

VAT NOTCHES, VOLUNTARY REGISTRATION, AND BUNCHING: THEORY AND U.K. EVIDENCE

Li Liu, Ben Lockwood, Miguel Almunia, and Eddy H. F. Tam*

Abstract—Using administrative tax records for U.K. businesses, we document both bunching in annual turnover below the VAT registration threshold and persistent voluntary registration by almost half of the firms below the threshold. We develop a conceptual framework that can simultaneously explain these two apparently conflicting facts. The framework also predicts that higher intermediate input shares, lower product-market competition, and a lower share of business to consumer sales lead to voluntary registration. The predictions are exactly the opposite for bunching. We test the theory using linked VAT and corporation tax records from 2004 to 2014, finding empirical support for these predictions.

I. Introduction

MOST countries use the value-added tax (VAT) as their primary indirect tax. It is standard to set a minimum registration threshold, usually based on annual turnover, below which businesses do not need to register for VAT. In the EU, a large majority of countries currently have a registration threshold, with the U.K. threshold being the highest at £85,000 (\$110,000).¹ As VAT rates are often quite high (in excess of 20% in many EU countries), this may create a large and salient tax notch for businesses whose turnover is around the threshold, depending on firm characteristics, as we explain below. The effects of these VAT notches on firm behavior have received little attention in the existing literature.

In this paper, we study the behavior of firms around the VAT registration threshold theoretically and empirically, using administrative data on U.K. corporations. We begin by documenting two stylized facts. First, we find substantial bunching below the registration threshold, as some firms restrict their

reported turnover to avoid having to register for the VAT. Second, we observe that a large fraction of firms with turnover below the threshold are registered for VAT (on average, 43% over the period 2004 to 2014). This behavior seems to be deliberate rather than accidental. For example, about half of the firms initially registered with turnover below the threshold in 2004/2005 were still registered and below the threshold three years later. We refer to this behavior as *voluntary registration*.

How can we explain the coexistence of voluntary VAT registration and bunching at the threshold? We develop a simple general equilibrium model that can explain both of these phenomena in a unified way. Our first observation is that for both behaviors to occur simultaneously within a given sector, firms in that sector must make sales to both final consumers (B2C sales) and sales to other VAT-registered businesses (B2B sales). Moreover, these firms must themselves use intermediate inputs in production. To see this, suppose that firms make only B2C sales. Then it is easily shown that irrespective of the degree of competition among firms, the cost of voluntary registration exceeds the benefit, because the burden of VAT charged on output when registered exceeds the burden of VAT paid on inputs when not registered.² Conversely, with only B2B sales, voluntary registration is always optimal, absent compliance costs, because the burden of output VAT can be passed on to the buyer, while the firm can claim back VAT paid on inputs.

We present the most parsimonious model that can explain both voluntary registration and bunching simultaneously. We show that this model must feature three stages of production, because the firms must sell some of their output to other VAT-registered firms and buy inputs that bear VAT. The model yields three key predictions. First, we show that voluntary registration by a firm is more likely when (a) the cost of inputs relative to sales is high; (b) the proportion of B2C sales is low; and (c) markets are less competitive, that is, firms have higher mark-ups. The intuition for prediction a is simply that when input costs are important, registration allows the firm to claim back a considerable amount of input VAT. The intuition for prediction b is that if most customers are VAT-registered, the burden of an increase in VAT can easily be passed on in the form of higher prices because more customers can claim back the increase. The intuition for prediction c is that in a more competitive market, it is more difficult to pass the burden of output VAT onto buyers. We then show that the determinants of bunching at the registration threshold are the same as for voluntary registration, with the signs of the effects reversed.

Received for publication June 19, 2019. Revision accepted for publication October 16, 2019. Editor: Rohini Pande.

*Liu: International Monetary Fund and Oxford University Centre for Business Taxation; Lockwood: University of Warwick, Oxford CBT, and CEPR; Almunia: Colegio Universitario de Estudios Financieros (CUNEF), CEPR, CAGE, and Oxford CBT; Tam: Oxford CBT.

The first version of this paper was circulated under the title “VAT Notches” (CEPR discussion paper 10606, May 2015). We thank the staff at Her Majesty’s Revenue and Customs (HMRC) for access to the data and their support of this project. This work contains statistical data from HMRC that is Crown Copyright. The research data sets used may not exactly reproduce HMRC aggregates. The use of HMRC statistical data 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 thank Michael Devereux, Judith Freedman, Chris Heady, James Hines, Louis Kaplow, Henrik Kleven, Tuomas Matikka, Joel Slemrod, and numerous seminar participants for helpful comments. We thank Dongxian Guo and Omiros Kouvas for excellent research assistance. The views expressed in this paper are our own and do not necessarily represent the views of the IMF. We acknowledge financial support from the ESRC under grant ES/L000016/1.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00884.

¹There is a positive registration threshold in all but five EU countries: Greece, Hungary, Malta, Spain, and Sweden. For details, see www.vatlive.com.

²This is proved very generally in online appendix A.

The main empirical contribution of the paper is to test the main predictions of our model. To do so, we use an administrative data set that links the population of VAT and corporation tax records in the United Kingdom for the period 2004 to 2014. One advantage of studying this VAT system is that businesses below the registration threshold are not subject to any other tax to replace the VAT, as is the case in many other countries.³ Thus, the United Kingdom allows us to perform a clean test of our theoretical predictions.

In the empirical analysis, we first show that the pattern of voluntary registration in the data is consistent with the theoretical predictions. We estimate that the probability that a firm voluntarily registers for VAT is increased by 4 percentage points for a 1 standard deviation (SD) decrease in the share of B2C sales, by 4.9 percentage points for a 1 SD increase in the input cost ratio (ICR), and by 3.9 percentage points for a 1 SD decrease in the Lerner index of competition. The results are robust to the use of either a linear probability model or a fixed-effects logit model and to the inclusion of additional firm-level control variables such as the distance to the VAT threshold.

We then look at bunching. In the aggregate, there is clear evidence of bunching at the VAT threshold. To test the predictions from the model, we partition the sample into two groups of firms based on their predicted likelihood of registering voluntarily, regardless of whether their turnover is above or below the threshold. Focusing on the subset of firms that are less likely to register voluntarily, we find that the bunching response is larger when (a) the proportion of B2C sales is high, (b) the cost of inputs relative to sales is low, or (c) the Lerner index of their industry is high. In contrast, the bunching patterns for firms that are highly likely to register voluntarily are unrelated to those three characteristics, as expected. Thus, we conclude that the heterogeneous bunching patterns are consistent with the theoretical predictions.

We further investigate some of the mechanisms underlying the registration decision. There is some suggestive evidence that part of the bunching response is driven by evasion through sales underreporting.⁴ We also analyze changes in registration status in a dynamic regression setting to address the possibility that voluntary registration may not be an optimizing choice of firms, but simply a failure to deregister due to inertia (or high deregistration costs). Our empirical findings suggest that while there is some persistence in firm behavior, the decision is not entirely driven by inertia.⁵

This paper contributes to the literature on the effects of tax and regulatory thresholds and, in particular, the effect of VAT thresholds on small business behavior. In an important paper, Keen and Mintz (2004) were the first to set up a model of VAT including a threshold, showing that there will be bunching below the threshold. However, there are a number of dif-

ferences between their approach and ours. First, their model only allows for final consumer sales, which cannot by itself explain voluntary registration, as argued above. Second, their main focus is on the optimal registration threshold, whereas our focus is on the coexistence and determinants of voluntary registration and bunching. A simplified version of our model without B2B is closely related to the framework in Keen and Mintz (2004).⁶

Second, there is a related literature on the relationship between the VAT and business informality in developing countries (Emran & Stiglitz, 2005; de Paula & Scheinkman, 2010). In particular, de Paula and Scheinkman (2010) present a model where firms can choose between formal and informal production, where the distinction is whether they must register for VAT. Firms can also choose to buy inputs from formal or informal suppliers. In their model, informality can be interpreted as producing below a VAT threshold. Moreover, because they model two stages in the chain of production (upstream and downstream firms), they allow for both B2C and B2B sales. Nevertheless, their model is not really suited to our task because voluntary registration cannot occur in equilibrium, as there are only two stages of production. This is demonstrated formally in online appendix A.

Our paper is also related to a small but growing empirical literature on the effects of VAT registration thresholds in different countries, such as Japan (Onji, 2009), Armenia (Asatryan & Peichl, 2016), and Finland (Harju, Matikka, & Rauhanen, 2019). The Finland study documents substantial bunching below the VAT threshold in that country and finds that compliance costs are the main driver. Such costs are likely to be less relevant in the U.K. setting because the threshold in Finland is set at a much lower level (€8,500, roughly \$9,800) and the VAT system in the United Kingdom is relatively simple. Moreover, our paper is the first in this literature to study the determinants of voluntary VAT registration. In terms of the estimation methodology, the paper relates to the broader literature on the behavioral responses to tax notches (Kleven & Waseem, 2013; Best & Kleven, 2018; Almunia & Lopez-Rodriguez, 2018; Bachas & Soto, 2018).

The rest of the paper proceeds as follows. Section II establishes the two stylized facts of bunching and voluntary registration. Section III sets up the conceptual framework. Section IV derives the main empirical predictions. Section V provides an overview of the VAT system in the United Kingdom and describes the data. Sections VIA and VIB present the empirical analysis for voluntary registration and bunching, respectively. Section VII discusses the implications of our analysis for the setting of the VAT threshold in practice. Finally, section VIII concludes.

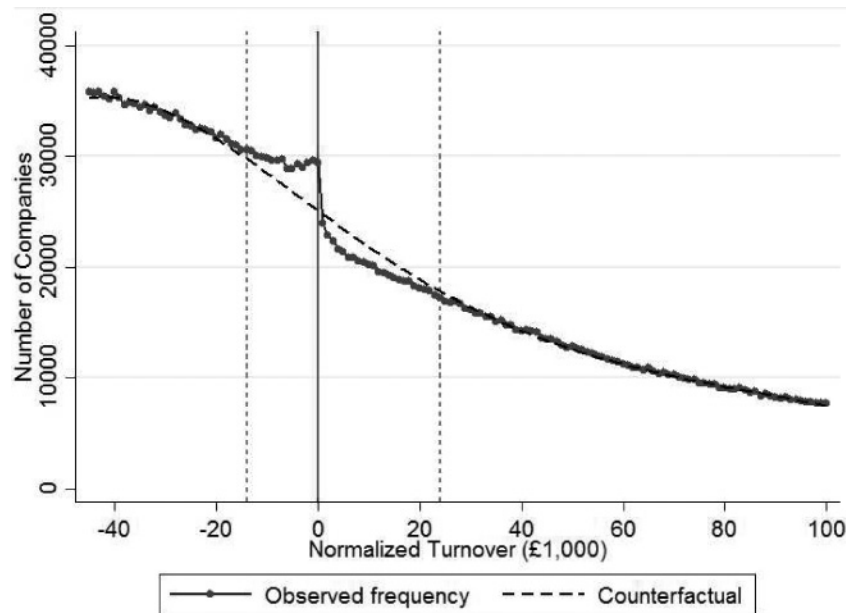
⁶The main differences are that we assume constant returns to scale and monopolistic competition, whereas they assume decreasing returns to scale, perfect competition, and price-taking firms. Under both sets of assumptions, when there are no B2B sales, the firms bear the burden of output VAT to a point where voluntary registration is not desirable. Also related is Kanbur and Keen (2014), who extend the Keen and Mintz (2004) framework to allow for VAT evasion and avoidance.

³These businesses may still be liable for corporation tax (if they are incorporated) or income tax (if they are sole proprietorships or other kinds of businesses).

⁴The details of this exercise are reported in online appendixes B and D.

⁵See online appendix E for details.

FIGURE 1.—TURNOVER DISTRIBUTION AROUND THE VAT REGISTRATION THRESHOLD



This figure shows the histogram of companies' turnover net of current-year VAT registration threshold (normalized VAT notch) by pooling data between 2004/2005 and 2014/2015. The bin width is £1,000 and the vertical red line denotes the normalized VAT notch. The dashed line is a counterfactual density fitted by excluding bins around the VAT notch.

II. Two Stylized Facts

To motivate the theoretical analysis, we present two key stylized facts from our U.K. administrative data set, which we describe in more detail in section VB. Here, we just note that we have more than 3.4 million turnover observations above and below the VAT threshold, for almost 1 million unique companies between April 2004 and April 2015.⁷

The first fact is bunching of firm turnover just below the VAT registration threshold. We normalize turnover by subtracting the threshold value so that the threshold is located at 0 in any year. The histogram of normalized turnover is shown in figure 1, where there is clear evidence of excess mass to the left of the registration threshold in an otherwise smooth distribution. The figure indicates that the VAT registration threshold is binding for at least a subset of U.K. firms.

It is worth noting that the bunching spike is not as sharp as in other bunching studies, in particular, those studying firms' responses to notches (Best et al., 2015; Harju, Matikka, & Rauhanen, 2019; Almunia & Lopez-Rodriguez, 2018). One possible explanation is that firms that benefit from voluntary registration do not respond in any way to the location of the threshold. Another potential reason is measurement error, because the registration threshold is set in terms of VAT taxable turnover, but we measure turnover from corporation tax records, where the definition is slightly different.⁸ The latter suggests that our measures of bunching should be interpreted as a lower bound of the true behavioral response.

⁷The U.K. fiscal year begins (and ends) in early April.

⁸VAT taxable turnover does not include sales of exempt goods and exports.

The second stylized fact is that in any given year, a significant number of firms are registered for VAT even though their turnover is below the threshold in the current year. On average, over our sample period, 43% of firms below the threshold are in this position. Possibly, part of this may be due to the rules of registration. In the United Kingdom, a business must register for VAT if its taxable turnover is likely to go over the threshold in the next thirty days, or if its taxable turnover in the previous twelve months was above the threshold. So, for example, a firm may register on the basis of previous year's turnover, and then its turnover may fall below the threshold in the current year.

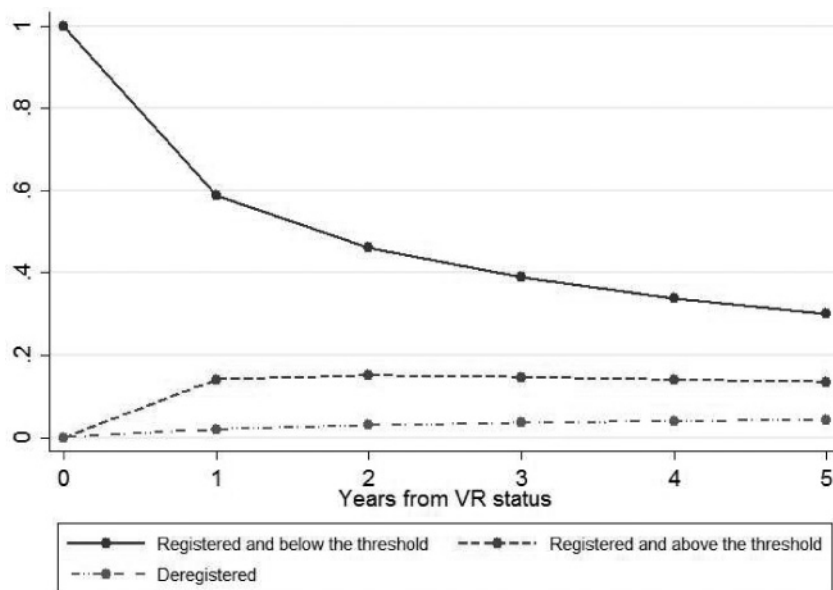
However, there is considerable persistence in registration below the threshold. Figure 2 shows what happens to firms initially registered and below the threshold during fiscal year 2004/2005. Almost half are still registered three years later, and over one-third are still registered five years later. So, it is likely that registration below the threshold is a conscious decision by firms, rather than just due to inability to forecast turnover one year in advance or inertia. We refer to this stylized fact as *voluntary registration*.

III. Conceptual Framework

A. Key Features of the Model

We aim to model the behavior of "small" firms (those with turnover around the threshold) selling to both final consumers and to businesses, where both voluntary registration and bunching can be equilibrium outcomes. The coexistence of these two behaviors requires that the small firms make both

FIGURE 2.—PERSISTENCE OF REGISTRATION BELOW THE THRESHOLD



This figure plots the transition probability for firms voluntarily registered during the period 2004/2005 to 2009/2010, to remain registered and below the threshold, registered and above the threshold or deregistered, in the following five years. Firms leave the sample when they are dissolved or become part of a larger VAT group in following years.

B2B and B2C sales and that they buy produced inputs. So, as already noted, the model must have (at least) three stages of production. Second, we wish to study the effect of the input cost ratio, share of B2C sales, and level of industry competition on voluntary registration and bunching, so the model must incorporate parameters measuring these.

We construct the simplest general-equilibrium model that has the required features. There is a single representative consumer that supplies labor and buys two kinds of goods: a differentiated good sold by the small firms and a single good produced by a large downstream firm. The large firm also buys inputs from the small firms, generating a B2B demand. We assume that the large firm is operating at a scale where nonregistration for VAT (i.e., operating so that the value of sales is below the VAT threshold) is never profitable. Finally, a homogeneous input to the small firm is produced by a third sector of upstream competitive firms from a labor input via a constant-returns technology. The behavior of this last sector is summarized by a zero-profit condition, implying that the price of the small-firm input is equal to the wage. Note that there are three stages of production, for reasons already discussed.

B. Consumers

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

$$U(X) + V(Y) + l, \quad (1)$$

where X is a CES index of differentiated goods with elasticity of substitution $e_C > 1$ and

$$U(X) = \lambda^{1/\phi} \frac{X^{1-1/\phi}}{1-1/\phi},$$

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

Here, λ is a measure of the final demand by households for the goods produced by the small firms relative to demand for the goods produced by the large downstream firm; as such, it will be the parameter that measures B2C demand in what follows.

Each differentiated good a is produced by a single small firm a , which can be registered for VAT or not. We also allow the firm to price-discriminate between final and intermediate consumers, so let $p_C(a)$, $p_B(a)$ be the prices charged to final consumers and the large firm, respectively, for good a , excluding VAT.⁹

The household faces a budget constraint,

$$P(1+t)Y + \int_a^{\bar{a}} [p_C(a)(1+I(a)t)]x(a)da = w(1-l) + \Pi, \quad (2)$$

where $1-l$ is labor supply, w is the wage, P is the price of the homogeneous good produced by the large firm, and

⁹This is without loss of generality, as we will assume $e_C = e_B$ when conducting comparative statics. The more general case with $e_C \neq e_B$ is covered in a previous version of this paper (Liu et al., 2019).

$I(a) \in \{0, 1\}$ is an indicator for VAT registration. If the firm is registered, the consumer price is grossed up by VAT, that is, $p(a)(1 + t)$. Finally, Π is aggregate profit. Because utility is linear in leisure l , there are no income effects in demand for $x(a)$, Y , and so this term plays no further role.

By standard arguments, maximization of equation (1) subject to equation (2) gives household demand for the homogeneous and differentiated goods:

$$Y = (1 - \lambda)(1 + t)^{-\gamma} P^{-\gamma}, \quad (3)$$

$$x(a) = \lambda \left(\frac{p_C(a)(1 + I(a)t)}{Q} \right)^{-\epsilon_C} Q^{-\phi}, \quad (4)$$

where Q is the CES price index corresponding to the quantity index X :

$$Q = \left[\int_{\underline{a}}^{\bar{a}} (p_C(a)(1 + I(a)t))^{1-\epsilon_C} da \right]^{1/(1-\epsilon_C)}. \quad (5)$$

We assume that in equilibrium, positive leisure is consumed. Then, from equation (1), the wage is fixed at the marginal utility of leisure, which is equal to 1.

C. The Large Firm

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

$$C = \left[\int_{\underline{a}}^{\bar{a}} (p_B(a))^{1-\epsilon_B} da \right]^{1/(1-\epsilon_B)}, \quad \epsilon_B > 1, \quad (6)$$

where $p_B(a)$ is the price of the input net of tax (as the large firm is VAT-registered, it can claim back any tax on inputs). So the large firm chooses P to maximize profit $Y(P - C)$ subject to (3). This gives the usual markup equation for price,

$$P = \frac{\gamma}{\gamma - 1} C, \quad (7)$$

and thus, combining equations (7) and (3), ultimately output is

$$Y = (1 - \lambda)(1 + t)^{-\gamma} \left(\frac{\gamma}{\gamma - 1} C \right)^{-\gamma}, \quad (8)$$

Finally, input demand for variety a , by Shephard's lemma and equation (8), can be calculated as

$$\begin{aligned} y(a) &= Y \frac{\partial C}{\partial p_B(a)} \\ &= (1 - \lambda)(1 + t)^{-\gamma} \left(\frac{\gamma}{\gamma - 1} C \right)^{-\gamma} \left(\frac{p_B(a)}{C} \right)^{-\epsilon_B}. \end{aligned} \quad (9)$$

D. The Small Firms

Following Keen and Mintz (2004), we assume that the production technology is fixed coefficients, with one unit of output for an a type firm requiring ω/a units of the input and $(1 - \omega)/a$ units of labor.¹⁰ The input costs are w , r , where r is the cost of the input. By assumption, $w = 1$, and we have also assumed, without loss of generality, that one unit of the intermediate good requires one unit of labor, so $r = 1$ also.

Let the unit cost function be denoted $c(I(a); a)$, where $I(a)$ is the variable recording registration status. Then, under the assumptions just stated, the unit cost function is

$$c(1; a) = \frac{1}{a}, \quad c(0; a) = \frac{1 + \omega t}{a}. \quad (10)$$

So, cost of the input is grossed up by the tax t if the firm is not registered, as the firm cannot claim the input tax back. Note that a is a measure of productivity, and ω is a measure of the firm's use of intermediate inputs relative to labor, independent of productivity.

Suppressing the dependence of p_C , p_B , I on a to lighten notation, the firm's profit is

$$\pi(p_C, p_B, I; a) = (p_C - c(I; a))x + (p_B - c(I; a))y, \quad (11)$$

where, from equations (4) and (9),

$$x = \lambda A_C (p_C(1 + I.t))^{-\epsilon_C}, \quad y = (1 - \lambda) A_B (p_B)^{-\epsilon_B}, \quad (12)$$

where

$$A_C = Q^{\epsilon_C - \phi}, \quad A_B = (1 + t)^{-\gamma} \left(\frac{\gamma}{\gamma - 1} \right)^{-\gamma} C^{\epsilon_B - \gamma} \quad (13)$$

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

The small firm chooses p_C , $p_B \in [0, \infty)$, $I \in \{0, 1\}$ to maximize equation (11) subject to equation (12) and the registration constraint. The latter says that if the firm chooses not to register ($I = 0$), the total value of sales $s \equiv p_C x + p_B y$ must be less than the VAT sales threshold s^* . This allows of course, for *voluntary registration*, which is defined by a choice $I = 1$ when $s < s^*$. The costs and benefits of registration are clear from equations (11) and (12). The benefit is that registration, $I = 1$, lowers the unit cost of production. The cost is that at a fixed price, registration lowers B2C sales because demand by the household is reduced by the tax.

E. Equilibrium

An *equilibrium* is (a) a price P for the homogeneous product given by equation (7); (b) for each $a \in [\underline{a}, \bar{a}]$, prices $p_C(a)$, $p_B(a)$ and a registration decision $I(a)$ that maximizes

¹⁰All the results that follow generalize if the small firm production function is assumed as a constant-returns CES function of labor and the intermediate input.

equation (11) subject to equation (12) and A_C, A_B fixed; and (c) demand shifts A_C, A_B in equation (13) that are functions of $p_C(a), p_B(a), I(a)$ via equations (6) and (5).

It should be noted that this describes a general equilibrium for the whole economy. Note that the equilibrium is conditional on fixed values of the share of B2C sales, λ ; the intensity of input use, ω ; and demand elasticities e_C, e_B . In the analysis that follows, we allow these parameters to vary in order to study comparative statics.

As a final note, A_C, A_B generally depend on equilibrium prices, and this considerably complicates the analysis. For the remainder of the theoretical section, we assume that $e_C = \phi, e_B = \gamma$. This ensures that the demand parameters are exogenous that is, from equation (13), $A_C = 1, A_B = (1 + t)^{-e_B} \left(\frac{e_B - 1}{e_B} \right)^{e_B}$.

IV. Analysis

A. Voluntary Registration

We first consider the condition under which a small firm will register voluntarily in equilibrium and how this condition is affected by our key parameters $\omega, \lambda, e_C, e_B$. We begin by defining two crucial cost and demand changes. First, from equation (10), the percentage increase in the firm's unit costs due to nonregistration, because of input VAT, can be defined independent of firm productivity a as

$$\Delta_c = \frac{c(0; a)}{c(1; a)} - 1 = \omega t. \quad (14)$$

Call this the *input VAT effect* on cost.

Second, if $e_C = e_B = e$, we can define a similar kind of *output VAT effect* on demand. Specifically, it is easy to calculate that for any fixed price $p_C = p_B = p$, the percentage reduction in overall demand for the firm's product due to the charging of output VAT on B2C sales is¹¹

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

This is because when output VAT is charged, at a fixed price p , all B2C sales (which count for λ of the total) are reduced by a factor $(1 + t)^{-e}$; call this the *output VAT effect*. In what follows, we now restrict attention to the case $e_C = e_B = e$; we dealt with the more general case in an earlier version of this paper (Liu et al., 2019). We can then show:

¹¹To see this, note that from equation (12), at any fixed price p , the ratio of total demand with VAT registration to without is

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

Using the fact that $A_B = (1 + t)^{-e} \left(\frac{e-1}{e} \right)^e$, we see that this expression is equal to $1 - \Delta_d$.

Proposition 1. *A firm of type a will register voluntarily if and only if*

$$T = (1 - \Delta_d)(1 + \Delta_c)^{e-1} \geq 1. \quad (16)$$

Moreover, condition (16) is more likely to hold the higher the input cost ratio, ω and the lower the share of B2C sales λ .

This proposition is proved in the online appendix and can be interpreted as follows. First, condition (16) implies that T is a sufficient statistic that captures the entire effect of the VAT system on voluntary registration. We will see later that it is also a sufficient statistic for the degree of bunching. Also, condition (16) says that if the input VAT effect on cost due to nonregistration, Δ_c , is large relative to the output VAT effect, Δ_d , there will be voluntary registration.

Finally, the last part of the proposition gives us two of our empirical predictions regarding voluntary registration. The intuition for these results is the following. Generally voluntary registration occurs when output effect Δ_d is small and/or when the input VAT effect Δ_c is large. The first observation is that, other things equal, the larger that λ is, the bigger is the output VAT effect Δ_d ; this explains the fact that T falls with λ . Second, other things equal, the larger ω , is the bigger is the input VAT effect Δ_c . This explains why T rises with ω .

It is also of interest to study how the level of competition, measured by e , affects voluntary registration. We see from equations (14), (15), and (16) that there are two effects of a higher e , working in opposite directions. First, the input effect, $(1 + \omega t)^{e-1}$, is increasing in e , which captures the fact that the input VAT burden from nonregistration rises with e because the higher cost (due to embedded VAT) is harder to pass on to both B2C and B2B consumers when demand becomes more elastic. Second, the output effect in equation (15) is decreasing in e and captures the fact that the output VAT burden from registration rises with e , because the tax on output (due to embedded VAT) is harder to pass on to B2C consumers when demand becomes more elastic.

We can show that as e becomes large, the output VAT effect becomes dominant. Specifically, as $e \rightarrow \infty$, $1 - \Delta_d$ is proportional to $1/(1 + t)^e$, which dominates the input effect $(1 + \omega t)^{e-1}$. Hence, eventually $T \rightarrow 0$. In fact, we can prove that in the competitive limit, as $e \rightarrow \infty$, voluntary registration is never optimal.¹² However, away from the competitive limit, the effect of e on the sufficient statistic, T , can be nonmonotonic, as shown by numerical simulations in online appendix A.3.

B. Bunching

If $T < 1$, a small firm will bunch—that is, restrict sales in order to stay below the threshold—because in this case, registration leads to a drop in profit at any fixed value of sales.

¹²The proof is simple. From equations (14), (15), and (16) we see that for e large, T behaves like $\left(\frac{1 + \omega t}{1 + t} \right)^e$. But as $\omega < 1$, this term goes to 0 as $e \rightarrow \infty$, and so $T \rightarrow 0$ as $e \rightarrow \infty$.

This implies that there must be an interval of bunching firms, $a \in [a^*, a^* + \Delta a^*]$. As demand is elastic by assumption, $e > 1$, they bunch by cutting price to keep sales low. The firm at the bottom of this bunching interval, a^* , is the one that has a profit-maximizing total value of sales of exactly s^* when not registered.

The firm at the top of the interval, $a^* + \Delta a^*$, is indifferent between restricting the value of sales to s^* and not registering, and registering and choosing price and thus sales without any restriction. If $\pi(I; a)$ denotes optimized profit, conditional on the registration decision $I \in \{0, 1\}$, this indifference condition can be written as

$$\pi(1; a^* + \Delta a^*) = \pi(0; a^* + \Delta a^*). \tag{17}$$

So the amount of bunching in the space of firms is measured by Δa^* .

We do not observe Δa^* , but we do observe firm sales. Let $s^* + \Delta s^*$ be the value of sales of the firm $a^* + \Delta a^*$, assuming that this firm does not have to register for VAT. So Δs^* is the difference in sales between the VAT threshold s^* and what the value of sales for the firm at the top of the bunching interval, $a^* + \Delta a^*$, would have been had it been unconstrained by the threshold. By well-known arguments, Δs^* measures the amount of bunching we expect to see empirically.¹³ This means that our empirical predictions need to be about the determinants of Δs^* . Using indifference condition (17), we can then show the following.¹⁴

Proposition 2. (a) *The amount of bunching at the VAT threshold Δs^* is given by the implicit relationship*

$$\frac{e}{(1 + \Delta s^*/s^*)} - (e - 1) \left[\frac{1}{1 + \Delta s^*/s^*} \right]^{e/(e-1)} - T = 0. \tag{18}$$

(b) *The amount of bunching Δs^* rises as the fraction of B2C sales, λ , increases and as the share of inputs in total cost, ω , falls. Moreover, if T is decreasing in e , the amount of bunching Δs^* increases as e rises.*

This is proved in the online appendix. Note that the entire effect of VAT on bunching is captured by the sufficient statistic T . The intuition for part b of proposition 2 is very

¹³Following Saez (2010), the fraction of firms bunching, B , in the space of sales is given by $B = \int_{s^*}^{s^* + \Delta s^*} h(s) ds$, where $h(s)$ is the distribution of firms in the space of sales, assuming that firms do not have to register. Moreover, because each variety a is produced by a single firm, the distribution of firms on the space of varieties is uniform on $[0,1]$, and so $h(s) = 1/\bar{s}$, where \bar{s} is the sales of the highest-productivity firm, $a = 1$. Thus, we can write $B = \frac{\Delta s^*}{\bar{s}}$.

¹⁴Note that equation (18) is closely related to the Kleven and Waseem (2013) formula relating bunching at a notch of the personal income tax schedule to the elasticity of the labor supply. In particular, in their formula, the tax notch is measured by the term $\Delta t/(1 - t)$, where t is the lower rate of income tax and Δt is the increase in the tax rate at the notch. In fact, it is easily verified: if we take equation (18) and substitute $e_L = e - 1$, where e_L is the elasticity of labor supply, replace $\Delta s^*/s^*$ by $\Delta z^*/z^*$, and replace $T^{1/e}$ by $1 - \frac{\Delta t}{1-t}$, we get equation (5) in their paper.

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

Finally, increased competition increases bunching as long as T is decreasing in e . While we cannot establish analytically that T is decreasing in e , the simulation results reported in online appendix A.3 indicate this is the case for a wide range of parameter values. The intuition here is again related to the intuition with voluntary registration: increased product market competition makes it harder for a firm to pass on output market VAT and thus increases the incentive to bunch.

C. Evasion and Compliance Costs

Here, we briefly explain how our theoretical results extend to evasion and compliance costs. Regarding evasion, the total VAT gap in the United Kingdom is around 10% of theoretical revenues, and most of this is probably due to sales underreporting and cost overreporting (HM Revenue and Customs, 2015). In online appendix B, we model the simplest and most common form of VAT evasion, underreporting of sales. We allow both nonregistered and registered firms to hide sales, for example, by using cash transactions, but we suppose, realistically, that it is more difficult for registered firms to hide sales. With evasion, it turns out that the qualitative effects of λ , ω , and e on T do not change, and so our predictions about the determinants of voluntary registration and bunching do not change. Therefore, our key empirical predictions are robust to the presence of evasion.

However, as already mentioned in section I, we do not measure evasion directly. Nor do we have any obvious way of decomposing the total bunching effect into an evasion response and a real response, although this can be done plausibly for business taxes in some other countries, using special features of national tax systems.¹⁵ Our empirical strategy therefore focuses primarily on identifying the effects of changes in the B2C ratio, ICR, and level of competition as predicted by the theory without taking a view on how much of this effect works through evasion. In online appendix D, we do present some suggestive evidence that firms underreport turnover to stay below the threshold.

We now turn to VAT compliance costs. These costs are relatively small for the United Kingdom (Federation of Small Businesses, 2010), but the model can easily be extended in this direction, by introducing a fixed cost of VAT registration. The details are available in online appendix B. The basic

¹⁵For example, Best et al. (2015) study a minimum tax scheme for corporations in Pakistan that has a kink point where the real incentive for bunching is small but the evasion incentive is large, and they find large bunching around the minimum tax kink.

conclusion is that proposition 2 continues to hold, and if the fixed cost is small enough, proposition 1 continues to hold.

V. Context and Data

A. The Value-Added Tax System in the United Kingdom

Approximately 2 million registered businesses remit VAT in the United Kingdom every fiscal year. The VAT is the third largest source of government revenue following income tax and national insurance contributions, raising 21.1% of total tax revenue and 6.1% of GDP in 2017/2018.¹⁶ VAT is levied on most goods and services sold domestically, on imports from other EU countries, and on goods and some services imported from non-EU countries.

VAT-registered businesses pay VAT on their purchases and charge VAT on the full sale price of their taxable supplies. Businesses with a turnover below the registration threshold may choose to register voluntarily to recover the VAT paid on their intermediate inputs. Businesses cannot charge output VAT on sales of zero-rated or exempt goods. Firms can claim back the VAT paid on inputs for zero-rated supplies but not if these are exempted.

VAT rates. The standard VAT rate in the United Kingdom was 17.5% between April 2004 and January 2011, except for a temporary reduction to 15% between December 2008 and January 2010. The standard rate was further raised to 20% in January 2011 and has not been modified since then. A small number of goods and services are liable to a reduced rate of 5%, and there are also goods and services that are zero-rated or exempt from VAT, as is standard in most VAT systems.

VAT registration threshold. All U.K. businesses must register for VAT if their taxable turnover is above the threshold, updated annually to keep up with inflation. The registration threshold increased from £58,000 in 2004/2005 to £81,000 in 2014/2015, making it the highest registration threshold in the EU.

In practice, two rules govern VAT registration: a forward-looking and a backward-looking rule. The forward-looking rule requires a business to register if its taxable turnover is likely to go over the threshold in the next thirty days. The backward-looking rule requires a business to register if its taxable turnover in the previous twelve months was above the threshold. Our static theoretical model is more consistent with the forward-looking decision. In our data, 67.4% of first-time registers have a previous-year turnover lower than the VAT threshold, suggesting that the forward-looking decision is the most relevant in practice.

Full details on the annual evolution of VAT rates and registration thresholds throughout the period of analysis are provided in table A.2 in the online appendix.

¹⁶www.gov.uk/government/statistics/hmrc-tax-and-nics-receipts-for-the-uk.

B. Data

We link two administrative data sets: the universe of VAT returns and the universe of corporation tax records in the United Kingdom (called CT600). The first data set provides detailed information on VAT-registered businesses, which may take a variety of legal forms, including sole proprietorships, partnerships, and companies. To obtain information on businesses not registered for the VAT, we link the VAT records to the population of corporation tax records based on a common anonymized taxpayer reference number. The linked data set allows us to identify whether *companies* are registered in the VAT or not and contains rich information on VAT and corporation tax for each business and year.¹⁷

We further merge the linked tax data set with two additional data sources: the Financial Analysis Made Easy (FAME) annual company account database, which contains additional firm characteristics and accounting information, and the annual sector-level statistics on the share of sales to final consumers based on the Office of National Statistics' (ONS) input-output tables. The latter gives us an empirical proxy for λ , the share of B2C sales at the two-digit SIC industry level.

The final data set contains 3,461,247 observations for 968,353 unique companies between fiscal years 2004/2005 and 2014/2015.¹⁸ For each company-year observation, we have information on the VAT-exclusive turnover taken from the corporate tax records and whether the company is registered for VAT.¹⁹

We now discuss the construction of our input cost ratio and industry competitiveness measures in a little more detail. The CT600 data contain an aggregate measure of input costs that includes both salaries and other inputs. Therefore, the input cost ratio derived from this data set is higher than the magnitude relevant in our setting. The FAME data set does report salaries and other inputs separately, but only 7% of the firms in our study sample have nonmissing salaries in FAME, severely limiting our sample size. To obtain a measure of the

¹⁷Note that this linked data set does not include sole proprietorships or partnerships that are below the VAT threshold and have chosen to not register for VAT. These types of business are taxed through the individual income tax.

¹⁸We take several steps to refine the sample to study the VAT registration decisions of individual companies. First, we eliminate companies which are part of a larger VAT group and focus only on stand-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, we drop all observations with partial-year tax or accounting records because the registration decision can be based on turnover in the previous twelve months. We further eliminate companies that mainly engage in overseas activities, based on the HMRC trade classification. This is because the taxable VAT turnover excludes exports. We also exclude nonprofit organizations. Online appendix F provides more details on how we construct our sample.

¹⁹Our empirical analysis is based on turnover reported in the CT600 for two reasons. The first is data availability, as we only observe VAT-liable turnover for firms that are registered for VAT. 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.

TABLE 1.—SUMMARY STATISTICS

Variable	Mean	SD	p10	p50	p90	Observations
Total Turnover	74.69	48.88	19.88	62.70	151.80	3,461,247
Trading profit	21.88	27.25	0.00	11.82	59.71	3,461,247
Total Input Costs (CT600)	52.81	44.27	12.04	36.55	123.46	3,461,247
Intermediate Input Costs (FAME)	31.34	33.79	2.00	18.00	82.00	238,838
Input Cost Ratio (CT600)	0.71	0.28	0.28	0.78	1.00	3,461,247
Input Cost Ratio (FAME)	0.38	0.25	0.04	0.37	0.73	238,838
Input Cost Ratio (Adjusted)	0.48	0.24	0.11	0.53	0.77	3,024,673
Share of B2C Sales	0.55	0.24	0.29	0.45	0.91	3,461,247
Lerner Index	0.75	0.11	0.58	0.77	0.90	3,461,247
VAT Registered	0.630	0.483	0	1	1	3,461,247
VAT Registered (below threshold)	0.429	0.495	0	0	1	2,405,144

This table shows the mean, standard deviation, and various percentiles for the key variables used in the empirical analysis. The top four variables are expressed in thousands of pounds (GBP), where GBP 1 = USD 1.29 as of September 2018. The rest of variables are defined to be in the interval [0,1]. Note that we only have data on salary-exclusive input costs for a subset of companies from the FAME data set. The input cost ratio (adjusted) is constructed by normalizing input-cost ratio (CT600) to match the mean and standard deviation of input-cost ratio (FAME) at the industry level. The share of B2C sales denotes the proportion of turnover that comes from sales to final consumers, as opposed to sales to other VAT-registered businesses.

input cost ratio closer to the theory, we extrapolate from the subset of firms in the FAME data set. Specifically, we rescale the ICR reported in the CT600 data to match the mean and standard deviations observed for each industry in the FAME data.

For the industry competitiveness measure, we use the Lerner index of competition, which is defined as 1 minus the average ratio of trading profit to value of sales for firms in a given industry. If demand is iso-elastic at e , for all firms in an industry, as it is in our theoretical model, the Lerner index is simply $(e - 1)/e$. This means that the Lerner index is an ideal measure for testing our predictions about the effect of competition as measured by e .²⁰

Notice that the input cost ratio varies at the firm level, while the share of B2C sales and the Lerner index vary at the two-digit and four-digit industry level, respectively. All three variables have annual variation, allowing us to include them in the panel regressions with fixed effects that we present in the next section.

We focus on two different subsamples to test the hypotheses developed in section IV. When studying the voluntary registration decision, we study only firms that are voluntarily registered. A firm is defined as such if it is registered in the current year and (a) has never registered before and has a turnover below the VAT threshold, or (b) if it was registered in the previous year and had a turnover below the VAT deregistration threshold. The idea behind imposing condition b is to exclude firms that are just registered below the threshold due to inertia. For the bunching analysis, we include all firms with turnover in the range between £50,000 below the current-year registration threshold and £100,000 above. In this larger sample, 69.5% of firms have a turnover below the VAT threshold, of which 42.9% are registered for VAT. So, overall, 29.8% of firms in the main sample of companies are voluntarily registered.

²⁰Other commonly used measures of competition are the four-firm concentration ratio, or the Herfindahl index, but they measure the relative importance of the largest firms in an industry and are not closely related to the demand elasticity faced by small firms close to the VAT registration threshold.

C. Summary Statistics

Table 1 reports summary statistics for companies in the neighborhood of the current-year VAT notch, that is, those with nominal turnover of between £10,000 and £200,000. We report the mean, standard deviation, various percentiles (10th, 50th, and 90th), and the number of nonmissing observations for the key variables used in empirical analysis. Firms in the final data set have £74,690 of average turnover and £21,880 of trading profit. The average salary-*inclusive* input cost ratio (using data from CT600) is 71% of total turnover, while the average salary-*exclusive* input cost ratio (using data from FAME for a subsample of firms) is 38%. The input cost ratio calculated with the extrapolation procedure explained above yields an average of 48%, which is in between but closer to the FAME subsample, as expected. The average share of B2C sales is 55%, and the average Lerner index is 0.75.²¹

VI. Results

We present two sets of empirical results. For voluntary registration, we estimate a linear probability model with firm and year fixed effects focusing on firms with turnover below the VAT registration threshold. The regression equation includes the share of B2C, the input cost ratio (ICR), the Lerner index as a proxy for the competitiveness of the industry, and the distance from the registration threshold. In the bunching analysis, we use graphical evidence and standard nonparametric techniques to estimate the excess bunching mass just below the threshold. We then investigate whether the amount of bunching varies with the three key variables mentioned above in the way predicted by the theory.

A. Voluntary Registration

We examine whether the decision to voluntarily register for the VAT is consistent with the three theoretical predictions stated in proposition 1: a firm is more likely to register

²¹Note that the share of B2C sales and the Lerner index vary at the industry-year level, but here we report the firm-year level averages.

TABLE 2.—DETERMINANTS OF VOLUNTARY VAT REGISTRATION

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of B2C Sales	-0.233*** (0.003)			-0.167*** (0.003)	-0.025** (0.010)			0.008 (0.011)
Input Cost Ratio		0.153*** (0.002)		0.204*** (0.002)		0.064*** (0.001)		0.064*** (0.001)
Lerner Index			-0.417*** (0.006)	-0.356*** (0.007)			-0.195*** (0.015)	-0.214*** (0.016)
Observations	2,405,144	2,143,833	2,405,144	2,143,833	2,405,144	2,143,833	2,405,144	2,143,833
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	Yes	Yes	Yes	Yes

This table presents estimation results from the binary choice model of VAT registration based on equation (19). The dependent variable is the binary indicator of VAT registration status that takes on the value 1 if a firm is voluntarily registered for VAT and 0 otherwise. Columns 1 to 4 present results from the linear probability model without firm fixed effects, and columns 5 to 8 present results by adding firm fixed effects. The input cost ratio is the adjusted measure: input cost ratio (CT600) normalized to match the mean and standard deviation of input cost ratio (FAME) at industry level. Additional firm-level control variables include distance to the registration threshold. *, **, and *** denote significance at 10%, 5%, and 1%, respectively. Standard errors are clustered at the firm level.

voluntarily for VAT if it sells mostly to other VAT-registered businesses (as opposed to final consumers), has a larger share of intermediate input costs (relative to total costs), or operates in a more competitive industry.

We evaluate these relationships more formally using a panel regression framework. We model the decision of voluntary registration as a binary choice model of the following form,

$$R_{it} = \alpha_i + \alpha_t + \gamma_1 B2C_{it}^j + \gamma_2 ICR_{it} + \gamma_3 L_{it}^j + \gamma_4 D_{it} + v_{it}, \quad (19)$$

where R_{it} is a dummy indicator that takes the value 1 if the firm is voluntarily registered and 0 otherwise. $B2C_{it}^j$ denotes the share of B2C sales in industry j where firm i operates in year t , ICR_{it} denotes the ICR for firm i in year t , and L_{it}^j is the Lerner index of competitiveness for industry j in year t . Additionally, we control for the distance to the VAT threshold, D_{it} , defined as the difference between total turnover and the registration threshold in year t . The time-invariant firm fixed effects and year dummies are denoted by α_i and α_t , respectively, and v_{it} is a random error term.

We estimate equation (19) using a linear probability model, which allows us to include firm fixed effects without a bias due to the incidental parameters problem. The estimation sample includes all firms with turnover below the current-year VAT registration threshold. According to proposition 1, we expect to obtain $\gamma_1 < 0$, $\gamma_2 > 0$ and $\gamma_3 < 0$.

Table 2 reports the estimation results from the linear probability model. The first four columns include year dummies but not firm fixed effects, which allows us to examine the total effect of the industry-level variation in the B2C sales ratio and the Lerner index on the probability of voluntary registration. We first include each of the three key variables, one at a time (columns 1 to 3), and then include them all together in column 4. The coefficients in the latter column are -0.17 for B2C sales, 0.20 for the ICR, and -0.36 for the Lerner index, all statistically significant at the 1% level. These coefficients are consistent with the predictions from our theoretical framework, and similar to those in columns 1 to 3.

In columns 5 to 8, we include firm fixed effects and follow the same progression as before. The fixed effects absorb a substantial part of the cross-sectional variation in the industry-level variables and reduce the size of their coefficient estimates. While all coefficients are statistically significant and have the expected signs in columns 5 to 7, the coefficient on the share of B2C sales becomes essentially 0 in column 8. In that last specification, the coefficient on the input cost ratio is 0.064, and the one on the Lerner index is -0.214 , both statistically significant.

One advantage of including firm fixed effects is that they partially control for inertia in registration status by controlling for whether a firm has previously been above the registration threshold. In addition, our main sample includes some firms that are zero-rated, which are more likely to register and benefit from input tax credit. Inclusion of firm fixed effects thus controls for the firm-specific net benefit of registration and identifies the effects of key variables from within-firm changes.²² However, including firm fixed effects also absorbs part of the variation underlying the predictions in our theoretical framework. This is because some of the characteristics that affect the incentives to register voluntarily, in particular the share of B2C sales and the input cost ratio, are fairly stable over short periods of time. Thus, it is not surprising that the coefficients decrease in size in the fixed-effects specification. While neither specification (with or without fixed effects) is flawless, we think the regression without fixed effects represents the best possible test of our theoretical predictions.

Therefore, we evaluate the size of the effects focusing on column 4, our preferred specification. Given these results, the likelihood of being registered voluntarily is on average 4.0 percentage points higher as the B2C ratio decreases by 1 SD, 4.9 percentage points higher as the ICR increases by 1 SD, and 3.9 percentage points higher as the Lerner index decreases by 1 SD. These are sizable effects that confirm the

²²It also implies that inclusion of zero-rated firms would lead to downward bias in our estimated coefficients. In contrast, firms that are exempted are indifferent about registration and would thus add noise in our estimation. To examine the robustness of our results to this, online appendix G presents the results on voluntary registration and bunching using a slightly smaller sample that excludes firms whose primary product or service is zero-rated or exempted. The results are very similar.

importance of these three variables in the firms' decision to register voluntarily for VAT.

Table A.3 in the online appendix reports similar specifications using the two alternative measures of the ICR: the measure from CT600 and the measure from FAME for the subsample of firms observed. All the coefficient estimates are qualitatively similar to those in table 2, and they are all statistically significant except for the one on the share of B2C sales in the fixed-effects specifications. We conclude that the results are robust to the use of alternative measures of the input cost ratio.

Dynamic behavior. One potential limitation of the above analysis is that we do not explicitly consider the dynamic behavior of firms. A change in the registration status involves some costs to firms, raising the possibility that firms that are initially above the registration threshold and later fall below may stay registered simply to avoid the cost of deregistration. Hence, some of the firms that seem to be voluntarily registered may just be behaving in this way because of inertia. As noted above, the firm fixed effects partially control for this type of behavior. As a further robustness check, we conduct additional regressions taking into account these dynamic effects. Specifically, we estimate a probit model with random effects where we include a lag of the dependent variable (i.e., whether the firm was registered the previous year), the initial registration status, and the averages of the key explanatory variables. These results from these regressions are reported in online appendix E.

B. Bunching Evidence

Estimation method. As explained in section III, the VAT registration threshold at the cutoff turnover value s^* will induce excess bunching at the threshold by companies for which voluntary registration is not optimal. Following the literature (Kleven & Waseem, 2013), we can write excess bunching as $B = \Delta s^* h(s^*)$, where $h(s^*)$ is the counterfactual density of firms over the bunching interval, assuming that this is constant. We can express this as a fraction of the counterfactual density of firms at the notch, so our empirical measure of bunching is

$$b = \frac{\sum_{j=s_-^*}^{s_+^*} (c_j - \hat{c}_j)}{\frac{1}{N} \sum_{j=s_-^*}^{s_+^*} \hat{c}_j} \tag{20}$$

Here, c_j is the actual number of firms in each £1,000 turnover bin, and \hat{c}_j is the counterfactual bin counts without the notch. The range (s_-^*, s_+^*) specifies turnover bins around the notch where bunching occurs and are therefore excluded from predicting the counterfactual distribution. Specifically, the lower bound of the excluded turnover region, s_-^* , is set at the point where excess bunching starts. The upper bound of the excluded region, s_+^* , is estimated with an iterative procedure to ensure that the excess mass below the VAT notch is equal to

the missing mass above (for details on this estimation method, see Kleven, 2016). Finally, N is the number of bins in the excluded range (s_-^*, s_+^*) .

To summarize, equation (20) says that the excess mass is empirically measured by the difference between the predicted and actual mass of firms in the excluded range, divided by the average counterfactual density of firms in that range.

Graphical evidence. This section presents evidence of bunching below the VAT notch using the main sample of companies with turnover in a range between £45,000 below and £100,000 above the registration threshold. Figure 1 shows the distribution of turnover for all companies in that range, pooling together data from fiscal year 2004/2005 through 2014/2015. Using standard bunching estimation methods (Kleven, 2016), we estimate the counterfactual distribution by fitting a flexible polynomial of order 5 to the empirical distribution, excluding a range around to the VAT notch.²³

Two points are worth noting in figure 1. First, the VAT notch creates evident bunching below the threshold. The bunching estimate is 1.361 (SE: 0.202), meaning that the total excess bunching mass is almost 1.4 times as large as the average height of the counterfactual over the excluded range.²⁴ Second, in contrast with the large spike at the threshold, there is only a small hole in the distribution above the VAT notch. We do not attempt to estimate the magnitude of optimization frictions implied by the missing mass to the right of the notch for the various reason discussed in section III.

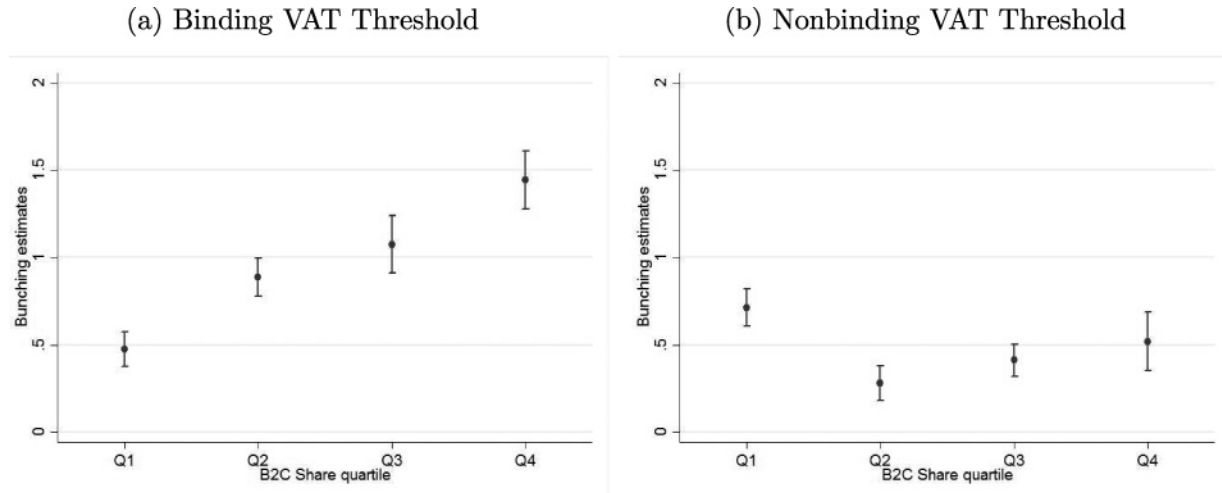
We do not attempt to decompose observed bunching into real and evasion responses because there is no variation that allows us to do that. In online appendix D, we show some suggestive evidence that the bunching behavior may partly be due to turnover misreporting.

Heterogeneity in bunching. We now explore potential heterogeneity in bunching to see whether the empirical patterns are consistent with the predictions set out in proposition 2. Implementing this analysis is challenging because some firms have incentives to voluntarily register for VAT and therefore are indifferent to the existence of the VAT threshold. To address this issue, we leverage the fact that we observe which firms choose to register voluntarily among those below the threshold. Specifically, we partition the sample into two groups of firms based on their predicted likelihood of registering voluntarily (regardless of their turnover) following three steps. First, we regress voluntary registration status on the three key variables (share of B2C sales, input cost ratio, and Lerner index), including only firms below the turnover

²³The excluded range goes from -£14,000 to £24,000, which ensures that the excess bunching mass to the left of the notch is almost identical to the missing mass to the right.

²⁴Bunching is sharp and significant every year, as shown in figure A.2 in the online appendix. Unlike studies analyzing bunching in the taxable income of individuals (Kleven & Waseem, 2013) and corporations (Devreux, Liu, & Loretz, 2014), we do not find evidence of bunching at round numbers.

FIGURE 3.—BUNCHING ACROSS QUANTILES OF THE B2C SHARE DISTRIBUTION



This figure shows the bunching estimates around the VAT notch across four different quartiles of the distribution of the share of B2C sales. (a) The point estimates and 95% confidence intervals for the subset of firms not predicted to register voluntarily. (b) The estimates for the subset of firms predicted to register voluntarily.

threshold. Second, we use the estimated coefficients to obtain a predicted probability of being voluntarily registered for each firm i in year t . Third, we divide firms into two groups depending on whether their predicted probability is above or below the median.

First, we explore how companies with different shares of B2C sales respond to the same VAT notch. We divide companies into four quartiles of the B2C share distribution and estimate bunching at the VAT registration threshold for the subsamples of firms more and less likely to register voluntarily. The left panel of figure 3 shows the bunching estimates and 95% confidence intervals for each quartile of the B2C share distribution, for firms predicted to not register voluntarily (i.e., the subgroup for which the VAT notch is binding). The bunching estimate is positively correlated with the share of B2C sales, taking a value of 0.5 for the first quartile (Q1) and about 1.4 for the fourth quartile (Q4). The right panel of figure 3 shows the estimates for the subgroup of firms predicted to register voluntarily, for which the VAT threshold is not binding. In this case, the bunching estimates are consistently low, between 0.3 and 0.6, and they do not follow any clear pattern across quartiles.

Second, we examine the extent of bunching depending on the degree of competition in the product market, measured by the Lerner index at the four-digit industry level. Since this index is defined as 1 minus the average profit margin in the industry, higher values of the index indicate that the industry is more competitive. As in the previous cases, we examine how bunching varies across quartiles of the Lerner index distribution for firms predicted to register versus those not predicted to register. The left panel of figure 4 shows a strong, positive correlation between the bunching estimates and the degree of competition for firms predicted to not register, with an estimate of 1.7 for firms in the top quartile (Q4). When studying firms predicted to register voluntarily in the bottom panel,

we observe consistently low bunching estimates at all four quartiles without any specific pattern.

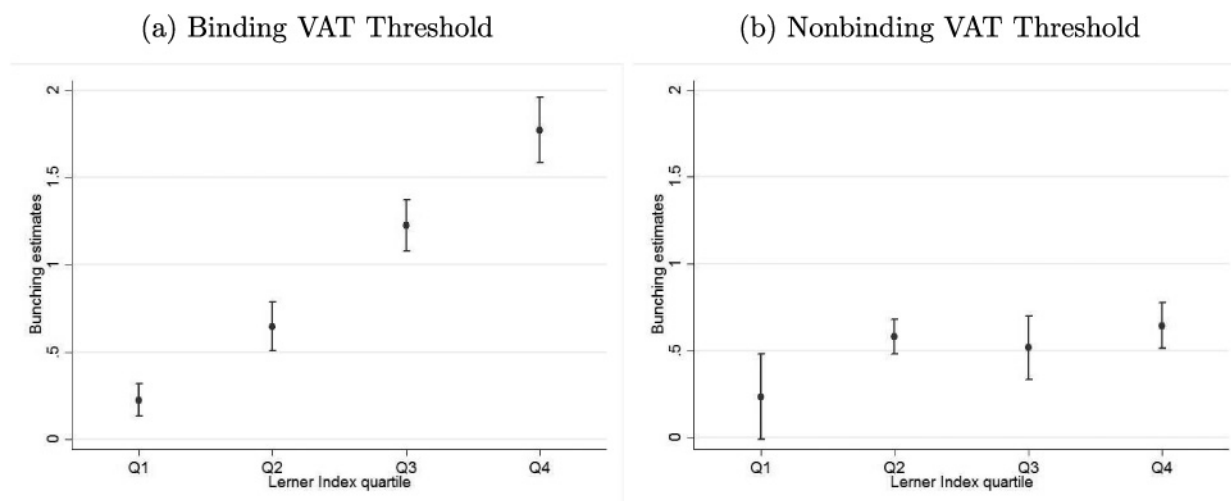
Finally, we examine how companies with different ICRs respond to the VAT notch. Again, we divide the sample into quartiles of the distribution of this variable and look separately at firms predicted to register voluntarily versus those not predicted to register. For this test, we use the ICR constructed using information from the FAME subsample. The left panel of figure 5 shows that the degree of bunching generally decreases with the ICR for firms predicted to not register voluntarily, although the relationship is not fully monotonic because the estimate for the first quartile is relatively low. In the right panel, we observe that the pattern of bunching estimates is flat for firms predicted to register voluntarily, confirming that the model's predictions do not apply to that group.

VII. Implications for VAT Thresholds in Practice

We conclude by discussing some of the implications of our work for the setting of the VAT threshold. The well-known work of Keen and Mintz (2004) makes clear that the basic trade-off in choosing the threshold is between minimizing administration costs for the revenue authority and compliance costs for businesses (implying a high threshold) and raising VAT revenue (implying a low threshold). On top of this, they show that behavioral responses to the threshold also affect threshold design.

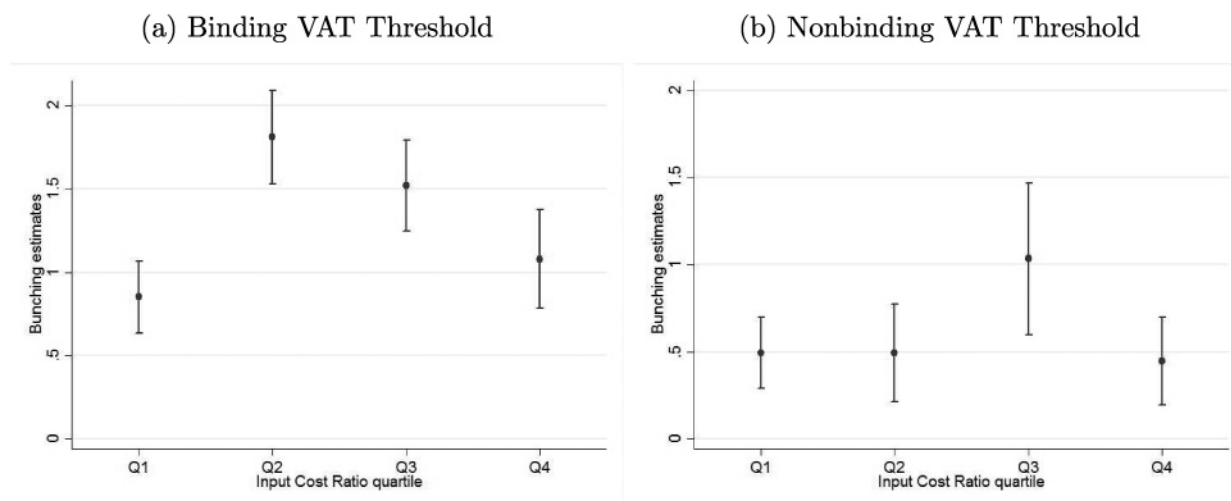
In the online appendix, we develop a formula for the optimal threshold in our model that refines their basic insight in two ways. First, it allows for a behavioral response that is specific to our model. Second, it allows for B2B sales, a feature not present in Keen and Mintz (2004). We find that the optimal threshold in the presence of B2B sales ($\lambda < 1$) is higher than in the absence of B2B sales.

FIGURE 4.—BUNCHING ACROSS QUANTILES OF THE LERNER INDEX DISTRIBUTION



This figure shows the bunching estimates around the VAT notch across four different quartiles of the distribution of Lerner index. (a) The point estimates and 95% confidence intervals for the subset of firms not predicted to register voluntarily. (b) The estimates for the subset of firms predicted to register voluntarily, for which the threshold is nonbinding.

FIGURE 5.—BUNCHING ACROSS QUANTILES OF THE INPUT COST RATIO DISTRIBUTION



This figure shows the bunching estimates around the VAT notch across quartiles of the distribution of the ICR. (a) The point estimates and 95% confidence intervals for the subset of firms not predicted to register voluntarily. (b) The estimates for the subset of firms predicted to register voluntarily, for which the threshold is nonbinding.

The intuition for this result is that only B2C sales by small firms above the threshold are taxed in our model, whereas the value of B2B sales to the large firm is eventually taxed, as all of the large firm's sales are taxed. This implies that raising the threshold is less costly in terms of forgone tax revenue when more of the sales of the small firm are B2B. This intuition is further developed in online appendix C.

While our formula is obviously specific to our model, the mechanism at work is likely to be more general. For a variety of different market structures, B2B sales are more likely to eventually be taxed than B2C sales.²⁵ In order to test this,

²⁵There is also a second, more indirect mechanism at work in our model with B2B sales. An increase in the threshold increases the prices that the small firms that are bunching charge for inputs to the large firm, and this is passed on to prices by the large firm, increasing the VAT base. See online appendix C for details.

we have compiled some cross-country empirical evidence. Figure A.13 in the online appendix shows the cross-country relationship between the average ratio of B2C sales in a country and the ratio of the VAT threshold in year 2017 over GDP per capita.²⁶ We focus on this ratio to adjust for the relative size of turnover by small businesses in countries at different levels of development. Moreover, the size of the informal sector might also influence the choice of threshold. We proxy the latter by the share of the agricultural sector in GDP, following Keen and Lockwood (2010).

²⁶Specifically, the B2C ratio is calculated as the final sales to consumers, including sales to households, nonprofit institutions serving households, and governments, relative to total sales to industries and consumers in each of the 103 countries. The cross-country input/output data are from the multi-region input-output table (MRIO) of the Eora global supply chain database, available at <https://worldmrio.com/eora26/>.

The empirical pattern in figure A.13 is quite intriguing and consistent with the model prediction in online appendix C. There is a positive correlation between the average B2C ratio and the VAT threshold as a fraction of per capita GDP (panel a). The estimated slope coefficient is 0.06 and is significant at the 10% level (with a p -value of 0.089). However, when controlling for the size of the informal sector, the relation between the average B2C ratio and the VAT threshold normalized by GDP per capita becomes negative, with the slope coefficient estimated to be -0.03 and highly significant (with a p -value of 0.005). As a large, informal sector (proxied by the share of agriculture sector) is typically associated with higher compliance and administration costs, the patterns suggest that the VAT threshold tends to be higher with higher compliance and administration costs and less direct selling to consumers.

VIII. Conclusion

In this paper, we first developed a conceptual framework that can explain the coexistence of voluntary VAT registration and bunching at the registration threshold. We showed that this required (at least) three stages of production, with firms at the intermediate stage selling to both final consumers and other firms. This framework predicts that voluntary registration is more likely, and bunching is less likely, when (a) the cost of inputs relative to sales is high, (b) the proportion of B2C sales is low, or (c) the level of product market competition is low. We then brought these predictions to an administrative data set that was created by linking the population of corporate income and value-added tax records in the United Kingdom. We found that patterns of voluntary registration and bunching in the data are consistent with the theoretical predictions. Finally, we provided a discussion and some cross-country evidence of the implications of our results for the optimal design of the VAT threshold.

REFERENCES

- Almunia, Miguel, and David Lopez-Rodriguez, "Under the Radar: The Effects of Monitoring Firms on Tax Compliance," *American Economic Journal: Economic Policy* 10:1 (2018), 1–38.
- Asatryan, Zareh, and Andreas Peichl, "Bunching and Non-Bunching of Firms at VAT and Admin Thresholds," University of Mannheim working paper (2016).
- Bachas, Pierre, and Mauricio Soto, "Not(ch) Your Average Tax System: Corporate Taxation under Weak Enforcement," World Bank Policy Research working paper (July 2018).
- Best, Michael C., Anne Brockmeyer, Henrik J. Kleven, Johannes Spinnewijn, and Mazhar Waseem, "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan," *Journal of Political Economy* 123 (2015), 1311–1355.
- Best, Michael C., and Henrik J. Kleven, "Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK," *Review of Economic Studies* 85 (2018), 157–193.
- de Paula, Aureo, and Jose Scheinkman, "Value-Added Taxes, Chain Effects, and Informality," *American Economic Journal: Macroeconomics* 2 (2010), 195–221.
- Devereux, Michael P., Li Liu, and Simon Loretz, "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records," *American Economic Journal: Economic Policy* 6 (2014), 19–53.
- Emran, M. Shahe, and Joseph E. Stiglitz, "On Selective Indirect Tax Reform in Developing Countries," *Journal of Public Economics* 89 (2005), 599–623.
- Federation of Small Businesses, *Impact of Increasing VAT Registration Threshold for Small Businesses* (London: Centre for Economics and Business Research, June 2010).
- Harju, Jarkko, Thomas Matikka, and Timo Rauhanen, "Compliance Costs vs. Tax Incentives: Why Do Entrepreneurs Respond to Size-Based Regulations?" *Journal of Public Economics* 173 (2019), 139–164.
- HM Revenue and Customs, *Measuring Tax Gaps* (March 2015).
- Kanbur, Ravi, and Michael Keen, "Thresholds, Informality, and Partitions of Compliance," *International Tax and Public Finance* 21 (2014), 536–559.
- Keen, Michael, and Ben Lockwood, "The Value Added Tax: Its Causes and Consequences," *Journal of Development Economics* 92 (2010), 138–151.
- Keen, Michael, and Jack Mintz, "The Optimal Threshold for a Value-Added Tax," *Journal of Public Economics* 88 (2004), 559–576.
- Kleven, Henrik, "Bunching," *Annual Review of Economics* 8 (2016), 435–464.
- Kleven, Henrik J., and Mazhar Waseem, "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *Quarterly Journal of Economics* 128 (2013), 669–723.
- Liu, Li, Ben Lockwood, Miguel Almunia, and Eddy Tam, "VAT Notches Voluntary Registration, and Bunching: Theory and UK Evidence," CEPR discussion paper 13983 (2019).
- Onji, Kazuki, "The Response of Firms to Eligibility Thresholds: Evidence from the Japanese Value-Added Tax," *Journal of Public Economics* 93 (2009), 766–775.
- Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (2010), 180–212.