VAT Notches, Voluntary Registration, and Bunching: Theory and UK Evidence *

Li Liu and Ben Lockwood†

This version: July 2016

Abstract

We develop a conceptual framework that allows simultaneously for (i) voluntary registration for VAT by firms below the registration threshold; and (ii) bunching at the registration threshold. This framework also generates predictions about how voluntary registration and bunching are related to intensity of input use, the share of B2C transactions for a firm, opportunities for evasion via under-reporting of sales, and the competitiveness of the market in which the firm is located. We bring the theory to the data, using linked administrative VAT and corporation tax records in the UK from 2004-2009. Consistently with the theory, we find that voluntary registration is positively related to the intensity of input use and negatively to the share of B2C transactions, and the amount of bunching is related to these variables in the opposite way. There is some evidence that product market competition leads to more voluntary registration, and less bunching. In addition, we find some suggestive evidence that firms are bunching by under-reporting sales.

*This paper was previously circulated under the shorter title, "VAT Notches". We thank the staff at Her Majesty's Revenue & Customs' (HMRC) Datalab for access to the data and their support of this project. This work contains statistical data from HMRC which is Crown Copyright. The research datasets used may not exactly reproduce HMRC aggregates. The use of HMRC statistical data in this work does not imply the endorsement of HMRC in relation to the interpretation or analysis of the information. All results have been screened by HMRC to ensure confidentiality is not breached.

†Liu: Centre for Business Taxation, University of Oxford (li.liu@sbs.ox.ac.uk). Lockwood: CBT, CEPR and Department of Economics, University of Warwick (b.lockwood@warwick.ac.uk). We would like to thank Steve Bond, Michael P. Devereux, Judith Freedman, Chris Heady, James R. Hines, Louis Kaplow, Henrik Kleven, Tuomas Matikka, Joel Slemrod, and participants at the 2014 International Institute of Public Finance Meeting, 2014 CEPR Public Economics Conference, Oxford-Michigan Tax Systems Conference, 2nd MaTax Conference in Mannheim, HM Treasury, Oxford University and University of Exeter for helpful comments. We would also like to thank Dongxian Guo and Omiros Kouvavas for excellent research assistance. The usual disclaimer applies. We acknowledge financial support from the ESRC under grant ES/L000016/1.

1 Introduction

Most countries around the world use the value-added tax (VAT) as their primary indirect tax, and most countries that adopted the VAT have thresholds, usually based on turnover, below which businesses do not need to register for VAT.[1] As VAT rates are often quite high (in excess of 20% in many EU countries), this creates a large and salient tax notch for small businesses whose turnover is around the threshold.[2] So far, the effect of these VAT notches on firm behavior has not received much attention.

In this paper, we make several contributions on this front, both theoretically, and empirically, using administrative data on UK corporations. First, what are the stylized facts to be explained? One striking feature of our data is that over the time period 2004-2009, approximately 44% of firms with turnover below the threshold were registered for VAT. It seems unlikely that this is entirely due to inertia - and indeed we present evidence below to show that this is not the case - and so we conclude that some firms actively choose to register, even if they are below the threshold. This is a striking phenomenon which deserves further study; to our knowledge, voluntary payment of any tax is very rare.

Also, in our UK administrative data, there is strong evidence that some firms are restricting their turnover to stay just below the registration threshold, i.e. they bunch. Apart from intrinsic interest, these features are also important because they impact on production inefficiencies induced by the VAT; when firms do not register, they face so-called embedded VAT on inputs, which distorts input choices, and also cascades through the production chain (Ebrill, Keen and Perry, 2001).

Our first contribution is to develop a simple general equilibrium model that can explain both these phenomena in a unified way. Our first observation is that the coexistence of both voluntary registration and bunching requires that firms make *both* sales to final consumers (B2C sales) and sales to other VAT-registered businesses (B2B sales). To see this, suppose first that firms make only B2C sales. Then, it is easily shown that irrespective of the degree of competition between firms, the cost of voluntarily registering exceeds the benefit, because the burden of VAT paid on output when registered exceeds the burden of VAT paid on inputs when not registered. Conversely, with only B2B sales, the voluntary registration is always optimal, because the burden of output VAT can be passed on to the buyer.[3]

Our second assumption is that firms have some market power, for reasons explained

[1] In the EU, all but two countries (Spain and Sweden) currently have positive thresholds, with the UK threshold being the largest at £81,000. The thresholds in the EU are generally low compared with those in countries that have more recently introduced a VAT, such as Singapore, which currently has a threshold of about 540,000 Euro (http://www.vatlive.com).

[2] A notch arises when the tax liability changes discontinuously.

[3] Both these points are made formally below.

in detail below. Finally, to study the effect of input costs on voluntary registration and bunching, we must of course also assume that small firms use intermediate inputs as well as labour in varying proportions. So, we construct the simplest general equilibrium model that has the required features. The model can also be extended to allow for VAT evasion via concealment of some fraction of B2C sales.

In this setting, we show the following. First, under some assumptions, the effect of the VAT system on the registration decision can be captured by a *VAT sufficient statistic*, which combines the effects of both input and output VAT. We then show that voluntary registration by a firm is more likely when either (i) the cost of inputs relative to sales is high, or (ii) when the proportion of B2C sales by the firm is low.[4] The intuition for (ii) is simply that if most customers are VAT-registered, the burden of an increase in VAT can easily be passed on in the form of a higher price, because the customer itself can claim back the increase. The intuition for (i) is that when input costs are important, registration allows the firm to claim back a considerable amount of input VAT.

Second, we show that the determinants of bunching at the registration threshold are the same as for voluntary registration, with the signs of the effects reversed. Specifically, bunching is more likely when (i) the cost of inputs relative to sales is low, or (ii) when the proportion of B2C sales is high. Third, we study the effect of product market competition on voluntary registration and bunching; the effects of increased competition are generally ambiguous, but we identify the features that make it positive or negative.

Fourth, with evasion, the qualitative results obtained so far are unchanged. We show that opportunities for evasion will increase voluntary registration and have an ambiguous effect on bunching. The latter result is somewhat surprising, as it is usually assumed in the empirical literature that bunching is facilitated by evasion opportunities. The explanation here is that while concealment of sales does indeed make it easier for firms to stay below the VAT threshold, at the same time, voluntary registration becomes more attractive, as output VAT is less of a burden.

We then bring these predictions to an administrative data-set created by linking the population of corporation and VAT tax records in the UK from 2004-2009. We first show that the pattern of voluntary registration in the data is consistent with the theory. In particular, voluntary registration is more likely with a low share of B2C sales or a high share

[4]Both these characteristics clearly differ widely across small firms that are close to the registration threshold. For example, a small tradesperson such as a plumber or electrician may typically have mostly B2C sales of his services to householders, and make relatively light use of intermediate inputs. So, they would face a low effective VAT rate when not registered, but a high rate when registered. Conversely, a small specialist engineering firm, such as a car component firm, may make mostly B2B sales with heavy use of intermediate inputs, and so will be in the reverse position.

of input costs. Quantitatively, the probability that a firm voluntarily registers for VAT is increased by 5 percentage points for a one standard deviation increase in the share of B2C sales and by 2-5 percentage points for a one standard deviation increase in the input cost ratio. The results are robust to use of either a linear probability model or fixed-effects logit model, and to the inclusion of additional firm-level control variables.

We then look at bunching. In the aggregate, there is clear evidence of bunching at the VAT threshold. Investigating further, we find that firms are more likely to bunch at the threshold when either (i) the cost of inputs relative to sales is high, or (ii) when the proportion of B2C sales is low, consistently with the theory. So, there is a clear pattern of heterogeneity in bunching.

We also investigate the mechanisms at work in bunching. We provide some suggestive evidence that part of bunching is driven by evasion, in the form of under-reporting of sales. Specifically, we find that the average input cost ratio moves in parallel between the registered and non-registered firms outside the bunching region but starts to increase substantially for the non-registered companies just below the threshold.

Finally, in the last section, we study the dynamic movement of firms in and out of registration status. Specifically, we address the concern that voluntary registration may not be an optimizing choice of firms, but simply a failure to deregister, once having been above the threshold and having registered. Our empirical findings suggest that while there is a considerable amount of persistence in firm behavior, the decision is not entirely driven by inertia; firms change their registration decisions in a way that is consistent with profit maximizing behavior.

The rest of the paper proceeds as follows. The next section reviews related literature. Section 3 develops the conceptual framework to analyze VAT bunching and voluntary registration. Section 4 derives the main empirical predictions. Section 5 provides an overview of the VAT system in the UK and describes the data. Sections 6 and 7 present the empirical analysis for voluntary registration and VAT bunching, respectively. Section 8 studies the dynamic movement of firms in and out of registration status. Finally, Section 9 concludes.

2 Related Literature

Our work contributes to several strands of literature, other than those already discussed. First, our work relates to the literature on the effect of tax and regulatory thresholds, and in particular, the effect of VAT thresholds on small business behavior. In an important paper, Keen and Mintz (2004) were the first to set up a model of VAT including a threshold; they show that there will be bunching below the threshold, and a "hole" above, where firms do not locate. However, there are a number of differences between their approach and ours.[5] First, their model has only final consumers (i.e. only B2C sales), and so, as argued above, cannot explain voluntary registration. Second, their main focus is on the optimal registration threshold, whereas ours is on the coexistence and determinants of voluntary registration and bunching. Kanbur and Keen (2014) extend the Keen and Mintz (2004) framework to allow for evasion, as well as avoidance, of VAT. Brashares et al. (2014) use a calibrated formula from Keen and Mintz (2004) to infer that for a 10 percent VAT rate, the optimal level for the threshold in the United States is $200,000.[6]

There is a small empirical literature on the effects of VAT thresholds. Onji (2009) documents the effects of the VAT threshold in Japan, focusing on the incentives for a large firm to split by separately incorporating. A comparison of the corporate size distributions before and after the VAT introduction of 1989 shows a clustering of corporations just below the threshold. Recent papers following our work document bunching of small firms at the VAT registration threshold in Finland (Harju, Matikka and Rauhanen (2016)) as well as lack of bunching in response to the VAT threshold in Armenia (Asatryan and Peichl (2016)). In particular, Harju, Matikka and Rauhanen (2016) provide strong evidence for Finland that bunching below the threshold occurs, and that compliance costs can explain a major part of bunching. However, neither of these papers studies the determinants of voluntary registration in detail, nor do they develop a theoretical framework specific to the VAT.[7]

There are several reasons that compliance costs may be more important in Finland than for our study. First, the VAT threshold in Finland is very low, at 8,500 Euro, compared to well over 100,000 Euro in the UK over our period. Other things equal, larger firms find compliance less costly. Second, in the UK, all active companies are required to file company accounts and corporation tax returns, so they already have the information required for the VAT return, and the VAT return itself is short and simple. Finally, a simplified VAT scheme,

[5] The main focus of their paper is to study the optimal VAT threshold, a topic beyond the scope of this paper.

[6] There is also a literature on VAT and the choice of informality in developing countries (Emran and Stiglitz (2005), de Paula and Scheinkman (2010)). In particular, de Paula and Scheinkman (2010) present a model where firms can choose between formal production, where they must register for VAT, and informal production; they can also choose between buying inputs from a formal and informal supplier. The focus of the paper is on the informality choice with an application to Brazil.

[7] Harju, Matikka and Rauhanen (2016) present formulae for the reduced form elasticity implied by observed bunching which are taken directly from Kleven and Waseem (2013). However, these formulae are originally developed for a labour supply model, and care must be shown when applying them to VAT. In particular, in an earlier version of this paper Liu and Lockwood (2015), we show that to apply the Kleven-Waseem formulae directly, it must be assumed that all sales are B2C, firms are price-takers, and the production function is fixed-coefficients must be assumed. Moreover, the elasticity estimated is an output supply elasticity, taking the price of output as given, not the elasticity of the tax base, because the latter will also be determined by the elasticity of demand.

the Flat-Rate Scheme (FRS), was introduced in the UK in 2002 to reduce compliance costs for small businesses; however, the FRS had a low take-up rate of around 3% among all eligible VAT traders, and there is no bunching at the FRS threshold above which firms are no longer eligible for FRS, suggesting that firms do not try to bunch to avoid the additional compliance costs of normal VAT registration (Vesal, 2013).[8]

More broadly, there are a few papers on the effects of other kinds of thresholds. The most relevant is Almunia and Lopez Rodriguez (2014), who study a threshold in Spain where firms with turnover above 6 million Euro face increased tax enforcement. Almunia and Lopez Rodriguez (2014) show that firms bunch at this threshold, and that bunching is more pronounced, the *lower* the fraction of B2C sales in the total for the sector. This is of course, the reverse to what we find. Their explanation is that if sales are mostly to other firms, if audited, a firm will have a "paper trail" that will make it relatively easy to cross-check tax returns to detect misreported intermediate input sales. The contrast between the results of Almunia and Lopez Rodriguez (2014) and ours indicates that the purpose of the turnover threshold is crucial i.e. whether it relates to a *change* in tax liability - as in our case- or *enforcement* of a given tax liability, as in theirs.[9]

Our work also relates to the literature on tax notches (Slemrod (2010), Kleven and Waseem (2013), Best and Kleven (2013), Kopczuk and Munroe (2015), and Kleven (2016)). In particular, Kleven and Waseem (2013) emphasize that if individuals behave fully rationally, notches give rise to bunching below the threshold, and "holes" above the threshold where maximizing agents will not locate, and then uses bunching to estimate both the elasticity of labour supply, and the degree of optimization frictions. As shown below, the "bunching equation" in this paper, which relates the amount of bunching to the elasticity of the tax base and parameters of the tax system, is mathematically very similar to the equation of (Kleven and Waseem (2013)). However, for reasons discussed below, it is much more difficult to back out credible elasticity estimates in the VAT case, and so we do not attempt this in the paper.

Also, because we study evasion, our paper further relates to recent literature using special features of tax systems to identify evasion from kinks and notches. For example, Best et al. (2015) study a minimum tax scheme for corporations in Pakistan which has a kink point where the real incentive for bunching is small, but the evasion incentive is large, and they

[8]Consistently with this, direct evidence on compliance costs for the UK put costs at around 1.5% of turnover for firms around the registration threshold (Federation of Small Businesses, 2010), which is relatively small compared to the burden of VAT at 20%.

[9]It may be possible that being registered for VAT increases the probability of audit. Howoever, in our data, this effect may be attenuated by the fact that all firms, VAT-registered or not, are also companies and thus file a corporate tax return.

find large bunching around the minimum tax kink.[10] However, we do not measure evasion directly, nor, given the UK VAT system, do we have any obvious way of decomposing the total bunching effect into bunching due to evasion, and that due to a real response. Rather, we show that our theoretical predictions are robust to evasion.

3 Conceptual Framework

Key Features of the Model. As described in the introduction, we aim to model the behavior of "small" firms selling to both final consumers and to businesses, where both voluntary registration and bunching can be equilibrium outcomes. The first point is that the coexistence of both voluntary registration and bunching requires that the "small" firms make *both* B2B and B2C sales. To see this, suppose first that firms make only B2C sales. Then, it is easily shown that irrespective of the degree of competition between firms, the cost of voluntarily registering exceeds the benefit, because the burden of VAT paid on output when registered exceeds the burden of VAT paid on inputs when not registered. Conversely, with only B2B sales, the voluntary registration is always optimal, because the burden of output VAT can be passed on to the buyer.[11]

Second, in order to study the effect of the input cost ratio transactions on voluntary registration and bunching, we must of course also assume that the "small" firms use intermediate inputs as well as labour in varying proportions, so this must be a feature of the model.

A final assumption is that the small firms have some market power; this is for two reasons. First, because as explained in the not-for-publication appendix in Section B, the simplest perfectly competitive model which allows for both B2B and B2C sales has some undesirable features; specifically, there is complete sorting,[12] which is not observed in practice, and the structure of the equilibrium is more complex than with monopolistic competition. Second, in practice, markets are not perfectly competitive, and it is interesting to ask what the effect of the degree of competition are on voluntary registration and bunching.

So, we construct the simplest general equilibrium model that has the required features. There is a single representative consumer that supplies labour and buys two kinds of goods; a differentiated good sold by the small firms, and a single good produced by a large firm. The large firm also buys inputs from the small firms, generating a B2B demand. We assume that the large firm is operating at a scale where non-registration for VAT (i.e. operating so that

[10] More recently, Waseem (2016), again for Pakistan, shows very large responses when reforms cut the rate of income tax to zero for the self-employed, which he interprets as being largely an evasion response.

[11] Both these points are made formally below in the context of our model, and are shown formally in the not-for-publication appendix.

[12] That is, registered firms sell only B2B, and non-registered firms sell only B2C.

the value of sales are below the VAT threshold) is never profitable. Finally, a homogenous input to the small firm is produced by a third sector of competitive firms from a labour input via a constant returns technology. The behavior of this last sector is summarized by a zero-profit condition that says that the price of the small-firm input is equal to the wage. The structure of this economy is illustrated in Figure 1.

Consumers. There is a representative household that has preferences over the homogenous good, consumed at level Y, a set of differentiated goods $a \in [\underline{a}, \overline{a}]$, consumed at levels $x(a),\ a \in [\underline{a}, \overline{a}]$, and leisure l. These preferences are of the following form:

$$U(X) + V(Y) + l,\ X = \left[\int_{\underline{a}}^{\overline{a}} (x(a))^{(e_C - 1)/e_C}\, da \right]^{e_C/(e_C-1)},\ e_C > 1 \tag{1}$$

where X is a CES index of differentiated products, and

$$U(X) = \lambda^{1/\phi} \frac{X^{1-1/\phi}}{1-1/\phi},\ V(Y) = (1-\lambda)^{1/\gamma} \frac{Y^{1-1/\gamma}}{1-1/\gamma},\ \phi > 0,\ \gamma > 1$$

Each differentiated good a is produced by a single small firm a, which can be either registered for VAT or not. For reasons further discussed below, we also allow the homogenous good Y to be subject to VAT or not. So, the household faces a budget constraint

$$P(1 + J.t)Y + \int_{\underline{a}}^{\overline{a}} (p(a)(1 + I(a)t)\, (x(a))^{(e_C - 1)/e_C}\, da = w(1 - l) \tag{2}$$

where and $1 - l$ is labour supply, w is the wage, P is the price of the homogenous good produced by the large firm, $p(a)$ is the price excluding VAT of the small firm $a's$ output, and where $I(a) \in \{0, 1\}$ is an indicator recording whether the firm registers for VAT or not, with a "1" indicating registration. So, if the firm is registered, the consumer price is grossed up by VAT i.e. $p(a)(1 + t)$. Finally, $J \in \{0, 1\}$ records whether Y is subject to output VAT. Our baseline case is $J = 1$, but $J = 0$ allows an interpretation of the model as a small open economy where Y is exported; see Section 3.1 below. Consistently with this interpretation, we assume that if $J = 0$, good Y is zero-rated for VAT, rather than exempt, so that the large firm can still claim back input VAT.[13]

So, by standard arguments, maximization of (1) subject to (2) gives household demand

[13] If the large firm were exempt, there would be no difference, from a VAT point of view, between B2B and B2C demand.

for the homogenous and differentiated goods:

$$Y = (1-\lambda)(1+J.t)^{-\gamma}P^{-\gamma} \tag{3}$$
$$x(a) = \lambda\left(\frac{(p(a)(1+I(a)t))}{Q}\right)^{-e_C} Q^{-\phi} \tag{4}$$

where Q is the CES price index corresponding to the quantity index X.[14] We assume that in equilibrium, positive leisure is consumed, so that from (1), the wage is fixed at unity.

The Large Firm. The large firm combines inputs $y(a),\ a \in [\underline{a}, \overline{a}]$ bought from the small firms via a constant returns CES production technology to produce output Y. This production technology is characterized by a CES cost function per unit of output of

$$C = \left[\int_{\underline{a}}^{\overline{a}} (p(a))^{1-e_B}\, da\right]^{1/(1-e_B)},\ e_B > 1$$

where $p(a)$ is the price of the input net of tax (as the large firm is VAT-registered, it can claim back any tax on inputs). So, the large firm chooses p to maximize $(1-\lambda)P^{-\gamma}(P-C)$. This gives the usual mark-up equation for price i.e.

$$P = \frac{\gamma}{\gamma-1}C \tag{5}$$

and thus, combining (5),(3), ultimately, output is

$$Y = (1-\lambda)(1+J.t)^{-\gamma}\left(\frac{\gamma}{\gamma-1}C\right)^{-\gamma} \tag{6}$$

Finally, input demand for variety a is, by Shepard's Lemma and (6):

$$y(a) = \frac{\partial C}{\partial p(a)}Y = (1-\lambda)(1+J.t)^{-\gamma}\left(\frac{\gamma}{\gamma-1}\right)^{-\gamma} C^{-\gamma}\left(\frac{p(a)}{C}\right)^{-e_B} \tag{7}$$

The Small Firms. A small firm's technology is assumed constant returns and CES, like the large firm, it is described by the unit cost function, which is specified as:

$$c(I(a);a) = \frac{1}{a}\left(\omega r(1+I(a)t)^{1-\sigma} + (1-\omega)w\right)^{1/(1-\sigma)} \tag{8}$$

[14] That is,

$$Q = \left[\int_{\underline{a}}^{\overline{a}} ((1+I(a)t)p(a))^{1-e_C}\, da\right]^{1/(1-e_C)}$$

where r, w are the prices of the intermediate input and labor. By assumption, $w = 1$, and we have also assumed, w.l.o.g., that one unit of the intermediate good requires one unit of labor, so $r = 1$ also. Note that the cost of the input is grossed up by the tax t if the firm is not registered, as the firm cannot claim the input tax back.

To interpret (8), note that a is a measure of productivity, and ω is a measure of the firm's use of intermediate inputs relative to labour, independently of productivity. To see this, note that setting $I(a) = 0, 1$ in (8), and using $w = r = 1$, we get

$$c(1;a) = \frac{1}{a} < \ c(0;a) = \frac{1}{a}\left(\omega(1+t)^{1-\sigma} + 1 - \omega\right)^{1/(1-\sigma)} \tag{9}$$

So, the firm has a unit cost that only depends on a unless it is not registered for VAT, in which case it pays tax t on its input, the cost burden of which is obviously increasing in ω.

Generally, as long as the elasticities of demand from the household and the large firm are different, i.e. $e_C \neq e_B$, the small firm prefers to price discriminate if it can. Moreover, ruling out price discrimination is not very tractable, as the profit-maximising price cannot then be solved for in closed form, unless it is assumed that $e_C = e_B$.

So, we assume for now that small firms can price-discriminate, setting prices $p_C(a), p_B(a)$ for B2C and B2B customers respectively.[15] Then, suppressing the dependence of p_C, p_B, I etc. on a to lighten notation, the firm's profit is

$$\pi(p_C, p_B,\ I; a) = x(p_C - c(I;a)) + y(p_B - c(I;a)) \tag{10}$$

where from (4), (7):

$$x = \lambda A_C \left(p_C(1 + I.t)\right)^{-e_C},\ y = (1-\lambda)A_B(p_B)^{-e_B} \tag{11}$$

and where

$$A_C = Q^{e_C - \phi},\ A_B = (1 + J.t)^{-\gamma}\left(\frac{\gamma}{\gamma - 1}\right)^{-\gamma} C^{e_B - \gamma} \tag{12}$$

are parameters that the small firms take as given, but are determined in industry equilibrium.

The small firm chooses $p_C, p_B \in [0, \infty), I \in \{0, 1\}$ to maximize $\pi(p_C, p_B,\ I; a)$ subject to the registration constraint, which says that if the firm chooses not to register ($I = 0$), the value of sales $s \equiv p_C x + p_B y$ must be less than the VAT sales threshold s^*. This allows of course, for *voluntary registration*, which is defined by a choice $I = 1$ when $s < s^*$.

The costs and benefits of registration are clear from equations (10) and (11). The benefit is that registration, $I = 1$, lowers the unit cost of production. The cost is that at a fixed

[15] Of course, if $e_B = e_C$ is assumed, the price-discrimination assumption becomes trivial.

price, registration lowers B2C sales, because demand by the household is reduced by the tax.

3.1 Discussion

Here, we discuss some other modelling choices. First, in practice, there are compliance costs to VAT registration. These could be introduced to the model, at the cost of some additional complexity. However, it should be noted that these costs are relatively small for the UK: for example, a recent literature review found that at the registration threshold, these costs were around 1.5% of turnover, declining to 0.1% or less for large companies (Federation of Small Businesses, 2010), and so we do not feel it necessary to include these in our analysis[16].

Second, it has been argued that the amount of output exported is a determinant of registration, because in practice, exports are exempt from VAT, and so firms exporting more of their output are more likely to register (Brashares et al., 2014). Our model could be interpreted to cover this case. This is because in the case of exports, the exporter does not bear any of the burden of the output VAT, and so from the supplier's point of view, domestic B2B sales and exports are equivalent in this respect. So, we could interpret good Y as being purchased by foreigners, rather than domestic consumers.[17] Finally, in our baseline model, we do not allow for evasion, to avoid overloading the analysis. We introduce evasion in Section 4.4.

4 Analysis

4.1 Necessary Conditions for Voluntary Registration and Bunching

As already argued above, the coexistence of both voluntary registration and bunching requires that the "small" firms make *both* B2B and B2C sales. This is relatively easy to show formally in our framework[18].

Proposition 1 *If all sales are B2C i.e.* $\lambda = 1$, *there is no voluntary registration. If all sales are B2B, i.e.* $\lambda = 0$, *all firms register voluntarily, whatever their turnover, so there can be no bunching. So, to observe both bunching and voluntary registration simultaneously, we require that there are both B2C and B2B sales i.e.* $0 < \lambda < 1$.

[16] Also, empirically, there are no major changes in compliance requirements duering our sample period.

[17] In that case, technically, we would have to replace the term $1 + J.t$ in (3) by $1 + J.t^*$ where t^* is the foreign rate of VAT.

[18] All proofs are in the Appendix.

Proposition 1 is in fact completely general; as shown in the not-for-publication appendix, it does not rely on our technical assumptions such as iso-elastic demand or constant marginal cost. For the remainder of the theoretical section of the paper, we thus focus on the general case where $0 < \lambda < 1$. Here, we face a problem; generally, A_C, A_B are endogenous to λ, ω because these are industry-level variables. So, we assume that:

A1: $e_C = \phi,\ e_B = \gamma$.

A1 ensures that the demand parameters are exogenous i.e. from (12), $A_C = 1,\ A_B = (1 + J.t)^{-\gamma} \left(\frac{\gamma}{\gamma-1}\right)^{-\gamma}$.

4.2 Voluntary Registration

We now turn to the registration decision, which is a choice $I \in \{0,1\}$, with the firm then maximizing profit given its registration decision. To lighten notation, we define the following parameters

$$B_C = \left(\frac{e_C - 1}{e_C}\right)^{e_C} \frac{1}{e_C - 1},\ B_B = (1 + J.t)^{-\gamma} \left(\frac{\gamma - 1}{\gamma}\right)^{\gamma} \left(\frac{e_B - 1}{e_B}\right) \frac{1}{e_B - 1} \tag{13}$$

Also, we define two crucial cost and demand changes. First, the increase in the firm's unit costs due to non-registration, because of input VAT, can be defined independently of firm productivity a as

$$\Delta_c = \frac{c(0;a)}{c(1;a)} - 1 = \big(\omega(1+t)^{1-\sigma} + (1-\omega)\big)^{1/(1-\sigma)} - 1 > 0. \tag{14}$$

Call this the *input VAT effect* on cost. Second, assuming $e_B = e_C = e$, it can be shown that at any fixed price $p_C = p_B = p$, the reduction in overall demand for the firm's product due to the charging of output VAT on B2C sales is[19]

$$\Delta_d = \frac{\lambda(1-(1+t)^{-e})}{\lambda + (1-\lambda)A_B} = \frac{\lambda(1-(1+t)^{-e})}{\lambda + (1-\lambda)(1+J.t)^{-e}\left(\frac{e-1}{e}\right)^{e}} > 0. \tag{15}$$

This is because when output VAT is charged, at a fixed price p_C, all B2C sales (which count for λ of the total) are reduced by a factor $(1+t)^{-e}$; call this the *output VAT effect*. We can then show:

[19] To see this, note that from (11), the ratio of demand wih registration to without is

$$\frac{\lambda(p(1+t))^{-e} + (1-\lambda)A_B p^{-e}}{\lambda p^{-e} + (1-\lambda)A_B p^{-e}} = \frac{\lambda(1+t)^{-e} + (1-\lambda)A_B}{\lambda + (1-\lambda)A_B} = 1 - \Delta_d$$

Proposition 2 *(a) A firm of type a will register voluntarily iff*

$$a^{e_B - e_C} \geq \frac{\lambda B_C}{(1-\lambda)B_B} \frac{(1+\Delta_c)^{1-e_C} - (1+t)^{-e_C}}{1-(1+\Delta_c)^{1-e_B}} \tag{16}$$

(b) If the voluntary registration condition (16) holds at λ, ω, it also holds at $\omega' \geq \omega$, $\lambda' \leq \lambda$; that is, voluntary registration is more likely, the higher the input cost ratio, ω, and the lower the share of B2C sales, λ.

(c) If $e_B = e_C = e$, (16) holds independently of a iff

$$T = (1 - \Delta_d)(1 + \Delta_c)^{e-1} \geq 1 \tag{17}$$

(d) Assuming $e_B = e_C = e$, if the large firm is zero-rated i.e. $(J = 0)$, in the the competitive limit, as $e \to \infty$, voluntary registration is always optimal i.e. (17) *holds for e high enough, but if it is not zero-rated, $(J = 1)$, in the competitive limit, voluntary registration is never optimal.*

The voluntary registration condition is most easily interpreted when it is in the form (17). There, it says that the if the input VAT effect on cost, Δ_c due to non-registration is large relative to the output VAT effect Δ_d, there will be voluntary registration. Note also when $e_C = e_B$, *T is a sufficient statistic that captures the entire effect of the VAT system on voluntary registration.* We will see later that it is also a sufficient statistic for the degree of bunching. Note also that T depends on the parameters of demand and cost functions, $\lambda, \omega, \sigma, e$, as well as the tax rate t.

Finally, the comparative statics results in Proposition 2 provide testable predictions. The predictions regarding the effect of the share of B2C sales λ, and the input cost ratio, ω, on voluntary registration are perfectly general. The effect of the level of competition, e, on voluntary registration can only be established in the competitive limit, however.

To get some further insights on the effect of e on the sufficient statistic, T, we report some numerical simulations. A calibration of ω described in Appendix A implies $\omega = 0.419$, assuming only a fixed-coefficients production function. The parameter λ is harder to calibrate from our data, because to do so, we need to make assumptions about e, which we also want to vary as a parameter. So, we simply allow λ to take on values 0.1, 0.5, and 0.9. Table 1 Panel A shows the value of T for different values of λ and e, for both the case of zero-rating of the large firm $(J = 0)$, and the case where it is subject to VAT $(J = 1)$. Voluntary registration generally occurs when the elasticity e is relatively high and the large firm is zero-rated.

The intuition for these analytical and numerical results is the following. Generally, voluntary registration occurs when output effect Δ_d is small, and/or when the input VAT effect

Δ_c is large. The first observation is that other things equal, the larger λ, the bigger is the output VAT effect Δ_d; this explains the fact that T falls with λ. Second, other things equal, the larger ω, the bigger is the input VAT effect Δ_c; this explains why T rises with ω.

As regards the effect of competition, measured as an increase in e, we can note the following. First, if the large firm is zero-rated i.e. $J = 0$, then the term $\left(\frac{e-1}{e}\right)^e (1 + J.t)^{-e} = \left(\frac{e-1}{e}\right)^e$ tends to 0.36, so $1 - \Delta_d$ tends to a strictly positive number of $\frac{0.36(1-\lambda)}{\lambda+0.36(1-\lambda)}$ i.e. the output VAT effect on demand is not too strong[20]. On the other hand, Δ_c is independent of e, for large e, so the term $(1 + \Delta_c)^{e-1}$ dominates in (17), and eventually $T > 1$. That is, for large e, the input VAT effect dominates the output VAT effect.

On the other hand, note that when the large firm is not zero-rated ($J = 1$), then the output VAT effect becomes very strong; in fact, as $e \to \infty, 1 - \Delta_d$ is proportional to $1/(1 + t)^e$, which dominates the input effect $(1 + \Delta_c)^{e-1}$. In this case, eventually $T \to 0$. That is, for large e, the output VAT effect dominates the input VAT effect.

4.3 Bunching

To get a precise characterization of the determinants of bunching, we assume $e_C = e_B = e$. Now consider what happens if $T < 1$. Consider a firm a^* which is just generating a total value of both B2C and B2B sales s^* when not registered i.e. has a value of $a = a^*$ such that

$$p(a^*)(x(a^*) + y(a^*)) = s^* \tag{18}$$

where $p(a)x(a)$ and $y(a)$ denote the optimal price and sales of a non-registered firm, so from (11);

$$p(a) = \frac{e}{e-1}c(0; a), \;\; x(a) + y(a) = (\lambda + (1 - \lambda)A_B)(p(a))^{-e} \tag{19}$$

Now consider a firm slightly more productive than firm a^*; that is, with an a' slightly above a^*. Then, as $T < 1$ (i.e. voluntary registration is never attractive), this firm is discretely worse off registering than not registering; so, it would be willing to change its price a little away from the profit-maximizing optimum, in order to stay below the registration threshold. For example, as demand is elastic, it can do this by cutting price a little. This implies that there must be an interval of firms, $a \in [a^*, a^* + \Delta a^*]$, who *bunch* by restricting sales in order to stay at the VAT threshold. The value of $a^* + \Delta a^*$ is given by the indifference condition that firm $a^* + \Delta a^*$ has the same profit whether it is registered or not. If $\pi(I; a)$ denotes optimized profit, conditional on the registration decision $I \in \{0, 1\}$, this condition

[20] Note that $\lim_{e\to\infty} \left(\frac{e-1}{e}\right)^e = \lim_{e\to\infty} \left(1 - \frac{1}{e}\right)^e \simeq 0.36$ by standard arguments.

is

$$\pi(1; a^* + \Delta a^*) = \pi(0; a^* + \Delta a^*) \tag{20}$$

So, the amount of bunching is measured by Δa^*.

Now, we do not observe a, but we do observe firm sales. Following Saez (2010) and Kleven and Waseem (2013), we reason as follows. First, note from the fact that (that $(a^* + \Delta a^*)$, which is unobservable, maps into $s^* + \Delta s^*$, via the relationship

$$s^* + \Delta s^* = (\lambda + (1-\lambda)A_B)\left(\frac{e(1+\Delta_c)}{e-1}\right)^{1-e}(a^* + \Delta a^*)^{e-1} \tag{21}$$

Then, combining (20),(21), we can eventually show:

Proposition 3 *The level of bunching Δs^* is given by the implicit relationship*

$$\frac{e}{(1+\Delta s^*/s^*)} - (e-1)\left[\frac{1}{1+\Delta s^*/s^*}\right]^{e/(e-1)} - T = 0 \tag{22}$$

where $T < 1$ is the VAT sufficient statistic.

Note that the entire effect of VAT on bunching is captured by the sufficient statistic T. Note that here, $T < 1$, otherwise there is voluntary registration. We can now use (22) to look at some of the determinants of bunching. We have:

Proposition 4 *The amount of bunching Δs^* rises (i) as λ, the fraction of B2C sales increases, and (ii) as the share of inputs in total cost, ω, falls. Moreover, if the derivative of T with respect to e is greater than 1, the amount of bunching Δs^* falls as e rises.*

The intuition for (i) and (ii) is very similar to the case of voluntary registration. That is, factors that make voluntary registration less attractive also provide incentives for staying under the VAT threshold by bunching. Specifically, this will be the case when most customers are not VAT-registered, so that the burden of an increase VAT can not easily be passed on to the buyer, and/or when input costs are relatively unimportant relative to labour costs. We will bring these predictions to the data below.

Finally, the effect of increased competition on bunching is more subtle and cannot be established analytically. Some simulation results are reported in Panel B of Table 1, where each cell reports the solution value $\Delta s^*/s^*$ to (22) for the given parameters. For some values in Table 1, $T > 1$, and in this case, (22) does not have a solution i.e. there is no bunching, in which case, we record a zero.

Table 1 Panel B shows first that the level of bunching always increases with λ, consistently with Proposition 4. Second, other things equal, the amount of bunching is higher if good Y is taxed, rather than zero-rated. This is consistent with Proposition 2, where voluntary registration is found to be less likely when good Y is taxed. Finally, the relationship between the elasticity e and bunching is generally not monotonic. As long as $T < 1$, the amount of bunching is generally increasing with e, but then, in the case where good Y is zero-rated, when T rises above 1, the amount of bunching falls to zero because all firms register voluntarily.

4.4 VAT Evasion

VAT in the UK has been susceptible to fraud and avoidance, as in many other countries. According to HMRC estimates, the VAT tax gap, which is defined as the difference between theoretical VAT liabilities and total VAT receipts on a timely basis, is currently around 10% of theoretical VAT liability. Some of this gap can be accounted for by VAT debt owed by firms, and by sophisticated fraud schemes, but most of the VAT gap is probably due to sales under-reporting and cost over-reporting.[21]

Here, we model the simplest and most common form of VAT evasion, where the registered seller does not charge VAT on some proportion of B2C sales, for example by taking a cash payment for this fraction. If we let this fraction be ν, the total cost of x units of the good to the household will be

$$\nu xp + (1-\nu)xp(1+t) = xp(1+(1-\nu)t).$$

That is, the household faces an average price of $p(1+(1-\nu)t)$, and the firm continues to get revenue p on every unit sold to the household. We will assume that ν is exogenously fixed, both for simplicity, and also because there are some analytical issues in endogenizing it[22]. The main qualitative points will extend to the endogenous case.[23]

[21] For example, HM Revenue and Customs (2015) estimates that the total VAT gap in 2014-14 was 13.1 bilion. Of this, at most £1.0 billion is due to Missing Trader Intra-Community fraud, £1.2 billion is VAT debt, and finally £0.2 billion due to VAT avoidance. This means that over 80% of the gap is due to other factors.

[22] Suppose that the firm chooses ν to maximize profit minus evasion cost $g(\nu)$. It is easily verified that optimized profit is convex in v, as it only depends on v via the term $(1+t(1-\nu))^{-e}$, which is a convex function of ν. So, to have an interior solution, $g(v)$ also has to be sufficiently convex in ν. But then, a closed-form solution for ν cannot be found.

[23] For example, suppose the cost of evasion is linear in ν, up to a limit $\overline{\nu} < 1$. Then, as profit is convex in ν, as explained in the previous footnote, and the evasion cost is small, the firm will always choose $\overline{\nu}$, so that

It is then easily verified that as regards voluntary registration, the analysis proceeds much as before except that the VAT sufficient statistic becomes

$$T(\nu) = (1 - \Delta_d(\nu))(1 + \Delta_c)^{e-1}, \ \Delta_d(\nu) = \frac{\lambda(1 - (1 + (1 - \nu)t)^{-e})}{\lambda + (1 - \lambda)A_B} \tag{23}$$

Thus, with evasion, the output VAT effect depends on ν and is smaller than without evasion i.e. $\Delta_d(\nu) < \Delta_d(0)$. This is intuitive; with some VAT evaded on sales, output VAT becomes less of a burden. It then follows that T is increasing in ν i.e. voluntary registration is more likely, the greater the opportunities for evasion, as measured by ν.

As regards bunching, evasion has two opposing effects. First, evasion allows relaxes the constraint imposed by the VAT threshold, as the tax authority only observes $1 - \nu$ of B2C sales, and so the firm can in fact produce over the threshold without registering. Second, as just discussed, evasion makes registration less costly, output VAT becomes less of a burden as explained above.

Both of these effects appear formally as follows. With evasion, we show in the not-for-publication Appendix **B** that the term T in the bunching equation (22) is replaced by

$$\hat{T}(\nu) = T(\nu)\frac{\lambda(1 - \nu) + (1 - \lambda)A_B}{\lambda + (1 - \lambda)A_B} \tag{24}$$

and in particular, positive bunching will occur when $\hat{T}(\nu)$ is less than 1.[24] So, an increase in ν has two opposing effects, as just described. First, $\hat{T}(\nu)$ falls, making bunching more likely, via the second term in (24), $\frac{\lambda(1-\nu)+(1-\lambda)A_B}{\lambda+(1-\lambda)A_B}$; this captures the constraint relaxation effect. Second, $\hat{T}(\nu)$ rises, via the fact that $T(\nu)$ rises; this captures the effect that evasion reduces the burden of output VAT.

Note that with evasion, the qualitative effects of λ and ω on T do not change, and so our predictions about the determinants of voluntary registration do not change; this is clear by inspection from (23). This is also true of bunching; it is seen by inspection that $\hat{T}(\nu)$ is decreasing in λ, and increasing in ω, as is T. So, our key empirical predictions are robust to the presence of evasion. We can summarize as follows:

Proposition 5 *If evasion is possible, the likelihood of voluntary registration rises with evasion ν, whereas the effect of evasion on bunching is ambiguous. Moreover, evasion does not affect our qualitative predictions about the effects of λ, the fraction of B2C sales, and as the share of inputs in total cost, ω, on voluntary registration and bunching.*

---

it is effectively exogenous.

[24]For a formal proof, see the not-for-publication Appendix in Section B.

However, as discussed in Section 1, we do not measure evasion directly. Nor do we have any obvious way of decomposing the total bunching effect into bunching due to evasion, and that due to a real response, although this can be done plausibly for business taxes in some other countries, using special features of national tax systems.[25] Our empirical strategy focuses primarily on identifying the effects of changes in the B2C ratio and input cost ratio as predicted by the theory, without taking a view on how much of this effect works through evasion.

4.5 The Elasticity of the Tax Base

Note that (22) is closely related to the Kleven-Waseem formula relating bunching at a notch of the personal income tax schedule to the elasticity of the labour supply, e_L. In their equation, the tax notch is measured by the term $\Delta t/(1-t)$, where t is the lower rate of income tax, and Δt is the increase in the tax rate at the notch. In fact, it is easily verified if we take (22) and substitute $e_L = e - 1$, replace $\Delta s^*/s^*$ by $\Delta z^*/z^*$, and replace $T^{1/e}$ by $1 - \frac{\Delta t}{1-t}$, we get equation (5) in their paper.

In the context of the personal income tax, Kleven and Waseem (2013) uses (22) in a reverse way to us, to make inferences about the elasticity of taxable income e_L given an estimate of $\Delta z^*/z^*$ from the data. The question then arises as to whether we can use (22) in the same way. The first question is whether it is worthwhile i.e. what interpretation we can give to e. It is possible to show that if good Y is taxed ($J = 1$), and production is fixed-coefficients ($\sigma = 0$), then the value-added of a registered small firm i.e. the base of the VAT for that firm is proportional to $(1+t)^{-e}$, so the elasticity of the tax base with respect to $1+t$ is e (see Appendix A). So, there is some interest in trying to recover e from (22). However, there are two complications here.

First, the unlike the personal income tax case, VAT sufficient statistic T does not depend just on the tax code. In particular, also depends on model parameters, for example, λ and ω even if $\sigma = 0$, and also on ν if evasion is assumed. This means in turn that recovery of e from (22) means making assumptions about these parameters. As shown in the not-for-publication Appendix B, ω can be calibrated from our data, but λ and ν are harder to specify.

Second, as shown in Lockwood (2016), in the presence of a notch, the elasticity of the tax base (in the case of the personal income tax, taxable income), is no longer a sufficient

[25] For example, Best et al. (2015) use a minimum tax scheme for corporations in Pakistan which has a kink point where the real incentive for bunching is small, but the evasion incentive is large, and they find large bunching around the minimum tax kink. Waseem (2016), again for Pakistan, shows very large responses when reforms cut the rate of income tax to zero for the self-employed, which he interprets as being largely an evasion response.

statistic for the marginal deadweight loss of the tax, so elasticity estimates are of less interest from a normative point of view. So, for these reasons, we do not attempt elasticity estimates.

4.6 Summary of Theoretical Results

We now summarize the theoretical results in Figure 2, which shows how firms of productivity type a behave as T varies. The figure is drawn under the simplifying assumption that $e_C = e_B$ and that there is no evasion. In this case, we know that all firms register for VAT if $T > 1$. So, all firms with $a > a^*$ are registered with a turnover above the threshold, but all firms with $a < a^*$ are voluntarily registered, where a^* is defined in (18) above.

When T falls below 1, firms $a \in [a^*,\ a^*+\Delta a^*]$ start to bunch at the registration threshold. as T falls, this bunching interval becomes larger. So, for any fixed value of T, firms can now be in one of three regimes: when a is low, the firm will be unconstrained but below the threshold, when a is high, the firm will be unconstrained and above the threshold and thus registered, and when a is intermediate, the firm will be just at the threshold.

It is also clear from Figure 2 how voluntary registration and bunching may coexist in any given industry. Some fraction of firms may have a $T > 1$, so the lower-productivity ones in this group will be voluntarily registered. At the same time, some other fraction may have a $T < 1$, so some of this latter group will be bunching.

5 Context and Data

5.1 The Value-Added Tax System in the UK

The Value-Added tax in the UK is remitted by approximately 2 million registered businesses in each fiscal year. It is the third largest source of government revenue following income tax and national insurance contributions. In 2011/12, VAT raised £98.2 billion, accounting for 21.1% of total tax revenue and 6.5% of GDP in the UK.[26]

VAT is levied on most goods and services provided by registered businesses in the UK, goods and some services imported from countries outside the European Union (EU), and brought into the UK from other EU countries. All businesses must register for VAT if their taxable turnover is above a given threshold.[27] A VAT registered business pays VAT on its purchases (input tax), and charges VAT on the full sale price of the taxable supplies (output

[26] See http://www.hmrc.gov.uk/stats/tax receipts/tax-receipts-and-taxpayers.pdf.

[27] VAT taxable turnover includes the value of any goods or services a business supplies within the UK, unless they are exempt from VAT. Any supplies that would be zero-rated for VAT are included as part of the taxable turnover.

tax). Businesses can also choose to register voluntarily with a turnover below the threshold in order to recover the input taxes.

The default VAT rate is the standard rate, which was 17.5% between April 1, 2004 and December 1, 2008 and was temporarily reduced to 15% before January 1, 2010. The standard rate was then reverted to 17.5% until 4 January 2011 when it was increased to 20% and has been at that rate since. A small number of goods and services have VAT levied at a reduced rate of 5% and there are also goods and services that are charged at a zero rate or are exempt from VAT altogether.[28]. Neither businesses that make zero-rate or exempt supplies charge output VAT to the customers, and the key difference between them is that input tax cannot be claimed against output tax on exempted supplies. The registration thresholds and standard rate of VAT over our sample period are shown in Table 2

The increase in registration threshold tracks the rate of inflation.[29] The UK has the highest registration threshold in the EU, which is perceived as a way for the government to reduce the compliance costs of small businesses not wishing to register for VAT.[30]

There are two rules governing registration, a forward-looking rule and a backward-looking one. Under the forward-looking rule, a firm must also register for VAT if its VAT taxable turnover is likely to go over the threshold in the next 30 days. Under the backward-looking rule, a firm must register if its VAT-taxable turnover in the previous 12 months was above the threshold. Strictly speaking, our theoretical model is static and applies to the forward-looking decision; that is, the firm must register if turnover in the current year is expected to exceed the threshold. In our sample, around 68% of first-time registers have a previous-year turnover less than the VAT threshold. This suggests that the forward-looking decision is more important.

5.2 Data

We construct our dataset by linking the universe of VAT returns to the universe of corporation tax records in the UK. The first data set provides detailed information for businesses in

[28] A reduced rate of 5% is charged on a small number of supplies under schedule 7A of the *Value Added Tax Act (VATA) 1994*. Principally, they include the supply of domestic fuel and power, the installation of energy saving materials, women's sanitary products, children's car seats and certain types of construction work.

[29] Specifically, under Article 24(2)(c) of the sixth EC VAT directive (77/388/EEC 17 May 1977). These provisions are now consolidated in the principal VAT directive (2006/112/EC); article 287 allows for States to increase the registration threshold in line with inflation. As far as we are aware, there is no interaction between the VAT registration threshold with other major taxes in the UK, not does it change the probability of being audited for firms around the threshold.

[30] Small firms with annual taxable turnover of up to £150,000 can use a simplified flat-rate VAT scheme, but are subject to the same registration threshold. The flat-rate scheme is not widely used. In 2007, only 16% of eligible firms were registered under the flat-rate scheme (Vesal, 2013).

different legal forms including sole traders, partnerships, and companies that are registered for VAT. To obtain information on non-VAT registered businesses, we link the VAT records to the population of corporation tax records based on a common anonymised taxpayer reference number. The linked dataset allows us to identify both registers and non-registers for the population of UK *companies*, and contains rich information on VAT and corporation tax for each company and year.

We further merge the linked tax dataset with two additional data sources: (1) the FAME (Financial Analysis Made Easy) annual company account database for additional firm characteristics and accounting information and (2) the annual sector-level statistics on the share of sales to final consumers based on the Office of National Statistics (ONS) Input-Output Tables. The last data source gives us an empirical proxy for λ, the share of sales that are B2C at the 2-digit SIC industry level.

The final dataset contains 731,706 observations for 435,688 companies between April 1, 2004 and March 30, 2010.[31] For each company-year observation, we have information on the VAT-exclusive turnover taken from the corporate tax records, and whether the company is registered for VAT.[32] We examine a few key factors that drive firms' decisions about voluntary registration, including the share of input cost relative to total turnover (input-cost ratio), the share of sales to final consumers (B2C sales ratio), firm-specific history of registration, and the degree of industry-level competitiveness (measured by four-firm concentration ratio, or CR4)[33].

[31] We take several steps to refine the sample to study the VAT registration decisions of individual companies. First, we eliminate companies which are part of a larger VAT group and focus only on standalone independent companies. This is because companies under common control–for example, subsidiaries of a parent company–can register as a VAT group and submit only one VAT return for all companies in a VAT group.

Second, we drop all observations with partial-year tax or accounting records because the registration decision can be based on turnover in the previous 12 months. We further eliminate companies that mainly engage in overseas activities based on the HMRC trade classification. This is because the taxable VAT turnover is based on sales of goods and services within the UK. Finally, we drop companies with an effective rate of VAT that is less than 10%, where the effective rate is the output VAT relative to VAT-eligible sales for registered companies, to form the main sample for empirical analysis. Empirical results based on the full sample that include all firms with an effective VAT rate below 10% are quantitatively similar to those obtained using the main sample and are available upon request.

[32] Our empirical analysis is based on turnover reported in the CT600 for two reasons. The first is mechanical: we only observe turnover liable for VAT for firms that are registered. The second is related to salience given that firms that are not registered for VAT are more likely to base their registration decision on the overall amount of turnover, instead of computing a separate measure of turnover that is subject to VAT. To see whether this is true, we predict (out-of-sample) the amount of turnover liable for VAT for unregistered firms, by regressing the amount of turnover liable for VAT on the amount of total turnover and a full set of industry and year dummies. We then plot a similar histogram of turnover as in Figure 2 Panel B based on actual/predicted turnover liable for VAT for registered/unregistered firm. Bunching below the VAT notch is still present, but much more noisy and imprecise comparing to bunching based on total turnover reported in CT600. The empirical differences suggest that for unregistered firms, they are more likely to rely on the overall turnover figure for their VAT registration decisions.

[33] We calculate the concentration ratio based on the market shares of the four largest firms in an industry,

We use two different data sets to test related hypotheses developed in Section 3. First, we use *all* the firms with turnover below the current-year VAT registration threshold to examine the choice of voluntary registration. We say that a firm is voluntarily registered if it has never registered before and has a turnover below the VAT threshold, or that if it was registered in the previous year and has a turnover below the VAT deregistration threshold. In the main sample, 62.5% of firms have a turnover below the VAT threshold, and of these, 44.1% of them are registered for VAT. So, overall, 27.6% of firms in the main sample of companies are voluntarily registered. Second, we focus on firms with turnover within the neighboring of the registration threshold, i.e. between £10,000 and £200,000, to analyze the extent of bunching below the VAT notch.

5.3 Summary Statistics

Figure 3 presents convincing evidence that the VAT registration threshold is binding in the UK, by showing a histogram of the normalized turnover by pooling data between 2004/05 and 2009/10, where the normalized turnover is defined as firms' nominal turnover net of the current-year VAT notch. There is an evident excess of mass just below the notch and a small missing mass above, in the otherwise smooth distribution of turnover. However, it is worth noting that relative to some other studies, the excess mass below the threshold is not sharply bunched at the notch. A plausible explanation is that firms have less control over their turnover than individuals do over their earnings for example. Alternatively, firms that benefit from voluntary registration can also stay below the registration threshold.

Table 3 reports summary statistics for companies in the neighborhood of current-year VAT notch, i.e. those with nominal turnover of between £10,000 and £200,000. The first three columns report the mean, standard deviation and the number of non-missing observations for the key variables used in empirical analysis. Companies in this turnover region account for around 52.94% of all companies in the linked dataset. Columns (4)-(6) focus on the registered companies while columns (7)-(9) focus on the non-registered. The last two columns test whether there is any significant difference between the means of the two groups, by reporting the $t-$statistic and the corresponding p-value in columns (9) and (10), respectively.

based on the larger dataset of population of corporate tax records.

6 Voluntary Registration

6.1 Evidence on Voluntary Registration

In this section, we examine whether the decision of voluntary registration is consistent with the theory in the three key aspects as predicted by Proposition 2, including whether a firm is more likely to voluntarily register for VAT if it mainly sells to final consumers, has a larger share of inputs in cost, or in a more competitive industry.

We first note in Table 4 that the empirical pattern is broadly consistent with Proposition 2. As the share of B2C sales falls, i.e. when moving from the fourth (Q4) to the first quartile (Q1), the share of voluntarily registered firms tends to rise. Similarly, as the input cost ratio rises, the share of voluntarily registered firms tends to increase.

Next, we model the decision of voluntary registration in a binary choice model of the following form:

$$R_{it} = \gamma_1 + \gamma_2 B2C_{j(i)} + \gamma_3 ICR_{it} + \gamma_4 CR4_{j(i)} + \gamma_5 \mathbf{X}_{it} + \rho_t + \phi_i + \upsilon_{it}, \tag{25}$$

where R_{it} is a dummy indicator that takes value 1 if the firm is voluntarily registered and zero otherwise. The key variables of interest are $B2C_{j(i)}$, the industry-level B2C ratio for firm i (that is, firm i in industry $j(i)$), ICR_{it}, the input cost ratio for firm i in year t, and $CR4_{j(i)}$, the four-firm concentration ratio to measure competition in industry j in which firm i is located. $\mathbf{X}_{it}$ are other firm-level controls, ϕ_i and ρ_t are time-invariant firm fixed effects and year dummies, and υ_{it} is the error term. We estimate equation (25) in a linear probability framework based on the standard OLS assumptions and in a fixed-effect logit model. The regression sample includes all firms with turnover below the current-year VAT registration threshold. Consistent with Proposition 2, we expect that $\gamma_2 > 0$, $\gamma_3 > 0$, whereas the sign of γ_4 is uncertain.

Table 5 reports the estimation results from the linear probability model.[34] Columns (1)-(3) includes each of the key variables at a time, and column (4) includes all three key variables in the regression. Without inclusion of firm fixed effects, we can examine the effect of industry-level B2C sales ratio and CR4 ratio on the probability of voluntary registration in the first four columns. The coefficient estimates are negative and statistically significant, indicating that the likelihood to voluntarily register for VAT is reduced by around 4 percentage points given a one-standard-deviation increase in the B2C sales ratio, and by around 0.2 to 1 percentage point given a one-standard-deviation increase in the CR4 ratio.

[34] The estimation results from the fixed-effects logit model are very similar and available from the authors upon request.

The rest of columns add firm fixed effects and the coefficients on the B2C and CR4 ratios are often imprecisely estimated due to limited time-series variation at the industry level. For comparison, column (5) does not include any additional controls, whereas column (6) adds firm-level trading profit and age since incorporation. Columns (7) checks the robustness of the results by replacing the salary-*inclusive* input cost ratio calculated from the corporation tax records with the salary-*exclusive* input cost ratio calculated from FAME. The latter improves precision in the measurement of input cost, but substantially reduces the sample size given that few firms report the direct cost of sales. The coefficient estimate for the input cost ratio remains positive and highly significant throughout the different specifications. Moreover, the coefficient estimates for the B2C sales ratio and CR4 ratio are negative and significant at 10% level in column (7). Focusing on results in columns (4) and (7), the likelihood of voluntarily registering for VAT is increased by around 1 percentage point given a one standard deviation increase in the input cost ratio.

7 Evidence on Bunching

7.1 Estimation Methodology

As set out in the conceptual framework in Section 3, the VAT registration threshold at the cutoff turnover value s^* will induce excess bunching at the threshold by companies for which voluntary registration is not optimal. We are interested in constructing the empirical equivalent of $\Delta s^*/s^*$. First, we measure Δs^*, the amount of excess bunching, as the difference between the observed and predicted bin counts over the excluded range that falls below the VAT notch:

$$\hat{\Delta} s^* = \sum_{i=s^*_-}^{s^*} (c_j - \widehat{c}_j).$$

Here, c_j is the actual number of firms in each £100 turnover bin, and $\widehat{c}_j$ is the counterfactual bin counts without the notch. The range $\left(s^*_-, s^*_+\right)$ specifies turnover bins around the notch where bunching occurs and are therefore excluded from predicting the counterfactual distribution. Specifically, the lower bound of the excluded turnover region, s^*_-, is set at the point where excess bunching starts. The upper bound of the excluded region, s^*_+, is estimated in an iteration procedure to ensure that the excess mass below the VAT notch is equal to the missing mass above.[35] We then measure the amount of bunching by $b = \hat{\Delta} s^*/s^*$, where s^* is simply the VAT threshold for that year.

[35] We follow the standard procedure to estimate the counterfactual distribution. By grouping companies into small turnover bins of £100, we estimate the counterfactual distribution around the VAT notch s^* in

7.2 Bunching Evidence

7.2.1 Baseline Estimates

This section presents evidence of bunching below the VAT notch using the main sample of companies with turnover between £10,000 and £200,000. Figure 4 presents bunching around the threshold in each financial year between 2004/05 and 2009/10. Panel A shows the empirical distribution of turnover (blue dots) as a histogram in £1,000 bins and the estimated counterfactual distribution (red line) in 2004-05. Each dot denotes the upper bound of a given bin and represents the number of companies in each turnover bin of £1,000. Similar to Chetty et al. (2011) and Kleven and Waseem (2013), we estimate the counterfactual distribution by fitting a flexible polynomial of order 3 to the empirical distribution, excluding firms in the excluded range close to the VAT notch.[36] The lower bound of the excluded turnover range is demarcated by the vertical dashed line and the VAT notch demarcated by the vertical solid line.[37] The next five panels focus on subsequent years during which the VAT notch was increased annually to track inflation. Each panel reports the estimated bunching ratio (b) and its standard error in in parenthesis.

Three main findings are worth noting in Figure 4. First, the VAT notch creates evident bunching below the threshold. Excess bunching ranges from 0.82 to 1.29 times the height of the counterfactual distribution, and is strongly significant in all years during the sample period.[38] Second, excess bunching tracks precisely the annual change in the nominal VAT notch due to adjustment to inflation. In each year the excess bunching is concentrated within £2,000 below the VAT. Third, in contrast with the large bunching below the threshold, there is a small hole in the distribution above the VAT notch. The range of the hole spans from £8,500 to £15,000 above the cutoff, although we do not attempt to estimate the magnitude of optimization frictions implied by the hole given the various reason discussed in section 3.

the following regression:

$$c_j = \sum_{l=0}^{q} \beta_i (s_j)^l + \sum_{i=s_-^*}^{s_+^*} \gamma_i I\{j = i\} + \varepsilon_j, \tag{26}$$

where c_j is the number of companies in turnover bin j, s_j is the distance between turnover bin j and the VAT notch s^*, q is the order of the polynomial, and $I\{\cdot\}$ is an indicator function. The error term ε_j reflects misspecification of the density equation.

[36] As a robustness check we have tried values between 3 and 5 for the order of the polynomial and our results are not significantly changed.

[37] The upper bound of the excluded turnover region is estimated in an iteration procedure to ensure that the area under the estimated counterfactual density is equal the area under the observed density.

[38] Unlike to studies analysing bunching in the taxable income of individuals (Kleven and Waseem (2013)) and corporations (Devereux, Liu and Loretz (2014)), we do not find any evident bunching at round numbers. The absence of round-number bunching suggests that firms have less control over their turnover than reported taxable earnings.

In Kleven (2016), it is argued that in the context of the personal income tax, bunching is much more likely to be due evasion, rather than to real earnings responses. Here, as already noted, both evasion and real responses may be driving observed bunching. Our main approach is not to try to decompose observed bunching into a real and evasion part, but just to note that our theory, which includes an evasion element, makes predictions about how bunching will vary with the share of B2C sales and the input cost ratio, and we turn to this next. However, in Section 7.2.3 below, we do address the evasion issue, and show some evidence that bunching may partly be due to turnover misreporting.

7.2.2 Heterogeneity in Bunching

We have shown a stable distribution of turnover throughout the entire period 2004/05-2009/10, with an evident and persistent bunching of companies below the VAT notch in each year. We now explore potential heterogeneity in bunching to see whether the empirical pattern is consistent with the predictions set out in Proposition 3, that firms are more likely to bunch below the VAT notch if the share of B2C sales is high, or the share of input costs is low. We will also investigate the effect of the level of competition in the industry where the firm is located, although that is predicted to have an ambiguous effect theoretically.

First, we explore how companies with different B2C sales ratio respond to the same VAT notch by dividing companies into four groups according to the quartiles of B2C sales ratio. We estimate annual bunching ratios separately for each quartile, and plot the point estimate of the bunching ratio with the corresponding 95% confidence intervals in Figure 5. There are two interesting findings in Figure 5. First, all bunching estimates are positive and highly significant, even in the lowest B2C quartile where on average between 0.3% and 25.4% of sales are B2C. Second, there is a clear pattern that the estimated bunching ratio increases with quartiles of the B2C sales ratio. In particular, the estimated bunching ratio for firms in the top quartile is significantly larger than for firms in the bottom quartile. The observed strong aggregate bunching is mainly driven by the behavioral responses of companies in the 3rd and 4th quartile of the B2C sales ratio.

To explore how companies with different shares of direct input cost respond to the same VAT notch, we construct a firm-specific measure of average input-cost ratio during the sample period and divide all companies into four groups according to the quartiles of input-cost ratio. We obtain information on direct cost of sales *excluding* salary from company accounts in FAME and since it is optional for small and medium-sized companies to disclose this information, only 12.52% of companies in the estimation sample report a non-missing direct cost of sales. To increase efficiency of the empirical test, we pool observations with non-missing input cost in all years and present bunching evidence with respect to the normalized

VAT notch in Figure 6.

Panel A compares the empirical distributions of companies around the normalized VAT notch across quartiles of the input-cost ratio. It presents clear evidence that the degree of bunching decreases with the input-cost ratio. The distribution of companies in the top quartile is quite smooth around the normalized VAT notch, while distributions of companies in the lower quartiles all exhibit some degree of bunching just below the VAT notch. Panel B quantifies the difference in the extent of bunching by plotting the estimated bunching ratio with the corresponding 95% confidence interval for each input-cost ratio quartile. Quantitatively, the bunching estimate is very small and insignificant for companies in the top quartile of the input-cost ratio distribution. In contrast, the bunching estimates for companies in the lower quartiles of the input-cost ratio are positive and highly significant.

Finally, we examine the extent of bunching depending on the degree of competition in the product market. We measure competition at the industry level using the four-firm concentration ratio (CR4), so a high CR4 is associated with a lower level of competition. As in the previous cases, we examine how bunching varies across quartiles of the CR4 ratio in Figure 7. Panel A demonstrates that bunching clearly increases as the CR4 increases, whereas Panel B quantifies the difference in the extent of bunching by plotting the bunching estimates with the corresponding 95% confidence interval. All the bunching estimates are significantly different from zero, and there is an substantial increase in the degree of bunching at the third and fourth quartile of the CR4 ratio where there is less competition in the product market.

7.2.3 Bunching via Turnover Misreporting

In this section, we provide some suggestive evidence on the extent of bunching due to turnover misreporting. When bunching is due to a decrease in real output, we expect companies to reduce their input costs in proportion, so that the distribution of input-cost ratio for non-registered companies should be smooth around the VAT notch. When bunching is due to turnover misreporting, we conjecture that the non-registered companies are less likely to under-report their input costs and wage expenses. Both costs are deductible for corporation taxes and the latter is subject to third-party reporting. In other words, the gain from under-reporting the deductible costs is considerably smaller than the gain from under reporting the turnover to avoid VAT registration. If the majority of companies bunch via turnover misreporting, we would expect to see a higher average input-cost ratio for the non-registered group just below the VAT notch, relative to that for the registered group.

Figure 8 pools all observations in the sample period and plots the distribution of average input-cost ratio for registered and non-registered companies in £1,000 turnover bins, respectively. In Panel A, the input-cost ratio is salary exclusive and represents the share of direct cost of sales relative to total turnover. The solid blue line shows the average input cost relative to sales for registered companies within each turnover bin of £1,000 normalized by the current-year VAT notch, and the dashed blue line shows the average input cost ratio for the unregistered companies. Consistent with the theory, voluntary registers incur a much larger input cost as indicated by their average input-cost ratio which is consistently larger than that for the non-registered companies below the VAT notch. On the other hand, there is no evident increase in the average input-cost ratio just below the VAT notch for the non-registered group. The distribution is relatively smooth and continues to increase with turnover above the VAT notch.

In comparison, Panel B plots the distribution of average input-cost ratio *inclusive* of salary, for registered and non-registered companies, respectively. There is striking difference between the two input-cost ratio series just below the VAT notch. The two series move in parallel directions until the average input-cost ratio for the non-registered companies starts to increase drastically just below the VAT notch. The sharp increase in the salary-inclusive cost ratio can be partly attributed to the fixed nature of salary cost which takes longer to adjust than variable costs of input. On the other hand, the sharp increase is also consistent with the fact that salary is subject to third-party reporting and thus it is more costly/difficult for small businesses to underreport salary expenses. Overall, Panel A and B in Figure 8 provide suggestive yet not conclusive evidence that part of bunching is due to turnover misreporting.

8 Evidence on Dynamics

One limitation of our study so far is that we do not consider dynamic behavior of firms. Changing registration status involves some cost to the firm. This raises the possibility, in particular, that firms who were initially above the registration threshold and fall below may stay registered, simply because of the cost of deregistration. So, some of the firms who seem to be voluntary registered may just be behaving in this way because of inertia.

In this section, we investigate the importance of inertia in driving VAT registration by analyzing the dynamic behavior of firms when they cross registration and deregistration thresholds. First, we compute a transition probability matrix for firms moving between registration states in Table C.1, which shows the probability of being registered or not registered t years after initially being in a given state.[39] For example, the entry in the first cell of the matrix, 89.3%, shows that of all the firms that where initially registered in

[39] Changes in the transition probability could also driven by attrition, therefore we focus on a balanced sample of firms that we observe throughout the sample period.

2004/05, 89.3% remain registered a year later.

Table C.1 shows that there is considerable persistence in registration status. On the other hand, comparing to the registered firms, this persistence does decline substantially over time for non-registered firms. For example, only 88% of initially non-registered firms remain unregistered after 5 years, whereas 97% of registered firms remain registered after 5 years. The difference in persistence is due to the fact that the majority of firms are growing over our sample, and so will tend to stay above the registration threshold once they cross it.

Next, we investigate to what extent is the persistence in registration status driven by persistence in turnover versus the costs of changing registration status. To answer this question, we augment equation (5) as follows:

$$\begin{aligned} R_{it} &= \gamma_0 + \gamma_1 R_{i,t-1} + \gamma_2 (1 - R_{it-1}) IR_{it} + \gamma_3 R_{it-1} ID_{it} + \gamma_4 B2C_{j(i_t)} + \gamma_5 ICR_{it} \\ &\quad + \gamma_6 CR4_{it} + \gamma_7 \mathbf{X}_{it} + \rho_t + \phi_i + \nu_{it}, \end{aligned} \tag{27}$$

where R_{it} is a dummy indicator that takes value 1 if the firm is currently registered and zero otherwise, as defined previously in Section 6. In addition,

$$IR_{it} = \begin{cases} 1, & Y_{it} \geq Z_t \\ 0, & Y_{it} < Z_t \end{cases}, \quad ID_{it} = \begin{cases} 1, & Y_{it} \geq Z'_t \\ 0, & Y_{it} < Z'_t \end{cases},$$

where Z_t, Z'_t are the registration and deregistration thresholds at time t, Y_{it} denotes the current-period turnover, so IR_{it} and ID_{it} are dummy indicators recording whether the firm is above the registration and deregistration thresholds respectively at time t. All the other variables are defined as before, and υ_{it} is the error term. We estimate equation (27) in a fixed-effect Probit model, and augment the estimation equation with the initial registration status R_{i0} and the mean characteristics of all the time-varying regressors (Wooldridge, 2005).

So, if firm registration decision was entirely backward-looking and ignores its current turnover relative to Z_t, Z'_t, we would expect the coefficients γ_2 to γ_5 to be insignificant. However, we expect most firms to comply with the VAT law. Specifically, we expect firms to register when they are initially not registered and their turnover passes above the threshold. Such a firm has a value 1 for the term $(1 - R_{it-1})IR_{it}$ and so we expect to find a positive γ_2.

The VAT legislation also requires firms to stay registered if they are registered in the previous year and their current turnover is above the deregistration threshold. Such a firm will have a value of 1 for $R_{it-1}ID_{it}$. So, we also expect to find that $\gamma_3 > 0$. On the other hand, if the firm remains registered simply due to the cost of deregistration such that whether crossing the deregistration threshold plays no role in the registration behavior, we would expect to find that $\gamma_3 = 0$.

Finally, we already know from our analysis of voluntary registration that the registration decision is significantly affected by the industry B2C ratio $B2C_{j(i_t)}$, the firm input cost ratio ICR_{it}, and the industry-level four-firm concentration ratio $CR4_{j(i_t)}$, so we expect γ_4, γ_5 and γ_6 to be positive.

Table 6 summarizes the results from estimation of (27), by reporting the partial effects of the variables of interest using coefficient estimates from the fixed-effect Probit model.[40] Column (1) shows the mean predicted probability of VAT registration at fixed value of R_{t-1}, ID_t, IR_t, the mean predicted probability across all firms in the sample and that for one-standard-deviation increase in the $B2C$, ICR, and $CR4$ ratios. Column (2) shows the average partial effects of these variables by taking the difference in the mean predicted probabilities given the change in their value. For example, for firms that are registered in the previous year, falling below the deregistration threshold lowers the probability of being currently registered by 10.9 percentage points. Alternatively, for firms that are not registered in the previous year, going above the registration threshold increases their probability of registration by 19.7 percentage points. These findings suggest that the registration decision is not entirely driven by the cost of deregistration or inertia.

Finally, the short-run partial effects of the B2C ratio and input cost ratio in the dynamic model are considerably smaller than the static estimates in section 6. A one standard deviation increase in the B2C ratio and CR4 ratio reduces the probability of registration by 0.2 and 0.5 percentage points, respectively, where there is no change in the probability of registration for one standard deviation increase in the input cost ratio.

Overall, we see that while there is a considerable amount of persistence in firm behavior, the registration decision is not entirely driven by inertia due to fixed cost of deregistration. Firms respect the legal registration requirement, and at the same time change their registration decisions in a way that is consistent with profit maximization behavior depicted in Section 3. The probability of registration is also affected significantly by the more fundamental determinants of VAT registration. The positive coefficient estimates and partial effects of $R_{it-1}ID_{it}$, the B2C ratio and input cost ratios provides supportive evidence that

[40] The coefficient estimates are reported in the Appendix Table C.2. We calculate the estimated average partial effects (APEs) of the variables of interests following Wooldridge (2005), where the partial effect refers to the effect on the mean probability of registration after averaging the unobserved heterogeneity across all firms in the sample. For example, to calculate the APE of a discrete change of ID from 0 to 1, we first compute the average predicted probability of registration at fixed values of $R_{t-1} = 1, ID_t = 0$ and $R_{t-1} = 1, ID_t = 1$, respectively, across all firms in the sample. We then take the difference between the two average probabilities to obtain the average partial effect of ID. We use a similar procedure to compute the average partial effect of a one-standard-deviation increase in $B2C$, ICR, and $CR4$, noting that the one-standard-deviation increase applies to their mean characteristics in addition to the time-varying values.

the VAT registration decision is rational and relates to the fundamental determinants of VAT registration as suggested by the theory in Section 3.

9 Conclusions

In this paper, we first developed a conceptual framework which can explain the co-existence of VAT voluntary registration and bunching. We showed that this required that firms sell both to final consumers (B2C) and to other firms (B2B) for this to happen. This framework predicts that voluntary registration is more likely, and bunching is less likely, when either (i) the cost of inputs relative to sales is high, or (ii) when the proportion of B2C sales is low. Additionally, evasion opportunities will generally make voluntary registration more likely, while having an ambiguous effect on bunching. Also, the effect of product-market competition on both voluntary registration and bunching is generally ambiguous.

We then brought these predictions to an administrative data-set that is created by linking the population of corporation and VAT tax records in the UK, and showed that the pattern of voluntary registration in the data is consistent with the theory. In particular, voluntary registration is more likely with a low share of B2C sales high share of inputs in cost, and more competition in the industry. Moreover, there is clear evidence of bunching at the VAT threshold. Investigating further, we saw that, consistently with the theory, there is a clear pattern of heterogeneity in bunching; the amount of bunching is increasing in the B2C sales ratio, and decreasing in share of ratio of input costs to sales and the amount of competition in the industry.

We also investigated the dynamic behavior of firms around the VAT notch. Specifically, we addressed the concern that voluntary registration may not be an optimizing choice of firms, but simply a failure to deregister, once having been above the threshold and having registered. Our empirical findings suggest that while there is a considerable amount of persistence in firm behavior, perhaps due to the fixed compliance costs of deregistration, the decision is not entirely driven by inertia; firms change their registration decisions in a way that is consistent with profit maximizing behavior.

References

Almunia, Miguel, and David Lopez Rodriguez. 2014. "Heterogeneous Responses to Effective Tax Enforcement: Evidence from Spanish Firms." Banco de Espana Working Paper 1419.

Asatryan, Zareh, and Andreas Peichl. 2016. "Bunching and Non-bunching of Firms at VAT and Admin Thresholds." ZEW memeo.

Best, Michael, and Henrik Jacobsen Kleven. 2013. "Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK." London School of Economics Memeo.

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

Brashares, Edith, Matthew J. Knittel, Gerald Silverstein, and Alex Yuskavage. 2014. "Calculating the Optimal Small Business Exemption Threshold for a U.S. VAT." *National Tax Journal*, 67(2): 283 – 320.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *The Quarterly Journal of Economics*, 126(2): 749–804.

de Paula, ÃĄureo, and Jose A. Scheinkman. 2010. "Value-Added Taxes, Chain Effects, and Informality." *American Economic Journal: Macroeconomics*, 2(4): 195–221.

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

Ebrill, Liam P., Michael Keen, and Victoria P. Perry. 2001. *The Modern VAT.* International Monetary Fund.

Emran, M. Shahe, and Joseph E. Stiglitz. 2005. "On selective indirect tax reform in developing countries." *Journal of Public Economics*, 89(4): 599–623.

Federation of Small Businesses. 2010. *Impact of Increasing VAT Registration Threshold for Small Businesses.* Centre for Economics and Business Research.

Harju, Jarkko, Tuomas Matikka, and Timo Rauhanen. 2016. "The Effects of Size-Based Regulation on Small Firms: Evidence from Vat Threshold." VATT Institute for Economic Research Working Paper 75.

HM Revenue and Customs. 2015. *Measuring Tax Gaps.* HM Revenue and Customs.

Kanbur, Ravi, and Michael Keen. 2014. "Thresholds, Informality, and Partitions of Compliance." *International Tax and Public Finance*, 21(4): 536–559.

Keen, Michael, and Jack Mintz. 2004. "The Optimal Threshold for a Value-Added Tax." *Journal of Public Economics*, 88(3-4): 559–576.

Kleven, Henrik. 2016. "Bunching." *Annual Review of Economics*, 8.

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

Kopczuk, Wojciech, and David Munroe. 2015. "Mansion Tax: The Effect of Transfer Taxes on the Residential Real Estate Market." *American Economic Journal: Economic Policy*, 7(2): 214–57.

Liu, Li, and Ben Lockwood. 2015. "VAT Notches." CEPR Discussion Papers 10606.

Lockwood, Ben. 2016. "Malas Notches." University of Warwick Unpublished 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.

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

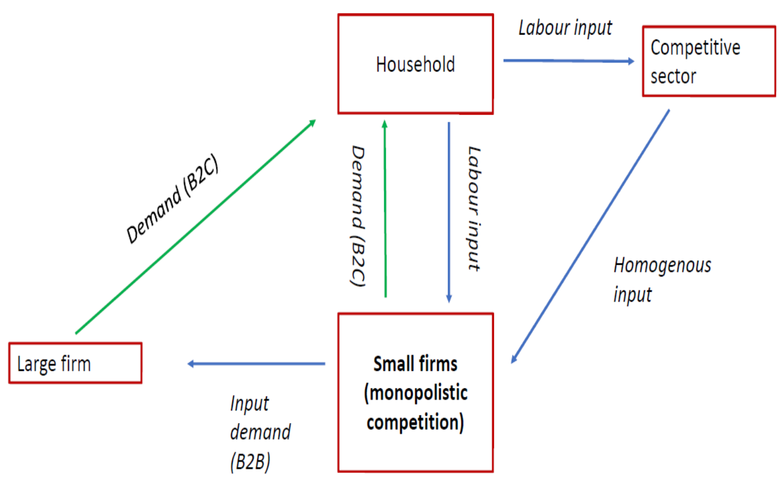
Slemrod, Joel. 2010. "Buenas Notches: Lines and Notches in Tax System Design." University of Michigan Working Paper.

Vesal, Mohammad. 2013. "Optimization Frictions in Choice of the UK Flat Rate Scheme of VAT." London School of Economics Memeo.

Waseem, Mazhar. 2016. "Using Movement of Exemption Cutoff to Estimate Tax Evasion: Evidence from Pakistan." University of Manchester Memeo.

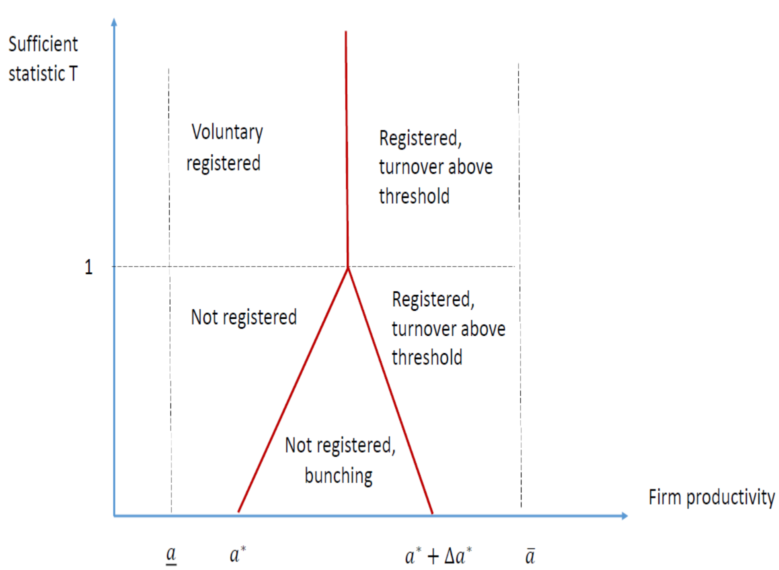
Wooldridge, Jeffrey M. 2005. "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics*, 20(1): 39–54.

Figure 1. STRUCTURE OF THE ECONOMY



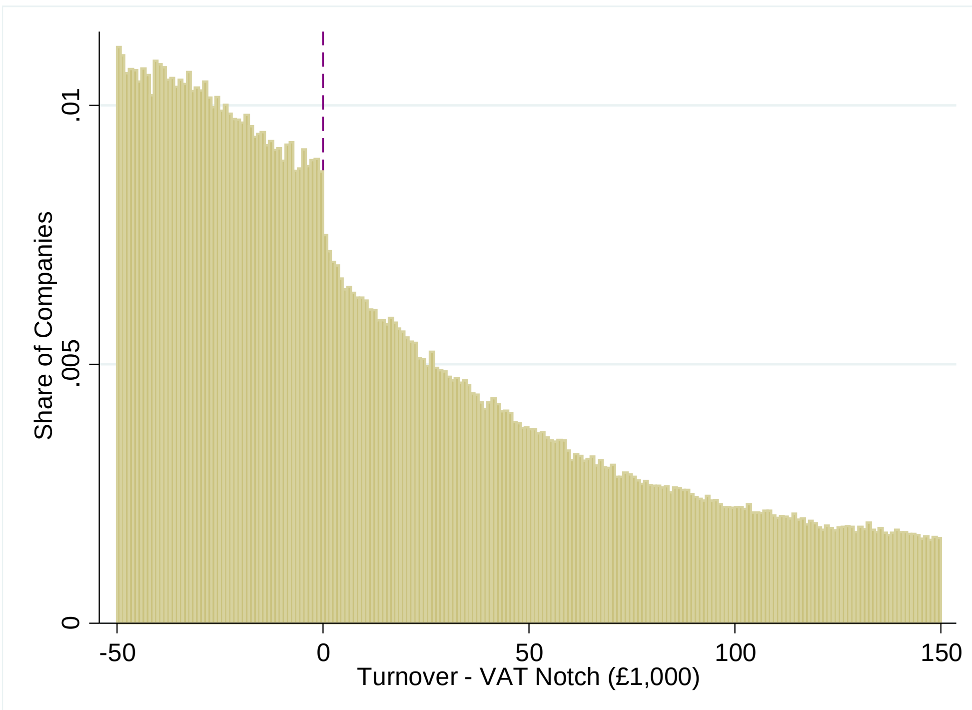
Notes: this figure shows the key features of the economy in which we model behaviourial responses of firms with respect to the VAT threshold, including voluntary registration and bunching.

Figure 2. SUMMARY OF THEORETICAL RESULTS



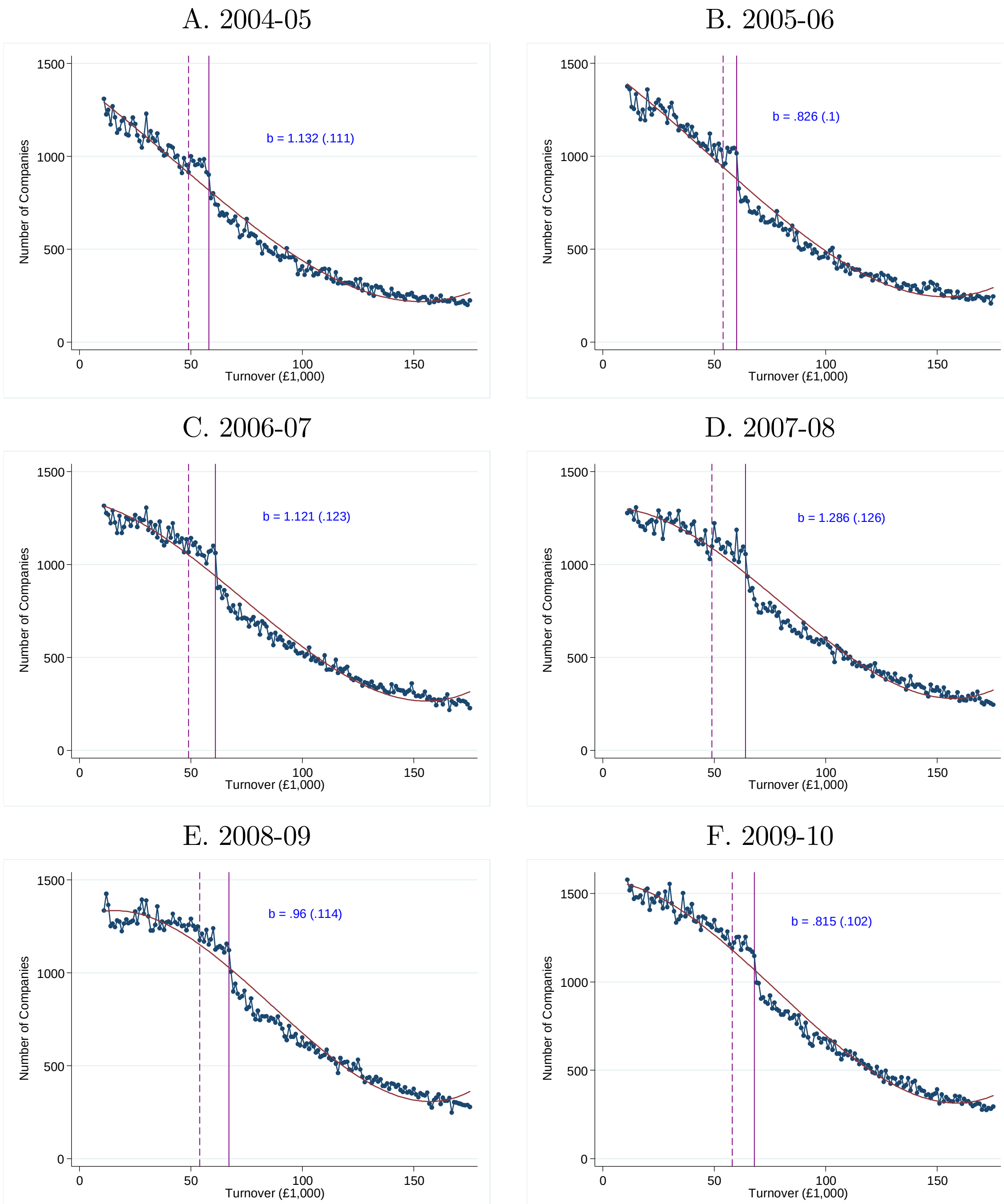
Notes: this figure shows how firms of productivity types a behave as the VAT sufficient statistic T varies.

Figure 3. A BINDING VAT NOTCH



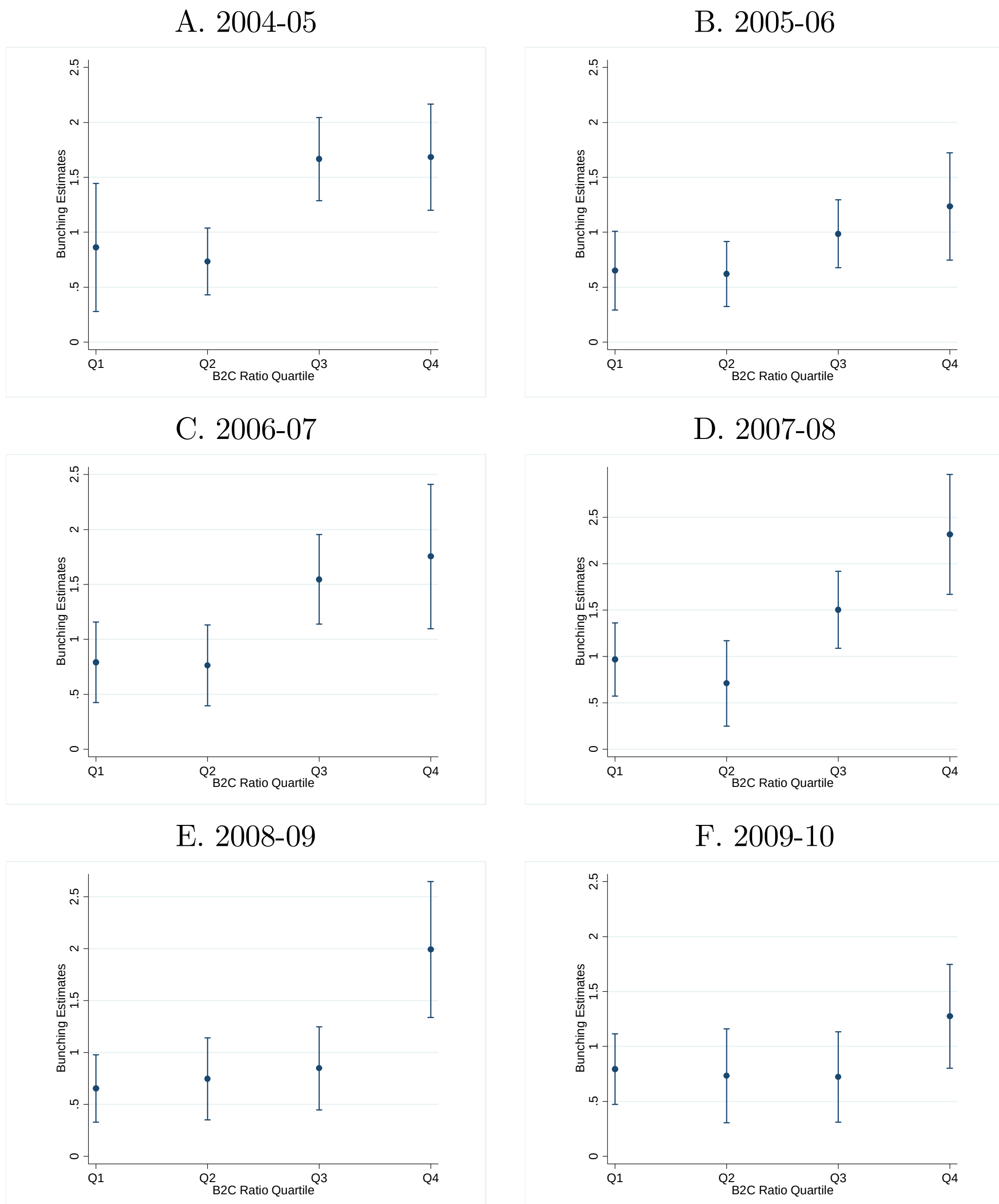
Notes: this figure shows the histogram of companies within the neighbourhood of turnover net of current-year VAT registration threshold (normalized VAT notch) by pooling data between 2004/05-2009/10. The bin width is £1,000 and the dashed line denotes the normalized VAT notch.

Figure 4. BUNCHING AT VAT NOTCH



Notes: this figure shows the observed distribution (solid-dotted line) and the estimated counterfactual distribution (solid-smooth line) of turnover for each year in 2004/05-2009/10. The counterfactual is a three-order polynomial estimated as in eq. (26). The excluded ranges around the VAT notch are demarcated by the vertical-dashed lines, and the VAT notch is demarcated by the vertical solid line. Bunching b is excess mass in the excluded range around the VAT notch relative to the average counterfactual frequency in this range. Standard errors are shown in parentheses.

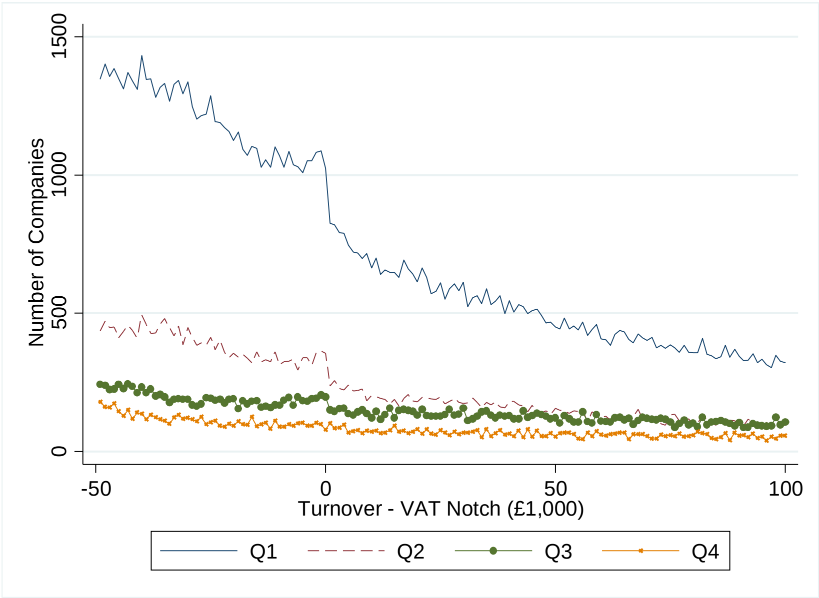
Figure 5. BUNCHING ACROSS B2C SALES RATIO QUARTILES



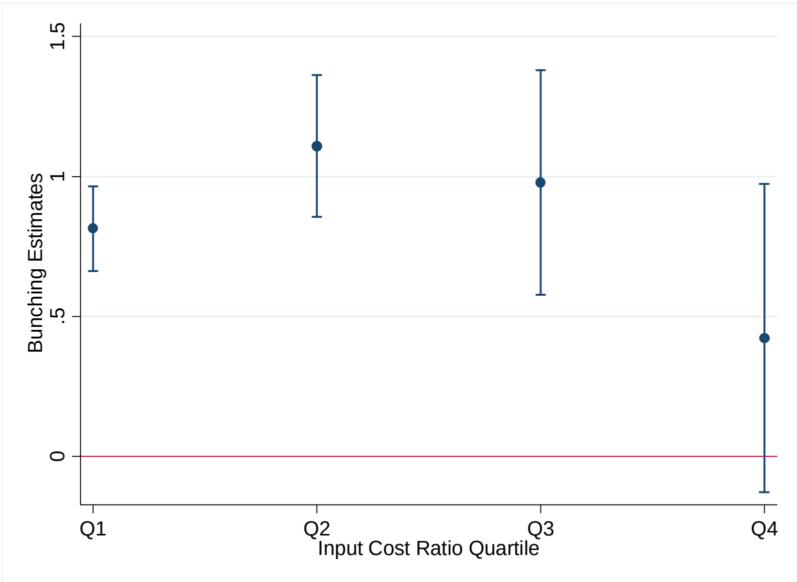
Notes: the figure plots the point estimate of the bunching ratio b with the corresponding 95% confidence intervals across four different quartiles of industry-level B2C sales ratio in each year during 2004/05-2009/10.

Figure 6. BUNCHING ACROSS INPUT-COST RATIO QUARTILES

A. Bunching Evidence



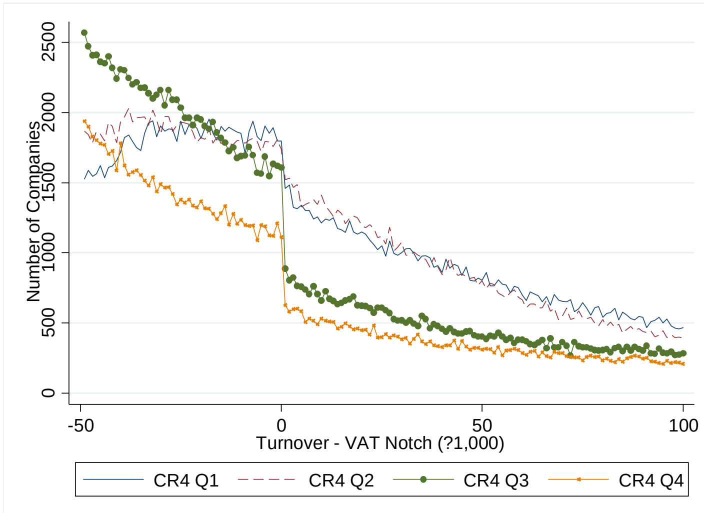
B. Bunching Estimates



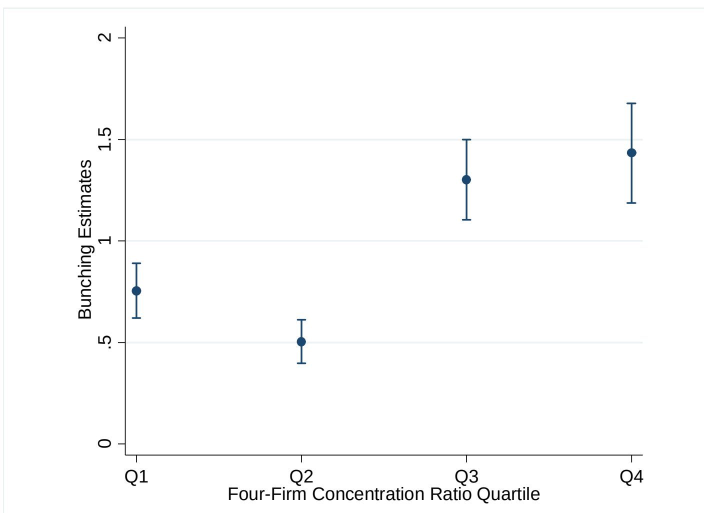
Notes: the figure shows the observed distribution of turnover across four different quartiles of input cost ratio within the neighbourhood of normalized VAT notch in 2004/05-2009/10 in Panel A. Panel B then plots the point estimate of the bunching ratio b and the corresponding 95% confidence intervals across the four quartiles of input cost ratio by pooling all the data in the sample years.

Figure 7. BUNCHING ACROSS FOUR-FIRM CONCENTRATION RATIO QUARTILES

A. Bunching Evidence



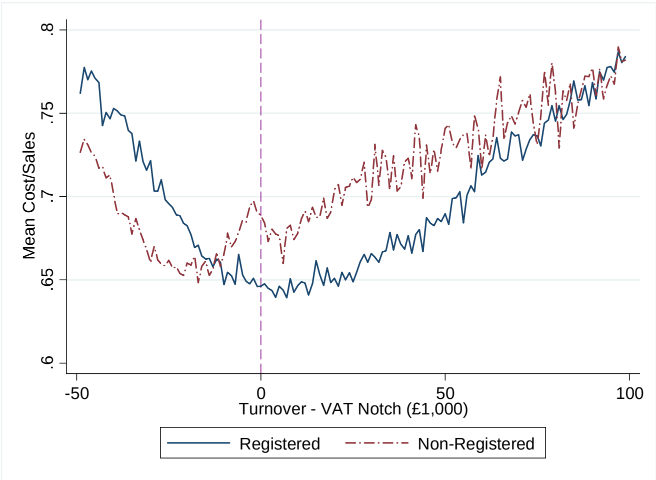
B. Bunching Estimates



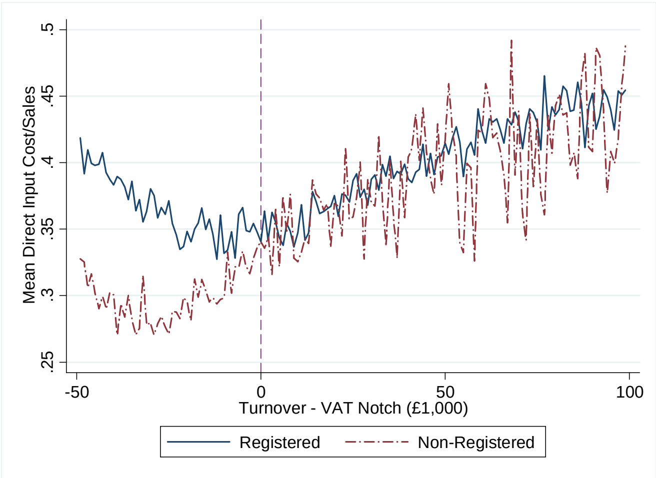
Notes: the figure shows the observed distribution of turnover across four different quartiles of four-frim concentration ratio (CR4) within the neighbourhood of normalized VAT notch in 2004/05-2009/10 in Panel A. Panel B then plots the point estimate of the bunching ratio b and the corresponding 95% confidence intervals across the four quartiles of CR4 ratio by pooling all the data in the sample years.

Figure 8. BUNCHING VIA TURNOVER MISREPORTING

A: Distribution of Direct Input-Cost Ratio



B: Distribution of Salary-Inclusive Cost Ratio



Notes: the figure plots separately the average input cost ratio for registered and non-registered firms with a turnover in the neighbourhood of normalized VAT notch during 2004/05-2009/10. Panel A uses the input cost ratio calculated from FAME and exclude the salary expenses. Panel B uses the input cost ratio calculated from the corporation tax records and includes salary expenses in the overall cost.

Table 1. EFFECTS OF COMPETITION ON VOLUNTARY REGISTRATION AND BUNCHING

e :	2	5	10	20	50
Panel A: the effect of Competition on the sufficient statistic T:					
Good Y zero-rated ($J = 0$)**:**					
$\lambda = 0.1$	0.98	1.17	1.65	3.55	39.53
$\lambda = 0.5$	0.82	0.76	0.78	1.31	13.77
$\lambda = 0.9$	0.76	0.58	0.40	0.29	2.01
Good Y taxed ($J = 1$)**:**					
$\lambda = 0.1$	0.95	1.00	0.92	0.47	0.02
$\lambda = 0.5$	0.80	0.65	0.43	0.16	0.01
$\lambda = 0.9$	0.76	0.57	0.34	0.13	0.01
Panel B: the effect of Competition on bunching $\Delta s^*/s^*$:					
Good Y zero-rated ($J = 0$)**:**					
$\lambda = 0.1$	0.16	0.00	0.00	0.00	0.00
$\lambda = 0.5$	0.74	1.32	1.30	0.00	0.00
$\lambda = 0.9$	0.95	2.59	5.88	10.17	0.00
Good Y taxed ($J = 1$)**:**					
$\lambda = 0.1$	0.27	0.00	0.58	4.66	212.75
$\lambda = 0.5$	0.80	2.02	5.25	23.31	302.75
$\lambda = 0.9$	0.97	2.75	7.43	32.63	302.75

Notes: Panel A shows how the value of the sufficient statistic T changes with respect to the values of e and λ, when good Y is zero-rated ($J = 0$ in the upper panel) and taxed ($J = 1$ in the lower panel), respectively. Panel B shows how the extent of bunching changes with respect to the level of competition, when good Y is zero-rated (upper panel) and taxed (lower panel), respectively.

Table 2. VALUE-ADDED TAX SCHEDULE IN THE UK

Fiscal Year	Registration Threshold (£)	Deregistration Threshold (£)	Standard Rate (%)	Flat-Rate Scheme Rate (%)
2001-02	54,000	52,000	17.5	100,000
2002-03	55,000	53,000	17.5	100,000
2003-04	56,000	54,000	17.5	150,000
2004-05	58,000	56,000	17.5	150,000
2005-06	60,000	58,000	17.5	150,000
2006-07	61,000	59,000	17.5	150,000
2007-08	64,000	62,000	17.5	150,000
Apr 1, 2008 - Nov 30, 2008	67,000	65,000	17.5	150,000
Dec 1, 2008 - Mar 30, 2009	67,000	65,000	15	150,000
Apr 1, 2009 - Jan 1, 2010	68,000	66,000	15	150,000
Jan 1, 2010 - Mar 30, 2010	68,000	66,000	17.5	150,000

Notes: the table shows changes in the registration threshold, deregistration threshold, Flat-Rate scheme threshold, and VAT standard rate over recent fiscal years. For more information on the UK VAT tax system, see http://www.hmrc.gov.uk/vat/forms-rates/rates/rates-thresholds.htm.

Table 3. SUMMARY STATISTICS

	Full Sample			**Registered**			**Non-Registered**			**Test of Equal Means**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**
Corporation Tax Variables											
Turnover (£000)	73.57	49.00	731,706	86.65	49.21	461,311	51.25	39.69	270,395	-336.38	0.00
Trading Profit (£000)	21.46	32.77	731,706	25.16	31.57	461,311	15.16	33.81	270,395	-125.11	0.00
Cost Variables											
Input Cost: CT600 (£000)	52.11	48.90	731,706	61.50	48.88	461,311	36.09	44.59	270,395	-226.92	0.00
Cost of Sales: FAME (£000)	31.08	33.79	82,420	36.44	35.47	54,327	20.72	27.45	28,093	-70.32	0.00
Other Variables											
Employment	7.84	198.76	10,592	8.63	234.53	6,181	6.73	133.40	4,411	-0.53	0.60
Firm Age	8.56	9.53	731,706	8.73	9.05	461,311	8.26	10.28	270,395	-19.45	0.00
Industry-level B2C Ratio (%)	0.45	0.28	694,530	0.44	0.27	449,156	0.46	0.28	245,374	18.14	0.00
Industry-level CR4 (%)	0.19	0.19	730,953	0.18	0.18	460,682	0.20	0.20	270,271	41.89	0.00

Notes: this table shows the mean, standard error, and number of non-missing observations for companies with turnover between £10,000 and £200,000 in the sample. All monetary values are in nominal £1,000 GBP, with 1 GBP = 1.42 USD as of April 2016. Columns (1)-(3) focus on the entire sample; columns (4)-(6) focus on the registered companies and columns (7)-(9) focus on the non-registered sample. Columns (10)-(11) show t-statistic and associated p-value from a test of equal means for each variable between the registered and non-registered sample.

Table 4. SHARE OF FIRMS THAT VOLUNTARILY REGISTERED FOR VAT (%)

	Input Cost Ratio Quartile			
B2C Sales Ratio Quartile	Q1	Q2	Q3	Q4
Q1	47.05	47.52	45.99	46.98
Q2	56.35	51.80	52.01	55.43
Q3	24.11	29.01	32.87	36.70
Q4	32.93	34.28	36.04	46.77

Notes: this table shows the share of voluntarily registered firms at different quartiles of B2C sales and input cost ratio. The share of voluntarily registered firms is calculated as the number of firms that are voluntarily registered for VAT relative to the total number of firms at each given quartile of B2C sales ratio and input cost ratio. Each column depicts the share of firms that are voluntarily registered for VAT at different quartiles of B2C sales ratio at a given input cost ratio quartile. Each row depicts the share of firms that are voluntarily registered for VAT at different quartiles of input cost ratio at a given B2C sales ratio quartile.

Table 5. DETERMINANTS OF VAT VOLUNTARY REGISTERATION

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of B2C Sales	-0.137***			-0.146***	0.004	0.004	-0.038**
	(0.003)			(0.003)	(0.004)	(0.004)	(0.016)
Input Cost Ratio		0.023***		0.047***	0.060**	0.069***	0.034***
		(0.003)		(0.003)	(0.002)	(0.003)	(0.009)
CR4 Ratio			-0.068***	-0.050***	0.002	0.002	-0.010*
			(0.004)	(0.004)	(0.002)	(0.002)	(0.005)
Firm FE	N	N	N	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
Firm-level Controls	N	N	N	N	N	Y	Y
R^2	0.006	0.001	0.001	0.007	0.003	0.007	0.004
N	478973	478973	473998	473,998	473,998	473,998	53,407

Notes: this table presents estimation results from the binary choice model of VAT registration based on equation (25). The dependent variable is the binary indicator of VAT registration status that takes on the value 1 if a firm is voluntarily registered for VAT and zero otherwise. Columns (1)-(4) present results from the linear probability model and columns (5)-(8) present results from the fixed-effects logit model. Additional firm-level controls include trading profits and company age. *,**,*** denotes significance at 10%, 5% and 1%, respectively. Standard errors presented in columns (1)-(4) are clustered at firm level.

Table 6. DETERMINANTS OF VAT VOLUNTARY REGISTERATION: AVERAGE PARTIAL EFFECTS

Evaluated at:		Mean $\widehat{\Pr(R_t = 1)}$ (1)	Average Partial Effect (2)
$R_{t-1} = 1$	$ID_t = 1$	0.926	
$R_{t-1} = 1$	$ID_t = 0$	0.817	0.109
$R_{t-1} = 0$	$IR_t = 1$	0.443	
$R_{t-1} = 0$	$IR_t = 0$	0.246	0.197
Average in the sample:		0.592	
$B2C + \sigma_{B2C}$		0.590	-0.002
$ICR + \sigma_{ICR}$		0.592	0
$CR4 + \sigma_{CR4}$		0.588	-0.005

Notes: this table presents the partial effects of the key variables of interest from the dynamic estimation of VAT registration in equation (27). Partial effects are based on coefficient estimates from fixed-effect probit model, which are reported in Table A 1.

A Appendix

Proof of Proposition 1. First, assume only B2C sales i.e. $\lambda = 1$. in this case, the firm chooses p_C, p_B to maximize (10) subject to (11),(12). This is easily solved to give price as a mark-up over cost i.e. $p_C = \frac{e_C}{e_C-1}c(I;a)$ and thus profit for non-registered and registered firms respectively is

$$\pi(0;a) = B_C(c(0,a))^{1-e_C},\ \pi(1;a) = B_C(c(1,a))^{1-e_C}\frac{1}{(1+t)^{e_C}} \tag{A.1}$$

where $B_C = A_C\left(\frac{e_C-1}{e_C}\right)^{e_C}\frac{1}{e_C-1}$. So, we only need show that $\pi(0;a) > \pi(1;a)$. But, from (A.1), (8), this holds iff

$$\left(\frac{c(0,a)}{c(1,a)}\right)^{1-e_C} = \left(\omega(1+t)^{1-\sigma} + (1-\omega)\right)^{(1-e_C)/(1-\sigma)} > \frac{1}{(1+t)^{e_C}}$$

Now note that as $e_C > 1$,

$$\left(\omega(1+t)^{1-\sigma} + (1-\omega)\right)^{(1-e_C)/(1-\sigma)} \geq (1+t)^{1-e_C} > (1+t)^{-e_C}$$

as required.

Next, assume only B2B sales i.e. $\lambda = 0$. Then, it is easily checked profit is $(p - c(I;a))A_Bp^{-e_B}$. So, here, the argument is even simpler; cost is decreasing, and therefore profit is increasing, if $I = 1$. □

Proof of Proposition 2. (a) If registered, the firm chooses p_C, p_B to maximize (10) subject to (11),(12) when $I = 1$. This is easily solved to give prices as a mark-up over cost i.e. $p_C = \frac{e_C}{e_C-1}c(I;a),\ p_B(a) = \frac{e_B}{e_B-1}c(I;a)$, and consequently, maximized profit of

$$\pi(1;a) = (\lambda B_C(1+t)^{-e_C}c(1;a)^{1-e_C} + (1-\lambda)B_Bc(1;a)^{1-e_B}) \tag{A.2}$$

where as before, $B_C = A_C\left(\frac{e_C-1}{e_C}\right)^{e_C}\frac{1}{e_C-1}$, and also $B_B = A_B\left(\frac{e_B-1}{e_B}\right)^{e_B}\frac{1}{e_B-1}$.

Next, if not registered, the firm chooses p_C, p_B to maximize (10) subject to (11),(12) when $I = 0$, subject also to the constraint $p_Cx + p_By \leq s^*$ that the total value of sales is below the threshold. Now consider voluntary registration. A necessary and sufficient condition for this is that profit with registration is greater than profit without, ignoring the constraint that sales be below the threshold. We will assume that a firm will register if indifferent, to break the tie. Following the case for the registered firm, we can show that the profit of an unregistered firm, ignoring the constraint that sales be below the threshold is

$$\pi(0;a) = (\lambda B_Cc(0;a)^{1-e_C} + (1-\lambda)B_Bc(0;a)^{1-e_B}) \tag{A.3}$$

Then, the voluntary registration condition is $\pi(1;a) \geq \pi(0;a)$. After some simple rearrangement, using (A.2),(A.3), this reduces to (16).

(b) We only need show that the RHS of (16) is increasing in λ and decreasing in ω. The statement for λ is obvious by inspection. For ω, note first that $1+\Delta_c \equiv (\omega(1+t)^{1-\sigma}+1-\omega)^{1/(1-\sigma)}$ is increasing in ω. So, both $(1+\Delta_c)^{1-e_C}$, $(1+\Delta_c)^{1-e_B}$ are decreasing in ω as $e_C, e_B > 1$. Finally, note that as $\Delta_c > 0$, $e_B > 1$, $1-(1+\Delta_c)^{1-e_B} > 0$. So, the RHS of (16) T is decreasing in ω as required.

(b) If $e_B = e_C = e$, and using $A_C = 1$, (16) rearranges to

$$\frac{\lambda(1+t)^{-e}+(1-\lambda)A_B}{\lambda+(1-\lambda)A_B}(1+\Delta_d)^{e-1} > 1 \tag{A.4}$$

But then

$$\frac{\lambda(1+t)^{-e}+(1-\lambda)A_B}{\lambda+(1-\lambda)A_B} = 1 - \frac{\lambda(1-(1+t)^{-e})}{\lambda+(1-\lambda)A_B} = 1-\Delta_d \tag{A.5}$$

Combining (A.4),(A.5) gives (17).

(c) Note that $\lim_{e\to\infty}\left(1-\frac{1}{e}\right)^e = 1/\exp \simeq 0.37$, and $\Delta_c > 0$, so if $J = 0$, it is clear that as $e \to \infty$, the LHS of (A.4) tends to infinity, so it always holds for e high enough. On the other hand, if $J = 1$, (A.4) becomes

$$\frac{\lambda+(1-\lambda)\left(\frac{e-1}{e}\right)^e}{\lambda(1+t)^e+(1-\lambda)\left(\frac{e-1}{e}\right)^e}(1+\Delta_c)^{e-1} > 1$$

which behaves for e large, like $\left(\frac{(\omega(1+t)^{1-\sigma}+1-\omega)^{1/(1-\sigma)}}{1+t}\right)^e$. But the number in the bracket is strictly less than $1+t$, independently of e, so in this case, $T \to 0$ as $e \to \infty$. □

Proof of Proposition 3. The indifference condition that determines $a^*+\Delta a^*$ is (20). To use this condition, we proceed as follows. To lighten notation, and using $e_C = e_B = e$, define

$$A_0 = \lambda+(1-\lambda)A_B, \;\; A_1 = \lambda(1+t)^{-e}+(1-\lambda)A_B$$

First, from (A.2), and using the restriction that $e_C = e_B = e$, a firm that registers has maximized profit:

$$\pi(1;a) = A_1\left(\frac{e}{1-e}\right)^{-e}\frac{a^{e-1}}{1-e} \tag{A.6}$$

Next, the profit from being just at the VAT threshold for an $a-$type when constrained is

$$\pi(0;a) = s^* - \frac{1+\Delta_c}{a}\frac{s^*}{p} \tag{A.7}$$

But solving for p from the constraint $A_0 p^{1-e} = s^*$, we get $p = \left(\frac{s^*}{A_0}\right)^{-1/(e-1)}$. Substituting this back into (A.7), and using $a = a^*+\Delta a^*$, we get

$$\pi(0;a^*+\Delta a^*) = s^* - \frac{1+\Delta_c}{A_0^{1/(e-1)}(a^*+\Delta a^*)}(s^*)^{e/(e-1)} \tag{A.8}$$

Also combining (21) and (A.8), we get:

$$\pi(0;a^*+\Delta a^*) = s^* - \frac{1+\Delta_c}{A_0^{1/(e-1)}(a^*+\Delta a^*)}(s^*)^{e/(e-1)} \tag{A.9}$$
$$= s^* - (s^*)^{e/(e-1)}\left(\frac{e-1}{e}\right)(s^*+\Delta s^*)^{1/(1-e)}$$

Now using (21) in (A.6), we get:

$$\pi(1;a^*+\Delta a^*) = A_1\left(\frac{e}{1-e}\right)^{-e}\frac{1}{1-e}(a^*+\Delta a^*)^{e-1} \tag{A.10}$$
$$= \frac{A_1}{A_0}\left(\frac{e}{1-e}\right)^{-e}\frac{1}{1-e}\left(\frac{e\,(1+\Delta_c)}{e-1}\right)^{e-1}(s^*+\Delta s^*)$$
$$= \frac{T}{e}(s^*+\Delta s^*)$$

So, using (A.10),(A.9), the indifference condition $\pi(1;a^*+\Delta a^*)=\pi_0(0;a^*+\Delta a^*)$ becomes

$$s^* - (s^*)^{e/(e-1)}\left(\frac{e-1}{e}\right)(s^*+\Delta s^*)^{1/(1-e)} - \frac{T}{e}(s^*+\Delta s^*) = 0 \tag{A.11}$$

After some simplification of (A.11) (divide through by s^*, then $1+\frac{\Delta s^*}{s^*}$, and multiply by e) we get (22) as required. □

Proof of Proposition 4. First, (22) can be rewritten as

$$f(x,e) - T(\lambda,\omega,e)= 0,\ f(x)\equiv ex-(e-1)x^{e/(e-1)},\ x =\frac{1}{(1+\Delta s^*/s^*)} \tag{A.12}$$

So, from (A.12):

$$\frac{dx}{d\lambda}=\frac{T_\lambda}{f_x},\ \frac{dx}{d\omega}=\frac{T_\omega}{f_x},\ \frac{dx}{de}=\frac{T_e-f_e}{f_x} \tag{A.13}$$

Moreover, note that

$$f_x = e(1-x^{1/(e-1)})>0 \tag{A.14}$$

because $x<1$, and $e>1$, so $x^{1/(e-1)}<1$. Also, we know that $T_\lambda<0,\ T_\omega>0$ and so from (A.13), (A.14), we conclude that $\frac{dx}{d\lambda}<0,\ \frac{dx}{d\omega}>0$. As x is an inverse measure of bunching, it follows that as λ increases, Δs^* rises, and ω rises, Δs^* falls. □

The Value-Added of A Registered Small Firm. If $\sigma=0$, a firm uses $\frac{\omega}{a}$ units of the input per unit of output. So, the value added of a registered "small" firm is the value of output minus the value of inputs used i.e.;

$$VA = p_C x + p_B y - \frac{\omega}{a}(x+y)$$

Now assuming A1 and $e_C=e_B$, and the mark-up expressions for prices, i.e. $p_C=p_B=p=$

$\frac{e}{e-1}\frac{1}{a}$, the firm's value-added is

$$\begin{aligned} VA &= \frac{1}{a}\left(\frac{e}{e-1}-\omega\right)(x+y) \\ &\propto \lambda(1+t)^{-e}+(1-\lambda)\left(\frac{e-1}{e}\right)^{e}(1+J.t)^{-e} \end{aligned}$$

So, if $J=1$, the elasticity of VA with respect to $1+t$ is e, as claimed.
Calibration of ω. The average share of costs that is non-labour is

$$ICR = \frac{\omega(1+\phi t)}{\omega(1+\phi t)+1-\omega}$$

where ϕ the fraction of overall sales that are by registered firms. Now $\phi = 0.74$, and $ICR = 0.496$, and $t = 0.175$, so

$$0.496 = \frac{1.123\omega}{1.123\omega+1-\omega}$$

which implies $\omega = 0.419$.

B Not-For-Publication Appendix

General Proof that Voluntary Registration is not Possible with Only B2C Sales. Consider a firm facing a residual demand function from final consumers of $x(q)$, where q is the consumer price. This covers both the cases of monopoly, where $x(.)$ is also the actual demand curve, and monopolistic competition, where $x(.)$ is demand for that firm's product, taking the prices of all other firms as fixed. Assume all sales are to final consumers i.e. B2C. If a firm is registered for VAT, profit is then

$$\pi_R(p) = px(p(1+t)) - c(x(p(1+t)), w, r)$$

where p is the producer price, and $c(x,\ w, r)$ is the cost function given output x, and prices of labour and the intermediate input w, r. This is completely general cost function that includes the constant returns CES cost function in the paper as a special case. If the firm is not registered for VAT, profit is

$$\pi_N(p) = px(p) - c(x(p), w, r(1+t))$$

Then, we have

$$\begin{aligned}
\pi_R &= \max_p \{px(p(1+t)) - c(x(p(1+t)), w, r)\} \\
&= \max_q \left\{ \frac{q}{1+t} x(q) - c(x(q), w, r) \right\} \\
&= \frac{1}{1+t} \max_q \{qx(q) - (1+t)\, c(x(q), w, r)\} \\
&< \max_q \{(q - c(x(q), w, r(1+t)))\} \\
&= \pi_N
\end{aligned}$$

So, with only B2C sales, no firm would every wish to register voluntarily.

The Competitive Economy. here, we present a competitive variant of the model and show (i) that without segmented markets i.e. different markets for B2B and B2C transactions, the equilibrium is trivial, with all firms registering for VAT; (ii) with segmented markets, there is generically complete sorting, with registered (non-registered) firms only selling B2B (resp. B2C). Moreover, in the second case, the equilibrium actually has a more complex structure,being defined by three equations, than in the monopolistic case, which makes generating empirical predictions more difficult.

Consider a "competitive" version of the model in the paper, where the small firms sell a homogenous product, with the model unchanged in all other respects. In particular, the small firms face fixed prices p_B, p_C in the B2B and B2C markets respectively. Moreover, B2B and B2C demand for the product of the small firms is now given by the perfect competition analogues of (4) and (13) i.e. $\lambda\,(p_C)^{-\phi}$, $(1-\lambda)\left(\frac{\gamma}{1-\gamma}\right)^{-\gamma}(p_B)^{-\gamma}$. So, the revenue per unit sold by registered and non-registered firms is as in the Table below:

Table B1: Revenue Per Unit

	registered firms	non-registered firms
B2C	$\frac{p_C}{1+t}$	p_C
B2B	p_B	p_B

This is because a registered firm perceives perfectly elastic demand at consumer price p_C in the B2C market, and so must bear all the burden of output VAT. The next steps to specify firms' cost and profit. With perfect competition, firm's cost function need to be strictly convex in order for firm scale to be well-defined. Without much loss of generality, and for easy comparison to the main model in the paper, we assume that the firm production functions are CES and homogeneous of degree $1/\delta$, $\delta > 1$, so the firm' cost function is $c(I(a);a)y^\delta$, where $c(I(a);a)$ is defined in (8).

Consider first the case where markets are not segmented; then arbitrage by buyers implies $p_C = p_B = p$. But then if a firm registers, by Table A1, it will prefer to sell B2B, and can thus get maximum profit

$$\pi_R(a) = \max_y \left\{ py - c(1;a)y^\delta \right\}$$

If a firm does not register, it is indifferent to whom it sells, and makes profit

$$\pi_N(a) = \max_y \left\{ py - c(0;a)y^\delta \text{ s.t. } py \leq s^* \right\}$$

where as in the main paper, s^* is the VAT threshold. So, clearly, as $c(0;a) > c(1;a)$, all firms will wish to register, as claimed. [It is not clear that this is an equilibrium, however, as final consumers will receive no supply, and they are willing to pay for supply].

If markets are segmented, then (ignoring non-generic cases) for an equilibrium to exist, by the above table the two prices must satisfy:

$$\frac{p_C}{1+t} < p_B < p_C \tag{B.1}$$

For any other price configuration, both types of firms would not wish to supply either final consumers or businesses, and that could not be an equilibrium, because from iso-elastic demand, the buyer without supply would be willing to pay an arbitrarily high price for the good. Note that for this price configuration, there is complete sorting *i.e.registered firms only sell B2B, and non-registered firms only sell B2C*[41].

But then, using (B.1), the profits to registering and not are

$$\pi_R(a) = \max_y \left\{ p_B y - c(1;a)y^\delta \right\}, \ \pi_N(a) = \max_y \left\{ p_C y - c(0;a)y^\delta \text{ s.t. } p_C y \leq s^* \right\}$$

If firms differ in the input cost ratio, there may be both voluntary registration and bunching, as in the main model. To see this, first note that a firm of type a will register voluntarily

[41] In can be shown that in the model of the main paper, there is partial sorting i.e. registered firms sell relatively more B2B than non-registered firms do, and that if $e_C = e_B = e$, this sorting effect is stronger, the higher is e i.e. the more competitive is the market for the differentiated good. Moreover, in the competitive limit as $e \to \infty$, there is complete sorting. A proof of this available on request.

iff $\pi_R(a)$ is greater than the payoff from not registering, ignoring the constraint $p_C y \leq s^*$. After performing the maximisations, the condition reduces to

$$(p_B)^{\delta/(\delta-1)}(c(1;a))^{-1/(\delta-1)} \geq (p_C)^{\delta/(\delta-1)}(c(0;a))^{-1/(\delta-1)}$$

and finally to a condition independent of a;

$$\frac{p_B}{p_C} \geq \left(\frac{c(1;a)}{c(0;a)}\right)^{1/\delta} \tag{B.2}$$

This condition says that voluntary registration will occur for those firms where the loss of revenue from a lower price, measured inversely by p_B/p_C, is less than the cost gain from being registered, measured inversely by $c(1;a)/c(0;a)$. On the face it it, this is a simpler condition than (17) in the paper. In particular, it is clear that conditional on $\frac{p_B}{p_C}$, the higher the input cost ratio, the more likely this is to be satisfied. However, recall that p_B, p_C are endogenous and remain to be determined.

Assume now that there are some firms with an input -cost ratio such that (B.2) is not satisfied. Then, by an argument as in the main paper, some of these firms will bunch i.e. there will be a critical $\hat{a}$ such that $\pi_R(\hat{a}) = \pi_N(\hat{a})$, so that only firms above $\hat{a}$ will register. This can be manipulated into a condition similar to (22). Finally, the two prices are determined in equilibrium by the two conditions

$$\lambda\,(p_C)^{-\phi} = \int_{\underline{a}}^{\hat{a}} y_N(p_C;a)da, \quad (1-\lambda)\left(\frac{\gamma}{1-\gamma}\right)^{-\gamma}((1+Jt)p_B)^{-\gamma} = \int_{\hat{a}}^{\overline{a}} y_R(p_B;a)da \tag{B.3}$$

where the supply functions are defined as;

$$\begin{aligned} y_R(p_B;a) &= \arg\max_y \left\{p_B y - c(1;a)y^\delta\right\} \\ y_N(p_C;a) &= \arg\max_y \left\{p_C y - c(0;a)y^\delta \text{ s.t. } p_C y \leq s^*\right\} \end{aligned}$$

So, in equilibrium, $\hat{a}$ and the prices (p_B, p_C) are determined by three simultaneous conditions $\pi_R(\hat{a}) = \pi_N(\hat{a})$, and (B.3). The effects of changes in ω and λ on voluntary registration and bunching could no doubt be calculated, but this may be a complex exercise.

Derivation of the Bunching Equation with Evasion. The proof follows the proof of Proposition 4 in the paper, with the following changes. First, we define $A_0,\ A_1$ as

$$A_0 = \lambda(1-\nu) + (1-\lambda)A_B,\ A_1 = (\lambda(1+(1-\nu)t)^{-e} + (1-\lambda)A_B)$$

As in (A.2) in the paper, any firm that registers has maximized profit:

$$\pi(1;a) = A_1\left(\frac{e}{1-e}\right)^{-e}\frac{a^{e-1}}{1-e} \tag{B.4}$$

Next, the payoff from being on the VAT threshold for an $a-$type when constrained is now

$$\pi(0;a) = (s^* + \nu px) - \frac{1+\Delta_c}{a}\frac{(s^* + \nu px)}{p} \tag{B.5}$$

This is because the firm can actually produce and sell $s^* + \nu px$ with a threshold s^* because sales νpx are "cash" and thus not observable by the tax authority. Solving for p from the definition that non-concealed sales must be equal to s^* i.e. $((1-\nu)\lambda + (1-\lambda)A_B)p^{1-e} = s^*$, we get:

$$p = \left(\frac{s^*}{(1-\nu)\lambda + (1-\lambda)A_B}\right)^{-1/(e-1)}$$

Combining this with the fact that $x = \lambda p^{1-e}$, we get

$$s^* + \nu px = s^* \frac{\lambda + (1-\lambda)A_B}{(1-\nu)\lambda + (1-\lambda)A_B} \equiv \mu s^*$$

Substituting this back into (B.5), and setting $a = a^* + \Delta a^*$, we get

$$\pi(0;a^* + \Delta a^*) = \mu\left(s^* - \frac{1+\Delta_c}{A_0^{1/(e-1)}(a^* + \Delta a^*)}(s^*)^{e/(e-1)}\right) \tag{B.6}$$

Also observed non-cash sales map into type by

$$s^* + \Delta s^* = A_0\left(\frac{e\,(1+\Delta_c)}{e-1}\right)^{1-e}(a^* + \Delta a^*)^{e-1},\ A_0 = \lambda(1-\nu) + (1-\lambda)A_B \tag{B.7}$$

Combining (B.7) and (B.6), we get:

$$\begin{aligned}
\pi(0;a^* + \Delta a^*) &= \mu\left(s^* - \frac{1+\Delta_c}{A_0^{1/(e-1)}(a^* + \Delta a^*)}(s^*)^{e/(e-1)}\right) \\
&= \mu\left(s^* - \frac{(1+\Delta_c)\,(s^*)^{e/(e-1)}}{A_0^{1/(e-1)}}(A_0)^{1/(e-1)}\left(\frac{e-1}{ec_0}\right)(s^* + \Delta s^*)^{1/(1-e)}\right) \\
&= \mu\left(s^* - (s^*)^{e/(e-1)}\left(\frac{e-1}{e}\right)(s^* + \Delta s^*)^{1/(1-e)}\right)
\end{aligned} \tag{B.8}$$

Now using (B.7) in (B.4), we get:

$$\begin{aligned}
\pi(1;a^*+\Delta a^*) &= A_1\left(\frac{e}{1-e}\right)^{-e}\frac{1}{1-e}(a^*+\Delta a^*)^{e-1} \\
&= \frac{A_1}{A_0}\left(\frac{e}{1-e}\right)^{-e}\frac{1}{1-e}\left(\frac{ec_0}{e-1}\right)^{e-1}(s^*+\Delta s^*) \\
&= \frac{A_1(1+\Delta_c)^{e-1}}{eA_0}(s^*+\Delta s^*) \\
&= T(\nu)\frac{s^*+\Delta s^*}{e}
\end{aligned} \tag{B.9}$$

So, using (B.8),(B.9), the indifference condition $\pi(1;a^*+\Delta a^*) = \pi_0(0;a^*+\Delta a^*)$ becomes

$$s^* - (s^*)^{e/(e-1)}\left(\frac{e-1}{e}\right)(s^*+\Delta s^*)^{1/(1-e)} - \frac{T(\nu)}{\mu e}(s^*+\Delta s^*) = 0 \tag{B.10}$$

After some simplification of (B.10) (divide through by s^*, then $1+\frac{\Delta s^*}{s^*}$, and multiply by e) we get (22) in the paper with the tax term $\frac{T(\nu)}{\mu} = T(\nu)\frac{\lambda+(1-\lambda)A_B}{(1-\nu)\lambda+(1-\lambda)A_B}$ as required. □

C Supplementary Tables

Table C.1. TRANSITION MATRIX OF VAT REGISTRATION STATUS

	$R_t = 1$ $ID_t = 1$ (1)	$R_t = 1$ $ID_t = 0$ (2)	$R_t = 0$ $IR_t = 1$ (3)	$R_t = 0$ $IR_t = 0$ (4)
$t = 1$				
$R_0 = 1, ID_0 = 1$	89.34%	10.61%	0.03%	0.02%
$R_0 = 1, ID_0 = 0$	20.50%	79.30%	0.01%	0.19%
$R_0 = 0, IR_0 = 1$	4.18%	1.14%	78.46%	16.22%
$R_0 = 0, IR_0 = 0$	1.45%	1.59%	6.15%	90.80%
$t = 2$				
$R_0 = 1, ID_0 = 1$	85.35%	14.45%	0.07%	0.12%
$R_0 = 1, ID_0 = 0$	25.49%	73.55%	0.03%	0.93%
$R_0 = 0, IR_0 = 1$	6.30%	1.63%	71.25%	20.82%
$R_0 = 0, IR_0 = 0$	3.23%	2.71%	7.22%	86.83%
$t = 3$				
$R_0 = 1, ID_0 = 1$	81.30%	18.03%	0.13%	0.54%
$R_0 = 1, ID_0 = 0$	26.71%	71.35%	0.07%	1.88%
$R_0 = 0, IR_0 = 1$	8.03%	2.19%	65.21%	24.58%
$R_0 = 0, IR_0 = 0$	4.85%	4.02%	6.90%	84.23%
$t = 4$				
$R_0 = 1, ID_0 = 1$	76.14%	21.71%	0.55%	1.60%
$R_0 = 1, ID_0 = 0$	26.34%	68.85%	0.32%	4.48%
$R_0 = 0, IR_0 = 1$	7.67%	3.13%	59.76%	29.44%
$R_0 = 0, IR_0 = 0$	5.65%	4.82%	5.99%	83.55%
$t = 5$				
$R_0 = 1, ID_0 = 1$	67.61%	29.57%	0.32%	2.50%
$R_0 = 1, ID_0 = 0$	22.50%	72.21%	0.18%	5.11%
$R_0 = 0, IR_0 = 1$	7.34%	4.60%	52.71%	35.35%
$R_0 = 0, IR_0 = 0$	5.37%	6.14%	5.34%	83.15%

Notes: this table shows in each cell the probability of changing from registration status in year t to year $t + 1$.

Table C.2. DETERMINANTS OF VAT VOLUNTARY REGISTERATION: COEFFICIENT ESTIMATES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
R_{t-1}	2.963***	2.952***	2.719***	3.256***	2.953***	2.958***	2.948***	2.957***	2.955***
	(0.027)	(0.027)	(0.029)	(0.025)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
R_0	2.282***	2.277***	2.127***	2.100***	2.142***	2.120***	2.161***	2.117***	2.125***
	(0.079)	(0.076)	(0.074)	(0.069)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)
Trading profit		0.019***	0.016***	0.012***	0.008***	0.008***	0.011***	0.008***	0.011***
		(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Average profit		0.001*	0.002***	0.001***	0.002***	0.002***	-0.001*	0.002***	-0.002*
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$R_{t-1} \times ID_t$			0.812***		0.937***	0.953***	0.941***	0.933***	0.944***
			(0.022)		(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
$(1 - R_{t-1}) \times IR_t$				0.874***	0.976***	0.988***	0.980***	0.979***	0.983***
				(0.019)	(0.021)	(0.021)	(0.022)	(0.021)	(0.022)
B2C Ratio						0.003***			0.003***
						(0.001)			(0.001)
Average B2C ratio						-0.005***			-0.005***
						(0.001)			(0.001)
Input Cost Ratio							0.004***		0.004***
							(0.001)		(0.001)
Average Input Cost Ratio							-0.005***		-0.004***
							(0.001)		(0.001)
CR4 Ratio								0.002***	0.002***
								(0.000)	(0.000)
Average CR4 Ratio								-0.007***	-0.006***
								(0.001)	(0.001)
Constant	-2.261***	-2.558***	-2.457***	-2.657***	-2.626***	-2.520***	-2.555***	-2.516***	-2.432***
	(0.027)	(0.031)	(0.030)	(0.031)	(0.031)	(0.032)	(0.049)	(0.031)	(0.049)
$ln\sigma_u$	-0.067	-0.100*	-0.221***	-0.192***	-0.176***	-0.192***	-0.169***	-0.197***	-0.195***
	(0.055)	(0.054)	(0.057)	(0.054)	(0.053)	(0.054)	(0.053)	(0.054)	(0.054)
N	480,025	480,025	480,025	480,025	480,025	480,025	480,025	480,025	480,025

Notes: this table presents the coefficient estimates from the dynamic estimation of VAT registration in equation (27) in a fixed-effect probit model.