

Online Appendix for "Malas Notches"

The Extended Model with Tax Avasion. Maximizing (24) in the paper with respect to z, s subject to (25) gives first-order conditions for the choice of z, s respectively;

$$1 - T'(z) - \psi_z(z + s; n) = 0 \quad (\text{N.1})$$

$$1 - g'(s) - a'(s) - \psi_z(z + s; n) = 0 \quad (\text{N.2})$$

Here, for convenience, we denote the derivative of ψ with respect to $z + s$ as ψ_z . Note from (N.1), (N.2) that the household equates the marginal cost of evasion $g' + a'$, to the marginal tax rate, T' as in Chetty (2009b). Now let $T' = t$. From (N.1), (N.2) we can then solve for reported and sheltered income as functions of the net of tax rate $1 - t$, which we write as $z(1 - t; n)$ and $s(1 - t; n)$ respectively. However, note that with evasion, even assuming the iso-elastic specification for disutility of income, (2), no longer gives a closed-form solution for z in terms of $1 - t$, as z, s are simultaneously determined via (N.1), (N.2).

Then, the behavior of the household faced with a kink or a notch is qualitatively the same as before. That is, under either type of tax schedule, there is a bunching interval $[n_L, n_H]$. In the case of a kink, n_L, n_H are characterized by (5), (6) as before. In the case of a notch, (7) is modified to

$$(1 - t_L)z_0 + \hat{s} - g(\hat{s}) - a(\hat{s}) - \psi(z_0 + \hat{s}; n_H) = v(1 - t_H; n_H) \quad (\text{N.3})$$

Here, $\hat{s}$ is the unique solution to (N.2) for type n_H , where also $z = z_0, T' = t_L$ and so $\hat{s}$ is the optimal amount of evaded income for this type at the threshold. Also,

$$v(1 - t; n) = \max_{z, s} \{ (1 - t)z + s - \psi(z + s; n) - g(s) - a(s) \}$$

So, (N.3) says that the marginal buncher at the top is the one for whom the utility from keeping reported income at z_0 , while optimizing sheltered income, is equal to the unconstrained indirect utility, given the higher rate of tax on all reported income.

Given this characterization of household behavior, it is then easily seen that much the analysis above generalizes to this case, with the mathematics being unchanged, while bearing in mind the new interpretation of $z(1 - t; n)$. In particular, it is still the case that the effect on revenue R of a change in n_H is zero in the kink case and negative in the notch case, as this simply follows from the (dis-)continuity of R in the kink (notch) case. So, Proposition 1 continues to hold.

As for Proposition 2, we can establish, using the envelope theorem, that

$$\frac{dW}{dt_H} = -B_H + A_H, \quad A_H = \int_{n_H}^{\bar{n}} a'(s(1-t_H; n)) s_t((1-t_H; n)) h(n) dn > 0 \quad (\text{N.4})$$

The intuition for $A_H > 0$ is as in Chetty (2009b); the audit cost is a transfer to government, which is recycled back to the households, and so reduces the welfare cost of the tax increase.

Using (N.4), and the formula (18) above, simple manipulation gives a formula for the excess burden

$$MEB = \frac{t_H \bar{e} + C - \frac{(1-t_H)A_H}{B_H}}{1 - t_H(1 + \bar{e}) - C} \quad (\text{N.5})$$

where C is given by the same formula as in Proposition 2. So, we can make two observations.

First, in the special case where there is no audit cost of evasion, i.e. $a \equiv 0$, Proposition 2 continues to hold. So, our basic insight that the MEB of a notch is equal to the MEB of a proportional tax plus a correction factor generalizes to the case of evasion. However, Proposition 3, and in particular, the explicit formula for C , depends heavily on the simple functional form $z(1-t; n) = (1-t)^e n$, and so does not generalize, even if ψ , g , and a are all assumed iso-elastic.

Second, in the more realistic case where there is an audit cost, the MEB is equal to the MEB of a proportional tax plus two correction factors, one for the notch, C , and the other $\frac{(1-t_H)A_H}{B_H}$, capturing the fact that the audit cost is a transfer. As the first is positive and the second is negative, they have offsetting effects on the MEB.

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, γ, 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

$$(1-t_N)z_0(1+e) - e(n_R)^{-1/e} (z_0)^{1+\frac{1}{e}} - n_R(1-t_R)^{1+e} = 0 \quad (\text{N.6})$$

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;

$$\begin{aligned} B_R &= (1-t_R)^e \int_{n_R}^{\bar{n}} n h(n) dn = (1-t_R)^e \frac{a}{a-1} (n_R)^{1-a} \\ B_N &= (1-t_N)^e \int_1^{n_N} n h(n) dn + z_0(H(n_R) - H(n_N)) \\ &= (1-t_N)^e \frac{a}{a-1} (1 - (n_N)^{1-a}) + z_0((n_N)^{-a} - (n_R)^{-a}) \end{aligned} \quad (\text{N.7})$$

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

$$\frac{\partial t_R}{\partial t} = \frac{1}{(1-\gamma)(1+t)^2}, \quad \frac{\partial t_R}{\partial t} = \frac{\gamma}{(1-\gamma)} \quad (\text{N.8})$$

So, plugging (N.8) into the formula for θ in the paper, we can write

$$\theta = \frac{\frac{B_R}{1-t_R}}{\frac{B_R}{1-t_R} + \frac{B_N}{1-t_N}\gamma(1+t)^2} \quad (\text{N.9})$$

Plugging (N.7) into (N.10) allows us to compute θ as a function of t, γ, 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;

$$\phi = \frac{\int_n^{n_N} z(1-t_N; n) h(n) dn}{B_N} = \frac{(1-t_N)^e \frac{a}{a-1} (1-(n_N)^{1-a})}{B_N} \quad (\text{N.10})$$

So, using (N.7), (N.10), ϕ can be computed as a function of t, γ, z_0 , and n_N, n_R .

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

$$\begin{aligned} C &= -\frac{\frac{\partial R}{\partial n_R} \left(\frac{\partial t_N}{\partial t} \frac{\partial n_R}{\partial t_N} + \frac{\partial t_R}{\partial t} \frac{\partial n_R}{\partial t_R} \right)}{\frac{B_N}{1-t_N} \frac{\partial t_N}{\partial t} + \frac{B_R}{1-t_R} \frac{\partial t_R}{\partial t}} \\ &= -\frac{\frac{\partial R}{\partial n_R} \left(\gamma \frac{\partial n_R}{\partial t_N} + \frac{1}{(1+t)^2} \frac{\partial n_R}{\partial t_R} \right)}{\frac{B_N}{1-t_N} \gamma + \frac{B_R}{1-t_R} \frac{1}{(1+t)^2}} \end{aligned} \quad (\text{N.11})$$

where in the second line, we use (N.8).

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.6), we have:

$$\begin{aligned} \frac{\partial n_R}{\partial t_R} &= \frac{(1-t_R)^e n_R (1+e)}{(1-t_R)^{1+e} - \left(\frac{z_0}{n_R} \right)^{1+1/e}} \\ \frac{\partial n_R}{\partial t_N} &= \frac{-z_0(1+e)}{(1-t_R)^{1+e} - \left(\frac{z_0}{n_R} \right)^{1+1/e}} \end{aligned} \quad (\text{N.12})$$

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

$$\frac{\partial R}{\partial n_R} = (t_N z_0 - t_R (1-t_R)^e n_R) h(n_R) \quad (\text{N.13})$$

Plugging (N.12), (N.13) into (N.11), 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, γ, 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) 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