VAT Notches*

Li Liu and Ben Lockwood†

Preliminary incomplete draft, please do not cite

2nd December 2014

Abstract

We develop a conceptual framework which captures the effect of the VAT system on profit by two effective taxes . This allows (i) predictions of the determinants of voluntary registration and bunching at the registration threshold; (ii) develops a formula for estimating the elasticity of value-added with respect to the statutory tax. We also show that the marginal excess burden of the tax is measured by this elasticity, extending Feldstein's analysis of the elasticity of taxable income to an indirect tax setting. 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, voluntary registration is positively related to the intensity of input use and negatively related to the share of B2C transactions. There is bunching at the VAT threshold, and the amount of bunching is negatively related to the intensity of input use and positively related to the share of B2C transactions, again consistently with the theory. Our estimate of the elasticity of the VAT tax base is XXX.

*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. We acknowledge financial support from the ESRC, under grant ES/L000016/1.

†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).

1 Introduction

Most countries around the world use the value-added tax (VAT) as their primary indirect tax, and most countries 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 has not been analyzed in the literature.[3]

There is a recent literature on income tax notches (Kleven and Waseem (2013)), and transactions tax notches in the housing market (Best and Kleven (2013) and Kopczuk and Munroe (2014)). These papers 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. They use bunching at notches to estimate both the elasticity of labour supply, and the degree of optimization friction.

However, the conceptual framework developed in these papers is not directly applicable to VAT, for several reasons. First, with VAT, unlike the personal income tax, the *effective* rate of VAT paid on the marginal unit of value-added is determined not just by the tax code, but also by firm characteristics[4]. First, even firms not registered for VAT pay a positive effective VAT rate, because they cannot recover tax paid on intermediate inputs. Second, if a firm registered for VAT sells to another registered firm, it will automatically simply pass on any change in the VAT charged on its outputs, because the buyer can claim the output VAT back. So, firms that have mostly so-called "B2B" sales have a lower effective tax.

Both these characteristics clearly differ widely across small firms that may be close to the registration threshold. For example, a small tradesperson such as a plumber or electrician may typically have "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, and make heavy use of intermediate inputs, and so will be in the reverse position.

Second, these different characteristics give rise to the important feature of *voluntary*

[1] In the EU, all but two countries (Spain and Sweden) currently have positive thresholds, with the UK threshold being the largest at £81000. The thresholds in the EU are very low compared with those in countries that have more recently introduced a VAT, such as Singapore, which currently has a threshold of about 540000 Euro. (Information on thresholds is from http://www.vatlive.com).

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

[3] See Slemrod (2010) for a general discussion of tax notches; the VAT registration threshold is an example of a quantity notch, in his terminology, which is relatively rare.

[4] In this respect, it is like the corporate tax, where it well-known that the effective marginal and average rates of tax depend on the characteristics of the investments firms make.

registration, where a firm registers for VAT even if it is below the turnover threshold, and thus not required to do so. This occurs when a firm has large purchases of intermediate inputs, and/or they can pass most of VAT on output onto the purchaser, as in the case of the car component firm; then, it may be profitable to *voluntarily* register for VAT so they can claim back input tax. In our data-set, over 35% of the firms with turnover below the threshold register voluntarily. Voluntary registration makes the VAT unique amongst all major taxes and thus is worthy of investigation.

In this paper, we first develop a conceptual framework for studying VAT voluntary registration and bunching. This framework is designed to be comparable to the framework first developed by Saez (2010) to study bunching at tax kinks, while capturing the distinctive features of VAT just mentioned. We consider a number of firms producing a homogenous product from a purchased input and the labor or managerial input of the firm's owner. These firms can vary in efficiency (the basic source of heterogeneity that is the analog of labour productivity in Saez (2010)), and also in the intensity with which they use the input, and the proportion of sales to non-VAT registered consumers, i.e. so-called B2C sales.

We show first in this setting that the effect of the VAT system on profit can be captured by a sufficient statistic, which we call the effective VAT rate, which combines the effects of both input and output VAT; this rate will be different for registered and non-registered firms. We then show that voluntary registration is more likely when either (i) the cost of inputs relative to sales is high, or (ii) when the proportion of B2C sales is low. The intuition for (ii) is simply that if most customers are VAT-registered, the burden of an increase VAT can easily be passed on the form of a higher price, because the customer himself 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. We also show that the elasticity of value-added of registered firms with respect to the effective VAT rate can be recovered from a formula very similar to that of Kleven and Waseem (2013). Finally, we show, that the elasticity of value-added can be related in a simple way to the deadweight loss of a small increase in the statutory rate of VAT, thus extending the well-known results of Feldstein (1999) and Chetty (2009) to an indirect tax setting. Ultimately, therefore, we can make an estimate of the deadweight loss of VAT for the UK.

We then bring these predictions to an administrative data-set that is created by linking the population of corporation and VAT tax records in the UK. 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 of inputs in cost. Quantitatively, the probability that a firm voluntarily registers for VAT is increased by 0.05 for a one standard deviation increase in the share of B2C sales and by 0.02-0.05 for a one standard deviation increase in the input cost ratio. The results are robust to use of linear probability model and fixed-effects logit model and 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. This is the first evidence, to our knowledge, that a VAT notch leads to bunching. Investigating further, we find that firms are less likely to "bunch" i.e. more likely to register voluntarily, 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. We also show, again consistently with the theory, that among firms who bunch, the amount of bunching is increasing in the B2C sales ratio, and decreasing in share of ratio of input costs to sales. So, there is a clear pattern of heterogeneity in bunching.

The next question is *how* it is that firms bunch; that is, what are the mechanism(s) at work? One possibility is that they genuinely restrict their sales to stay below the threshold. If so, the distribution of input-cost ratio should be smooth around the VAT notch. We provide some suggestive evidence that part of bunching is driven by under-reporting of sales. Specifically, we find that the salary-inclusive input cost ratio moves in the parallel direction between the registered and non-registered group outside the bunching region but starts to increase substantially for the non-registered companies just below the threshold. We interpret the large and sharp increase in the salary-inclusive input cost ratio to be partly driven by the fact that it is costly to underreport salary expenses due to third-party reporting.

Finally, we address the issue of the elasticity of the tax base. [TO BE COMPLETED]

Our work relates to the literature on the effect of tax and regulatory thresholds, in particular, the effect of VAT thresholds on small business behavior. The literature on VAT thresholds is small. In an important paper, Keen and Mintz (2004) are 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 is set up in such a way that none of the burden of output VAT can be passed on to purchasers (all sales are to final consumers) so it is never optimal for the firm to voluntarily register. Given the large amount of voluntary registration that we observe in the data, clearly, this is a limitation of

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

their model. Second, their main focus is on the optimal registration threshold, whereas our welfare analysis concerns the marginal deadweight loss of an increase in the statutory rate of VAT, following the literature on the elasticity of taxable income. Kanbur and Keen (2014) extend the Keen and Mintz (2004) framework to allow for evasion, as well as avoidance, of VAT. In our baseline model, we do not allow for evasion; the implications of doing so are discussed in Section 2.3.

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. More broadly, there is a small literature on firm bunching below non-VAT thresholds to avoid burdensome taxes and regulation; for example, in Spain, firms with turnover above a 6 million Euro threshold face increased tax enforcement; Almunia and Lopez Rodriguez (2014) show that firms bunch below this threshold to avoid increased scrutiny of their tax returns.

Our work also contributes to the literature on estimating the elasticity of the tax base for VAT. There are a small number of relevant contributions here. First, there are a number of studies (Carbonnier (2007) for France, Kosonen (2013) and Kosonen and Harju (2013) for Finland) that exploit large cuts in the rate of VAT on specific categories of goods (e.g. restaurant meals and haircuts in Finland) to estimate the percentage of the VAT cut passed on to consumers in the form of lower prices. The general finding is that there is less than full tax pass-through, with pass-through ranging from 80% to as low as 20%. We do not have price data and do not study pass-through; rather, we look directly at the elasticity of the tax base. But, pass-through is generally less than 100% in our model, because we allow for an upward-sloping marginal cost curve for the firm.

Second, there are a very few studies that estimate the effects of VAT cuts on quantities as well as prices, and thus on the VAT base. The two studies for Finland estimate the quantity responses to be very small, but do not quantify the overall effect of the VAT cuts on the VAT base. Blundell (2009) conjectures that the elasticity of tax base with respect to a temporary cut in the standard rate of VAT in the UK from 17.5% to 15% between 1 December 2008 and 1 January 2010 to be between 0.25 and 1.[6] But, this is a forward-looking estimate, assuming cost pass-though of between 75 and 100%, and an inter-temporal elasticity of substitution in consumption of 0.5 to 1. Our estimates of the elasticity of the tax base are closer to the Finnish studies than the Blundell estimate; one possible reason for this

[6] Blundell claims that in the UK, between 75-100% of the VAT cut would be passed on to the consumer, and based on the elasticity of inter-temporal subsitution, the elasticiy of real consumtion with respect to the VAT cut would be 0.5-1.0. This gives an overall elasticity of the tax base of between 0.25 and 1.

is that the structural approach gives a long-run elasticity that should be interpreted as the response to a permanent VAT change, whereas the Blundell calculation is for a temporary change, where the elasticity will of course be higher, due to inter-temporal substitution in consumption.

Finally, there is a policy literature on the design of the VAT that includes question of voluntary registration (e.g. Due (1990)). However, our work is the first, as far as we know, to quantitatively analyse the determinants of the voluntary registration decision.

The rest of the paper proceeds as follows. In the next section we develop the conceptual framework to analyse VAT bunching and voluntary registration. In section 3 we provide an overview of the VAT system in the UK and describe the data. Section 4 and 5 present the empirical analysis for voluntary registration and VAT bunching, respectively. Section 6 estimates the elasticity of the tax base, and section 7 concludes.

2 Conceptual Framework

2.1 The Set-Up

We consider a single industry with a fixed, large number of small traders producing a homogenous good, indexed by productivity parameter $a \in [\underline{a}, \overline{a}]$. Small trader a combines his own labour input l with an intermediate input x to produce output y via a fixed coefficients technology

$$y = a \min \left\{ l, \frac{x}{\beta} \right\}$$

where a measures the productivity of the trader and β the input requirement. Let t be the rate of VAT. If the trader is registered, he can claim back VAT on the input use x, so the price of the input is r. If not registered, the price of the input is $r(1+t)$.

There are also two types of buyers, those who are not registered for VAT (consumers) and those who are (businesses) in proportions λ and $1-\lambda$ respectively. It is assumed both types of buyers have perfectly elastic demand for the good at price p. This is analogous to the assumption made in the taxable income literature that labour demand is perfectly elastic at a fixed wage.

So, the profit for the non-registered trader is

$$(p - \beta r(1+t))y$$

For the registered trader, we reason as follows. The registered trader must charge VAT on his output. If he sells to a registered buyer, all the VAT can be passed on, as the buyer can

reclaim it. So, revenue per unit sold to a registered buyer is p. On the other hand, none of the output VAT can be passed on to the non-registered buyer, as he has perfectly elastic demand. So, revenue per unit sold to a registered buyer is $p/(1+t)$. So, overall, the profit for the registered trader is

$$\left(p\left(\frac{\lambda}{1+t}+1-\lambda\right)-\beta r\right)y$$

Finally, the trader has a disutility of labour $\phi(l)$, where ϕ is strictly increasing and strictly convex. For the most part, following Saez(2010) and Kleven and Waseem(2012), we assume that the disutility of labour is iso-elastic i.e.

$$\phi\left(l\right)=\frac{1}{1+\frac{1}{e}}\left(l\right)^{1+1/e} \tag{1}$$

So, the utility for the registered and nonregistered trader of productivity a can be written

$$\begin{aligned} u_R(y;a) &= py\left(\left(\frac{\lambda}{1+t}+1-\lambda\right)-s)\right)-\phi\left(\frac{y}{a}\right) \\ u_N(y;a) &= py\left(1-(1+t)s)\right)-\phi\left(\frac{y}{a}\right) \end{aligned}$$

where $s=\beta r/p$ is the share of inputs in total cost, and is an exogenous parameter in what follows. As p as been assumed fixed, we set it equal to 1 in what follows, so that y denotes both output and the value of sales.

2.2 Effective VAT Rates

We now proceed by defining the effective rate of VAT. Note that u_R, u_N can be written

$$\begin{aligned} u_R(y;a) &= y(1-s)(1-t_R)-\phi\left(\frac{y}{a}\right) \\ u_N(y;a) &= y(1-s)(1-t_N)-\phi\left(\frac{y}{a}\right) \end{aligned} \tag{2}$$

$$t_R=\frac{\lambda t}{(1+t)(1-s)}, t_N=\frac{st}{1-s} \tag{3}$$

That is, revenue net of input costs, $y(1-s)$, or value-added, is taxed at effective rate t_R if registered, and t_N if not. Note that t_R is increasing in the B2C ratio, λ, and increasing in s, whereas t_N is increasing in s. Obviously, both effective rates are increasing in the statutory rate, t.

Then, it is easily established that, given (1) , the outputs y_R, y_N that maximize $u_R(y;a)$,

$u_N(y;a)$ are

$$y_N(a) = a^{1+e}((1-s)(1-t_N))^e \quad (4)$$
$$y_R(a) = a^{1+e}((1-s)(1-t_R))^e$$

These are the output supplies of the trader. Thus, e measures the elasticity of output supply with respect to the effective taxes. Note also that the value-added of the trader is simply $v = y(1-s)$, so e measures the elasticity of the individual trader's value-added with respect to the effective taxes.

Finally, we cannot ignore the fact that there are significant compliance costs to VAT registration. It is well known that these costs, as a fraction of turnover, decline rapidly with turnover; 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[7]. We model these as a fixed cost K, so that net utility with registration is $u_R(y;a) - K$.

2.3 Discussion

One possible objection our analytical framework might seem very special; firms sell a homogenous product, and there is no substitution between inputs and the managerial labour input. We have two responses to this. First, both of these assumptions can be relaxed at the cost of some more analytical complexity. In a not-for-publication Appendix, we present a version of our model with differentiated products and a more general production function; then, it can be shown that the impact of the VAT system on the profit of the trader can no longer be measured just by an effective tax rate, but by a parameter that we call the discouragement index, which is itself a function of t, s, λ as here, but also of the firm-level elasticity of demand, and the elasticity of substitution between labour and the produced inputs. Many of the qualitative results extend to this case.

Second, while our model has some special features, it can be argued that it is in fact more general than the Saez (2010), also used by Kleven and Waseem (2013), framework used to study the personal income tax, where a worker with utility linear in consumption and iso-elastic labour supply faces a fixed pre-tax wage and a kinked or notched income tax schedule. In the Saez/Kleven-Waseem set-up, because the worker takes his pre-tax wage as given, he bears the full burden of the tax. In our setting, this corresponds to the assumption that no customer can reclaim VAT ($\lambda = 1$); then, the trader bears the full burden of VAT. Moreover, in a labour supply setting, there is no input tax; in our setting, this corresponds

[7] Federation of Small Businesses, "Impact of Increasing the VAT registration threshold for Small Businesses", June 2010.

to the case where $s = 0$. Finally, we also have a compliance cost of registration, K; in the Saez/Kleven-Waseem set-up, there are no compliance costs of moving over a tax notch, but there is a "pure notch" or lump-sum change in the tax liability, ΔT in their notation, which plays the same role. So, under the assumptions i.e. $\lambda = 1, s = 0$, our model reduces mathematically to the Kleven-Waseem model[8].

A limitation of our model is that we do not allow for evasion[9]. Estimated evasion of the VAT in the UK is currently around 11% of potential revenues[10]. Following Chetty(2010), one way to incorporate an evasion option would be to say that a firm with real turnover y can hide an amount h of turnover at cost $g(h)$, where $g(.)$ is increasing and convex. It is then easy to show that with an evasion option, there will be more bunching at the notch than without. This in turn implies that some of the observed bunching will be due to evasion, rather than the underlying elasticity of output supply, so that using bunching to infer e, as we do below, will tend to over-estimate e. However, without knowing something about the cost of evasion function, we cannot correct our estimates of e for this factor[11].

3 The Registration Decision

3.1 The Cut-Off

The VAT has a registration threshold; a firm must register if sales exceed y^*, but a firm can register below this threshold. There is also a compliance cost K of registering. The payoff from registration is thus

$$u_R(a) = \max_y \left\{ y(1-s)(1-t_R) - \phi\left(\frac{y}{a}\right) \right\}$$

The payoff from not registering is

$$u_N(a) = \max_y \left\{ y(1-s)(1-t_N) - \phi\left(\frac{y}{a}\right) \right\} \text{ s.t. } y \leq y^* \tag{5}$$

[8] Formally, with $\lambda = 1, s = 0$, our model is equivalent to a variant of their model where their higher rate of income tax above the notch, $\tau + \Delta\tau = \frac{t}{1+t}$, and where $\tau = 0$, $\Delta T = K$.

[9] The effects of an evasion or non-compliance option at tax notches are also discussed in Kanbur and Keen (2014), and where two forms of evasion are studied, total avoidance (bounders), and avoidance of some fraction of the tax (cads).

[10] https://www.gov.uk/government/statistics/vat-gap-estimates

[11] Chetty(2010) discusses some methods for estimating the marginal cost of evasion.

Also, let y_N, y_R be the optimal outputs in the two cases i.e.

$$\begin{aligned} y_R(a) &= a\phi'^{-1}(y(1-s)(1-t_R)) \\ y_N(a) &= \min\left\{a\phi'^{-1}(y(1-s)(1-t_N)),\ y^*\right\} \end{aligned} \tag{6}$$

Now, the net gain to registering is $\Delta(a) = u_R(a) - u_N(a) - K$, so a firms will register iff $\Delta(a) \geq 0$. We first provide a basic characterization of the registration decision.

Proposition 1. *Given fixed values for the other parameters, there is a critical $\tilde{a}$ such that all firms with $a \geq \tilde{a}$ register for VAT and all $a < \tilde{a}$ do not.*

The intuition is the following. First, the higher a, the higher is optimal output, and so the fixed cost of registration is less important overall. Second, the cost of meeting the registration turnover constraint $y \leq y^*$ is higher, the higher is a.

3.2 Voluntary Registration

The first aspect of the registration decision that we are interested in is voluntary registration. Say that a firm a chooses *voluntary registration* if it chooses to register, and has a turnover below the threshold i.e. $y_R(a) < y^*$. Our empirical predictions concern the share of the firms who produce below the threshold who register voluntarily.

From now on, we will also assume that ϕ is iso-elastic, as in (1) .Let a_R be the firm type which, if registered, just wants to produce at the registration threshold i.e. from (4), $a_R = \left(\frac{y^*}{(x_R)^e}\right)^{1/(1+e)}$. From Proposition 1, if $\tilde{a} < a_R$, all firms between $\tilde{a}$ and a_R register voluntarily; otherwise, none do. So, recalling that a is distributed uniformly, the share of firms producing less than y^* which are registered voluntarily is

$$\sigma = \max\left\{\frac{a_R - \tilde{a}}{a_R}, 0\right\} \tag{7}$$

Now we have:

Proposition 2. *(i) If $s/\lambda < \frac{t}{1+t}$, $\sigma = 0$. (ii) There is a summary statistic of the parameters, $\beta = ((x_R)^{1+e} - (x_N)^{1+e})/(x_R)^e$, such that $\sigma > 0$ iff $\beta > \frac{K(1+e)}{y^*} = \beta_0$, and σ is strictly increasing in β if $\beta > \beta_0$. (iii) The share of voluntary registrations σ is decreasing in the B2C ratio λ. (iv) There is a critical value $\widehat{\frac{1-t_N}{1-t_R}} < 1$ above which the share of voluntary registrations, σ, is increasing in the share of input costs in turnover s.*

3.3 Bunching

Now consider that group of firms for which λ, s are such that voluntary registration is not optimal i.e. for which $t_R > t_N$. Note that this group has the full range of productivity a.In this case, it is easy to show that there is bunching at the cutoff. In particular, let a^* be the firm which just produces at the threshold when non-registered i.e. $y_N(a^*)$. Then, all firms between a^* and $\tilde{a}$ will produce at the threshold, with any firm $a^* < a < \tilde{a}$ restricting its output to avoid paying the registration cost and the higher effective tax. So, now, $\tilde{a}$ will be defined by the condition

$$\max_y \left\{ y(1-s)(1-t_R) - \phi\left(\frac{y}{\tilde{a}}\right) \right\} - K = y^*(1-s)(1-t_N) - \phi\left(\frac{y^*}{\tilde{a}}\right) \qquad (8)$$

For consistency with Kleven and Waseem, now define $\Delta a^* \equiv \tilde{a} - a^*$, so that $\tilde{a} \equiv a^* + \Delta a^*$.So, all firms located between a^* and $a^* + \Delta a^*$ in the productivity distribution bunch at the threshold. However, we do not observe a directly, only y, so we need to map the bunching interval into the space of turnover. To do this, note that in the absence of bunching, the critical firm $a^* + \Delta a^*$ would have turnover $y^* + \Delta y^* = (a^* + \Delta a^*)((1-s)(1-t_R))^e$. So, the percentage turnover response to the notch is measured by $\Delta y^*/y^*$. Then we can show:

Proposition 3. *Given e, the level of bunching Δy^* is given by the implicit relationship*

$$\frac{1}{(1+\Delta y^*/y^*)}\left(1 + \frac{K/y^*}{(1-s)(1-t_N)}\right) - \frac{1}{1+1/e}\left[\frac{1}{1+\Delta y^*/y^*}\right]^{1+1/e} - \left(\frac{1-t_R}{1-t_N}\right)^{1+e}\frac{1}{1+e} = 0 \qquad (9)$$

Note also that (9) is very 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; the latter is given by equation (5) in their paper, which, in our notation, is

$$\frac{1}{(1+\Delta y^*/y^*)}\left(1 + \frac{\Delta T/y^*}{1-t}\right) - \frac{1}{1+1/e}\left[\frac{1}{1+\Delta y^*/y^*}\right]^{1+1/e} - \left(1 - \frac{\Delta t}{1-t}\right)^{1+e}\frac{1}{1+e} = 0 \qquad (10)$$

where t is the initial rate of income tax, and ΔT, Δt are the notches i.e. when pre-tax income goes above y^*, a fixed penalty ΔT is paid, and then *all* income is taxed at rate $t + \Delta t$. There are two differences between (9) and (10). First, with the VAT, the compliance cost, K takes the place of ΔT. Second, $\frac{1-t_R}{1-t_N}$ replaces $1 - \frac{\Delta t}{1-t}$.

We can now use (9) to look at some of the determinants of bunching. We have:

Proposition 4. *Assume the registration cost K is small. Then, the amount of bunching*

Δy^* rises (i) as λ, the fraction of B2C sales increases, and (ii) as the share of inputs in total cost, s, falls.

The intuition for (i) is simply that if most customers are VAT-registered, the burden of an increase VAT can easily be passed on the form of a higher price, because the customer himself can claim back the increase. The intuition for (ii) is that when input costs are important, registration allows the firm to claim back a considerable amount of input VAT. We will bring this prediction to the data below.

4 Welfare

In this section, we show how the elasticity of output, e, with respect to the effective taxes t_R, t_N can be related to the deadweight loss of the VAT. Assume that all firms have the same s, λ, so that they only vary in a. Following **Chetty(2009)**, our welfare criterion is $W = U + T$, where U is the average utility across all firms i.e.

$$U = \int_0^{\tilde{a}} u_N(a)da + \int_{\tilde{a}}^{a_{\max}} (u_R(a) - K)da$$

and where T is tax revenue. This has two components, the VAT collected from output on registered firms, and the VAT charged on the inputs of non-registered firms. So, overall,

$$\begin{aligned} T &= ts\int_0^{\tilde{a}} y_N(a)da + \frac{t\lambda}{1+t}\int_{\tilde{a}}^{a_{\max}} y_R(a)da \qquad (11) \\ &= t_N V_N + t_R V_R,\ V_N = \int_0^{\tilde{a}} (1-s)y_N(a)da,\ V_R = \int_{\tilde{a}}^{a_{\max}} (1-s)y_R(a)da \end{aligned}$$

where, in the second line, we write tax revenue in a more standard way as the sum of effective rates t_N, t_R for non-registered and registered forms respectively, times the corresponding tax bases i.e. value added of registered and non-registered firms V_R, V_N.

As in Chetty(2009), we measure the deadweight loss of an increase in the VAT rate by $\frac{dW}{dt}$. The first, and simplest, case is where there is voluntary registration, i.e. $t_R > t_N$.It is then possible to show the following:

Proposition 5. *If $t_R < t_N$, so that there is voluntary registration, then the deadweight loss of a small tax increase is*

$$\frac{dW}{dt} = \underbrace{\left(t_N\frac{\partial V_N}{\partial t} + t_R\frac{\partial V_N}{\partial t}\right)|_{\tilde{a}\text{ const}}}_{\text{intensive DWL}} \quad + \underbrace{\frac{\partial T}{\partial \tilde{a}}\frac{\partial \tilde{a}}{\partial t}}_{\text{extensive DWL}} \tag{12}$$

Moreover, the intensive DWL, as a fraction of the additional revenue raised, $\frac{\partial t_N}{\partial t}V_N + \frac{\partial t_R}{\partial t}V_R$, *can be written*

$$-e\left(\theta_N\frac{t_N}{1-t_N} + (1-\theta_N)\frac{t_R}{1-t_R}\right) \tag{13}$$

where $\theta_N = \frac{V_N\frac{\partial t_N}{\partial t}}{\frac{\partial t_N}{\partial t}V_N+\frac{\partial t_R}{\partial t}V_R}$. *Finally, the extensive DWL is proportional to* $K^{1/(1+e)}$, *and vanishes as* $K \to 0$.

(12) is a variant of the Feldstein-Chetty formula in for the deadweight loss of a proportional income tax, $\frac{dW}{dt} = t\frac{dTI}{dt}$, where TI is taxable income, and t is the proportional rate of income tax. It differs in two ways. First, there is also the effect of the tax on welfare via the registration threshold, measured by $\frac{\partial T}{\partial \tilde{a}}\frac{\partial \tilde{a}}{\partial t}$, which we call the deadweight loss at the extensive margin, or extensive DWL. Second, in this case, there are two tax bases V_N, V_R and two effective taxes, t_N, t_R, so the formula is more complex. The fact that the intensive DWL can be written proportional to e is again analogous to the Feldstein-Chetty formula, which can be written $\frac{dW}{dt}/TI = -e\frac{t}{1-t}$, where e is the elasticity of taxable income with respect to t.

Now consider the case with $t_R > t_N$, where there is bunching. Now, the main differences are twofold. First, as all non-registered firms between a^* and $\tilde{a}$ bunch, we have:

$$y_N(a) = \begin{cases} a^{1+e}(x_N)^e, & a \le a^* \equiv \frac{(y^*)^{1/(1+e)}}{(x_N)^{e/(1+e)}} \\ y^* & a^* < a \le \tilde{a} \end{cases}$$

Second, the formula for $\tilde{a}$ is now rather different. As a consequence of (i), we have a different formula for tax revenue i.e.

$$T = t_N(V_N + V_B) + t_R V_R,\ V_N = \int_0^{a^*}(1-s)y_N(a)da,\ V_B = \int_{a^*}^{\tilde{a}} y^* da \tag{14}$$

and V_R is as before. so V_B is the value-added of the bunchers. Note also (i) for a fixed $a^*, \tilde{a}$, V_B does not respond to t; (ii) from (14), the effect of a change in a^* on tax revenue is zero; $\frac{\partial T}{\partial a^*} = 0$. Then, we have;

Proposition 6. *If* $t_R > t_N$, *so that there is bunching, formula* (12) , *continue to hold. But*

now, the intensive DWL, as a fraction of the additional revenue raised, $\frac{\partial t_N}{\partial t}(V_N+V_B)+\frac{\partial t_R}{\partial t}V_R$, *can be written*

$$-e\left(\gamma_N \frac{t_N}{1-t_N}+\gamma_R \frac{t_R}{1-t_R}\right) \tag{15}$$

where $\gamma_N=\frac{V_N \frac{\partial t_N}{\partial t}}{\frac{\partial t_N}{\partial t}(V_N+V_B)+\frac{\partial t_R}{\partial t}V_R}$, $\gamma_R=\frac{V_R \frac{\partial t_R}{\partial t}}{\frac{\partial t_N}{\partial t}(V_N+V_B)+\frac{\partial t_R}{\partial t}V_R}$

So, now, there are two differences to Proposition 5. First, in (15) the weights on t_N, t_R are slightly different. Second, from the different definition of $\tilde{a}$ in (8) , the detailed formula for the extensive DWL is different, and that term does not vanish as $K \to 0$.

In principle, formulae (13) , (15) can be computed from our data. [TO BE COMPLETED]

5 Context and Data

5.1 The Value-Added Tax System in the UK

The Value-Added tax in the UK is paid by approximately 2 million registered businesses in each fiscal year.[12] It is the third largest source of government revenue following income tax and national insurance contributions. In 2011/12, VAT raised £98,228 million, accounting for 21.05% of total tax revenue and 6.54% of GDP in the UK.[13]

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, and brought into the UK from other EU countries.[14] All businesses must register for VAT if their taxable turnover is above a given threshold.[15] The current registration threshold is £79,000 in 2013/14. As permitted by the EU VAT law, increases in the registration threshold should be no greater than the rate of inflation.[16] The UK currently set 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.[17]

[12] Authors' estimates based on the universe of UK VAT records between 2004/05 and 2010/11.

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

[14] There are complex regulations for goods and services imported from within the EU.

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

[16] 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.

[17] See http://www.oecd.org/tax/tax-policy/tax-database.htm#vat. Among all OECD countries, Denmark has the lowest threshold, which requires businesses with sales of more than DKK 50,000 (GBP£4,308) to register. There is no VAT threshold in Mexico, Sweden, and Spain so that all businesses in these countries are required to register unless exempt otherwise.

VAT is charged on the additional value of each transaction, and is collected at each stage of production and distribution. A business pays VAT on its purchases–known as input tax, and charges VAT on the full sale price of the taxable supplies–known as output 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 are charged at a reduced rate of 5% and there are also goods and services that are charged at a zero rate or exempt from VAT altogether.[18] 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.

Small firms with annual taxable turnover of up to £150,000 can use a simplified flat-rate VAT scheme, which was introduced in 2002 and allows firms to pay VAT at a single rate on their total sales.[19] The flat rate, which varies between 4% and 14.5% depending on the industry, is intended to reflect the average VAT rate in each industry and reduce the compliance cost associated with keeping detailed records and calculating VAT for each transaction separately. In practice, the extent of such administrative savings is rather unclear, since firms must keep similar records to calculate and compare their VAT liability under both the standard scheme and the flat-rate scheme in order to decide whether to join or leave the flat-rate scheme. As discussed in a 2007 Public Accounts Committee report and in Vesal (2013), the take-up rate for the flat-rate scheme among eligible firms are extremely low and most eligible firms are registered under the standard scheme.[20]

There are two rules regarding registration, a forward-looking rule and a backward-looking one. First, a firm must also register for VAT if either (i) the VAT taxable turnover of the firm may go over the threshold in the next 30 days alone, or the firm takes over a VAT-registered business as a going concern. Second, a firm must register for VAT if its VAT-taxable turnover for the previous 12 months was more than the threshold. Strictly speaking, out theoretical

[18] 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.

[19] Under the flat-rate sheme, firms surrender the right to reclaim VAT on inputsThe turnover ceiling for FRS has been increased from £100,000 when it was introduced in 2002 to £150,000 since 2003.

[20] In October 2007, the Public Accounts Committee published a report on new business' tax obligations and found that out of 705,000 eligible businesses, only 16% of firms were registered under the flat-rate scheme. A more recent study Vesal (2013) also finds that twenty six percent of eligible VAT traders gain from the flat-rate scheme but very few join the scheme. Both studies attribute the low takeup rate to the lack of awareness of the flat-rate scheme scheme.

model applies to the forward-looking decision, as the model is static; that is, the firm must register if turnover in the current year is expected to exceed the threshold. In our sample, among firms that register for the first time, around 68% of them have turnover in the previous year lower than the VAT notch. This suggests that the forward-looking decision is he more important .

VAT compliance in the UK has been long susceptible to fraud and avoidance. According to HMRC estimates, the VAT tax gap, which is defined as the difference between net theoretical tax liabilities and total VAT receipts on a timely basis, is around 10.4% of theoretical VAT liability since 2010. This is considerably higher than the tax gap estimates for many other taxes in the UK except for tobacco duties and self assessment. The most recent estimate of the £11.4 billion VAT gap in 2011-12, is composed of (1) £0.5 – 1.0 billion of MTIC (Missing Trader Intra-Community) fraud,[21] (2) £1.8 billion of VAT debt,[22] (3) £0.2 billion due to VAT avoidance.[23]

Table 1 summarises the two sources of variation in the VAT tax system that we explore in empirical analysis. The first source of variation, as shown in Column 1, is the discrete jump in the tax rate and the overall VAT liability at the registration threshold. The registration threshold was £58,000 in 2004/05, has been increased annually to £68,000 in 2009/10, and is currently £79,000 since 2013/14. We analyze the excess number of firms bunching below the threshold to estimate the elasticity of the turnover with respect to the standard rate of VAT in a structural approach. The second source of variation is a temporary reduction in the main rate of VAT between December 1, 2008 and January 1, 2010, which was the main lever of a fiscal stimulus package to counter the recession. As shown in Column 3, the standard rate of VAT was temporarily reduced to 15 percent on 1 December 2008 and returned to 17.5 percent on 1 January 2010. Taking a reduced-form approach, we exploit variation in the main rate as a result of this temporary VAT reduction in combination with variation in treatment intensity for firms across distribution of B2C sales ratio and estimate the elasticity of value-added using a difference-in-differences approach.

[21] MTIC VAT fraud is an organised criminal attack on the EU VAT system in which fraudulent traders acquire goods and services VAT free from EU Member States by charging VAT on their onward sale and disappear to avoid paying the VAT charged to the relevant tax authorities.

[22] VAT debt is defined as the difference between new debts arising in the financial year and debt payments plus debt adjustments made in the financial year.

[23] A detailed description of the methodology and latest estimate of the VAT tax gap is available in "Measuring Tax Gaps - 2013" at *www.hmrc.gov.uk/statistics/tax-gaps/mtg-2013.pdf.*

5.2 Data

We construct the dataset by linking the universe of VAT returns to the universe of corporation tax records in the UK. The first data set provides VAT tax information for businesses in different legal forms including sole traders, partnerships, and companies but only for those who are registered. 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 VAT 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) annual company accounts from the FAME (Financial Analysis Made Easy) database for additional firm characteristics and accounting information[24] and (2) annual sector-level statistics on the share of sales to final consumers (i.e. a measure of the B2C sales ratio), which are derived from the Office of National Statistics (ONS) Input-Output Tables and are available at 2-digit SIC industry levels.

We take the following steps to refine the sample to better capture VAT registration of individual companies. First, we eliminate companies which are part of a larger VAT group and focus only on standard-alone 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, because the registration decision is based on turnover in the previous 12 months, we drop all observations with an accounting period less than 12 month in the corporation tax record. In addition, we eliminate companies that mainly engage in overseas business activities based on the trade classification since the taxable VAT turnover is based on sales of goods and services within the UK. The final dataset contains 2,120,000 observations for 550,604 companies between April 1, 2004 and March 30, 2010. For each company-year observation, we have information on the VAT-exclusive turnover taken from the corporate tax records, and whether it is registered for VAT[25]. We also observe a few key factors that drive firms' decision about

[24]FAME database is published by Bureau van Dijk and contains detailed financial and ownership information for more than 1.9 million companies in the UK and Ireland.

[25]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

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), and company-specific history of registration status.

5.3 Summary Statistics

Figure 1 presents convincing evidence that the VAT registration threshold is binding in the UK. In Panel A, there is a discrete jump in the share of registered companies at the normalized VAT notch during 2004/05-2009/10, with a substantial number of voluntary registers below the threshold. On average, around 40.93% of companies with a turnover below the current-year VAT notch are registered for VAT, which suggests that for these companies the benefits of being registered to reclaim the input taxes may well outweigh the costs. The share of registers increases considerably to around 85% once reaching the threshold, with non-VAT registered companies above the threshold consisting of three types: (1) those providing exempt supplies, (2) those providing primarily zero-rated supplies, and (3) those with turnover temporarily exceeding the threshold.[26] Panel B further shows a histogram of nominal turnover net of current-year VAT notch by pooling data between 2004/05 and 2009/10. There is an evident excess of mass just below the VAT notch that is normalized to zero in the otherwise smooth distribution of turnover.

Table 2 provides summary statistics for companies in the neighborhood of current-year VAT notch, which include all companies with a nominal turnover between £10,000 and £200,000 over the sample period. Column 1-3 shows 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 51.44% of all companies in the linked dataset. Column 4-6 focus on the registered companies while column 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, with column 9 and 10 reporting the t-statistics and the corresponding p-value, respectively. In total, there are 805,351 observations for 289,352 companies in the sample, with around 62% of them registered for VAT. On average, registered companies have a significantly higher turnover and trading profit comparing to non-registered companies. Consistent with Proposition 1, registered companies on average have a significantly higher input-cost ratio and B2C sales ratio. Companies in the registered sample are

overall turnover figure for their VAT registration decisions.

[26] (a footnoteon the turnover variables: we have three turnover variables: CT600, FAME and VAT; Around 90% of companies report turnover in Box 1 of CT600. Only % of them report turnover in FAME. Turnover variable in VAT is only available for registered companies. So the turnover variable that we use is the CT600 one—make sure that I only include companies with a narrow turnover difference window).

When to register: http://www.hmrc.gov.uk/vat/start/register/when-to-register.htm

significantly older, which is consistent with the correlation that younger firms also tend to be small in size.

6 Evidence on Voluntary Registration

In this section, we examine whether the empirical pattern of voluntary registration is consistent with the theory in two key aspects as predicted by Proposition 1, i.e. whether voluntary registration is more likely if the share of B2C sales is low, or the share of inputs in cost is high. We define that a firm is voluntarily registered if (i) it has a turnover below the annual VAT notch in that year if it has never registered before, or (ii) its current-year turnover is below the annual VAT deregistration threshold if it was registered in the previous year. In the full dataset, 63% of firms have a turnover below the VAT threshold, and of these, 35% of them are registered for VAT. So, overall, 22% of firms in the entire sample are voluntarily registered for VAT.

We first note in table 3 that voluntary registration varies with the share of B2C sales and with the share of inputs in cost in a way that is consistent with the theory. As the share of B2C sales falls, i.e. when moving from Q4 to Q1, the share of voluntarily registered firms tends to rise. Similarly, as the input cost ratio rises, share of voluntarily registered firms tends to rise. The empirical pattern is broadly consistent with Proposition 1. To investigate further, we plot the distribution of the B2C sales ratio and the input cost ratio by registration status, for all firms below the threshold.[27] The empirical pattern is again consistent with Proposition 1, as for all firms with a turnover below the VAT notch, those who are voluntarily registered tend to have a lower B2C sales ratio and a higher input cost ratio comparing to their non-registered counterparts.

Finally, we model the decision of voluntary registration as a function of the B2C sales ratio and the input cost ratio in a binary choice model of the following form:

$$R_{it} = \gamma_1 + \gamma_2 B2C_{j(i)} + \gamma_3 ICR_{it} + \gamma_4 \mathbf{X}_{it} + \rho_t + \phi_i + \upsilon_{it}, \tag{16}$$

where R_{it} represents the binary voluntary registration variable which takes on the value 1 if a firm is voluntarily registered for VAT and the value 0 otherwise. The key variables of interest are $B2C_{j(i)}$, the industry-level B2C ratio for firm i (that is, firm i in industry $j(i)$), and ICR_{it}, the input cost ratio for firm i in year t. $\mathbf{X}_{it}$ are other firm-level controls, and ϕ_i and ρ_t are time-invariant firm fixed effects and year dummies. υ_{it} is the error term. We first estimate equation (16) in a linear probability framework based on the standard OLS

[27] We are waiting for HMRC's approval to disclose these figures.

assumptions. To check the robustness of the estimation results, we reestimate equation (16) in a fixed-effect logit model which assumes that the error term follows a logistic distribution.

The results are shown in table 4. Column (1)-(4) present estimation results from the linear probability model and column (5)-(8) present estimation results from the fixed-effects logit model. While the magnitude of the coefficients are not directly comparable between the two models, it is assuring that they have the same sign and significance level.[28] Column (1) and (5) do not include firm fixed effects and allow us to examine the effect of industry-level B2C sales ratio on the probability of voluntary registration. The coefficient estimates are negative and statistically significant, indicating that the likelihood for a firm to voluntarily register for VAT is reduced by around 0.04 given a one standard deviation increase in the B2C sales ratio.

The rest of the specifications add firm fixed effects and the coefficient on the B2C sales ratio becomes often imprecisely estimated due to its limited variation at the industry level over time. For comparison, column (2) and (6) do not include any additional firm-level controls while column (3) and (7) include firm-level trading profit and age as additional control variables. Column (4) and (8) check the robustness of the results by replacing the salary-inclusive input cost ratio with the salary-exclusive input cost ratio calculated from FAME. Given that few firms report the direct cost of sales, the sample size is dramatically decreased but nevertheless the coefficient estimate for the input cost ratio remains positive and highly significant. Moreover, the coefficient estimate for the B2C sales ratio is negative and significant at 10% level. Focusing on results in column (3) and (4), the likelihood to voluntarily register for VAT is increased by around 0.01-0.05 given a one standard deviation increase in the input cost ratio.

To further investigate the robustness of our results to the limited variation in the B2C ratio roughly at the 2-digit SIC level, we compute the share of firms that are voluntarily registered in each year, and regress it against the industry-level B2C sales ratio and input cost ratio. The results are presented in table 5 and are fairly consistent with findings from the firm-level regression analysis. The coefficient estimate for the B2C sales ratio is negative and highly significant in pooled regressions in column (1)-(4) which do not include industry fixed effects, and becomes positive and imprecisely estimated in column (5)-(8) with inclusion of industry fixed effects. Similarly, the coefficient estimate for the average input cost ratio is positive and highly significant in pooled regressions without inclusion of industry fixed

[28]Following the rule of thumb as suggested in (Wooldridge, 2001, p. 465-468), we divide the logit estimates by four to make them roughly comparable to the LPM estimates. The scaled logit estimates are quite similar to the liner probability model (LPM) estimates. We use the LPM estimates to infer the average partial effects of our key variables of interest on the response probability since the fixed effects logit estimator does not allow for estimation of partial effects.

effects. The loss of significance is due to limited variation in the two variables of interest at the industry level across time.

7 Evidence on Bunching

7.1 Estimation Methodology

As set out in the conceptual framework in Section 2, the VAT registration threshold at the cutoff turnover value z^* will induce excess bunching at the threshold by companies for which voluntary registration is not optimal. The bunching is driven by the productivity parameter a, and will generate an excess mass by companies who would have reported a turnover between z^* and $z^* + \Delta z^*$ absent the notch of

$$B(z^*) = \int_{z^*}^{z^*+\Delta z^*} g(z)d(z) \simeq g(z)\Delta z^*,$$

where $B(z^*)$ is the excess mass at the threshold and $g(z)$ is the counterfactual density distribution of turnover had there been no registration threshold. The approximation is accurate to the extent that $g(z)$ is uniform around the notch.

By grouping companies into small turnover bins of £100, we estimate the counterfactual distribution around the VAT notch z^* in the following regression:

$$c_j = \sum_{l=0}^{q} \beta_i (z_j)^l + \sum_{i=z_-^*}^{z_+^*} \gamma_i I\{j = i\} + \varepsilon_j, \tag{17}$$

where c_j is the number of companies in turnover bin j, z_j is the distance between turnover bin j and the VAT notch z^*, q is the order of the polynomial, and $I\{\cdot\}$ is an indicator function. The range (z_-^*, z_+^*) in the second term specifies turnover bins around the notch where bunching occurs and are therefore excluded from the regression. The lower bound of the excluded turnover region, z_-^*, is set at the point where excess bunching starts. The upper bound of the excluded region, z_+^*, is estimated in an iteration procedure to ensure that the area under the estimated counterfactual density is equal the area under the observed density. In other words, the estimation procedure ensures that the excess mass below the VAT notch is equal to the missing mass above. The error term ε_j reflects misspecification of the density equation.

The estimated counterfactual distribution is defined as the predicted bin counts $\widehat{c}_j$ from (17) omitting the contribution of the dummies in the excluded region (z_-^*, z_+^*), and excess

bunching is estimated as the difference between the observed and predicted bin counts over the excluded range that falls below the VAT notch:

$$\widehat{B} = \sum_{i=z_{-}^{*}}^{z^{*}} (c_j - \widehat{c}_j).$$

We use the excess mass $B(z^*)$ and counterfactual distribution $g(z)$ to recover the bunching ratio $b(z) = B(z^*)\ /g(z)$, which denotes the fraction of companies that bunch at the notch relative to the counterfactual density and approximates Δz^* under the assumption of no optimization frictions.

In the empirical application, we observe that there is a very small hole in the observed distribution above the threshold, suggesting that many companies are not able to adjust their turnover due to optimization frictions. To examine the extent of non-response given frictions, we follow Kleven and Waseem (2013) in first defining a dominated turnover region $(z^*, z^* + \Delta z^D)$, where no optimizing firm will locate, whatever the parameter values. Kleven and Waseem (2013) show that the Δz^* solving (10) for any $e \in (0, \infty)$ is at least as great as

$$\Delta z^D = \frac{1 - \lim_{e \to 0} \left(1 - \frac{\Delta t}{1-t}\right)^{1+e}}{\lim_{e \to 0} \left(1 - \frac{\Delta t}{1-t}\right)^{1+e}} z^* = \frac{\Delta t}{1 - t - \Delta t} z^*$$

So, by the same argument, the dominated region in our case is

$$\Delta y^D = \frac{1 - \left(\frac{1-t_R}{1-t_N}\right)}{\left(\frac{1-t_R}{1-t_N}\right)} y^* = \left(\frac{\frac{\lambda t}{(1+t)} - st}{1 - s - \frac{\lambda t}{(1+t)}} \right) y^*$$

We know that for the full sample**,** λ is approximately 0.49, which gives a range of values of Δz^D of between $0.08z^*$ and $0.17z^*$.

We then estimate the proportion a^* of companies with large adjustment costs locating in the strictly dominated region between z^* and z^D relative to the counterfactual density $g(z)$ as:

$$a^* = \frac{\int_{z^*}^{z^* + \triangle z^D} g(z) d(z)}{g(z)}.$$

Finally, we take account of the fact that some firms who voluntarily register will be in the dominated region, even if they are fully rational. The corresponding excess bunching

accounting for optimization frictions, and voluntary registrations, is therefore estimated as

$$\widehat{B}^* = \frac{\widehat{B}}{1 - a^*(1 - v)}.$$

where v is the share of firms who register voluntarily with a turnover in the neighbourhood of the notch. We interpret estimates of $\widehat{B}^*$ as an upper bound of the firms' response to the VAT notch, which represents the amount of bunching had all companies overcome adjustment costs. We use this adjusted bunching estimate to evaluate the structural elasticity. Standard errors on all estimates are calculated using a residual-based bootstrap procedure as in Chetty et al. (2011) and Devereux, Liu and Loretz (2014).

7.2 Bunching Evidence

7.2.1 Baseline Estimates

This section presents evidence of bunching below the VAT notch. Figure 2 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. The excluded turnover range is demarcated by the vertical dashed lines and the VAT notch demarcated by the vertical solid line.[29] The next five panels focus on subsequent years during which the VAT notch was increased annually to track inflation. Each panel shows estimates of excess bunching below the VAT notch scaled by the counterfactual frequency at the notch (b) and the share of companies in the dominated range who are unresponsive (a^*) to the VAT notch.

Three main findings are worth noting in Figure 2. First, the VAT notch creates evident and sharp bunching below the threshold. Excess bunching ranges from 0.81 to 1.20 times the height of the counterfactual distribution, and is strongly significant in each year during the sample period. 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 and sharp bunching below the threshold, the VAT notch is associated with a small hole in the distribution above the

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

cutoff. The range of the hole spans from £8,500 to £15,000 above the VAT notch and the share of non-responsive companies stays consistently between 0.82 and 0.89 during the sample period.

Given the large and sharp bunching at the VAT notch, an immediate implication is that non-registered companies bunch below the threshold should on average, (1) sell mainly to consumers and non-VAT businesses so that they are less able to pass the cost of the increased tax on to their customers, and/or (2) incur lower input cost relative to the final output so that the gain from reclaiming the input VAT is not particularly large. To verify these empirical regularities, we split companies with a turnover *below* the VAT notch to voluntary registers and non-registers, and first test whether there is any significant difference in the ratio of input cost relative to turnover between the two groups. As suggested by a negative and highly significant t statistic of -29.662 (with p-value of 0.000), on average voluntary registers incur a higher input cost than the non-registers. A second test of two-group mean difference in the B2C sales ratio suggests that on average, voluntary registers engage more in B2B sales than the non-registers, implying that it is easier for the former group to pass the tax onto the next stage of the production chain.

To examine whether bunching is primarily driven by fixed compliance cost, we separately examine bunching behaviour of growing and shrinking companies. Figure 3 pools all data over the sample period and presents a histogram of turnover (net of current-year VAT notch) for growing and shrinking companies in panel A and B, respectively. While there is bunching below the VAT notch in both panels, it is evident that the excess mass that we observe in Figure 2 is mainly due to behavioural responses of growing companies in panel A, with shrinking firms responding in a much smaller extent to the VAT notch in panel B. These patterns suggest that as small firms grow and approach the threshold, a non-trivial proportion of them slow down their growth to avoid crossing the threshold for registration, for which the saving in tax and compliance costs exceeds the reduction in sales volume.[30]

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 2, that firms are more likely to bunch below the VAT notch if (1) the share of B2C sales is high, and (2) the share of input costs is low.

[30] An early SBRT Quarterly Survey of Small Business in Britain, (citation), reports that almost one-fifth of firms close to the VAT-threshold turn away business to avoid registration.

Following Proposition 2 in Section 2, the VAT notch should have little effect on small businesses selling to registered traders since the tax is simply passed onward in the production chain with the only effect of registration being to receive a credit from VAT imposed on purchases. Therefore, conditional on having a reported turnover in the neighborhood of the VAT notch, firms that mainly sell to non-registered traders and final consumers should benefit the most from bunching below the VAT notch. To explore how companies with different B2C sales ratio respond to the same VAT notch, we divide the sample into four groups according to the quartiles of B2C sales ratio and estimate annual bunching ratios separately for each quartile.

Figure 4 plots the point estimate of the bunching ratio with the corresponding 95% confidence intervals in each year and quartile and suggests two interesting findings. 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 behavioural responses of companies in the 3rd and 4th quartile of the B2C sales ratio.

Because VAT registration allows companies to reclaim input taxes, the incentives for them to bunch below the VAT notch are likely to decrease with the share of direct inputs in cost. This is formally shown in proposition 4 as the amount of bunching is decreasing in the share of inputs in cost β. 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 11.87% 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 normalized VAT notch in Figure 5.

Panel A compares the empirical distributions of companies around the normalized VAT notch at four different quartiles of input-cost ratio. It presents clear evidence that the degree of bunching decreases with the share of input costs relative to output. The distribution of companies in the top quartile is considerably 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 further quantifies the difference in the extent of bunching by plotting the estimated bunching ratio with the corresponding 95% confidence interval for

companies in 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. For companies in the lower quartiles of the distribution, the bunching estimates are positive and highly significant, with some suggestive evidence that the largest bunching occurs for companies in the 2nd and 3rd quartiles of the distribution.

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 the 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 6 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 6 provide suggestive yet not conclusive evidence that part of bunching is due to turnover misreporting.

8 Estimating e

To get a numerical estimate of e, from (9), we need to set several parameters. First, λ is measured by the ratio of sales that are B2C in the sample, which, from Table 2, is 0.49. Second, s is measured by the input to sales ratio in the whole sample, which is from Table 2, is 0.38. Third, the bunching ratio $\Delta z^*/z^*$ is taken from Table 1; we use the average across all years, adjusted for optimisation frictions [TO BE COMPLETED]

9 Conclusions

In this paper, we provide a conceptual framework to study VAT bunching. We derive theoretical predictions that bunching at the registration threshold is increasing in the B2C share of sales, and is decreasing in input cost share. The theoretical predictions are largely supported by the data. We extend the conceptual framework developed by Saez, Slemrod and Giertz (2012) and Kleven and Waseem (2013) to deal with "special characteristics" of VAT, which allows us to estimate both the elasticity of turnover w.r.t. the VAT "discouragement index" and the elasticity of value-added of registered firms w.r.t. the VAT rate. Although our empirical estimates of the two elasticities are still preliminary, they suggest that both elasticities might be small.

References

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

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 Working Paper.

Blundell, Richard. 2009. "Assessing the Temporary VAT Cut Policy in the UK." *Fiscal Studies*, 30(1): 31–38.

Carbonnier, Clément. 2007. "Who Pays Sales Taxes? Evidence from French VAT Reforms, 1987-1999." *Journal of Public Economics*, 91(5Ű6): 1219 – 1229.

Chetty, Raj. 2009. "Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance." *American Economic Journal: Economic Policy*, 1(2): 31–52.

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.

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.

Due, John F. 1990. "Some Unresolved Issues in Design and Implementation of Value Added Taxes." *National Tax Journal*, 43: 383–394.

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

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 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 J. Munroe. 2014. "Mansion Tax: The Effect of Transfer Taxes on the Residential Real Estate Market." National Bureau of Economic Research Working Paper 20084.

Kosonen, Tuomas. 2013. "More Haircut after VAT Cut? On the Efficiency of Service Sector Consumption Taxes." Government Institute for Economic Research Finland (VATT) Working Papers 49.

Kosonen, Tuomas, and Jarkko Harju. 2013. "Restaurant VAT Cut: Cheaper Meal and More Service?" Government Institute for Economic Research Finland (VATT) Working Papers 52.

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.

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

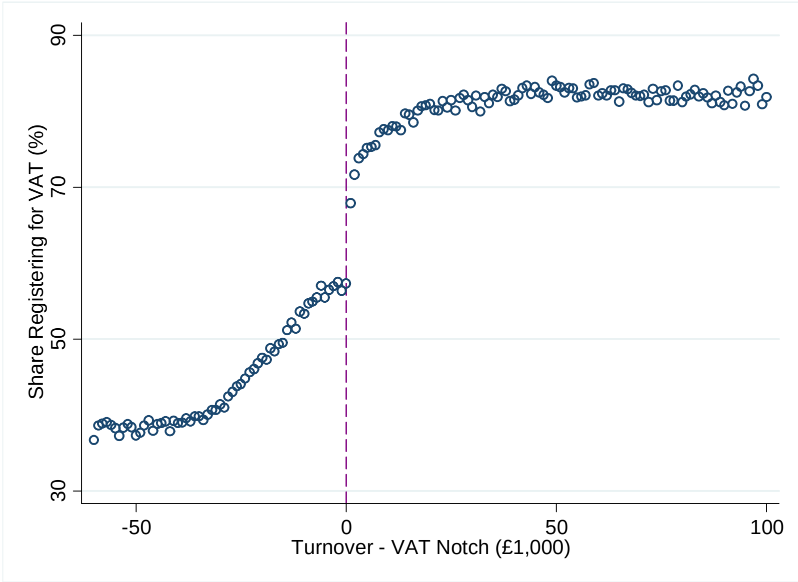
Slemrod, Joel. 2010. "Buenas Notches: Lines and Notches in Tax System Design." University of Michigan Working Paper.

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

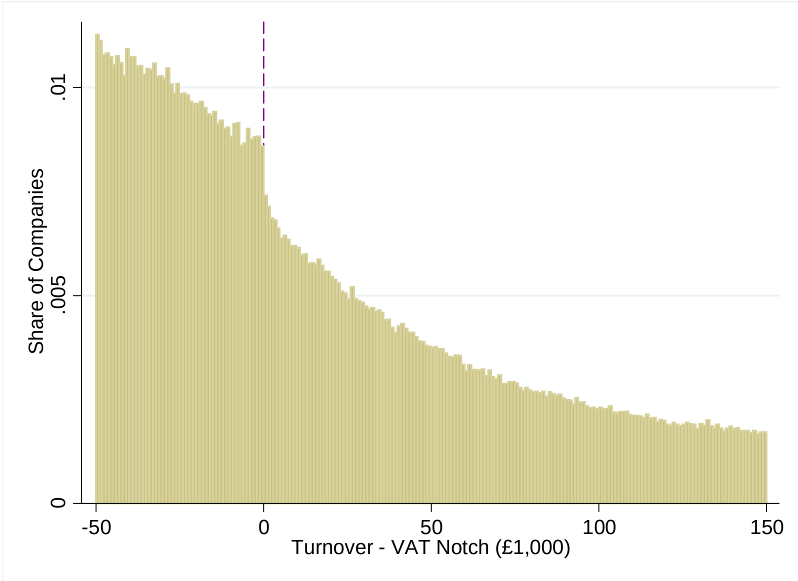
Wooldridge, Jeffrey M. 2001. *Econometric Analysis of Cross Section and Panel Data.* Vol. 1 of *MIT Press Books*, The MIT Press.

Figure 1. A BINDING VAT NOTCH

Panel A. Share of Registered Companies

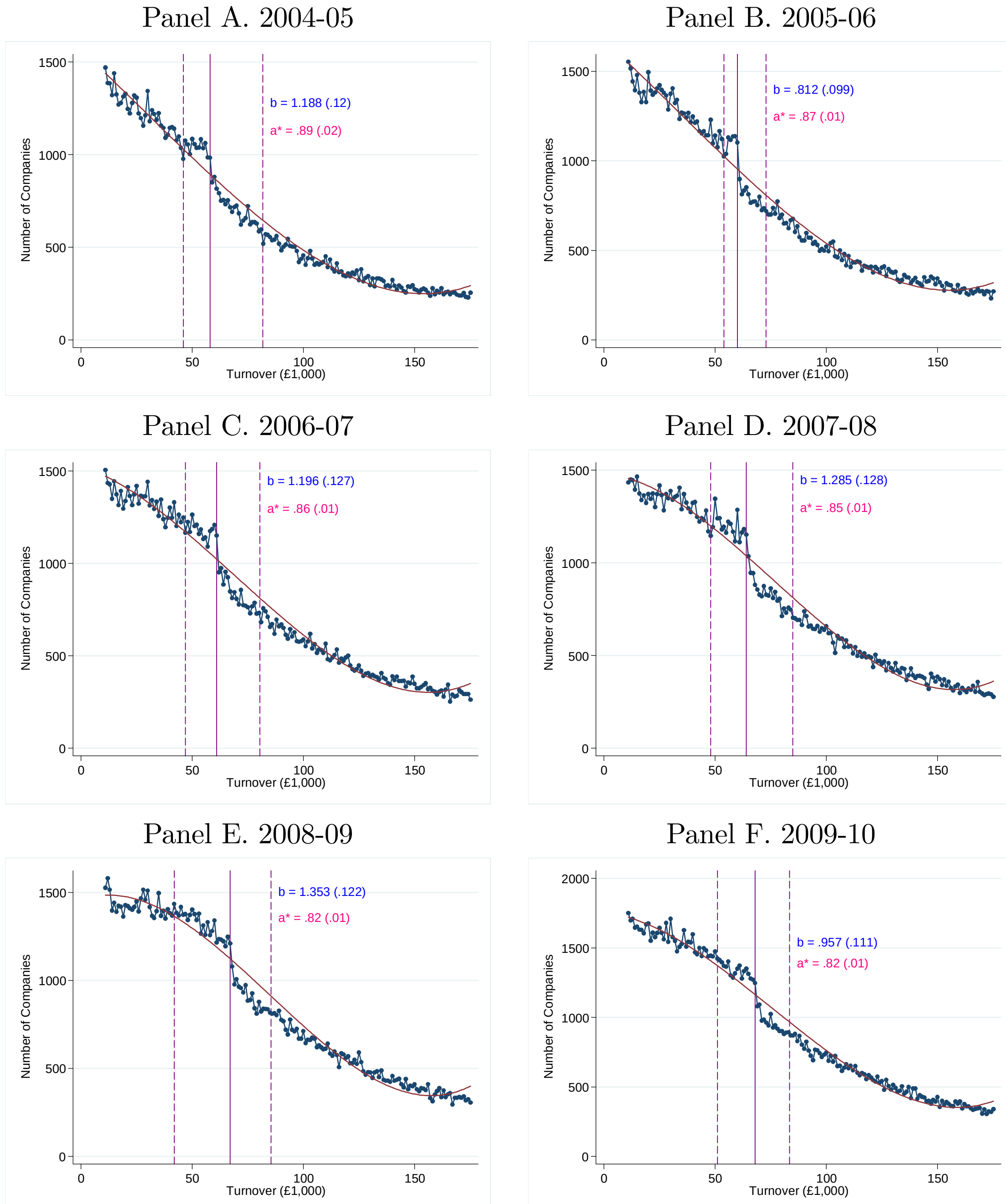


Panel B. Bunching Below the VAT Notch



Notes: This figure shows a binding VAT registration threshold in the UK. Panel A shows the share of VAT-registered companies in the neighbourhood of normalized VAT notch during 2004/05-2009/10. Each observation represents the share of registered companies relative to the total number of companies within each turnover bin of £1,000, net of current-year VAT threshold. The dashed line indicates the normalized VAT notch. Panel B shows the histogram of companies within the neighbourhood of 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.

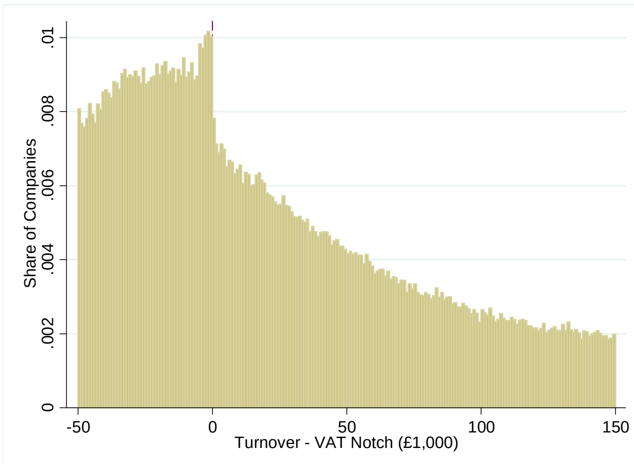
Figure 2. BUNCHING AT VAT NOTCH



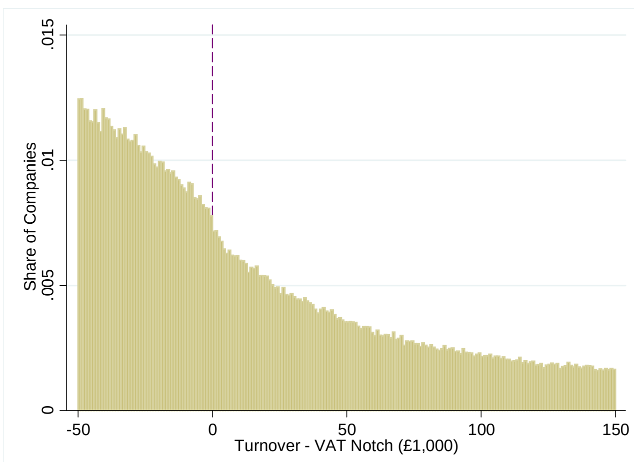
Notes: this figure shows the observed distribution (solid-dotted line) and the estimated counterfactual distribution (solid-smooth line) of turnover for each year in 2004/05-2009/10. The counterfactual is a three-order polynomial estimated as in eq. (17). The excluded ranges around the VAT notch are 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, and $a*$ is the proportion of companies with large adjustment costs locating in the strictly dominated region. Standard errors are shown in parentheses.

Figure 3. HETEROGENEITY IN BUNCHING

Panel A. Growing Companies



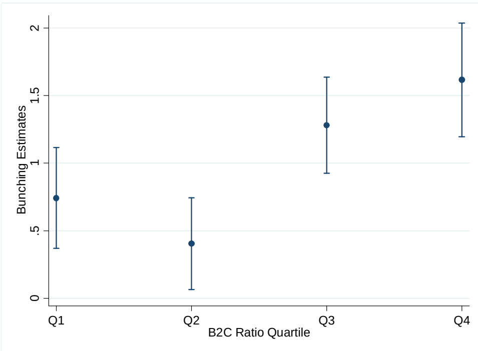
Panel B. Declining Companies



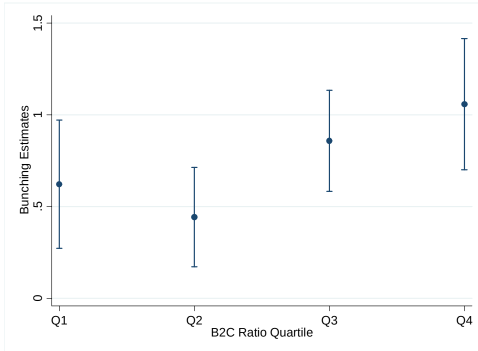
Notes: The figure shows the histogram of growing companies within the neighbourhood of normalized VAT notch between 2004/05-2009/10 in the top panel and that of declining companies in the bottom panel. The bin width is £1,000 and the dashed line denotes the normalized VAT notch.

Figure 4. BUNCHING ACROSS B2C SALES RATIO QUARTILE

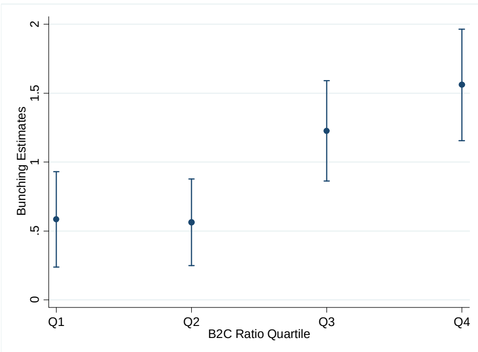
Panel A. 2004-05



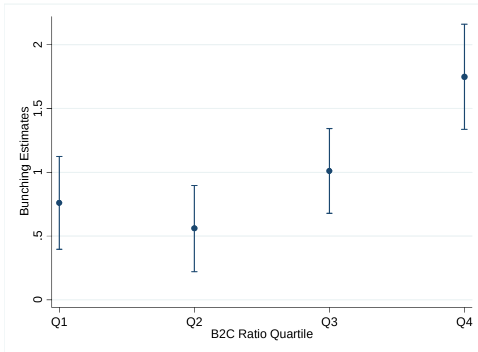
Panel B. 2005-06



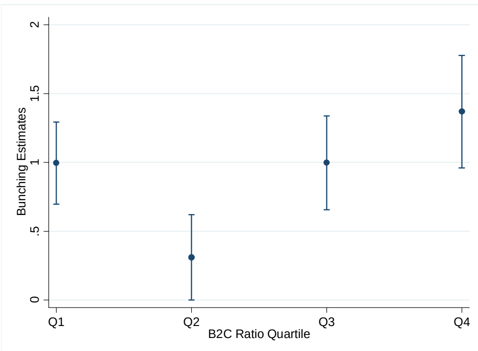
Panel C. 2006-07



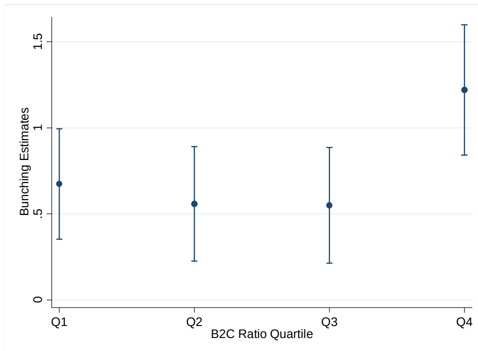
Panel D. 2007-08



Panel E. 2008-09



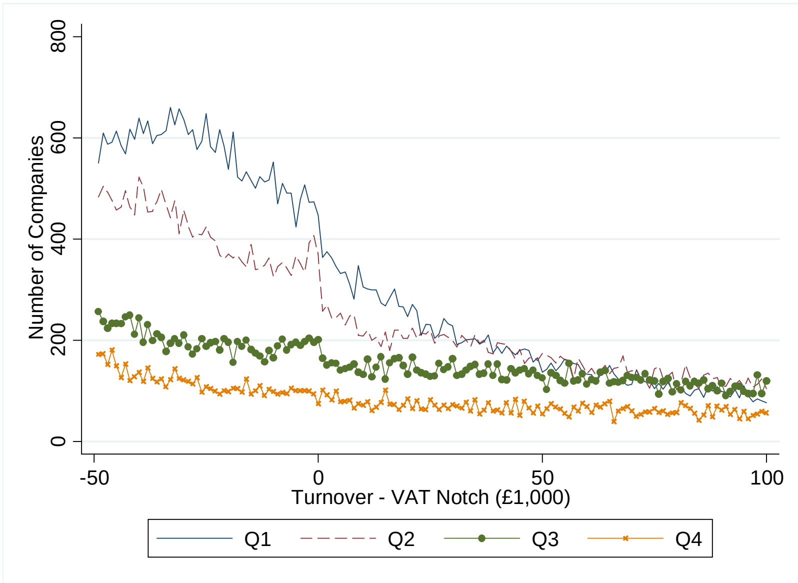
Panel F. 2009-10



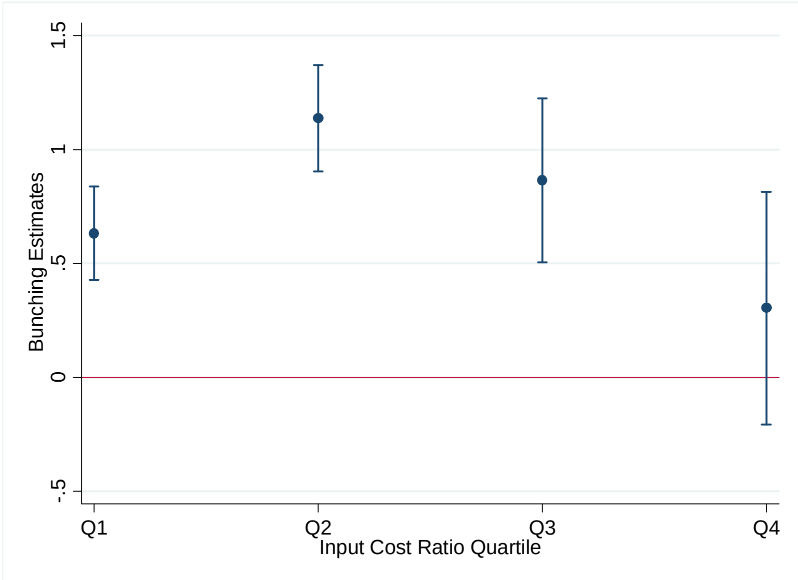
Notes: The figure plots the point estimate of the bunching ratio b with the corresponding 95% confidence intervals across four different quartiles of industry-level B2C sales ratio in each year during 2004/05-2009/10.

Figure 5. BUNCHING ACROSS INPUT-COST RATIO QUARTILE

Panel A. Bunching Evidence



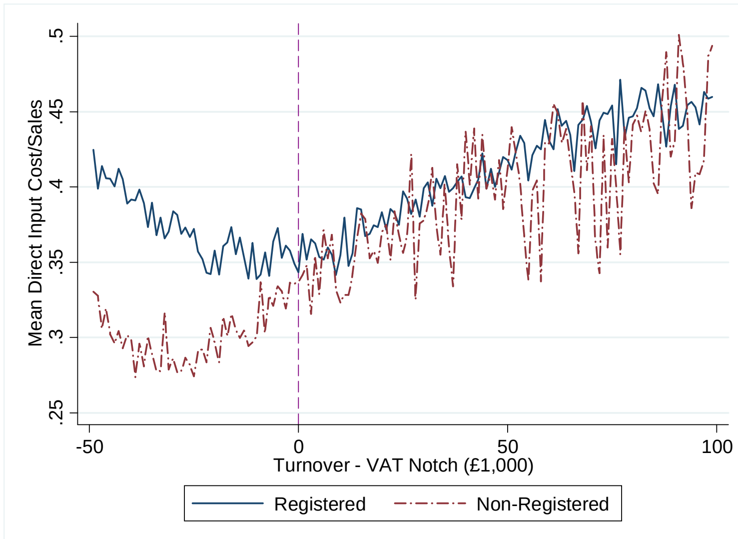
Panel B. Bunching Estimates



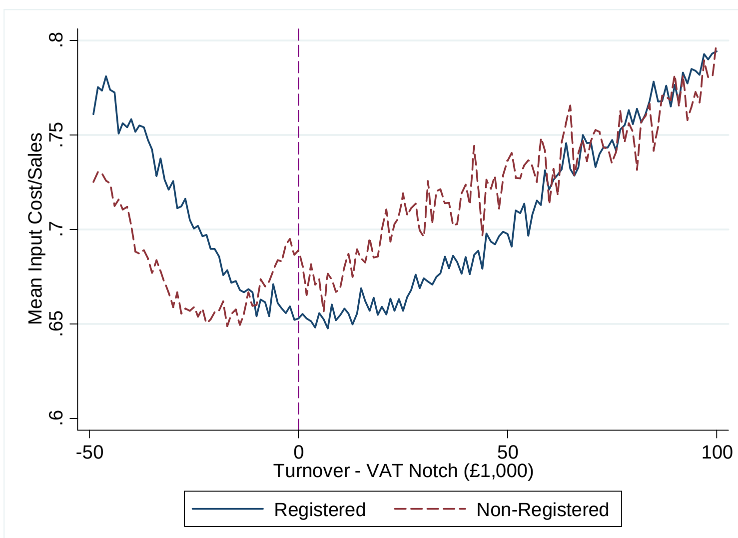
Notes: The figure shows the observed distribution of turnover across four different quartiles of input cost ratio within the neighbourhood of normalized VAT notch in 2004/05-2009/10 in Panel A. Panel B then plots the point estimate of the bunching ratio b and the corresponding 95% confidence intervals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

Figure 6. BUNCHING VIA TURNOVER MISREPORTING

Panel A: Distribution of Direct Input-Cost Ratio



Panel 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 while Panel uses the input cost ratio calculated from the corporation tax records and includes salary expenses in the overall cost.

Table 1. VALUE-ADDED TAX SCHEDULE IN THE UK

Fiscal Year	Registration Threshold (£)	Dereg. Threshold (£)	Standard Rate (%)	FRS Threshold (£)
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
2008-09				
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
2009-10				
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/deregistration threshold, FRS 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 2. SUMMARY STATISTICS

	All Compaines			Registered Sample			Non-Registered Sample			Test of Equal Means	
	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N	t statistics	*p*-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
CT600 Variables											
Turnover	73.86	49.33	805,351	86.80	49.58	500,355	52.64	40.84	304,996	-340	0
Trading Profit	21.03	32.47	805,351	24.33	31.72	500,355	15.62	32.97	304,996	-120	0
Cost Variables											
Input Cost: CT600	52.83	49.27	805,351	62.47	49.61	500,355	37.01	44.36	304,996	-240	0
Cost of Sales: FAME	34.46	116.80	95,629	39.59	62.61	62,575	24.76	178.62	33,054	-1.45	0.15
Other Key Variables											
Employment	7.46	187.21	12,467	7.89	218.51	7,121	6.89	134.66	5,346	-0.31	0.75
Industry-level B2C Ratio (%)	0.46	0.28	767,761	0.45	0.28	487,888	0.47	0.29	279,873	-28.51	0
Firm Age	8.77	9.92	805,351	9.02	9.57	500,355	8.36	10.45	304,996	22.53	0

Notes: This table shows the mean, standard error, and number of non-missing observations for companies with turnover between č10,000 and č200,000 in the sample. All monetary values are in nominal č1,000 GBP, with 1 GBP = 1.69 USD as of May 2014. Column 1-3 focus on the entire sample; column 4-6 focus on the registered companies and column 7-9 focus on the non-registered sample. Column 10-11 show t-statistic and associated p-value from a test of equal means for each variable between the registered (column 4) and non-registered (column 7) sample.

Table 3. SHARE OF FIRMS THAT VOLUNTARILY REGISTERED FOR VAT (%)

Ratio of B2C Sales\Input Cost (%)	Q1	Q2	Q3	Q4
Q1	46.49	46.7	45.12	46.03
Q2	51.59	47.79	48.76	49.24
Q3	26.23	31.83	35.08	42.55
Q4	29.15	33.67	36.46	47.27

Notes: This table shows the share of voluntarily registered firms at different quartiles of B2C sales ratio and input cost ratio, respectively. 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 for each input cost ratio quartile. Each row depicts the share of firms that are voluntarily registered for VAT at different quartiles of input cost ratio for each B2C sales ratio quartile.

Table 4. DETERMINANTS OF VAT VOLUNTARY REGISTERATION: FIRM-LEVEL REGRESSION

	Linear Probability Model					Fixed-Effects Logit		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of B2C Sales	-0.149***	0.004	0.003	-0.028*	-0.612***	0.253	0.208	-1.948***
	(0.003)	(0.004)	(0.004)	(0.015)	(0.011)	(0.176)	(0.178)	(0.839)
Input Cost Ratio	0.068***	0.006***	0.067***	0.034***	0.399***	0.416***	3.032***	1.836***
	(0.003)	(0.002)	(0.003)	(0.008)	(0.012)	(0.090)	(0.156)	(0.394)
Firm FE	N	Y	Y	Y	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Firm-level Controls	N	N	Y	Y	N	N	Y	Y
R-squared	0.007	0.003	0.006	0.004	0.005	0.039	0.072	0.057
N	515,333	515,333	515,333	60,070	515,333	30,364	30,364	2,115

Notes: *,**,*** denotes significance at 10%, 5% and 1%, respectively. Standard error presented in column (1)-(4) are clustered at firm level. Time fixed effects are always included.

Table 5. DETERMINANTS OF VAT VOLUNTARY REGISTERATION: INDUSTRY-LEVEL REGRESSION

	Pooled Regression				Fixed-Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of B2C Sales	-0.249***	-0.232***	-0.205***	-0.226*	0.033	0.030	0.063	0.085
	(0.048)	(0.049)	(0.041)	(0.043)	(0.076)	(0.077)	(0.058)	(0.061)
Input Cost Ratio	0.059***	1.372***	0.273***	0.178***	-0.206	0.024	-0.021	0.019
	(0.146)	(0.246)	(0.090)	(0.087)	(0.388)	(0.152)	(0.057)	(0.042)
Industry FE	N	N	N	N	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry-level Controls	N	Y	Y	Y	N	Y	Y	Y
R-squared	0.156	0.189	0.107	0.134	0.068	0.085	0.102	0.169
N	352	352	344	344	352	352	344	344

Notes: *,**,*** denotes significance at 10%, 5% and 1%, respectively. Heteroskadasticity-robust standard error are clustered at industry level. Industry-level controls include industry averages of trading profit and firm age. The industry-average input cost ratio in column (4) and (8) are shares of direct costs of sales in total costs reported in company accounts in FAME.

Table 6. Structural Elasticity Estimates

δ	σ	elasticity of turnover w.r.t d_R,d_N	elasticity of the VAT base w.r.t. t
2	0	0.351	0.099
2	1	0.341	0.098
2	∞	0.231	0.090
10	0	0.954	0.143
10	1	1.002	0.146
10	∞	0.370	0.100

Notes: This table shows estimates of the elasticity of turnover with respect to the VAT "discouragement index" and the elasticity of value-added with respect to the VAT rate at varying value of δ and σ. δdenotes the elasticity of demand for varieties and takes value of 2 or 10. σ denotes the elasticity of substitution between labour and inputs and takes value of 0, 1, or ∞. d_R and d_N are the VAT "sufficient statistic" or "discouragement index" and t is the VAT rate for all registered firms.

A Appendix

Proof of Proposition 1. (i) Let $\Delta(a) = u_R(a) - u_N(a)$. By the envelope theorem,

$$\frac{d\Delta}{da} = \frac{1}{a^2}\left(\phi'\left(\frac{y_R}{a}\right)y_R - \phi'\left(\frac{y_N}{a}\right)y_N\right)$$

Next, note that if $y_R \leq y_N$, this implies $t_R \geq t_N$ from (6). So, in this case, $u_R(a) - u_N(a) < 0$; if a firm has lower output due to a higher tax, and pays a registration cost, it must be worse off registering. So, if $\Delta(a) = u_R(a) - u_N(a) \geq 0$, it must be the case that $y_R > y_N$. But then

$$\frac{d\Delta}{da} > 0 \Longleftrightarrow \phi'\left(\frac{y_R}{a}\right)y_R > \phi'\left(\frac{y_N}{a}\right)y_N \Longleftrightarrow \phi'\left(\frac{y}{a}\right)y \text{ increasing in } y \Longleftrightarrow \phi'' > 0$$

So, we conclude that $\Delta(a) \geq 0$ implies that $\frac{d\Delta}{da} > 0$. This in turn, plus continuity of $\Delta(a)$ in a, implies that there is exactly one root $\Delta(\tilde{a}) = 0$, with $\Delta(a) > 0, a > \tilde{a}, \Delta(a) < 0, a < \tilde{a}$. $\square$.

Proof of Proposition 2. (i) Assume first that $s/\lambda < \frac{t}{1+t}$. It is easily verified that $s/\lambda < \frac{t}{1+t} \Longleftrightarrow t_R > t_N$. Assume now to the contrary that $\sigma > 0$; from (7), this requires $\tilde{a} < a_R$. Then, there is a firm $\tilde{a} < a' < a_R$ that when registered, produces strictly less than $y_R(a') < y^*$, and who would produce more if non-registered. Then, for this firm

$$\begin{aligned} u_R(a') &= y_R(a')(1-s)(1-t_R) - \phi\left(\frac{y_R(a')}{a'}\right) \\ &< y_R(a')(1-s)(1-t_N) - \phi\left(\frac{y_R(a')}{a'}\right) \\ &\leq \max_{y \leq y_R(a')}\left\{y(1-s)(1-t_N) - \phi\left(\frac{y}{a'}\right)\right\} \\ &\leq \max_{y \leq y^*}\left\{y(1-s)(1-t_N) - \phi\left(\frac{y}{a'}\right)\right\} = u_N(a') \end{aligned}$$

So, this firm is better off not registering, a contradiction.

(ii) Now suppose that $s/\lambda > \frac{t}{1+t} \Longleftrightarrow t_R < t_N$. Suppose that $\sigma \geq 0$ i.e. $\tilde{a} \leq a_R$. Then, as $t_R < t_N, \tilde{a} \leq a_R$;

$$y_N(\tilde{a}) < y_R(\tilde{a}) \leq y_R(a_R) = y^*$$

So, then, the constraint $y \leq y^*$ in the definition of $u_N(\tilde{a})$ does not bind, and so by straightforward computation, we have

$$u_N(\tilde{a}) = \max_y\left\{y(1-s)(1-t_N) - \phi\left(\frac{y}{a}\right)\right\} = \frac{\tilde{a}^{1+e}(x_N)^{1+e}}{1+e}$$

and, in the same way, $u_R(\tilde{a}) = \frac{\tilde{a}^{1+e}(x_R)^{1+e}}{1+e}$. So, then $\tilde{a}$ is characterized by

$$\frac{\tilde{a}^{1+e}(x_R)^{1+e}}{1+e} - \frac{\tilde{a}^{1+e}(x_N)^{1+e}}{1+e} - K = 0$$

or, solving:

$$\tilde{a} = \left(\frac{K(1+e)}{(x_R)^{1+e} - (x_N)^{1+e}}\right)^{1/(1+e)} \tag{18}$$

Substituting (18) in (7), we get

$$\begin{aligned}
\sigma &= \max\left\{\frac{\left(\frac{y^*}{(x_R)^e}\right)^{1/(1+e)} - \left(\frac{K(1+e)}{(x_R)^{1+e}-(x_N)^{1+e}}\right)^{1/(1+e)}}{\left(\frac{y^*}{(x_R)^e}\right)^{1/(1+e)}}, 0\right\} \qquad (19)\\
&= \max\left\{\frac{(y^*)^{1/(1+e)} - \left(\frac{(x_R)^e K(1+e)}{(x_R)^{1+e}-(x_N)^{1+e}}\right)^{1/(1+e)}}{(y^*)^{1/(1+e)}}, 0\right\}\\
&= \max\left\{\frac{(y^*)^{1/(1+e)} - \left(\frac{K(1+e)}{\beta}\right)^{1/(1+e)}}{(y^*)^{1/(1+e)}}, 0\right\}
\end{aligned}$$

Inspection of (19) reveals that $\sigma \geq 0$ for $\beta \geq \beta_0 = \frac{K(1+e)}{y^*}$, and that σ is increasing in β, as required.

(iii) First consider an increase in λ. We require $\frac{\partial\beta}{\partial\lambda} < 0$ for the result. But

$$\begin{aligned}
\frac{\partial\beta}{\partial\lambda} &= \left((1+e) - e\frac{(x_R)^{1+e} - (x_N)^{1+e}}{(x_R)^{1+e}}\right)\frac{\partial x_R}{\partial\lambda}\\
&= -\left((1+e) - e\left(1 - \left(\frac{x_N}{x_R}\right)^{1+e}\right)\right)\frac{t}{1+t}\\
&= -(1 + \left(\frac{x_N}{x_R}\right)^{1+e} e)\frac{t}{1+t} < 0
\end{aligned}$$

as required. Next, consider an increase in s. We need $\frac{\partial\beta}{\partial s} > 0$ for the result. But

$$\begin{aligned}
\frac{\partial\beta}{\partial s} &= \left((1+e) - e\frac{(x_R)^{1+e} - (x_N)^{1+e}}{(x_R)^{1+e}}\right)\frac{\partial x_R}{\partial s} - (1+e)\left(\frac{x_N}{x_R}\right)^e\frac{\partial x_N}{\partial s}\\
&= -(1 + e\left(\frac{x_N}{x_R}\right)^{1+e}) + (1+e)\left(\frac{x_N}{x_R}\right)^e(1+t)
\end{aligned}$$

Now note that $\frac{\partial\beta}{\partial s} = (1+e)t > 0$ at $\frac{x_N}{x_R} = 1$, and $\frac{\partial\beta}{\partial s} = -1 < 0$ at $\frac{x_N}{x_R} = 0$. Moreover,

$$\begin{aligned} \frac{\partial^2\beta}{\partial s \partial\left(\frac{x_N}{x_R}\right)} &= -e(1+e)\left(\frac{x_N}{x_R}\right)^e + e(1+e)\left(\frac{x_N}{x_R}\right)^{e-1}(1+t) \\ &> e(1+e)\left(\frac{x_N}{x_R}\right)^e\left(\frac{x_R}{x_N} - 1\right) > 0 \end{aligned}$$

So, given these two facts, there is a critical value of $\frac{x_N}{x_R} = \frac{1-t_N}{1-t_R}$, say $\widehat{\frac{1-t_N}{1-t_R}}$ above which $\frac{\partial\beta}{\partial s} > 0$, as required. □

Proof of Proposition 3. First, a^* is the type that is just willing to produce at y^* when not registered i.e. from (4),

$$y^* = (a^*)^{1+e}(x_N)^e$$

where $x_i = (1-s)(1-t_i),\ i = R, N$. Also, substituting (4) back in (2), we see that the trader's utility at $y_R(a)$, not including compliance costs, is

$$\begin{aligned} u_R(a) &= y_R(a)x_R - \frac{1}{1+\frac{1}{e}}\left(\frac{y_R(a)}{a}\right)^{1+1/e} \\ &= a^{1+e}(x_R)^{1+e} - \frac{1}{1+\frac{1}{e}}\left(a^e(x_R)^e\right)^{1+1/e} \\ &= \frac{\tilde{a}^{1+e}(x_R)^{1+e}}{1+e} \end{aligned} \tag{20}$$

Note that the payoff from being on the threshold for an $a-$type when constrained is

$$\tilde{u}_N(a) = y^* x_N - \frac{1}{1+\frac{1}{e}}\left(\frac{y^*}{a}\right)^{1+1/e} \tag{21}$$

Then the indifference condition that determines $a^*+\Delta a^*$ is $\tilde{u}_N(a^*+\Delta a^*) = u_R(a^*+\Delta a^*) - K$. But note that $(a^*+\Delta a^*)$, which is unobservable, maps into $y^*+\Delta y^*$, which is observable, as follows:

$$\begin{aligned} y^* + \Delta y^* &= y_R(a^*+\Delta a^*) = (a^*+\Delta a^*)^{1+e}(x_R)^e \rightarrow \\ a^* + \Delta a^* &= (y^*+\Delta y^*)^{1/(1+e)}(x_R)^{-e/(1+e)} \end{aligned} \tag{22}$$

So, using (22) in (20), we get:

$$u_R(a^*+\Delta a^*) = (y^*+\Delta y^*)x_R\left(\frac{1}{1+e}\right)$$

Next, we have

$$\begin{aligned}
\tilde{u}_N(a^* + \Delta a^*) &= y^* x_N - \frac{1}{1+\frac{1}{e}}\left(\frac{y^*}{a^* + \Delta a^*}\right)^{1+1/e} \\
&= y^* x_N - \frac{1}{1+\frac{1}{e}}(y^*)^{1+1/e}\left((y^* + \Delta y^*)^{1/(1+e)}\,(x_N)^{-e/(1+e)}\right)^{-(e+1)/e} \\
&= y^* x_N - \frac{1}{1+\frac{1}{e}}(y^*)^{1+1/e}(y^* + \Delta y^*)^{-1/e} x_N
\end{aligned}$$

Then, the indifference condition $\tilde{u}_N(a^* + \Delta a^*) = u_R(a^* + \Delta a^*)$ becomes

$$y^* x_N - \frac{1}{1+\frac{1}{e}}(y^*)^{1+1/e}(y^* + \Delta y^*)^{-1/e} x_N - (y^* + \Delta y^*) x_R \left(\frac{1}{1+e}\right) + K = 0$$

After some simplification (divide through by y^*, then $1 + \frac{\Delta y^*}{y^*}$, x_N) we get

$$\frac{1}{(1+\Delta y^*/y^*)}\left(1 + \frac{\kappa}{(1-s)(1-t_N)}\right) - \frac{1}{1+1/e}\left[\frac{1}{1+\Delta y^*/y^*}\right]^{1+1/e} - \frac{1-t_R}{1-t_N}\frac{1}{1+e} = 0,\ \kappa = \frac{K}{z^*}$$

as required. □

Proof of Proposition 4. Assume $K = 0$. First, (9) can be rewritten as

$$\frac{1-t_R}{1-t_N} = (1+e)\left(x - \frac{1}{1+1/e}x^{1+1/e}\right) \equiv f(x),\ x = \frac{1}{(1+\Delta y^*/y^*)}$$

Moreover, note that $f'(x) = (1+e)(1-x^{1/e})$. As $x < 1$, and $e > 0$, $x^{1/e} < 1$ and so, f is increasing in x and thus decreasing in Δy^* at a fixed y^*. Finally,

$$\frac{1-t_R}{1-t_N} = \frac{1 - s - \lambda t/(1+t)}{1 - s(1+t)}$$

By inspection, $\frac{1-t_R}{1-t_N}$ is decreasing in λ. Also, it can be shown to be increasing in s;

$$\begin{aligned}
\frac{\partial\left(\frac{1-t_R}{1-t_N}\right)}{\partial s} &= \frac{1}{(1-s(1+t))^2}\left(\left(1 - s - \frac{\lambda t}{(1+t)}\right)(1+t) - 1 + s(1+t)\right) \\
&= \frac{1}{(1-s(1+t))^2} t(1-\lambda) > 0
\end{aligned}$$

So, putting all this together, as λ increases, $\frac{1-t_R}{1-t_N}$ falls, x falls and thus Δy^* rises. In the same way, as s rises, Δy^* falls. The above arguments generalize to the case of a sufficiently

small K as then we still have $f'(x) > 0$. □

Proof of Proposition 5. (i) As firms are indifferent about registering or not at the cutoff, by the envelope theorem,

$$\begin{aligned}\frac{dU}{dt} &= -(1-s)\frac{\partial t_N}{\partial t}\int_0^{\tilde{a}} y_N(a)da - (1-s)\frac{\partial t_R}{\partial t}\int_{\tilde{a}}^{a_{\max}} y_R(a)da \\ &= -\frac{\partial t_N}{\partial t}V_N - \frac{\partial t_R}{\partial t}V_R\end{aligned} \tag{23}$$

Moreover, from (11), we have:

$$\frac{dT}{dt} = \frac{\partial t_N}{\partial t}V_N + \frac{\partial t_R}{\partial t}V_R + \underbrace{\left(t_N\frac{\partial V_N}{\partial t} + t_R\frac{\partial V_N}{\partial t}\right)|_{\tilde{a}\text{ const}}}_{\text{intensive response}} + \underbrace{\frac{\partial T}{\partial \tilde{a}}\frac{\partial \tilde{a}}{\partial t}}_{\text{extensive response}} \tag{24}$$

So, combining (23), (24), we get (12).

(ii) Using $y_i(a) = a^{1+e}(x_i)^e,\ i = N, R$, we get

$$V_N = \int_0^{\tilde{a}} (1-s)^{1+e}(1-t_N))^e a^{1+e}da,\ V_R = \int_{\tilde{a}}^{a_{\max}} (1-s)^{1+e}(1-t_R))^e a^{1+e}da$$

and so

$$\frac{\partial V_N}{\partial t} = -e\frac{V_N}{1-t_N}\frac{\partial t_N}{\partial t},\ \frac{\partial V_R}{\partial t} = -e\frac{V_R}{1-t_N}\frac{\partial t_R}{\partial t} \tag{25}$$

Substituting (25) into $t_N\frac{\partial V_N}{\partial t} + t_R\frac{\partial V_N}{\partial t}$ and dividing by $\frac{\partial t_N}{\partial t}V_N + \frac{\partial t_R}{\partial t}V_R$ gives (13).

(iii) Moreover, from the formula for $\tilde{a}$ in (18), it is easy to check that

$$\frac{\partial T}{\partial \tilde{a}}\frac{\partial \tilde{a}}{\partial t} = \tilde{a}\frac{\partial T}{\partial \tilde{a}}\frac{(x_R)^e\frac{\lambda}{(1+t)^2} - (x_N)^e s}{(x_R)^{1+e} - (x_N)^{1+e}} \propto K^{1/(1+e)}$$

as required. □