

# Malas Notches\*

Ben Lockwood<sup>†</sup>

First version: 4 July 2016

This version: 2 March 2018

## Abstract

This paper shows that the sufficient statistic approach to the welfare properties of income (and other) taxes does not easily extend to tax systems with notches, because with notches, changes in bunching induced by changes in tax rates have a first-order effect on tax revenues. In an income tax setting, we show that the marginal excess burden (MEB) of a change in the top rate of tax is given by the [Feldstein \(1999\)](#) formula for the MEB of a proportional tax, plus a correction term. Also, the optimal top rate of tax is given by a formula for the optimal proportional tax, minus a correction term. These correction terms cannot be calculated just from knowledge of the elasticity of taxable income and quantitatively, they can be large. An application to VAT is discussed; with a calibration to UK data, the MEB of the VAT is roughly three times what it would be if VAT was simply a proportional tax.

*JEL Classification:* H20,H21,H31

*Keywords:* tax kink, tax notch, excess burden, sufficient statistic

---

\*I would like to thank Omiros Kouvavas for outstanding research assistance, and Miguel Almunia, Steve Bond, Thomas Brosy, Michael Devereux, Michael Keen, and Joel Slemrod, and participants at the 2017 ASSA conference for helpful comments. I also acknowledge financial support from the ESRC under grant ES/L000016/1.

<sup>†</sup>CBT, CEPR and Department of Economics, University of Warwick, Coventry CV4 7AL, England; Email: B.Lockwood@warwick.ac.uk

# 1 Introduction

In a recent survey, [Chetty \(2009b\)](#) argues that an important new development in public economics is the so-called sufficient statistic approach, which "derives formulas for the welfare consequences of policies that are functions of high-level elasticities rather than deep primitives" ([Chetty \(2009b\)](#), p 451). In turn, this means that to assess the welfare properties of these policies, only these elasticities, rather than fully structural models, need to be estimated.<sup>1</sup>

The sufficient statistic approach originated in a seminal paper by [Feldstein \(1999\)](#), who showed that the marginal excess burden (MEB) of a proportional income tax only depends on the behavioral responses to the tax via a sufficient statistic, the personal elasticity of taxable income (ETI). Feldstein's paper has given rise to a large literature devoted to obtaining empirical estimates of the ETI ([Gruber and Saez \(2002\)](#), [Saez et al. \(2012\)](#), [Kleven and Schultz \(2014\)](#), [Weber \(2014\)](#)).

Subsequently, [Saez \(2001\)](#) and [Saez et al. \(2012\)](#) showed that the Feldstein formula for the MEB could be extended to the top rate of tax in a progressive piece-wise linear income tax system, and they also established formulae for the revenue and welfare-maximizing rate of tax. These formulae also have the sufficient statistic feature; specifically, they depend only on the elasticity of the ETI, a statistic of the income distribution, which is constant if the top tail of the income distribution is Pareto<sup>2</sup>, and possibly a welfare weight.

In this paper, we ask the question as to whether these sufficient statistic properties of key formulae also extend to tax systems with notches. Generally, a tax notch occurs when there is a discontinuous change in the tax liability as the tax base varies ([Slemrod \(2013\)](#), [Kleven \(2016\)](#)).

In practice, we do see notches in several major kinds of taxes, and these are being increasingly studied in the empirical literature. Significant notches in the personal income tax system are quite rare, although they do exist; for example, in Pakistan, there are notches where the tax on all income below the notch can rise by as much as 5% ([Kleven and Waseem \(2013\)](#)), and in Ireland, an emergency income levy after the financial crisis had a notch of up to 4% ([Hargaden \(2015\)](#))<sup>3</sup>. There are even small notches in the federal income tax in the US, and larger notches induced by income-dependent entitlement to tax credits ([Slemrod \(2013\)](#)).

Notches also exist in other major taxes. For example, notches are, or were until

---

<sup>1</sup>[Chetty \(2009a\)](#) also argues that this sufficient statistic approach is also valuable in several other contexts, such as evaluating the welfare gain from social insurance programs, and the welfare effects of changes in taxes with optimization frictions.

<sup>2</sup>The formula is that the marginal excess burden equals  $\frac{tea}{1-t-te}$ , where  $t$  is the rate of tax,  $e$  is the personal elasticity of taxable income with respect to the net of tax rate  $1-t$ , and  $a$  is the Pareto parameter.

<sup>3</sup>From Table 1 of [Hargaden \(2015\)](#), in 2010, earnings of above 26000 Euro incurred a charge of 1040 Euro.

recently, present in housing transactions taxes in the UK and the US (Best and Kleven (2013), Kopczuk and Munroe (2015)). They also arise in the corporate income tax in Costa Rica (Bachas and Soto (2015)). Slemrod (2013) notes that there are many examples of commodity tax notches, where a marginal change in some characteristic can change the product classification so as to produce a discrete change in the tax liability.<sup>4</sup> Finally, as argued by Liu and Lockwood (2015), a VAT threshold can be thought of as a tax notch; a firm’s VAT liability changes discontinuously when its sales go over the registration threshold. Indeed, given the importance and near-ubiquity of VAT, this is in fact the most important example of a tax notch.

We first study notches in the income tax setting of Saez (2010) and others, where households differ in ability or taste so that the disutility of generating taxable income varies across households. For simplicity, we assume a two-bracket tax i.e. a tax with a lower rate below a threshold, and a higher rate above. In this setting, our first contribution is to derive an exact formula for the marginal excess burden (MEB) of the higher rate of tax, This formula similar to Feldstein (1999)’s formula for the MEB of a proportional income tax, but includes a correction factor that captures the effect of the *bunching response* to an increase in the top rate tax on tax revenue.

The bunching response measures the change in the *number* of households bunching at the threshold to avoid paying the top rate of tax, and is thus distinct from the intensive margin response of taxable income of given household to the tax rate, the elasticity of taxable income, or ETI. With a notch, unlike the case of a kink, the bunching response affects tax revenue because with a notch, the tax schedule is discontinuous at the threshold. Specifically, an increase in the top rate of tax increases bunching, which - due to the notch - lowers tax revenue, and thus raises the MEB.

Our second finding is that the optimal (i.e. welfare-maximizing) top rate of tax in the presence of a notch is closely related to both the optimal proportional tax and the optimal top rate of tax in an ordinary kinked income tax system, as analyzed by Saez (2001). In particular, if the government has no redistributive objective, the optimal top rate of income tax with a notch is equal to the optimal proportional income tax rate, minus a correction factor.

Moreover, both these correction factors depend on an underlying variable  $C$  that *cannot be expressed as a simple function of the usual sufficient statistics* i.e. the ETI and the Pareto parameter of the upper tail of the income distribution. It does depend on these variables, but it also depends on the lower rate of tax, the position of the notch, and a counterfactual, i.e. the earnings that the individual at the top of the interval (the top buncher) would choose if faced with the higher rate of tax. So, the sufficient statistic approach seems to break down with tax notches.

However, all is not lost; we show that the counterfactual earnings of the top buncher

---

<sup>4</sup>For example, in the US, the Gas Guzzler Tax, under which high-performance cars are subject upon initial sale to a per-vehicle tax that is higher, the lower is the fuel economy of the car.

can either be computed theoretically, using the indifference condition that the top buncher is indifferent between bunching and being above the notch, or, in any empirical study of bunching, it can be computed empirically, using the estimate of excess mass at the notch (the parameter  $B$  in [Kleven and Waseem \(2013\)](#)). Thus, this paper is the first to show how bunching estimates at notches can be used to make welfare calculations.

Of course, if the correction factor turns out to be small, the Feldstein formula still provides a good approximation to the MEB. Our third contribution is to investigate whether this is the case. Calibrations show that the percentage error from using the Feldstein formula for the MEB can be very large. At baseline values, the marginal excess burden is underestimated by a factor of six <sup>5</sup>. So, the conclusion is that at least in the income tax setting, the sufficient statistic approach is not practical.

We then turn to apply our approach to the VAT, which is the most empirically important example of a tax notch. We present a simple model of small traders who differ in productivity, and are subject to VAT at rate  $t$  above a threshold level of sales. We show that this model is formally equivalent to our income tax model, in the sense that registered firms above the threshold face an effective rate of VAT  $t_R$  on value-added, and firms below the threshold face  $t_N$ , a lower rate <sup>6</sup>.

We then show that the MEB of an increase in the statutory rate of VAT is given by the Feldstein formula for a proportional tax plus a correction factor as in the income tax case. However, the details of the correction factor are more complex, because an increase in the statutory rate  $t$  increases *both* the effective rates  $t_R, t_N$ . A calibration of the model shows that the proportional tax formula for the MEB of the VAT underestimates the true MEB by a factor of up to three. This framework also allows us to evaluate the effect of increased compliance costs of VAT in the MEB via its impact on bunching; increased compliance costs increase bunching and thus increase the MEB, but the effect is quantitatively small relative to the effect of the notch.

The remainder of the paper is arranged as follows. After the literature review in Section 2, in Section 3, we set up the model. Section 4 has the main analytical results for the income tax, and Section 5 the simulations. Section 6 deals with the extension to the VAT, and Section 7 concludes.

## 2 Related Literature

This paper speaks to a number of related literatures. First, it is already known that due to externalities of one kind or another, the sufficient statistic approach has its limitations. [Saez et al. \(2012\)](#) give the examples of deductibility from income tax of charitable giving

---

<sup>5</sup>We also show that using the formula for the optimal proportional tax substantially overestimates the true optimal top rate of tax with a notch.

<sup>6</sup>It may seem counter-intuitive that non-registered firms face a positive rate of effective VAT; this is because non-registered firms cannot claim back VAT on inputs, so-called "embedded" VAT.

and mortgage interest payments for residential housing. In these cases, an increase in the marginal rate of tax will boost charity income and home ownership respectively, which may be valuable objectives in themselves. [Saez et al. \(2012\)](#) call these classical externalities<sup>7</sup>.

Fiscal externalities, where the actions of the household generate additional revenue for the government and thus benefits other households, can also cause the sufficient statistic approach to fail, or at least require adjustment, but in these cases a simple change to the formula is sometimes possible. The analysis of income tax evasion of [Chetty \(2009b\)](#) is a case in point<sup>8</sup>. As [Gillitzer and Slemrod \(2016\)](#) show, in this case the standard formula for the marginal efficiency cost of funds can be adjusted in the same way it must be adjusted for any fiscal externality, i.e. whenever a change in tax rates induces taxpayers to shift income to another tax. Our results are rather different to these cases of both classical and fiscal externalities. In our setting, there is no fiscal or other externality- rather, the sufficient statistic approach fails because the bunching response has a first-order effect on tax revenue.

A second related literature is on VAT. Here, there are two distinct sets of related papers. First, there is a growing literature on the effect of VAT thresholds on firm behavior. Theoretical contributions include [Keen and Mintz \(2004\)](#), [Kanbur and Keen \(2014\)](#) and [Liu and Lockwood \(2015\)](#), and empirical studies include [Liu and Lockwood \(2015\)](#) and [Harju et al. \(2016\)](#). The theoretical work of Kanbur, Keen and Mintz focusses on the optimal threshold of the VAT, holding the rate of tax fixed, and is thus complementary to this paper, which characterizes the MEB of an increase in the rate, holding the threshold fixed. In fact, we effectively ask the question of whether it is legitimate to ignore the threshold altogether when calculating the MEB of the VAT.

Therefore, our paper relates to a literature on the marginal excess burden of indirect taxes, including VAT (e.g. [Ballard et al. \(1985\)](#), [Rutherford and Paltsev \(1999\)](#)). In these papers, when the marginal excess burden of VAT is calculated, it is always assumed that the VAT is a proportional tax i.e. the VAT threshold is ignored. This paper shows that this simplifying assumption yields seriously biased estimates.

A third related literature is that on the MEB and welfare-maximizing taxes with kinks in the tax schedule. Here, we make a small contribution as a by-product of our main focus, which is on notches. In the case of kinks, it is generally understood that the marginal excess burden of the top rate of income tax, and the welfare-maximizing top rate depends via simple formulae, only on the elasticity of the ETI, and the Pareto statistic of the income

---

<sup>7</sup>See [Doerrenberg et al. \(2015\)](#) for a more formal statement of this argument, and estimates of how deductions respond to tax rate changes for the case of Germany.

<sup>8</sup>Chetty shows that when the household can evade the personal income tax at a cost, if that cost is a pure transfer payment i.e. a fine times a probability of detection, there is effectively a positive fiscal externality of evasion - it generates additional revenue for the government and thus benefit for all households. In this case, as we might expect, we see that the elasticity of taxable income over-estimates the excess burden of the tax.

distribution. However, there seems to be some confusion about the conditions required for this result. Saez et al. (2012) suggest that what is required is that assumption that "behavioral responses take place only along the intensive margin", or more precisely that the bunching response of an increase in the top rate of tax is of second order relative to the extensive margin response.<sup>9</sup> This assumption is very strong, as even with a kink, there is always a bunching response. Our Proposition 1 below shows that this assumption is *not* necessary, because no matter what the size of the bunching response, the response has no effect on tax revenue, to first order, as the tax schedule is continuous. All that is required is that the distribution of taxpayer types is continuous, a standard assumption.

### 3 The Model and Preliminary Results

#### 3.1 Set-Up

We follow Saez (2010) in our set-up. There are individual taxpayers indexed by a skill or taste parameter  $n \in [\underline{n}, \bar{n}]$ , assumed continuously distributed in the population with distribution  $H(n)$  and density  $h(n)$ . A type  $n$  individual has preferences over consumption  $c$  and taxable income  $z$  of the form

$$u(c, z; n) = c - \psi(z; n)$$

where  $\psi(z; n)$  is the disutility of earning income  $z$ . So, as utility is linear in  $c$ , we are assuming away income effects. We also assume:

**A1.**  $\psi_z, \psi_{zz} > 0$ ,  $\psi_n, \psi_{nz} < 0$ .

A1 says that a higher  $n$  represents a higher skill level (i.e. higher wage), or a lower taste for leisure. In particular, the higher  $n$ , the lower the total and marginal disutility of generating a given amount of taxable income. Assumption A1 is satisfied for example, by the iso-elastic specification of Saez (2010):

$$\psi(z; n) = \frac{n}{1 + \frac{1}{e}} \left( \frac{z}{n} \right)^{1 + \frac{1}{e}} \tag{1}$$

The budget constraint is  $c = z - T(z)$ , where  $T(\cdot)$  is the tax function. So, a household's utility over  $z$  is  $u(z; n) = z - T(z) - \psi(z; n)$ .

Finally, for future reference, define the optimal taxable income at tax rate  $t$  for a type

---

<sup>9</sup>Specifically, they say the following. "The change  $dt$  could induce a small fraction  $dN$  of the  $N$  taxpayers to leave (or join if  $dt < 0$ ) the top bracket. As long as behavioral responses take place only along the intensive margin, each individual response is proportional to  $dt$  so that the total revenue effect of such responses is second order ( $dN \cdot dt$ ) and hence can be ignored in our derivation."

$n$  taxpayer to be;

$$z(1-t, n) \equiv \arg \max_z \{(1-t)z - \psi(z; n)\}$$

Note from A1 that  $z_{1-t}, z_n > 0$ , where subscripts denote derivatives. So,  $z_{1-t}$  is the response of taxable income to the net-of-tax rate. Following [Saez et al. \(2012\)](#), we call this the *intensive margin* response to the tax.

### 3.2 Kinks and Notches

For simplicity, we focus on a two-bracket tax, although our arguments apply straightforwardly to the case of the highest tax in a piecewise-linear tax system with any number of brackets. We will assume that the tax system is progressive; that is, the tax rate on incomes in the higher income bracket is strictly greater than the tax on incomes in the lower income bracket.

So, with a two-bracket tax, for a kink, the tax function is

$$T_K(z) = \begin{cases} t_L z, & z \leq z_0 \\ t_L z_0 + t_H(z - z_0), & z > z_0 \end{cases} \quad (2)$$

for  $z_0 > 0$ ,  $t_H > t_L \geq 0$ ; that is, all income below the kink point  $z_0$  is taxed at the lower rate  $t_L$ , and all income in excess of the kink is taxed at the higher rate. For a notch, the tax function is

$$T_N(z) = \begin{cases} t_L z, & z \leq z_0 \\ t_H z, & z > z_0 \end{cases} \quad (3)$$

with  $t_H > t_L \geq 0$ . That is, when taxable income is below  $z_0$ , a tax at rate  $t_L$  is paid on all income, but when  $t_H$  is above  $z_0$ , a tax at rate  $t_H$  is paid on *all* income.

### 3.3 Bunching

With either a kink or a notch, all types in an interval  $n \in [n_L, n_H]$  will bunch at taxable income  $z_0$ . In both cases, the lowest type who bunches is the one who is just willing to earn taxable income  $z_0$  at the lower tax rate. So,  $n_L$  is defined by the condition

$$z(n_L, 1 - t_L) = z_0 \quad (4)$$

With a kink, the highest type who bunches,  $n_H$ , is defined by the condition that the optimal choice of taxable income at tax  $t_H$  is just  $z_0$  i.e.

$$z(1 - t_H; n_H) = z_0 \quad (5)$$

With a notch,  $n_H$  is defined by the condition that the  $n_H$  type must be indifferent between staying at the notch and paying tax  $t_L$ , and choosing  $z$  optimally, and paying  $t_H$  on *all income*. To write this indifference condition, we first define the indirect utility

function

$$v(t; n) \equiv \max_z \{(1 - t)z - \psi(z; n)\}$$

Then, the condition defining  $n_H$  can be written:

$$(1 - t_L)z_0 - \psi(z_0; n_H) = v(t_H; n_H) \quad (6)$$

The left-hand side of (6) is utility when taxable income is constrained to be at the notch value  $z_0$ . Note that this indifference condition implies  $z(1 - t_H, n_H) > z_0$ , because if  $z(1 - t_H, n_H) < z_0$ , the  $n_H$ -type could choose  $z$  optimally *and* stay below the notch. Note the difference between indifference condition (6) and the condition (5).

### 3.4 The Bunching Response

Here, we study the effect of a change in  $t_H$  on the mass of individuals who bunch i.e. on the size of the interval  $[n_L, n_H]$ . Note first from (4) that  $n_L$  is unaffected by  $t_H$  for both a kink and a notch. Next, in the kink case, we can calculate from (5) note that

$$\frac{\partial n_H}{\partial t_H} = \frac{z_{1-t_H}}{z_n} > 0 \quad (7)$$

So, we have a bunching response to an increase in  $t_H$ : i.e. an increase in the tax rate above the kink makes going above the kink less attractive, and so more people bunch below the kink.

In the notch case, note that  $v_t = -z$ , where  $v_t$  is the derivative of  $v$  with respect to  $t$ . Then, we can calculate from (6) that

$$\frac{\partial n_H}{\partial t_H} = \frac{z(1 - t_H, n_H)}{\psi_n(z_0; n_H) - \psi_n(z(1 - t_H, n_H); n_H)} \quad (8)$$

Also, as  $\psi_{nz}(z; n) < 0$  and  $z(1 - t_H, n_H) > z_0$ , we see that the denominator of (8) is positive, and consequently from (8):

$$\frac{\partial n_H}{\partial t_H} > 0 \quad (9)$$

So, again we see that there is a bunching response to a change in  $t_H$ ; an increase in the tax rate above the notch makes going above the notch less attractive, and so more people bunch at the notch.



## 4 Main Results

### 4.1 The Effect of the Bunching Response on Tax Revenue

Here, we establish a key result that the effect of the bunching response on tax revenue with a kink and a notch are qualitatively different, being zero and negative respectively. With a kink, revenue can be written

$$R = t_L \left( \int_{\underline{n}}^{n_L} z(1 - t_L; n)h(n)dn + \int_{n_L}^{\bar{n}} z_0 h(n)dn \right) + t_H \left( \int_{n_H}^{\bar{n}} (z(1 - t_H; n) - z_0)h(n)dn \right) \quad (10)$$

Note that all households with  $n \geq n_L$  pay tax at the lower rate on the first  $z_0$  of earnings.

In the kink case, the bunching effect on tax revenue i.e. the effect of a change in  $t_H$  on  $R$  via a change in  $n_H$  in  $t_H$  is, from (10):

$$\frac{\partial R}{\partial n_H} = -t_H(z(1 - t_H; n_H) - z_0)h(n_H) = 0 \quad (11)$$

So, overall, with a kink, the effect of the bunching response on tax revenue is zero. This is simply due to the fact that a kinked tax schedule is continuous in  $z$ .

With a notch, revenue is

$$R = t_L \left( \int_{\underline{n}}^{n_L} z(1 - t_L; n)h(n)dn + \int_{n_L}^{n_H} z_0 h(n)dn \right) + t_H \left( \int_{n_H}^{\bar{n}} z(1 - t_H; n)h(n)dn \right) \quad (12)$$

Comparing this to (10), we see a key difference. Because the higher rate applies to *all* income for those earning above  $z_0$ , the threshold  $z_0$  no longer enters into the tax base for  $t_H$ , and so the upper limit of integration on  $z_0$  in the tax base for  $t_L$  falls from  $\bar{n}$  to  $n_H$ , reflecting the fact that now only individuals below  $n_H$  pay any tax at the lower rate.

Note from (12) that;

$$\frac{\partial R}{\partial n_H} = (t_L z_0 - t_H z(1 - t_H; n_H))h(n_H) < 0 \quad (13)$$

This is strictly negative as  $t_H > t_L$ ,  $z(1 - t_H; n_H) > z_0$ . So, in contrast to the kink case, the bunching effect on tax revenue  $R$  from an increase in  $t_H$  is negative, as  $\frac{\partial n_H}{\partial t_H} > 0$  from (9). This is because a small increase in  $n_H$  has two effects on revenue that are both negative. First, there is a discontinuity in the tax *base*; the earnings of these who now locate at the notch fall discontinuously from  $z(1 - t_H; n_H)$  to  $z_0$ . Second, there is a discontinuity in the tax *rate* applying to that base; all these earnings are taxed at a lower rate,  $t_L$  rather than  $t_H$ .

So, we conclude:

**Proposition 1.** *The effect of the bunching response on tax revenue is zero for a kink, but strictly negative for a notch.*

This result is the key one that drives the rest of the paper. So, to fully appreciate the intuition, we consider the two following figures. Each Figure shows how tax revenue collected from an  $n$ -type household varies with  $n$ . To make the figures as clear as possible, we assume iso-elastic utility as in (1), in which case it is easily verified that tax revenue collected from an  $n$ -type household is linear in  $n$ . In both Figures 1 and 2, the tax revenue as a function of  $n$  before the change is shown by the red line. Note that for  $n$  between  $n_L$  and  $n_H$ , households are bunching and so revenue is constant.

Now suppose that  $t_H$  increases. In Figure 1, the green line shows the hypothetical revenue paid by households above  $n_H$  following the increase *if there were no behavioral response*; all households above  $n_H$  pay more, proportionally to  $n$ . The green arrow shows the bunching response; some households move from above the threshold to just below. From this figure, it is clear that this has only a *second-order effect* on tax revenue; the government loses just the small triangle shown.

Figure 1: The Effect of the Bunching Response on Tax Revenue with A Kink

Tax revenue

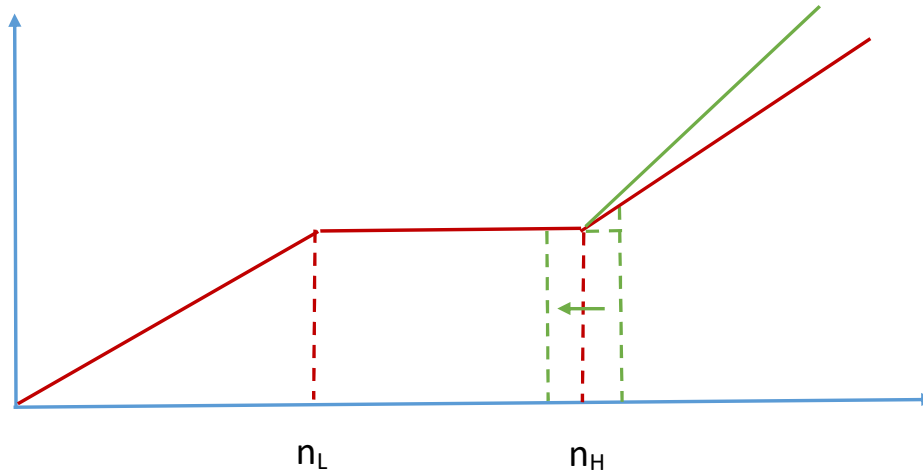
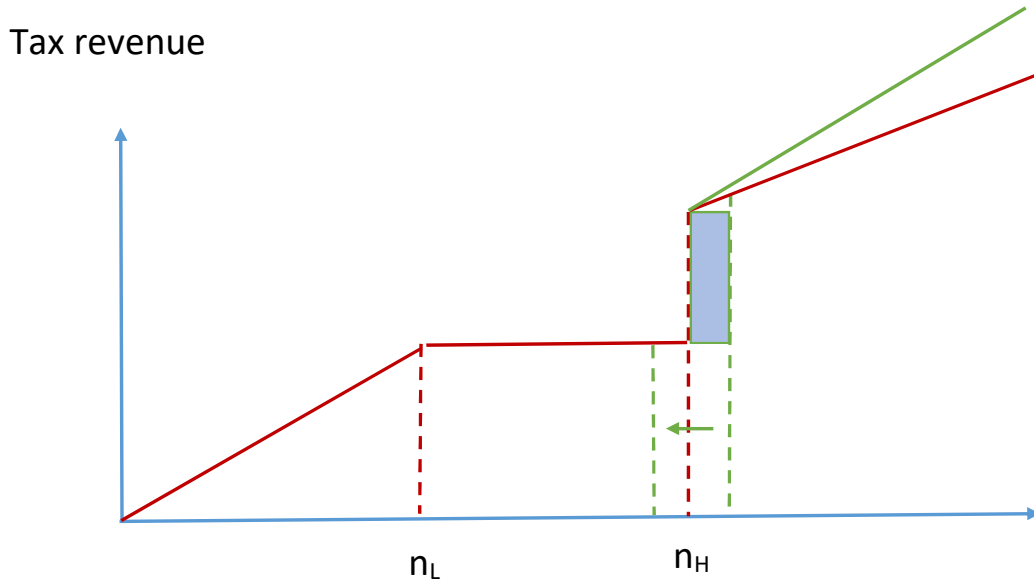


Figure 2 below shows the same change in the top rate of tax, but for a notched tax. With a notch, there is a discrete increase in the tax liability above the notch. When the top rate of tax  $t_H$  increases, as before, the green arrow shows the bunching response; some households move from above the threshold to just below. But now, it is clear that this change causes a *first-order drop* in tax revenue, as shown by the grey square; this is because tax revenue as a function of  $n$  is discontinuous at this point.

Figure 2: The Effect of the Bunching Response on Tax Revenue with A Notch



Finally, the result that the bunching response on tax revenue is zero for a kink also helps to clarify some confusion in the literature. As already noted, [Saez et al. \(2012\)](#) argue that for sufficient statistic formulae to apply in the kink case, what is required is that assumption that "behavioral responses take place only along the intensive margin", or more precisely that the bunching response of an increase in the top rate of tax is of second order relative to the extensive margin response. Proposition 1 shows that this assumption is not required, because no matter how large is  $\frac{\partial n_H}{\partial t_H}$ ,  $\frac{\partial R}{\partial n_H} = 0$  in the kink case.

## 4.2 The Marginal Excess Burden

Here, we derive a formula for the marginal excess burden (MEB) of  $t_H$  when there is a notch and show that it can be written as the MEB of a proportional tax plus a correction factor. To define the MEB, note that due to quasi-linearity, the natural measure of welfare is the integral of indirect utilities, say  $W$ , plus revenue  $R$ , which is assumed to be redistributed as a lump-sum back to households when calculating the MEB. So,

$$MEB = -\frac{d(W + R)/dt_H}{dR/dt_H} \quad (14)$$

The minus sign ensures that the marginal excess burden is measured as a positive number.

From (12), we see that the effect of an increase in  $t_H$  on tax revenue is:

$$\frac{dR}{dt_H} = B_H + \underbrace{t_H \frac{\partial B_H}{\partial t_H} \Big|_{n_H \text{ const}}}_{\text{intensive-margin}} + \underbrace{\frac{\partial R}{\partial n_H} \frac{\partial n_H}{\partial t_H}}_{\text{bunching}} \quad (15)$$

Here

$$B_H = \int_{n_H}^{\bar{n}} z(1 - t_H; n)h(n)dn \quad (16)$$

is the base in which the higher rate of tax is levied.

So, (15) is composed of three terms, the mechanical effect  $B_H$ , and two behavioral effects on tax revenue, the intensive-margin and bunching effects. The intensive-margin effect on tax revenue is standard; it describes how tax revenue changes because of changes in earnings, conditional on the taxpayer staying the same tax bracket. The bunching effect on tax revenue and its impact on the marginal excess burden is the focus of our investigation.

To compute  $dW/dt_H$ , note first that the integral of indirect utilities is

$$W = \int_{\underline{n}}^{n_L} v(1 - t_L; n)h(n)dn + \int_{n_L}^{n_H} (z_0(1 - t_L) - \psi(z_0; n))h(n)dn + \int_{n_H}^{\bar{n}} v(1 - t_H; n)h(n)dn \quad (17)$$

By definition, a small change in  $n_H$  has no effect on welfare, because  $n_H$  is defined by (6) above. So, using  $v_t = -z$ , we see that

$$\frac{dW}{dt_H} = - \int_{n_H}^{\bar{n}} z(1 - t_H; n)h(n)dn = -B_H \quad (18)$$

So, plugging (15), (18) back into (14), dividing through by  $B_H$ , multiplying by  $1 - t_H$ , and noting that holding  $n_H$  constant,  $\frac{\partial B_H}{\partial(1-t_H)} = -\frac{\partial B_H}{\partial t_H}$ , we see that

$$MEB = \frac{t_H e + C}{1 - t_H - t_H e + C}, \quad C = -\frac{1 - t_H}{B_H} \frac{\partial R}{\partial n_H} \frac{\partial n_H}{\partial t_H} \quad (19)$$

Here,

$$e = \left. \frac{1 - t_H}{B_H} \frac{\partial B_H}{\partial(1 - t_H)} \right|_{n_H \text{ const}} = \frac{1 - t_H}{B_H} \int_{n_H}^{\bar{n}} \frac{\partial z(1 - t_H; n)}{\partial(1 - t_H)} h(n)dn \quad (20)$$

is the intensive-margin elasticity of the tax base  $B_H$  with respect to the net of tax rate  $1 - t_H$ , and  $C$  is a correction factor, which captures the effect of a changing  $n_H$ , the bunching response, on the MEB, via its effect on revenue. Of course, given the specification (1),  $e$  is a constant independent of  $n_H$ .

We can then prove <sup>10</sup>;

---

<sup>10</sup>This and subsequent Propositions are proved in the Appendix.

**Proposition 2.** Assume iso-elastic utility (1), and that the distribution of  $n$  is Pareto, with shape and scale parameters  $a, \underline{n}$ . Then, the MEB with a notch is

$$MEB = \frac{t_H e + C}{1 - t_H - t_H e - C}, \quad (21)$$

where

$$C = \frac{(t_H - t_L z_0 / \tilde{z}_H)(a - 1)(1 + e)}{1 - \left(\frac{z_0}{\tilde{z}_H}\right)^{(1+e)/e}} > 0. \quad (22)$$

Moreover, in (22), and  $\tilde{z}_H = n_H(1 - t_H)^e$  and  $n_H$  is defined by (6).

Some comments are appropriate at this point. First, the MEB (21) is the formula for the marginal excess burden of a *proportional* income tax, as shown by Feldstein (1999), plus the correction factor  $C$ . This is intuitive; all households above  $n_H$  are paying tax at rate  $t_H$  on all their income, so for these households,  $t_H$  is indeed a proportional tax. So, as already remarked, the correction factor  $C$  just captures the effect of a changing  $n_H$ , the bunching response, on the MEB, via its effect on revenue.

Second, we can ask how the MEB compares to the MEB in a kinked tax system. As shown for example, by Saez (2001), the latter is

$$MEB_K = \frac{t_H e a}{1 - t_H - t_H e a}$$

Clearly,  $MEB_K$  depends only on simple sufficient statistics; other than the tax rate  $t_H$ , it depends only on  $e$ , the intensive-margin elasticity of taxable income, and  $a$ , the shape parameter of the income distribution.

By contrast, from (22), it is clear that  $C$  is a more complex object. It depends not only on sufficient statistics  $e, a$ , and the top rate of tax,  $t_H$ , but also on other parameters of the tax system  $t_L, z_0$ , and on  $\tilde{z}_H$ , which is the unconstrained earnings of the type  $n_H$ , given that they face the higher rate of tax.

So, there are two ways of solving for  $C$ . One is simply to compute  $C$  using the formulae (22), (6), choosing calibrated values for  $e, a, z_0$ , and that is what we do in this paper. Alternatively, as shown by Kleven and Waseem (2013), in any empirical study of a notch, the earnings  $n_H(1 - t_L)^e$  can be estimated. Specifically,  $n_H(1 - t_L)^e$  is simply  $z^* + \Delta z^*$  in the notation of their paper, where  $z^*$  is the earnings notch and as explained there,  $\Delta z^*/z^*$  can be estimated from excess bunching at the notch. Given this,  $\tilde{z}_H$  can be recovered simply by multiplying  $z^* + \Delta z^*$  by  $(1 - t_H)^e / (1 - t_L)^e$ .

### 4.3 The Welfare-Maximizing Rate of Tax

In his well-known article, Saez (2001) derived a formula for the welfare-maximizing rate of tax for a one-bracket tax system, where a tax is levied at rate  $t$  above some income

$z_0$ .<sup>11</sup> He showed that this tax depended only on  $e$  and  $a$ , plus a parameter he called  $\bar{g}$ , which is the "ratio of social marginal utility for top bracket taxpayers to the marginal value of public funds for the government." In the special case where  $\bar{g} = 0$ , his formula gives the revenue-maximizing rate of tax.

Here, we develop a similar formula for the optimal  $t_H$ . We will show that under some conditions, it is equal to the formula for the welfare-maximizing proportional tax, plus a term in the correction factor  $C$  above. To do this, we assume now, following Saez, that the government's objective is not the integral of indirect utilities as in (17), but the integral of a strictly increasing, concave transformation  $G(\cdot)$  of utilities. The function  $G$  captures social aversion to inequality in the usual way. We assume, following Saez, that the welfare weight  $g = G'$  is constant above  $n_H$  at some  $\bar{g}$ ; if this is not the case, the optimal tax has an additional term in the covariance of  $g$  and  $z$ . We suppose that the government has a fixed revenue requirement  $E$ , and we normalize the Lagrange multiplier on the government revenue constraint  $R \leq E$  to unity. Then we can show:

**Proposition 3.** *The welfare-maximizing level of  $t_H$  is;*

$$t^* = \frac{1 - \bar{g} - C}{1 - \bar{g} + e} \quad (23)$$

with  $C$  defined in (22) above.

To interpret this, note first that there is a direct connection of (23) to the formula derived by Saez (2001) for the optimal linear tax on top earners, which is equation (9) in his paper. He allowed for an income effect in labor supply in his setting, so setting this equal to zero, his equation, in our notation, reduces to

$$t^* = \frac{1 - \bar{g}}{1 - \bar{g} + ea} \quad (24)$$

Here,  $a$  is the shape parameter of the Pareto distribution of the upper tail of the income distribution, as above. So, there are two differences between (24) and (23), working in opposite directions. First, there is the correction factor  $-\frac{C}{1-\bar{g}+e}$  in (23), lowering  $t^*$  in the notch case. This reflects the fact that with a notch, there is an additional cost to taxation because of the effect of the bunching response on tax revenue. Second, an  $a > 1$  tends to lower  $t^*$  in the kink case.

To get a closer comparison, note that if  $z_0 = 0$ , so that there is no lower tax bracket, then we have just a proportional tax on all incomes. The optimal value of this tax is given by setting  $a = 1$  in (24), so

$$t^* = \frac{1 - \bar{g}}{1 - \bar{g} + e} \quad (25)$$

Note, however, that now the  $\bar{g}$  in (25) is the average welfare weight across *all* households,

---

<sup>11</sup>This is given in equation (9) of Saez (2001).

whereas the  $\bar{g}$  in (23) is the average welfare weight across all households above the notch.

So, to make a more precise comparison, we can assume that the government has no redistributive objective, so that  $g$  is constant across all households. Then, the terms  $\bar{g}$  in (25) and (23) are the same and equal to  $g$ . So, our final conclusion is that if the government has no redistributive objective, *the optimal top rate of tax with a notch is equal to the optimal proportional tax minus the correction factor  $\frac{C}{1-\bar{g}+e}$* . A special case of this is, of course, where the government is a tax revenue-maximizer, so  $\bar{g} = 0$ .

## 5 Simulations

We have seen that the MEB of an increase in  $t_H$  and the optimal  $t_H$  are given by the corresponding formulae for a proportional tax  $t_H$  plus a correction factor,  $C$ . Moreover, the formulae for a proportional tax are very simple, depending only on the intensive-margin elasticity  $e$ , and thus can easily be calculated.

So, a key question is whether we can get a good approximation to *MEB* by setting  $C = 0$  i.e. treating  $t_H$  as a proportional tax. In this section, we investigate whether the MEB and optimal tax, calculated assuming that  $t_H$  is a proportional tax, are good approximations to the true MEB and optimal tax.

To do this, we need to calibrate the model. In particular, we require values for  $e, a, t_H, t_L$ , and  $z_0$ . Our baseline parameter values are chosen as follows. Following [Piketty and Saez \(2013\)](#), we set  $a = 1.5$ , and following [Saez et al. \(2012\)](#) and [Kleven and Schultz \(2014\)](#), we set  $e = 0.25$ . Regarding the tax rates, we first set  $t_L = 0.2$ , which is broadly in line with the average income and payroll tax paid by US households<sup>12</sup>. It is also the basic rate of income tax in the UK. For the notch, we use the fact that notches in personal income tax, where they exist, are small. For example, [Kleven and Waseem \(2013\)](#) show that in the Pakistani income tax, the notch ranges between 2 and 5 percentage points. So, we will take our baseline notch  $t_H - t_L = \Delta t = 0.03$ .

To choose  $\underline{n}, z_0$  we assume that only the top 20% of the population pay a higher rate of income tax, roughly the proportion in the UK. Define  $n_0$  to be the skill level corresponding to taxable income just at the notch i.e.  $n_0(1 - t_L)^e = z_0$ . This requires that 80% of the population have skills below  $n_0$  i.e.  $H(n_0) = 1 - \left(\frac{\underline{n}}{n_0}\right)^\alpha = 0.8$ , or  $\frac{\underline{n}}{n_0} = (0.2)^{1/1.5} = 0.342$ . Given that only the ratio  $\frac{\underline{n}}{n_0}$  is determined, we set  $\underline{n} = 1$ , so  $n_0 = 2.924$ . But then  $z_0 = 2.924(0.8)^{0.25} = 2.168$ .

Finally, from (22), we need a value for  $n_H$ . Under the assumption (1), the indifference condition (6) reduces to

$$e(n_H)^{-1/e} (z_0)^{1+\frac{1}{e}} + n_H(1 - t_H)^{1+e} - (1 - t_L)z_0(1 + e) = 0 \quad (26)$$

---

<sup>12</sup>"Overview Of The Federal Tax System As In Effect For 2015", Joint Committee on Taxation, Congress of the United States.

Equation (26) has two roots, and we take the larger root to ensure that  $n_H(1 - t_L)^e > z_0$ . Finally, parameter values are chosen so that the denominator in (21) is positive, which is equivalent to  $dR/dt_H > 0$  i.e. that the tax rate is on the right side of the Laffer curve. This requires simply that the notch is greater than 0.0015.<sup>13</sup>

Figures 3, 4 show both the true MEB, as given by (21), and the approximation, treating  $t_H$  as a proportional tax i.e. setting  $C = 0$  in (21). The former is denoted by  $MEB$  in the Figures, and the latter by  $MEB_A$ .

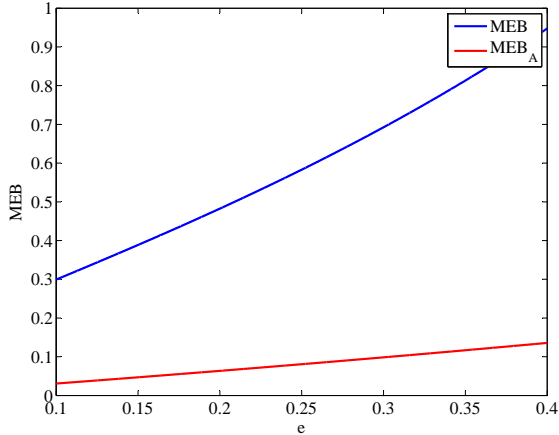


Figure 3: MEB as  $e$  varies

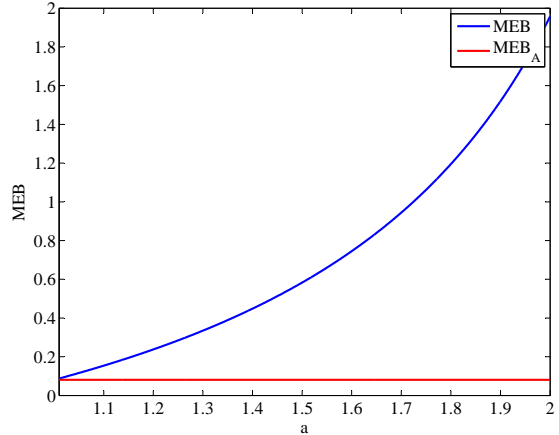


Figure 4: MEB as  $a$  varies

The error in using  $MEB_A$  at the baseline values can be read off from Figure 3, setting  $e = 0.25$ . It can be seen that true MEB is about 0.6, whereas the approximation is about 0.1. So, the error in using the proportional formula is about a factor of six. Figure 3 also shows that  $MEB$  is increasing in  $e$ , at a faster rate than  $MEB_A$ , so when  $e = 0.4$  for example, the error in using  $MEB_A$  is almost an order of magnitude.

Figure 4 shows that  $MEB$  is also increasing in  $a$ , the Pareto parameter which measures (inversely) the size of the tail of the income distribution. As  $MEB_A$  is independent of  $a$ , this means that the error in using  $MEB_A$  is increasing in  $a$ .

Finally, Figure 5 shows  $MEB$ ,  $MEB_A$  as the size of the tax notch varies. We can see that as the notch becomes very small, the true MEB becomes very large. This is because  $C \rightarrow \infty$  as  $t_H \rightarrow t_L$ . While this cannot be proved analytically, the intuition is clear from (22). As  $t_H \rightarrow t_L$ , then  $\tilde{z}_H = n_H(1 - t_H)^e \rightarrow z_0$ . So, both numerator and denominator in (22) tends to zero, but the denominator does so faster.

<sup>13</sup>For the denominator in (21) to be positive, we require  $1 - t_H(1 + e) > C$ , which is satisfied for  $t_H - t_L > 0.0015$ .



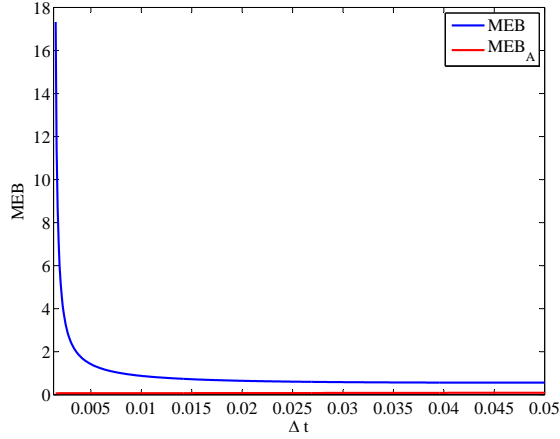


Figure 5: MEB as the notch  $\Delta t$  varies

We now turn to the question of how well (25) approximates (23), assuming that the government has no redistributive objective. Figures 6 and 7 below show the optimal tax  $t^*$  for the case where the welfare weight on all households is  $\bar{g} = 0.25 > 0$ , a similar assumption to the one made by Saez (2001)<sup>14</sup>. In each figure, we show  $t^*$  as defined in (23), along with the approximation setting  $C = 0$ , which we denote by  $t_A^*$ .

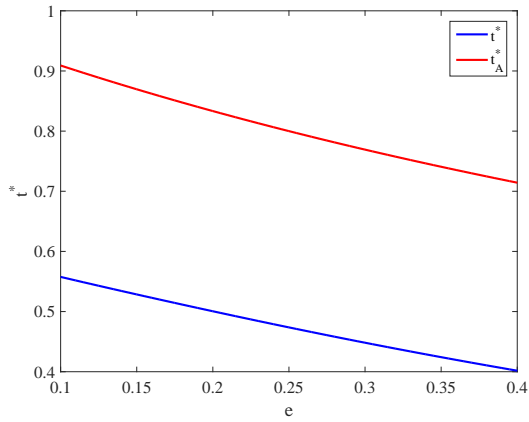


Figure 6:  $t^*$  as  $e$  varies

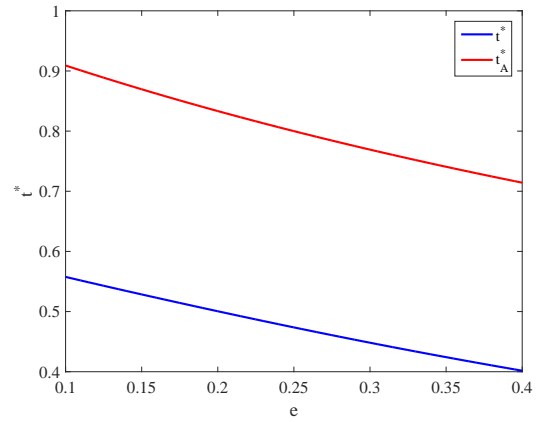


Figure 7:  $t^*$  as  $a$  varies

Both of these taxes are decreasing in  $e$ , as we might expect. Also, both taxes are decreasing in  $a$ . The error in using  $t_A^*$  as an approximation for  $t^*$  is generally smaller than for the MEB. For example, at baseline parameter values, the true welfare-maximizing tax is about 0.5, whereas the approximation is 0.8.

<sup>14</sup>In Saez (2001), numerical values of (24) are presented, assuming that the welfare weight on higher-rate taxpayers is 0.25

## 6 An Application to VAT

### 6.1 The Set-Up

As remarked in the introduction, perhaps the most important example of a tax notch is the value-added tax. In this section, we present a simple model of value-added tax, based on [Liu and Lockwood \(2015\)](#), which is mathematically equivalent to the model developed above. We then calibrate the model using UK data from [Liu and Lockwood \(2015\)](#), to estimate the MEB from the VAT, taking into account bunching at the threshold.

Consider a single industry with a fixed, large number of small traders producing a homogeneous good. Each small trader combines his own labor input  $l$  with an intermediate input  $x$  to produce output  $y$  via a fixed coefficients technology

$$y = \min \left\{ l, \frac{x}{\gamma} \right\}, \quad (27)$$

where  $\gamma$  measures the the input requirement per unit of output. In particular, for all traders, to produce one unit of output requires  $\gamma$  units of input.

Individual traders are indexed by a skill or taste parameter  $m \in [\underline{m}, \bar{m}]$ , assumed continuously distributed in the population with distribution  $H(m)$  and density  $h(m)$ . A trader of type  $m$  has an overall payoff of

$$u(l; m) = \pi - \psi(l; m), \quad \psi(l; m) = \frac{Am}{1 + \frac{1}{e}} \left( \frac{l}{m} \right)^{1 + \frac{1}{e}} \quad (28)$$

where  $\pi$  is profit and  $\psi(l; m)$  is the disutility of labour. So, traders are differentiated by disutility of labor. This assumption is not essential, but facilitates comparison to the income tax case.<sup>15</sup>

For simplicity, it is assumed that traders only sell to final consumers, who have perfectly elastic demand for the good at price  $p = 1$ . This is analogous to the assumption made in the taxable income literature that the wage is fixed, i.e. labor demand is perfectly elastic at a fixed wage. Finally, the intermediate input is produced only from labor supplied by non-trader households via a fixed-coefficients technology where one unit of labor is needed to produce one unit of the intermediate input. So, the tax-exclusive price of the output is  $w$ , the wage, which we also assume to be 1.

The traders and the producer of the intermediate inputs face a VAT system. It is assumed that the producer is VAT-registered. If the trader is registered, he must charge VAT on sales  $y$  at rate  $t$ , but can claim back any VAT paid on inputs. The trader must register for VAT if the value of sales  $y$  exceeds the threshold  $y_0$ , but can register voluntarily even if  $y < y_0$ .

Note that when not registered, the price of the input is  $1 + t$ . So, the profit for the

---

<sup>15</sup>For example,  $m$  could enter into the production function, (27) instead.

non-registered trader is

$$\pi_N = (1 - \gamma(1 + t))y. \quad (29)$$

where  $\gamma$  is the cost of inputs relative to revenue per unit sold. For the registered trader, we reason as follows. This trader must charge VAT on his output. None of the output VAT can be passed on to the buyer, as he has perfectly elastic demand. So, revenue per unit sold is  $p/(1 + t)$ . But, if the trader is registered, he can claim back VAT on the input use  $x$ , so the price of the input is  $w$ . So, overall, the profit for the registered trader is

$$\pi_R = \left( \frac{1}{1 + t} - \gamma \right) y. \quad (30)$$

We now assume, to make the analysis interesting, that  $1 > \gamma(1 + t)$ . From (29), this ensures that non-registered firms make a positive profit. Also, it ensures that for a given value of sales  $y$ ,  $\pi_N > \pi_R$ , so there is no voluntary registration. This is important because then the VAT threshold functions exactly like a tax notch.

## 6.2 Effective VAT Rates

Now define  $n \equiv m(1 - \gamma)$ . Then, substituting (29), (30) into (28), after some rearrangement, we can show that the payoff of trader  $n$  can be written as a function of value-added  $z = y(1 - \gamma)$  and the VAT system as follows;

$$u(z; n) = z - T(z) - \frac{A}{(1 - \gamma)} \frac{n}{1 + \frac{1}{e}} \left( \frac{z}{n} \right)^{1 + \frac{1}{e}} \quad (31)$$

where

$$T(z) = \begin{cases} t_N z, & z \leq z_0 \\ t_R z, & z > z_0 \end{cases}, \quad t_R = \frac{t}{(1 + t)(1 - \gamma)}, \quad t_N = \frac{\gamma t}{1 - \gamma}. \quad (32)$$

As  $A$  is a free parameter, we set it equal to  $1 - \gamma$ . Then, (31), (32) describe a utility function and a tax schedule as function of  $Z$  that are mathematically equivalent to the income tax model - although, obviously, the economic interpretation of  $z$  is different. Here,  $t_N, t_R$  are the *effective tax rates* faced by non-registered and registered traders respectively on the value-added they generate. Obviously, both effective rates are increasing in the statutory rate,  $t$ . Also, note that both rates are increasing in input intensity  $\gamma$ . Moreover, from our assumption  $1 > (1 + t)\gamma$ ,  $t_R > t_N$ .

So, faced with the tax schedule (32), all traders in the interval  $n \in [n_L, n_R]$  will bunch at the VAT threshold  $z_0$ . Moreover,  $n_L = z_0/(1 - t_N)^e$ , and  $n_R$  solves (26) with  $t_H, t_L$  replaced by  $t_R, t_N$ .

Finally, letting  $z(1 - t; n)$  be the value-added chosen by an unconstrained firm facing tax  $t$ , it can be shown that the revenue from the VAT is as in (12), with  $t_H, t_L$  replaced

by  $t_R, t_N$  i.e.

$$R = t_N \left( \int_{\underline{n}}^{n_N} z(1 - t_N; n)h(n)dn + \int_{n_N}^{n_R} z_0 h(n)dn \right) + t_R \left( \int_{n_R}^{\bar{n}} z(1 - t_R; n)h(n)dn \right) \quad (33)$$

In (33), the base on which  $t_N$  is levied is the value-added of non-registered traders, and the base of  $t_R$  is the value-added of registered traders.

### 6.3 The Marginal Excess Burden of the VAT

With the VAT, a change in the statutory rate  $t$  of VAT will change both effective tax rates  $t_N, t_R$  unless  $\gamma = 0$  i.e. no intermediate inputs are used. This is of course, analogous to a reform that changes both  $t_H$  and  $t_L$  in the income tax model. So, for the VAT, the formula for the MEB becomes somewhat more complex. To present the formula for the MEB in this case, we need a few more definitions. First, note from (33), using  $z(1 - t; n) = (1 - t)^e n$ , the effective bases of  $t_N$  and  $t_R$  are

$$B_N = \int_{\underline{n}}^{n_N} (1 - t_N)^e n h(n) dn + z_0 (H(n_R) - H(n_N)), \quad B_R = \int_{n_R}^{\bar{n}} (1 - t_R)^e n h(n) dn \quad (34)$$

Then, from (34), the intensive-margin elasticities of  $B_R, B_N$  with respect to the net-of-tax rate are

$$\frac{1 - t_R}{B_R} \frac{\partial B_R}{\partial t_R} \Big|_{n_R \text{ const}} = e, \quad \frac{1 - t_N}{B_N} \frac{\partial B_N}{\partial (1 - t_N)} \Big|_{n_N \text{ const}} = e\phi, \quad (35)$$

where

$$\phi = \frac{\int_{\underline{n}}^{n_N} z(1 - t_N; n)h(n)dn}{B_N} < 1 \quad (36)$$

The term  $\phi$  captures a new effect of bunching; with bunching, a mass  $H(n_R) - H(n_N)$  of the non-registered firms that are bunching are unresponsive to a change in the rate of VAT, which lowers the aggregate intensive-margin elasticity of the tax base  $B_N$  with respect to  $t_N$ .<sup>16</sup>

Moreover, recall that an increase in  $t$  causes both  $t_N$  and  $t_R$  to increase, so

$$\theta = \frac{\frac{B_R}{1 - t_R} \frac{\partial t_R}{\partial t}}{\frac{B_R}{1 - t_R} \frac{\partial t_R}{\partial t} + \frac{B_N}{1 - t_N} \frac{\partial t_N}{\partial t}} \quad (37)$$

measures the importance of a change in  $t_R$  on revenue relative to  $t_N$ . Armed with these new definitions, we can state our result.

---

<sup>16</sup>A similar point has been noted before by [Slemrod et al. \(1994\)](#) and [Apps et al. \(2014\)](#) who consider the design of a two-bracket income tax. Because the tax system studied is kinked, not notched, the formula for the optimal lower rate of tax depends only on the intensive margin elasticity, but this elasticity is dampened by the fact that taxpayers at the kink do not adjust their behavior in response to the tax.

**Proposition 4.** Assume that the distribution of sales (and pretax-income) is Pareto, with shape and scale parameters  $a, \underline{n}$ . Then, the MEB of the VAT is

$$MEB = \frac{\tau\varepsilon + C}{1 - \tau(1 + \varepsilon) - C} \quad (38)$$

where

$$\tau = (1 - \theta)t_N + \theta t_R, \quad \varepsilon = \frac{(1 - \theta)t_N\phi + \theta t_R}{(1 - \theta)t_N + \theta t_R} e \quad (39)$$

and finally the correction factor is

$$C = - \frac{\frac{\partial R}{\partial n_R} \left( \frac{\partial n_R}{\partial t_N} \frac{\partial t_N}{\partial t} + \frac{\partial n_R}{\partial t_R} \frac{\partial t_R}{\partial t} \right)}{\frac{B_R}{1-t_R} \frac{\partial t_R}{\partial t} + \frac{B_N}{1-t_N} \frac{\partial t_N}{\partial t}} \quad (40)$$

So, we note now that bunching impacts the calculation of the MEB in two ways. First, as before, there is a correction factor  $C$  in (38). The correction factor is more complex than in the income tax case. The reason for the additional complexity is clear from (40); an increase in  $t$  now increases both  $t_R, t_N$  and in turn, both of these effective taxes affect  $n_R$ , the top of the bunching interval, and thus revenue. An explicit formula for  $C$  in terms of parameters can be derived as in (22) above; this is done in the Appendix.

In addition, there is a second, new effect of bunching in (39). Bunching dampens the intensive-margin response to a change in  $t$ , because at a fixed  $n_N, n_R$ , firms in this interval will not adjust their sales in response to a change in  $t$ . This is captured by the term  $\phi > 0$ , which lowers the intensive margin response from  $e$  to  $\varepsilon$ .

An interesting special case is where the small traders do not use any intermediate input, so i.e.  $\gamma = 0$ . Then from (32),  $t_N = 0, t_R = \frac{t}{1+t} = \tau$ , so (38) simplifies to

$$MEB = \frac{\frac{t}{1+t}e + C}{1 - \frac{t}{1+t}(1 + e) - C} \quad (41)$$

It can be checked that in this case,  $C$  is given by the explicit formula (22), replacing  $t_H, t_L$  by  $t_R, 0$  respectively.<sup>17</sup>

## 6.4 Simulations

Here calibrate the VAT model, and plot the true MEB in (38) and an approximation to the MEB as parameters vary<sup>18</sup>. The approximation is the one treating VAT as a

<sup>17</sup>If there is no bunching i.e. if  $t_N = t_R$ , then  $\phi = 1, C = 0$ , so  $MEB = \frac{\tau e}{1 - \tau(1 + e)}$ . But this requires that  $\tau = 1$ , so in this case,  $z \rightarrow 0$ . Also, before this point,  $1 - \tau(1 + e) < 0$ , so this case is not interesting.

<sup>18</sup>The details of the calibration are described in the Online Appendix.

proportional tax i.e. setting  $C = 0$  in (41), which gives

$$MEB_A = \frac{\frac{t}{1+t}e}{1 - \frac{t}{1+t}(1+e)}$$

The parameters are calibrated as follows. In the UK, the statutory rate of VAT is 20%, so  $t = 0.2$ . Liu and Lockwood (2016) calculate that for the universe of firms in the UK that file a corporate tax return,  $\gamma = 0.45$ . This gives  $t_N = 0.16$ ,  $t_R = 0.30$ .

Next, define  $n_0$  to be the productivity level corresponding to turnover just at the threshold i.e.  $n_0(1 - t_N)^e = z_0$ . From Liu and Lockwood (2015), 62.5% of firms are below the threshold. So,  $\frac{n}{n_0}$  must satisfy  $H(n_0) = 1 - \left(\frac{n}{n_0}\right)^{1.2} = 0.625$ , or  $\frac{n}{n_0} = (0.375)^{1/1.2} = 0.442$ . Given that only the ratio  $\frac{n}{n_0}$  is determined, we set  $\underline{n} = 1$ , so  $n_0 = 2.26$ . But then  $z_0 = 2.53(0.84)^{0.25} = 2.164$ .

Finally, we need a value for  $a$ . A prior question is whether the "upper tail" of the distribution of firm sales  $y$  is well-described by a Pareto distribution. In the case of personal incomes, a Pareto distribution of the upper tail is widely accepted, but less is known about firms. In the US, there is evidence that the size distribution of firms as measured by sales is Pareto (Luttmer (2007)), and Luttmer estimates a value for the US of  $a = 1.06$ . In the Online Appendix, I provide evidence that this is also the case for the UK, using firm sales from administrative data on corporate tax returns. I show that for firms above the VAT threshold, the estimate  $a$  is about 1.2. So, this is the figure we will use in the simulations.

Our results are given in Figures 8-11 below. Figures 8 and 9 show the simpler case with no intermediate inputs i.e.  $\gamma = 0$ , in which case we know that formula (38) reduces to formula (41).

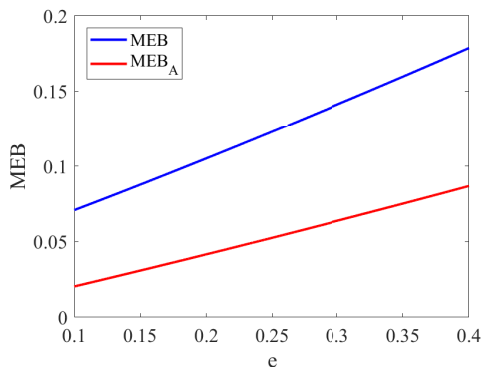


Figure 8: MEB of VAT as  $e$  varies,  
 $\gamma = 0.0$

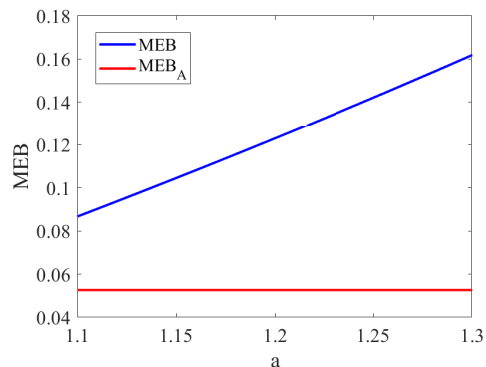


Figure 9: MEB of VAT as  $a$  varies,  
 $\gamma = 0.0$

We can see that at the baseline figures for the parameters e.g.  $e = 0.25$  in Figure 8, the true MEB is about twice as high as the approximation. This difference is much

smaller than in the income tax case, and is driven partly by the lower value of  $a$  in the VAT case. Indeed, we can see in Figure 11 that the accuracy of the approximation  $MEB_A$  falls rapidly as  $a$  rises, because  $MEB$  is increasing in  $a$  whereas  $MEB_A$  is independent of  $a$ .

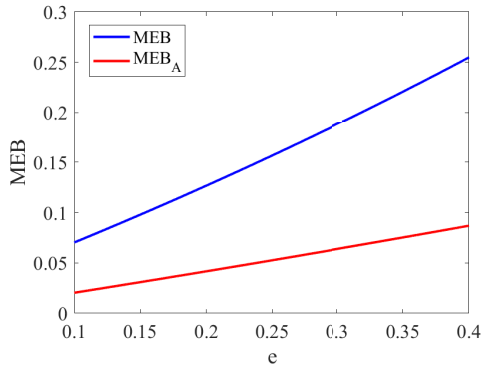


Figure 10: MEB of VAT as  $e$  varies,  $\gamma = 0.45$

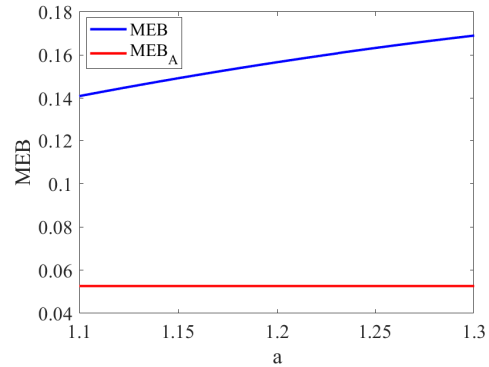


Figure 11: MEB of VAT as  $a$  varies,  $\gamma = 0.45$

Figures 10, 11 below show the more realistic case with  $\gamma = 0.45$ . Here, we see that the difference between the true MEB and the approximation is somewhat higher; the true MEB is about 3 times higher than the approximation. As in the case with no inputs, the true MEB is increasing in both  $e$  and  $a$ .

## 6.5 The Marginal Excess Burden and the Cost of VAT Compliance

In practice, there are significant compliance costs to being VAT-registered i.e. preparing and filing a tax return, and paying any tax owed. These costs will clearly induce additional bunching at the VAT notch. It is therefore of interest to ask how the additional bunching induced by compliance costs affects the MEB, in the presence of a VAT notch. In particular, we want to investigate whether compliance cost has quantitatively a similar effect on the MEB as the notch itself.

We can model compliance costs as follows. Let  $k$  be the cost of compliance as a fraction of sales at the threshold. We simplify slightly by assuming a fixed compliance cost  $ky_0$  if registered. This can be written as  $\frac{k}{1-\gamma}z_0$ ,  $z_0 = \frac{y_0}{1-\gamma}$ , so that net utility of the trader who registers is  $u_R(z; n) - \frac{k}{1-\gamma}z_0$ .

The MEB can then be calculated exactly as before, except that  $n_R$  now solves

$$(1 - t_N)z_0(1 + e) - e(n_R)^{-1/e} (z_0)^{1+\frac{1}{e}} - n_R(1 - t_R)^{1+e} + \frac{(1 + e)k}{1 - \gamma}z_0 = 0$$

To calibrate  $k$ , we note that in the UK, compliance costs as a proportion of turnover for

firms at the VAT registration threshold have been estimated at around 1.5% of turnover ([Federation of Small Businesses \(2010\)](#)). However, compliance costs can be much higher in other countries. For example, a report by PwC found that for a fictional small firm, the hours taken for compliance with VAT vary by region from an average 73 hours within the EU to 192 for Latin America, and even within the EU, there are substantial differences, with 22 hours in required in Finland to 288 in Bulgaria (PwC (2009)). To cover all these possibilities, we allow  $k$  to vary between 0% and 5% of sales.

The effect of variation in  $k$  on the MEB is shown in Figure 12. All other parameters are at their baseline values, with  $\gamma = 0.45$ . We expect that an increase in  $k$  will increase bunching and thus increase the correction factor and the MEB, and this is exactly what happens.

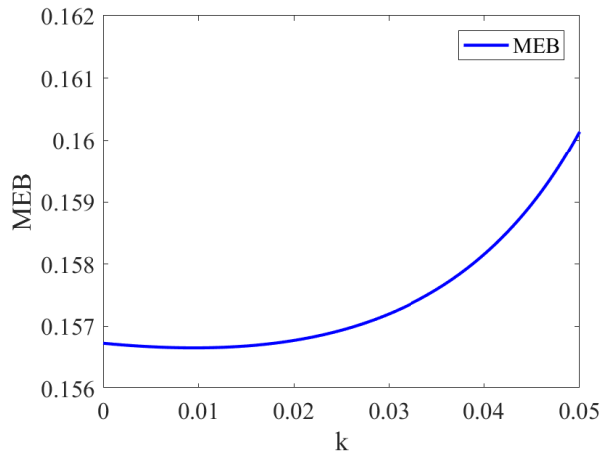


Figure 12: Compliance Costs and the MEB of a VAT

However, Figure 12 also shows that the effect is very small. An increase of compliance costs from zero to 5% of turnover at the threshold only increases the MEB from approximately 0.157 to 0.160, an increase of about 2%. So, an interesting observation is that the effect of the bunching response on the MEB of VAT is several orders of magnitude higher than that of compliance costs.

## 7 Conclusions

This paper shows that the sufficient statistic approach to the welfare properties of income (and other) taxes does not easily extend to tax systems with notches, because with notches, changes in bunching induced by changes in tax rates have a first-order effect on tax revenues. In an income tax setting, we showed that the MEB of a change in the top rate of tax is given by the [Feldstein \(1999\)](#) formula for the MEB of a *proportional* tax, plus a correction term. Also, under certain conditions, the optimal top rate of tax



is given by the formula for the optimal proportional tax, minus a correction term. These correction terms can be computed empirically, using an estimate of excess mass at the notch. Quantitatively, these correction terms can be very large.

An application to VAT was also discussed. A simple model of small traders who differ in productivity, and are subject to VAT at rate  $t$  above a threshold level of sales was shown to be formally equivalent to the income tax model. We showed that the MEB of an increase in the statutory rate of VAT is given by the Feldstein formula for a proportional tax plus a correction factor as in the income tax case. With a calibration to UK data, the MEB of the VAT is roughly three times what it would be if VAT was simply a proportional tax.

Finally, it should be noted that in this paper, we take all parameters of the tax system, including the notch, as given, and only vary the top rate of tax. A broader question, to be addressed in future work, is whether a notch can ever be part of an optimal tax system<sup>19</sup>.

---

<sup>19</sup>In the standard Mirrlees framework, where the tax is fully non-linear, this is not the case; the optimal tax schedule is always continuous in income. However, where skills are continuously distributed, and the government is restricted to a finite number of tax rates, the answer to this question is less obvious.

## References

- Apps, Patricia, Ngo Long, and Ray Rees**, “Optimal piecewise linear income taxation,” *Journal of Public Economic Theory*, 2014, 16 (4), 523–545.
- Bachas, Pierre and Mauricio Soto**, “Not (ch) your average tax system: corporate taxation under weak enforcement,” *Unpublished manuscript*, 2015.
- Ballard, Charles L, John B Shoven, and John Whalley**, “The total welfare cost of the United States tax system: a general equilibrium approach,” *National Tax Journal*, 1985, pp. 125–140.
- Best, Michael and Henrik Jacobsen Kleven**, “Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK,” Mimeo, London School of Economics 2013.
- Chetty, Raj**, “Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance,” *American Economic Journal: Economic Policy*, August 2009, 1 (2), 31–52.
- , “Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods,” *Annu. Rev. Econ.*, 2009, 1 (1), 451–488.
- Doerrenberg, Philipp, Andreas Peichl, and Sebastian Siegloch**, “The elasticity of taxable income in the presence of deduction possibilities,” *Journal of Public Economics*, 2015.
- Federation of Small Businesses**, *Impact of Increasing VAT Registration Threshold for Small Businesses*, Centre for Economics and Business Research, June 2010.
- Feldstein, Martin**, “Tax avoidance and the deadweight loss of the income tax,” *The Review of Economics and Statistics*, 1999, 81 (4), 674–680.
- Gillitzer, Christian and Joel Slemrod**, “Does Evasion Invalidate the Welfare Sufficiency of the ETI?,” *The BE Journal of Economic Analysis & Policy*, 2016, 16 (4).
- Gruber, Jon and Emmanuel Saez**, “The elasticity of taxable income: evidence and implications,” *Journal of public Economics*, 2002, 84 (1), 1–32.
- Hargaden, Enda Patrick**, “Taxpayer Responses over the Cycle: Evidence from Irish Notches,” 2015.
- Harju, Jarkko, Tuomas Matikka, and Timo Rauhanen**, “The Effects of Size-Based Regulation on Small Firms: Evidence from Vat Threshold,” Working Paper 75, VATT Institute for Economic Research June 2016.
- Kanbur, Ravi and Michael Keen**, “Thresholds, Informality, and Partitions of Compliance,” *International Tax and Public Finance*, 2014, 21 (4), 536–559.
- Keen, Michael and Jack Mintz**, “The Optimal Threshold for a Value-Added Tax,” *Journal of Public Economics*, 2004, 88 (3-4), 559–576.
- Kleven, Henrik**, “Bunching,” *Annual Review of Economics*, 2016, 8.

- Kleven, Henrik Jacobsen and Esben Anton Schultz**, “Estimating taxable income responses using Danish tax reforms,” *American Economic Journal: Economic Policy*, 2014, 6 (4), 271–301.
- **and Mazhar Waseem**, “Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan,” *The Quarterly Journal of Economics*, 2013, 128 (2), 669–723.
- Kopczuk, Wojciech and David Munroe**, “Mansion Tax: The Effect of Transfer Taxes on the Residential Real Estate Market,” *American Economic Journal: Economic Policy*, May 2015, 7 (2), 214–57.
- Liu, Li and Ben Lockwood**, “VAT Notches,” Discussion Papers 10606, CEPR May 2015.
- Luttmer, Erzo GJ**, “Selection, growth, and the size distribution of firms,” *The Quarterly Journal of Economics*, 2007, 122 (3), 1103–1144.
- Piketty, Thomas and Emmanuel Saez**, “Optimal Labor Income Taxation,” *Handbook of Public Economics*, 2013, 5, 391.
- Rutherford, Thomas and Sergey Paltsev**, “From an input-output table to a general equilibrium model: assessing the excess burden of indirect taxes in Russia,” *Draft, University of Colorado*, 1999.
- Saez, Emmanuel**, “Using Elasticities to Derive Optimal Income Tax Rates,” *The Review of Economic Studies*, 2001, 68 (1), 205–229.
- , “Do taxpayers bunch at kink points?,” *American Economic Journal: Economic Policy*, 2010, 2 (3), 180–212.
- , **Joel Slemrod, and Seth H. Giertz**, “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review,” *Journal of Economic Literature*, 2012, 50 (1), 3–50.
- Slemrod, Joel**, “Buenas notches: lines and notches in tax system design,” *eJournal of Tax Research*, 2013, 11 (3), 259.
- , **Shlomo Yitzhaki, Joram Mayshar, and Michael Lundholm**, “The optimal two-bracket linear income tax,” *Journal of Public Economics*, 1994, 53 (2), 269–290.
- Weber, Caroline E**, “Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences,” *Journal of Public Economics*, 2014, 117, 90–103.

## A Appendix

**Proof of Proposition 2.** It remains to derive a formula for  $C$ . From (8), noting that

$$\psi_n = -\frac{1}{e} \frac{n^{-(1+1/e)}}{1 + \frac{1}{e}} (z)^{1+\frac{1}{e}} = -\frac{1}{1+e} \left(\frac{z}{n}\right)^{1+1/e}. \quad (\text{A.1})$$

and  $z(1-t; n) = (1-t)^e n$ , we have

$$\frac{\partial n_H}{\partial t_H} = \frac{(1-t_H)^e n_H (1+e)}{(1-t_H)^{1+e} - \left(\frac{z_0}{n_H}\right)^{1+1/e}} \quad (\text{A.2})$$

Next, from (13) and (16), using the fact that  $z(1-t; n) = (1-t)^e n$ , we have

$$\begin{aligned} \frac{1}{B_H} \frac{\partial R}{\partial n_H} &= \frac{(t_L z_0 - t_H z(1-t_H; n_H)) h(n_H)}{\int_{n_H}^{\bar{n}} z(1-t_H; n) h(n) dn} \\ &= \frac{(t_L z_0 - t_H (1-t_H)^e n_H) h(n_H)}{(1-t_H)^e \int_{n_H}^{\bar{n}} n h(n) dn} \end{aligned} \quad (\text{A.3})$$

So, plugging (A.2), (A.3) into (N.6), we have:

$$\begin{aligned} C &= \frac{(1-t_H)(t_H(1-t_H)^e n_H - t_L z_0) h(n_H)}{(1-t_H)^e \int_{n_H}^{\bar{n}} n h(n) dn} \frac{(1-t_H)^e n_H (1+e)}{(1-t_H)^{1+e} - \left(\frac{z_0}{n_H}\right)^{1+1/e}} \\ &= \frac{(1-t_H)(t_H(1-t_H)^e n_H - t_L z_0)}{(1-t_H)^e E[n|n \geq n_H]} \frac{h(n_H)}{(1-H(n_H))} \frac{n_H(1+e)}{(1-t_H)^{1+e} - \left(\frac{z_0}{n_H}\right)^{1+1/e}} \end{aligned} \quad (\text{A.4})$$

where in the second line we have used  $\int_{n_H}^{\bar{n}} n h(n) dn = E[n|n \geq n_H] (1-H(n_H))$ .

Now, given that  $n$  follows a Pareto distribution with shape and scale parameters  $a, \underline{n}$ , we also know that

$$E[n|n \geq n_H] = \frac{a n_H}{a-1}, \quad \frac{h(n)}{1-H(n)} = \frac{a}{n} \quad (\text{A.5})$$

Plugging (A.5) into (A.4), we get:

$$C = \frac{(1-t_H)(t_H(1-t_H)^e - t_L z_0/n_H)(a-1)(1+e)}{\left((1-t_H)^{1+e} - \left(\frac{z_0}{n_H}\right)^{1+1/e}\right)} \quad (\text{A.6})$$

Then, using the definition  $\tilde{z}_H = n_H(1-t_H)^e$  in (A.6), and rearranging, we get (22) as required.  $\square$

**Proof of Proposition 3.** The government objective, written as a Lagrangean including the constraint  $R = E$ , is

$$\begin{aligned} W &= \int_{\underline{n}}^{n_L} G(v(1-t_L; n)) h(n) dn + \int_{n_L}^{n_H} G((z_0(1-t_L) - \psi(z_0; n))) h(n) dn \\ &\quad + \int_{n_H}^{\bar{n}} G(v(1-t_H; n)) h(n) dn + \lambda(R - E) \end{aligned}$$

So, the welfare-maximizing top rate of tax is defined by

$$\frac{\partial W}{\partial t_H} = - \int_{n_H}^{\bar{n}} g(n)z(1 - t_H; n)h(n)dn + \lambda \frac{\partial R}{\partial t_H} = 0 \quad (\text{A.7})$$

where  $g = G'$ . So, plugging (15), in (A.7), and rearranging, we get;

$$- \text{cov}(g, z)(1 - H(n_H)) - \bar{g}B_H + \lambda \left[ B_H + t_H \frac{\partial B_H}{\partial t_H} \Big|_{n_H \text{ const}} + \frac{\partial R}{\partial n_H} \frac{\partial n_H}{\partial t_H} \right] = 0 \quad (\text{A.8})$$

where

$$\text{cov}(g, z) = \int_{n_H}^{\bar{n}} (g - \bar{g})zh^*(n)dn, \quad h^* = h/(1 - H(n_H)), \quad \bar{g} = \int_{n_H}^{\bar{n}} gh^*(n)dn.$$

Dividing though (A.8) by  $B_H$  and  $\lambda$ , and using (20), we get

$$- \theta + 1 - \frac{t_H e}{1 - t_H} + \frac{1}{B_H} \frac{\partial R}{\partial n_H} \frac{\partial n_H}{\partial t_H} = 0 \quad (\text{A.9})$$

with

$$\theta = \frac{\text{cov}(g, z)(1 - H(n_H))}{B_H \lambda} + \frac{\bar{g}}{\lambda}$$

Multiplying (A.9) through by  $1 - t_H$  and using the definition of  $C$  in (19), we get  $(1 - t_H)(1 - \theta) - t_H e - C = 0$ . Rearranging this expression, assuming  $\text{cov}(g, z) = 0$  and normalizing  $\lambda$  to 1 gives (23).  $\square$

**Derivation of (31), (32), (33).** We first derive (31), (32). Trader utility is profit minus the disutility of labour. So, combining (A.1), (29), (30) and using  $n = m(1 - \gamma)$ ,  $l = y$ , get:

$$u_N = (1 - \gamma(1 + t))y - \frac{A}{1 - \gamma} \frac{n}{1 + \frac{1}{e}} \left( \frac{y(1 - \gamma)}{n} \right)^{1 + \frac{1}{e}} \quad (\text{A.10})$$

$$u_R = \left( \frac{1}{1 + t} - \gamma \right) y - \frac{A}{1 - \gamma} \frac{n}{1 + \frac{1}{e}} \left( \frac{y(1 - \gamma)}{n} \right)^{1 + \frac{1}{e}}$$

Now, using  $z = y(1 - \gamma)$  in (A.10), we get

$$u_N = \frac{1 - \gamma(1 + t)}{1 - \gamma} z - \frac{A}{1 - \gamma} \frac{n}{1 + \frac{1}{e}} \left( \frac{z}{n} \right)^{1 + \frac{1}{e}} \quad (\text{A.11})$$

$$u_R = \left( \frac{1}{(1 + t)(1 - \gamma)} - \frac{\gamma}{1 - \gamma} \right) z - \frac{A}{1 - \gamma} \frac{n}{1 + \frac{1}{e}} \left( \frac{z}{n} \right)^{1 + \frac{1}{e}}$$

Finally, we note that for (A.11) to imply (32), we require

$$1 - t_N = \frac{1 - \gamma(1 + t)}{1 - \gamma}, \quad 1 - t_R = \frac{1}{(1 + t)(1 - \gamma)} - \frac{\gamma}{1 - \gamma} \quad (\text{A.12})$$

But, solving (A.12) for  $t_N, t_R$ , we get (32) as required.

Now we derive (33). Let  $y(n)$  be the sales of an  $n$ -type trader. Then, revenue from

the from the VAT is

$$R = \frac{t}{1+t} \int_{n_R}^{\bar{n}} y(n)h(n)dn + t \int_{\underline{n}}^{n_R} \gamma y(n)h(n)dn \quad (\text{A.13})$$

The first term is revenue from VAT levied on the value of sales of registered firms, because the sale price is  $1/(1+t)$ , and the second term is revenue from inputs sold by the intermediate input producer to firms that do not register for VAT. Using  $z(n) = y(n)(1-\gamma)$ , we can write this as

$$R = \frac{t}{(1+t)(1-\gamma)} \int_{n_R}^{\bar{n}} z(n)h(n)dn + \frac{t\gamma}{1-\gamma} \int_{\underline{n}}^{n_R} z(n)h(n)dn \quad (\text{A.14})$$

Finally, replacing  $z(n)$  by  $z(1-t_N; n)$ ,  $z_0$ , or  $z(1-t_R; n)$  where appropriate, we get (33) as required.  $\square$

**Proof of Proposition 4.** Let  $B_N, B_R$  be the bases of the effective taxes  $t_N, t_R$  defined in (34). Then from (17),(33), and remembering that a change in the statutory rate of VAT  $t$  changes  $t_N, t_R$  via (32), we have:

$$\frac{dW}{dt} = - \left( \frac{\partial t_N}{\partial t} B_N + \frac{\partial t_R}{\partial t} B_R \right) \quad (\text{A.15})$$

$$\frac{dR}{dt} = \frac{\partial t_N}{\partial t} \left( B_N + t_N \frac{\partial B_N}{\partial t_N} \Big|_{n_R \text{ const}} \right) + \frac{\partial t_R}{\partial t} \left( B_R + t_R \frac{\partial B_R}{\partial t_R} \Big|_{n_R \text{ const}} \right) - C' \quad (\text{A.16})$$

where

$$C' = - \frac{\partial R}{\partial n_R} \left( \frac{\partial t_N}{\partial t} \frac{\partial n_R}{\partial t_N} + \frac{\partial t_R}{\partial t} \frac{\partial n_R}{\partial t_R} \right) \quad (\text{A.17})$$

So, plugging (A.15),(A.16) into (14), we have, after rearrangement

$$\begin{aligned} MEB &= - \frac{d(W+R)/dt}{dR/dt} \quad (\text{A.18}) \\ &= \frac{\frac{B_N}{1-t_N} \frac{\partial t_N}{\partial t} t_N \left( \frac{1-t_N}{B_N} \frac{\partial B_N}{\partial (1-t_N)} \Big|_{n_R \text{ const}} \right) + \frac{B_R}{1-t_R} \frac{\partial t_R}{\partial t} t_R \left( \frac{1-t_R}{B_R} \frac{\partial B_R}{\partial (1-t_R)} \Big|_{n_R \text{ const}} \right) + C'}{\frac{B_N}{1-t_N} \frac{\partial t_N}{\partial t} \left( 1 - t_N - t_N \frac{1-t_N}{B_N} \frac{\partial B_N}{\partial t_N} \Big|_{n_R \text{ const}} \right) + \frac{B_R}{1-t_R} \frac{\partial t_R}{\partial t} \left( 1 - t_R - t_R \frac{1-t_R}{B_R} \frac{\partial B_R}{\partial t_R} \Big|_{n_R \text{ const}} \right) - C'} \\ &= \frac{\frac{B_N}{1-t_N} \frac{\partial t_N}{\partial t} e\phi + \frac{B_R}{1-t_R} \frac{\partial t_R}{\partial t} t_R e + C'}{\frac{B_N}{1-t_N} \frac{\partial t_N}{\partial t} (1 - t_N(1 + e\phi)) + \frac{B_R}{1-t_R} \frac{\partial t_R}{\partial t} (1 - t_R(1 + e)) - C'} \end{aligned}$$

where in the last line, we have used (35). So, dividing top and bottom of (A.18) by  $\frac{B_R}{1-t_R} \frac{\partial t_R}{\partial t} + \frac{B_N}{1-t_N} \frac{\partial t_N}{\partial t}$  and using the definition of  $\theta$  from (37), and the definition of  $C$  from (40), we get

$$MEB = \frac{(1-\theta)t_N e\phi + \theta t_R e + C}{1 - (1-\theta)t_N(1 + e\phi) - \theta t_R(1 + e) - C} \quad (\text{A.19})$$

Finally, using the definitions of  $\tau = (1-\theta)t_N + \theta t_R$ ,  $\varepsilon = \frac{(1-\theta)t_N \phi + \theta t_R}{(1-\theta)t_N + \theta t_R} e$ , (A.19) can be rearranged to (39), as required.  $\square$

## Online Appendix

**Details of MEB Simulation for the VAT Case.** We need to express all the relevant elements of the *MEB* in terms of the parameters,  $t, \gamma, z_0,$  and  $n_R, n_N$ . In turn, we know that  $n_N = z_0/(1 - t_N)^e$  and that  $n_R$  is determined by

$$(1 - t_N)z_0(1 + e) - e(n_R)^{-1/e} (z_0)^{1+\frac{1}{e}} - n_R(1 - t_R)^{1+e} = 0 \quad (\text{N.1})$$

Assume that the distribution of firms is Pareto with shape and scale parameters  $a, \underline{n}$ . Without loss of generality, we assume  $\underline{n} = 1$ ; so, the distribution and density of  $n$  is  $H(n) = 1 - n^{-a}$ ,  $h(n) = \frac{a}{n^{a+1}}$ . So, using these formulae and  $z(1 - t; n) = (1 - t)^e n$ , we have by routine calculation;

$$\begin{aligned} B_R &= (1 - t_R)^e \int_{n_R}^{\bar{n}} nh(n)dn = (1 - t_R)^e \frac{a}{a - 1} (n_R)^{1-a} \\ B_N &= (1 - t_N)^e \int_1^{n_N} nh(n)dn + z_0(H(n_R) - H(n_N)) \\ &= (1 - t_N)^e \frac{a}{a - 1} (1 - (n_N)^{1-a}) + z_0((n_N)^{-a} - (n_R)^{-a}) \end{aligned} \quad (\text{N.2})$$

Moreover, from the formulae for  $t_N, t_R$  in the paper, we have:

$$\frac{\partial t_R}{\partial t} = \frac{1}{(1 - \gamma)(1 + t)^2}, \quad \frac{\partial t_R}{\partial t} = \frac{\gamma}{(1 - \gamma)} \quad (\text{N.3})$$

So, plugging (N.3) into the formula for  $\theta$  in the paper, we can write

$$\theta = \frac{\frac{B_R}{1-t_R}}{\frac{B_R}{1-t_R} + \frac{B_N}{1-t_N} \gamma (1+t)^2} \quad (\text{N.4})$$

Plugging (N.2) into (N.5) allows us to compute  $\theta$  as a function of  $t, \gamma, z_0,$  and  $n_N, n_R$ .

Next, using  $z(1 - t; n) = (1 - t)^e n$ , and the properties of the Pareto distribution, we have;

$$\phi = \frac{\int_{\underline{n}}^{n_N} z(1 - t_N; n)h(n)dn}{B_N} = \frac{(1 - t_N)^e \frac{a}{a-1} (1 - (n_N)^{1-a})}{B_N} \quad (\text{N.5})$$

So, using (N.2), (N.5),  $\phi$  can be computed as a function of  $t, \gamma, z_0,$  and  $n_N, n_R$ .

Finally, recalling the definition of  $C$  in the paper, we have:

$$\begin{aligned} C &= -\frac{\frac{\partial R}{\partial n_R} \left( \frac{\partial t_N}{\partial t} \frac{\partial n_R}{\partial t_N} + \frac{\partial t_R}{\partial t} \frac{\partial n_R}{\partial t_R} \right)}{\frac{B_N}{1-t_N} \frac{\partial t_N}{\partial t} + \frac{B_R}{1-t_R} \frac{\partial t_R}{\partial t}} \\ &= -\frac{\frac{\partial R}{\partial n_R} \left( \gamma \frac{\partial n_R}{\partial t_N} + \frac{1}{(1+t)^2} \frac{\partial n_R}{\partial t_R} \right)}{\frac{B_N}{1-t_N} \gamma + \frac{B_R}{1-t_R} \frac{1}{(1+t)^2}} \end{aligned} \quad (\text{N.6})$$

where in the second line, we use (N.3).

It remains to calculate  $\frac{\partial n_R}{\partial t_N}$ ,  $\frac{\partial n_R}{\partial t_R}$ ,  $\frac{\partial R}{\partial n_R}$ . From (N.1), we have:

$$\begin{aligned}\frac{\partial n_R}{\partial t_R} &= \frac{(1-t_R)^e n_R (1+e)}{(1-t_R)^{1+e} - \left(\frac{z_0}{n_R}\right)^{1+1/e}} \\ \frac{\partial n_R}{\partial t_N} &= \frac{-z_0(1+e)}{(1-t_R)^{1+e} - \left(\frac{z_0}{n_R}\right)^{1+1/e}}\end{aligned}\tag{N.7}$$

Moreover, from the formula for  $\frac{\partial R}{\partial n_R}$  in the paper, and the iso-elastic form of  $z(1-t, n)$ , we get

$$\frac{\partial R}{\partial n_R} = (t_N z_0 - t_R (1-t_R)^e n_R) h(n_R)\tag{N.8}$$

Plugging (N.7), (N.8) into (N.6), and using the formula for the density of the Pareto density to substitute out  $h(n_R)$ , we eventually get:

$$C = \frac{(t_R(1-t_R)^e n_R - t_N z_0) \left( (1-t_R)^e n_R \frac{1}{(1+t)^2} - z_0 \gamma \right) (1+e)}{\left( (1-t_R)^{1+e} - \left(\frac{z_0}{n_R}\right)^{1+1/e} \right) \left( \frac{B_R}{1-t_R} \frac{1}{(1+t)^2} + \frac{B_N}{1-t_N} \gamma \right)} \frac{a}{(n_R)^{a+1}}$$

This expression for  $C$  is computable knowing  $t$ ,  $\gamma$ ,  $z_0$ , and  $n_R, n_L$ . Thus, all the components of  $MEB$  in the paper can be calculated.  $\square$

**Calculation of the Pareto Parameter for UK firms.** We use the method of [Luttmer \(2007\)](#) and others to estimate the distribution of firm size for the UK using corporate tax return data. Firm size  $y$  is measured by sales. If the distribution of firm size is Pareto, the log of the size of the upper tail of the distribution of firm size is linear in  $y$ , with the coefficient on  $y$  being  $a$ .

We briefly describe the data here: a fuller description is given in [Liu and Lockwood \(2015\)](#). We have annual sales of firms, taken from the universe of corporation tax records (CT600) in financial years 2004/5 to 2009/10. The data is then refined by eliminating companies which are part of a larger VAT group i.e. using only standard-alone independent companies. We also drop all observations with partial-year corporation tax records. In addition, we eliminate companies that mainly engage in overseas activities. This yields a data-set with 731,706 observations for 435,688 companies between April 1, 2004 and March 30, 2010.

To analyze the data, we group firm sales into bins of size £10,000. A visual inspection of the data (available on request) indicates that log of the size of the upper tail of the distribution of firm size is near to linear in  $y$ . We then regress (by year) the size of the upper tail of the distribution, denoted  $1-F$ , on firm size as measured by sales. We define the upper tail as starting at the VAT threshold in any year. The coefficient on sales in this regression gives a value for  $a$ . Inspection of the results in [Table 1](#) below indicates that for our population of UK firms,  $a$  is approximately 1.2.



Table 1: Regression of Upper Tail of the Size Distribution on Firm Size

	Dependent Variable: log of 1-F					
	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10
Sales	-1.202*** (0.004)	-1.182*** (0.004)	-1.180*** (0.004)	-1.202*** (0.004)	-1.193*** (0.004)	-1.182*** (0.004)
Constant	3.353*** (0.029)	3.192*** (0.027)	3.285*** (0.028)	3.413*** (0.031)	3.344*** (0.029)	3.133*** (0.026)
R-squared	0.975	0.977	0.975	0.971	0.974	0.979
N	2119	2185	2336	2349	2421	2334