. Worldwide VAT, GST and Sales Tax Guide 2015.

GARICANO, LUIS, LELARGE, CLAIRE, & VAN REENEN, JOHN. 2016. Firm Size Distortions and the Productivity Distribution: Evidence from France. *American Economic Review*, **106**(11), 3439–79.

GEBRESILASSE, MESAY M, & SOW, SOULE. 2016. *Firm Response to VAT Registration Threshold in Ethiopia*. Tech. rept.

GOLLIN, DOUGLAS. 1995. *Do Taxes on Large Firms Impede Growth? Evidence from Ghana*. Economic Development Center, Department of Economic, Minneapolis Department of Agricultural and Applied Economics, St. Paul, University of Minnesota.

GUNER, NEZIH, VENTURA, GUSTAVO, & XU, YI. 2008. Macroeconomic Implications of Size-Dependent Policies. *Review of Economic Dynamics*, **11**(4), 721–744.

HARJU, JARKKO, MATIKKA, TUOMAS, & RAUHANEN, TIMO. 2016. The Effects of Size-Based Regulation on Small Firms: Evidence from VAT Threshold. *Working paper*.

INTERNATIONAL TAX DIALOGUE. 2007. Taxation of Small and Medium Enterprises. *Background paper for the International Tax Dialogue Conference*.

INTERNATIONAL TAX DIALOGUE. 2013. Key Issues and Debates in VAT, SME Taxation and the Tax Treatment of the Financial Sector. *Background paper for the International Tax Dialogue Conference*.

KEEN, MICHAEL, & MINTZ, JACK. 2004. The Optimal Threshold for a Value-Added Tax. *Journal of Public Economics*, **88**(3-4), 559–576.

KEEN, MICHAEL, & SMITH, STEPHEN. 2006. VAT Fraud and Evasion: What Do We Know and What Can Be done? *National Tax Journal*, 861–887.

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.

LIU, LI, LOCKWOOD, BEN, & ALMUNIA, MIGUEL. 2017. VAT Notches, Voluntary Registration and Bunching: Theory and UK Evidence. *Working paper*.

MARX, BENJAMIN. 2019. Dynamic Bunching Estimation with Panel Data. *Working paper*.

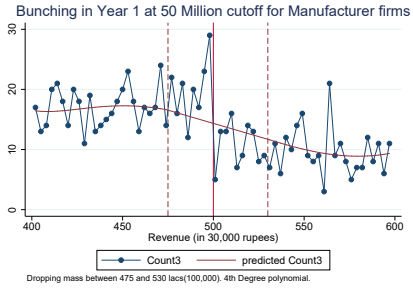
MITTAL, SHEKHAR, & MAHAJAN, APRAJIT. 2017. VAT in Emerging Economies: Does Third Party Verification Matter? *Working paper*.

ONJI, KAZUKI. 2009. The Response of Firms to Eligibility Thresholds: Evidence from the Japanese Value-Added Tax. *Journal of Public Economics*, **93**(5-6), 766–775.

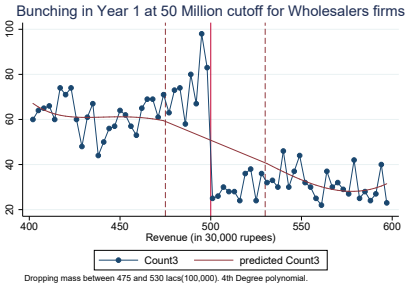
Appendix

Figure A.I: Bunching at the High Threshold by Firm Type

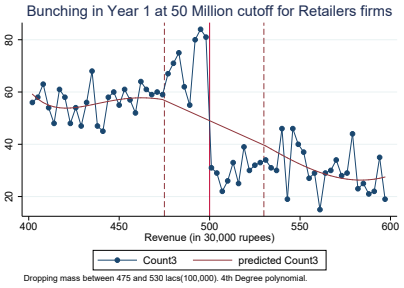
(a) Year 1, Manufacturers



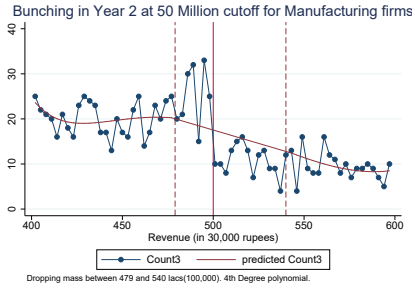
(b) Year 1, Wholesalers



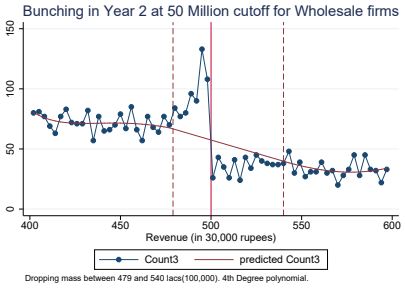
(c) Year 1, Retailers



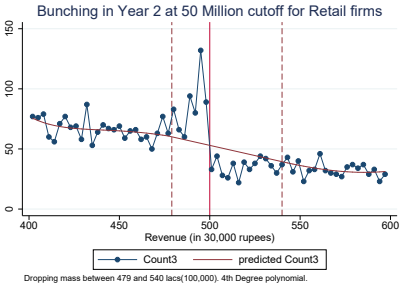
(d) Year 2, Manufacturers



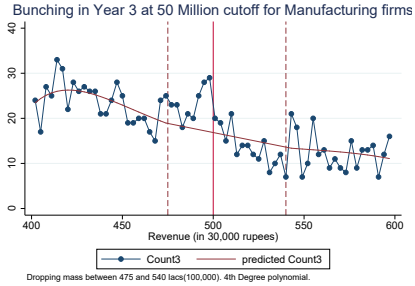
(e) Year 2, Wholesalers



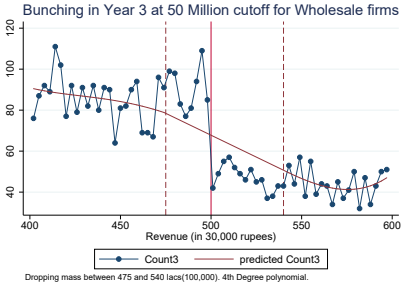
(f) Year 2, Retailers



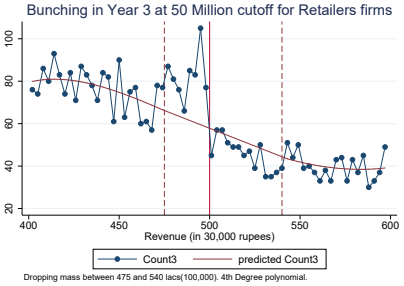
(g) Year 3, Manufacturers



(h) Year 3, Wholesalers



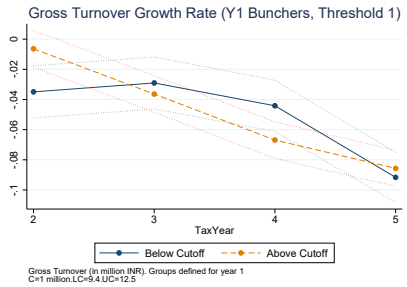
(i) Year 3, Retailers



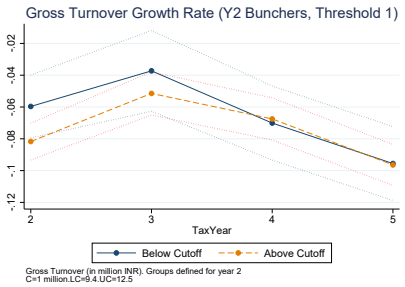
Notes: The figure shows the bunching figures at the high threshold (₹50 million) by the self-reported nature of the firms' business: for manufacturers, whosalers, and retailers, for each of the relevant years (years 1-3). The y-axis shows the frequency of the firms, while the x-axis shows the reported firm turnover in ₹30,000 bins.

Figure A.II: Annual Turnover Growth Rates for Bunchers and Non-Bunchers

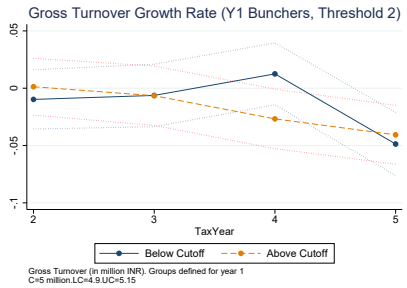
(a) Threshold 1, First Year's Bunchers



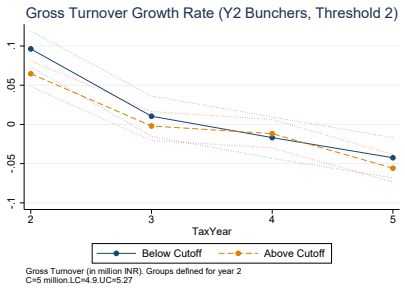
(b) Threshold 1, Second Year's Bunchers



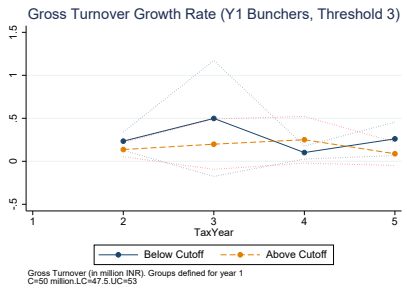
(c) Threshold 2, First Year's Bunchers



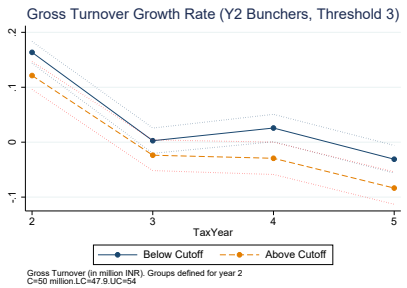
(d) Threshold 2, Second Year's Bunchers



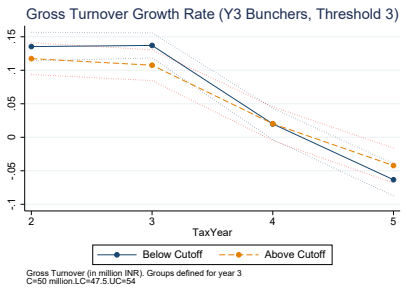
(e) Threshold 3, First Year's Bunchers



(f) Threshold 3, Second Year's Bunchers



(g) Threshold 3, Third Year's Bunchers



Notes: The figure shows annual turnover growth rates for the firms around the low threshold for years 1 and 2 (panels a) and b); ₹1 million), the middle threshold for years 1 and 2 (panels c) and d); ₹5 million), and the high threshold for years 1, 2, and 3 (panels e) f), and g); ₹50 million). The blue line represents the firms right below the high threshold cutoff in each of the years: these are the so-called "bunchers", located in the lower excluded area in the bunching estimations. The orange line represents the firms right above the high threshold cutoff in each of the years: these are the so-called comparable "non-bunchers", located in the upper excluded area in the bunching estimations.