Michael Carlos Best, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn & Mazhar Waseem

Production vs revenue efficiency with limited tax capacity: theory and evidence from Pakistan

Conference paper

Original citation:
Originally presented at Best, Michael Carlos, Brockmeyer, Anne, Kleven, Henrik Jacobsen, Spinnewijn, Johannes and Waseem, Mazhar (2013) *Production vs revenue efficiency with limited tax capacity: theory and evidence from Pakistan.* In: Growth week 2013, 23 - 25 Sep 2013, London, UK.

Production vs Revenue Efficiency With Limited Tax Capacity: Theory and Evidence From Pakistan*

Michael Carlos Best, Anne Brockmeyer, Henrik Jacobsen Kleven,
Johannes Spinnewijn & Mazhar Waseem

London School of Economics

September 2013

Abstract

This paper analyzes the design of tax systems under imperfect enforcement. A common policy in developing countries is to impose minimum tax schemes whereby firms are taxed either on profits or on turnover, depending on which tax liability is larger. This production inefficient tax policy has been motivated by the idea that the broader turnover tax base is harder to evade. Minimum tax schemes give rise to a kink point in firms' choice sets as the tax rate *and* tax base jump discontinuously when one tax liability surpasses the other. Using administrative tax records on corporations in Pakistan, we find large bunching around the minimum tax kink. We show that the combined tax rate and tax base change at the kink provides small *real* incentives for bunching, making the policy ideal for eliciting evasion. We develop an empirical approach allowing us to put (tight) bounds on the evasion response to switches between profit and turnover taxation, and find that turnover taxes reduce evasion by up to 60-70% of corporate income. Our analysis sheds new light on the use of production-inefficient tax tools in countries with limited tax capacity and can easily be replicated in other contexts as the quasi-experimental variation needed is ubiquitous.

*E-mail addresses of authors: m.c.best@lse.ac.uk; a.brockmeyer@lse.ac.uk; h.j.kleven@lse.ac.uk; j.spinnewijn@lse.ac.uk; m.waseem@lse.ac.uk. We thank Tony Atkinson, Michael Devereux, Jim Hines, Tim Schmidt-Eisenlohr, Agnar Sandmo, Joel Slemrod and numerous seminar participants for discussions and comments. Financial support from the Economic and Social Research Council (ESRC) Grant ES/J012467/1 is gratefully acknowledged.

1 Introduction

A central result in public economics, the *production efficiency theorem* (Diamond & Mirrlees 1971), states that tax systems should leave production undistorted even in second-best environments. This result permits taxes on consumption, wages and profits, but precludes taxes on intermediate inputs, turnover and trade. The theorem has been hugely influential in the policy advice given to developing countries, but a key concern with such advice is that the underlying theoretical assumptions are ill-suited to settings with limited tax capacity. In particular, the theorem considers an environment with perfect tax enforcement—zero tax evasion at zero administrative costs—which is clearly at odds with the situation in developing countries. Once we allow for tax evasion or informality, it may be desirable to deviate from production efficiency if this leads to less evasion and therefore larger revenue efficiency. While there is some theoretical work along these lines (e.g. Emran & Stiglitz 2005; Gordon & Li 2009), there is virtually no empirical evidence on the trade-off between production and revenue efficiency in the choice of tax instruments.

To address this question empirically, we need simultaneous variation in tax instruments that vary with respect to their production efficiency properties (such as switches between two instruments). This is more challenging than the usual search for variation in tax rates for a given tax instrument, because we are interested in comparing instruments that apply to different tax bases and often to very different taxpayer populations. A few studies have taken a macro cross-country approach focusing on trade vs domestic taxes (Baunsgaard & Keen 2010; Cage & Gadenne 2012). This paper proposes instead a micro approach that exploits a production inefficient tax policy commonly observed in developing countries. This is the imposition of *minimum tax schemes* according to which firms are taxed either on profits or on turnover (with a lower rate applying to turnover), depending on which tax liability is larger.[1] This policy has been motivated by the idea that the broader turnover base is harder to evade, an argument that seems intuitive but is so far untested. Crucially, these minimum tax schemes give rise to kink points in firms' choice sets: the tax rate *and* tax base jump discontinuously at a threshold for the profit rate (profits as a share of turnover), but tax liability is continuous at the threshold. We show that such kinks provide an ideal setting for estimating evasion responses to switches between profit and turnover taxes using a bunching approach, allowing us to evaluate the desirability of deviating from production efficiency to achieve more revenue efficiency. Compared to existing bunching approaches (Saez 2010; Chetty *et al.* 2011; Kleven & Waseem 2013) a conceptual contribution is to develop a method that exploits the

[1] Such minimum tax schemes have been implemented in numerous developing countries, including Argentina, Bolivia, Cambodia, Cameroon, Chad, Colombia, Democratic Republic of the Congo, Ecuador, El Salvador, Equatorial Guinea, Gabon, Guatemala, Guinea, Honduras, India, Ivory Coast, Kenya, Laos, Madagascar, Malawi, Mauritania, Mexico, Morocco, Nigeria, Pakistan, Panama, Philippines, Puerto Rico, Republic of the Congo, Rwanda, Senegal, Taiwan, Tanzania, Trinidad and Tobago, and Tunisia (see Ernst & Young 2013 for a description). Most of these minimum tax schemes are based on turnover, but a few of them are based on alternative bases such as total assets or broader taxable income measures in between profits and turnover.

simultaneous discontinuity in the tax rate and the tax base.

The basic empirical idea is that excess bunching at the minimum tax kink will be driven (mostly) by evasion or avoidance responses rather than by real production responses. To see this, consider first a stylized comparison between a turnover tax and a pure profit tax on an individual firm. Because turnover is a much broader base than profits, minimum tax schemes are always associated with very small turnover tax rates as compared to profit tax rates. For example, in our empirical application to Pakistan, the turnover tax rate is .5% while the profit tax rate is 35%. The low turnover tax rate implies that this tax introduces only a *small* distortion of real production at the *intensive* margin, while a profit tax levied on true economic profits would be associated with a zero distortion of real production at the intensive margin.[2] Hence, the simultaneous changes in tax base and tax rate at the kink offset each other to produce a very small change in *real* incentives for the individual firm. On the other hand, because the tax bases are completely different on each side of the kink, there will be a large change in *evasion* incentives if those bases are associated with different evasion opportunities. Hence, if we see large bunching at the minimum tax kink, this is difficult to reconcile with real output responses under reasonable elasticity parameters and provides *prima facie* evidence of an evasion response to the switch between turnover and profit taxation. We show in the paper that this basic argument is robust to a number of generalizations, including real distortions of turnover taxes driven by cascading[3] and distortionary profit taxes levied on bases that deviate from pure economic rent.[4] We develop a simple model allowing us to put bounds on the evasion response using bunching at the minimum tax kink under different assumptions about the real output elasticity. Due to the weak real incentive, the bounds on the evasion response are extremely tight under a very wide range of real output elasticities.

We use administrative data from the Federal Board of Revenue in Pakistan to analyze the responses by Pakistani firms to the minimum tax regime. The data contains all corporate tax returns between 2006 and 2010, which are predominantly filed electronically, contributing to the quality of the data in this context. Our main empirical findings are the following. First, we observe large and sharp bunching in reported profit rates around the threshold below which the turnover

[2]We provide two important clarifications to this small-distortion argument in the paper. First, because of the broadness of the turnover base, a small tax rate on turnover may create a large average tax rate on profits (tax liability as a share of profits) and therefore create significant distortions at the extensive margin (such as informality, sector, and location choices). However, since tax liability is continuous at the minimum tax kink, there will be no extensive responses to the kink and our bunching estimates are not affected by such responses. Second, a small turnover tax rate can create a large reduction in turnover when production technology is close to constant returns to scale (in which case the turnover elasticity with respect to the turnover tax rate becomes very large). However, with near-constant returns to scale changes in turnover have almost no impact on profit rates, and so real production responses are not able to generate much bunching around the minimum tax kink even in this case.

[3]While a small turnover tax introduces only a small *firm-level* distortion of real production at the intensive margin, there may still be significant *economy-wide* distortions because of cascading—taxing the same item multiple times—through the production chain. As we explain later, these general equilibrium cascading effects do not generate *bunching* at the minimum tax kink, which is a great advantage for our ability to identify evasion.

[4]In general, actual corporate income taxes do not correspond to taxes on pure economic profits, and so may be associated with significant real distortions (e.g. Hassett & Hubbard 2002; Devereux & Sørensen 2006; Auerbach *et al.* 2010). By itself, this effect makes a profit tax *more* distortionary compared to a turnover tax. We show that this creates real production incentives around the minimum tax kink that move firms *away* from the kink, and therefore reinforces our argument that bunching is not driven by real responses.

tax applies. We exploit variation in the minimum tax kink over time and across firms to confirm that the excess bunching is indeed a response to the tax system. The variation includes a temporary elimination of the minimum tax scheme as well as differences in the size and location of the kink for different populations of firms. These findings provide compelling non-parametric evidence that firms respond to the minimum tax incentives in the way that our theory predicts, and the presence of weak real incentives around the kink suggests that evasion is an important part of the story. To explore the role of evasion, we also show that firms with greater evasion opportunity—either smaller firms or firms with less activities subject to a paper trail—bunch more strongly at the minimum tax kink.

Second, we combine our non-parametric bunching evidence with a simple conceptual framework in order to bound the evasion response to switches between profit and turnover taxes under different assumptions on the real output elasticity. We find that turnover taxes reduce evasion by up to 60-70% of corporate income compared to profit taxes. The evasion estimates are very robust to the size of the real output elasticity, because the smallness of real incentives around the kink implies that real responses contribute very little to bunching even under very large elasticities. Third, we use our empirical estimates as sufficient statistics in an analysis of the optimal choice of tax base and tax rate in an environment with limited tax capacity. We find that welfare can be increased by moving away from a pure profit tax towards a much broader base that is closer (but not identical) to turnover, because the loss of production efficiency is more than compensated for by the increase in tax compliance. It is in general not optimal to go all the way to a pure turnover tax in our framework, but the administrative simplicity of a pure turnover tax could further tip the balance in practice. Overall, our findings demonstrate that governments with limited tax capacity face an important trade-off between production efficiency and revenue efficiency that has first-order implications for the choice of tax instruments.

Our paper contributes to several literatures. First, we contribute to an emerging empirical literature on public finance and development using administrative microdata (Kleven & Waseem 2013; Pomeranz 2013; Kumler *et al.* 2013). Second, a theoretical literature has studied the implications of limited tax capacity for optimal taxation (Emran & Stiglitz 2005; Keen 2008; Boadway & Sato 2009; Gordon & Li 2009; Kleven *et al.* 2009; Besley & Persson 2011; Dharmapala *et al.* 2011). While most of these papers study movements between the formal and informal sectors, our paper studies corporate tax evasion at the intensive margin and derives simple expressions for optimal tax policy that depend on parameters which we estimate.[5]

Third, a vast literature studies the determinants of tax evasion (see Andreoni *et al.* 1998 and Slemrod & Yitzhaki 2002 for surveys). This literature has used macroeconomic indicators (money supply, aggregate electricity demand etc.), survey data on consumption and income, or audit data to estimate the extent of tax evasion (see Slemrod 2007, Fuest & Riedel 2009 and Slemrod & Weber 2012 for surveys). However, with the exception of the rare occasions when randomised audits are available (Slemrod *et al.* 2001; Kleven *et al.* 2011), methodological limitations mean that

[5]Both the empirical and the theoretical literatures on public finance and development are surveyed in Besley & Persson (2013).

the credibility and precision of these estimates are questionable. Our paper contributes a novel methodology for the estimation of evasion using quasi-experimental variation created by tax policy. The approach generates robust estimates of evasion and can be easily replicated in other contexts as the tax variation needed is ubiquitous, especially in the developing world. Fourth, our paper is related to the literature on taxable income elasticities (Saez *et al.* 2012), and especially work that emphasizes the endogeneity of taxable income elasticities to the broadness of the tax base (Slemrod & Kopczuk 2002; Kopczuk 2005). Finally, we contribute to the large stream of literature studying responses by corporations to the tax code (see Auerbach 2002, Hassett & Hubbard 2002 and Auerbach *et al.* 2010 for surveys, and Gruber & Rauh 2007, Bach 2012, Dwenger & Steiner 2012, Kawano & Slemrod 2012 and Devereux *et al.* 2013 for recent estimates of the elasticity of corporate taxable income with respect to the effective marginal tax rate).

The paper is organized as follows. Section 2 presents our conceptual framework, which is used in section 3 to develop an empirical methodology based on minimum tax schemes. Section 4 describes the context and data and section 5 presents our results. Section 6 numerically analyzes optimal policy, and section 7 concludes.

2 Conceptual Framework

This section develops a stylized model of the firm to analyze the optimal design of a tax on the firm's activities. Our analysis focuses on the firm's responses to tax rate and tax base changes in environments with or without tax evasion. When tax enforcement is perfect, the optimal tax system leaves the firm's production decision undistorted by taxing profits. When tax enforcement is imperfect, it becomes optimal to move towards a distortionary tax on turnover/output if this discourages tax evasion by firms.[6] The stylized model allows us to identify sufficient statistics that capture this trade-off between production efficiency and revenue efficiency (compliance) and guides our empirical strategy in the next section. Our framework abstracts from general equilibrium effects of taxation, consistent with the empirical application (using bunching) which is also not affected by potential general equilibrium effects.

2.1 Firm Behavior and Tax Policy Without Evasion

A firm chooses how much output y to produce at a convex cost $c(y)$. The firm pays taxes $T[y, c(y)] = \tau [y - \mu c(y)]$, which depend on the tax rate τ and a tax base parameter μ. The tax base parameter equals the share of costs that can be deducted from a firm's revenues when determining the tax base. The tax base thus ranges from an output tax base to a pure profit tax base when increasing μ from 0 to 1. The firm's after-tax profits equal

$$\Pi(y) = (1 - \tau) y - c(y) + \tau \mu c(y). \tag{1}$$

[6]Normalizing the output price to 1, turnover and output are of course identical and so we will in general use the terms "output tax" and "turnover tax" to mean the same thing throughout the paper.

The profit-maximizing output level solves

$$c'(y) = 1 - \tau \frac{1-\mu}{1-\tau\mu} \equiv 1 - \omega, \tag{2}$$

where ω denotes the tax wedge between the social and private return to output. For a pure profit tax base ($\mu = 1$), the tax wedge disappears and the output choice is efficient, regardless of the tax rate. For an output tax base ($\mu = 0$), the tax wedge equals the tax rate. The impact of the tax rate τ and the base parameter μ on the firm's output choice depends on the implied change in the tax wedge ω, with $\frac{\partial \omega}{\partial \tau} \geq 0$ and $\frac{\partial \omega}{\partial \mu} \leq 0$. The change from a high tax rate on a profit tax base to a lower tax rate on a broader output tax base will only affect the firm's output choice if it affects the tax wedge ω.

The government sets tax parameters τ, μ to maximize welfare subject to an exogenous revenue requirement R. In this stylized framework, this amounts to maximizing after-tax profits (corresponding to aggregate consumption by firm owners) subject to the revenue requirement. Hence, the welfare objective of the government can be written as

$$W = \Pi(y) + \lambda \{T[y, c(y)] - R\}, \tag{3}$$

where the firm's output choice satisfies (2) and $\lambda \geq 1$ denotes the (endogenous) marginal cost of public funds. The welfare effect of changing the tax parameters τ, μ can be decomposed into a *mechanical* welfare effect from transferring resources from the firm to the government for a given output level and a *behavioral* welfare effect due to the response in output. While the behavioral response in y affects welfare through government revenue, it has only a second-order welfare effect through firm profits (envelope result following from $\Pi'(y) = 0$). From equations (1) and (3), the mechanical welfare effect of the tax rate τ (normalized by the marginal cost of funds λ) can be written as $M_\tau \equiv [y - \mu c(y)] \times [\lambda - 1]/\lambda \geq 0$. The mechanical welfare effect of the tax base parameter μ (again normalized by λ) can be written as $M_\mu \equiv -\tau c(y) \times [\lambda - 1]/\lambda \leq 0$. Both mechanical effects equal 0 if the marginal cost of public funds λ equals 1. The total welfare impact of τ and μ equal respectively

$$\frac{\partial W}{\partial \tau}/\lambda = M_\tau + \omega \frac{\partial y}{\partial \tau}, \tag{4}$$

$$\frac{\partial W}{\partial \mu}/\lambda = M_\mu + \omega \frac{\partial y}{\partial \mu}. \tag{5}$$

This allows us to establish a natural implication of the production efficiency theorem in this stylized model. With a pure profit tax base, the government can raise taxes without distorting the firm's output. Hence, if possible, it is optimal to tax pure profits.

Proposition 1 (**Production Efficiency**)**.** *With perfect tax enforcement, the optimal tax base is given by the firm's pure profit (i.e.,* $\mu = 1$*).*

Proof. For $\mu = 1$, the government can increase tax revenues by increasing the tax rate without affecting the production choice. Hence, the marginal cost of public funds $\lambda = 1$. The government sets $\tau = R/[y - c(y)]$. For any $\mu < 1$, we can increase μ by $d\mu$ and increase τ by $d\tau = \frac{\tau c(y)}{y - \mu c(y)} d\mu$ so that the mechanical welfare effects cancel out. Hence, the impact on welfare equals

$$\begin{aligned} dW &= \frac{\partial W}{\partial \tau} d\tau + \frac{\partial W}{\partial \mu} d\mu \\ &= \lambda \omega \frac{\partial y}{\partial \omega} \frac{\partial \omega}{\partial \mu} \left[\frac{\partial \omega / \partial \tau}{\partial \omega / \partial \mu} \frac{\tau c(y)}{y - \mu c(y)} + 1 \right] d\mu. \end{aligned}$$

The impact on welfare is positive if the wedge ω decreases in response to the change. This is true if and only if

$$-\frac{\partial \omega / \partial \tau}{\partial \omega / \partial \mu} \leq \frac{y - \mu c(y)}{\tau c(y)}.$$

Since $-\frac{\partial \omega / \partial \tau}{\partial \omega / \partial \mu} = \frac{1-\mu}{\tau(1-\tau)}$, the condition simplifies to the after-tax profits being positive,

$$(1 - \tau)(y - \mu c(y)) - (1 - \mu) c(y) \geq 0,$$

which is always satisfied for the firm's output choice. □

2.2 Firm Behavior and Tax Policy With Evasion

We now relax the assumption of perfect enforcement and analyze optimal tax policy in the presence of tax evasion. In particular, we incorporate in our model the notion that an output tax is harder to evade than a profit tax, the argument being that it is harder to evade a broader base and possibly also that the *level* of output is more visible than the *difference* between output and input. We capture the relative ease of evading profit taxes by allowing firms to declare costs $\hat{c} \neq c(y)$ at a convex cost of misreporting $g(\hat{c} - c(y))$ with $g(0) = 0$.[7] The key implications are similar if we also allow the output level y to be misreported, as long as it remains harder to evade an output tax than a profit tax.

The firm again maximizes its after-tax profits, but these now depend on both real output y (at real costs $c(y)$) and reported costs $\hat{c}$ for tax purposes,

$$\Pi(y, \hat{c}) = \hat{\Pi}(y, \hat{c}) - g(\hat{c} - c(y)) = (1 - \tau) y - c(y) + \tau \mu \hat{c} - g(\hat{c} - c(y)), \tag{6}$$

where $\hat{\Pi}(\cdot)$ denotes the reported after-tax profits (i.e., exclusive of costs of evasion $g(\cdot)$). At the firm's optimum,

$$c'(y) = 1 - \omega, \tag{7}$$

$$g'(\hat{c} - c(y)) = \tau \mu. \tag{8}$$

[7] The modelling of evasion (or avoidance) based on a convex and deterministic cost function $g(.)$ was first proposed by Slemrod (2001). This model is simpler to work with than a model incorporating the probability of audit and choice under uncertaintly as in the classic formulation of Allingham & Sandmo (1972).

The output level depends on the tax wedge ω in exactly the same way as before, and is therefore not affected by the presence of evasion.[8] The level of evasion is increasing in the base parameter μ and is thus higher for a profit tax base than for an output tax base. The level of evasion is also increasing in the tax rate τ. The latter result relies on the assumption that the cost of evasion $g(\cdot)$ depends on the difference between reported and true costs rather than on the difference between reported and true tax liability (Allingham & Sandmo 1972; Yitzhaki 1974), but this assumption is not key for the main analytical insights that we present below.

With evasion, the government's tax revenue can be decomposed into the revenue based on the true tax base and the foregone revenue due to misreporting the base,

$$T[y,\hat{c}] = \tau \times \underbrace{[y-\mu\hat{c}]}_{\text{reported base}} = \tau \times \{\underbrace{[y-\mu c(y)]}_{\text{true base}} - \underbrace{\mu[\hat{c}-c(y)]}_{\text{unreported base}}\}.$$

The government's welfare objective can be written as $W = \Pi(y,\hat{c}) + \lambda\{T[y,\hat{c}] - R\}$, where we assume that the private cost of evasion $g(\cdot)$ is also a social cost by including $\Pi(\cdot) = \hat{\Pi}(\cdot) - g(\cdot)$ in W.[9] The (normalized) mechanical welfare effects of τ and μ can be written as $M_\tau \equiv [y-\mu\hat{c}]\times[\lambda-1]/\lambda \geq 0$ and $M_\mu \equiv -\tau\hat{c}\times[\lambda-1]/\lambda \leq 0$. Hence, the total welfare effects of τ, μ equal

$$\frac{\partial W}{\partial \tau}/\lambda = M_\tau + \omega\frac{\partial y}{\partial \tau} - \tau\mu\frac{\partial(\hat{c}-c)}{\partial \tau}, \tag{9}$$

$$\frac{\partial W}{\partial \mu}/\lambda = M_\mu + \omega\frac{\partial y}{\partial \mu} - \tau\mu\frac{\partial(\hat{c}-c)}{\partial \mu}. \tag{10}$$

Both an increase in the tax rate ($\tau \uparrow$) and an increase in the tax base ($\mu \downarrow$) entail a positive mechanical welfare effect, but a negative revenue effect through a decrease in the firm's real output. However, while an increase in the tax rate increases the level of misreporting, an increase in the tax base decreases the level of misreporting. We may state the following key proposition:

Proposition 2 (**Production Inefficiency**). *With imperfect tax enforcement, the optimal tax base is interior, i.e.,* $\mu \in (0,1)$. *The optimal tax system satisfies*

$$\frac{\tau}{1-\tau}\cdot\frac{\partial\omega}{\partial\tau}(\mu) = G(\mu)\cdot\frac{\varepsilon_{\hat{c}-c}}{\varepsilon_y}, \tag{11}$$

where $\varepsilon_{\hat{c}-c} \equiv \frac{\partial(\hat{c}-c)}{\partial\tau\mu}\frac{\tau\mu}{\hat{c}-c} \geq 0$ *is the elasticity of evasion with respect to* $\tau\mu$, $\varepsilon_y \equiv \frac{\partial y}{\partial(1-\omega)}\frac{1-\omega}{y} \geq 0$ *is the elasticity of real output with respect to* $1-\omega$, *and* $G(\mu) \equiv [\hat{c}-c(y)]/\hat{\Pi}(y,\hat{c}) \geq 0$ *is evasion as a*

[8] The independence of real production and evasion relies on the assumption of additively separable evasion costs $g(.)$ that depend only on the evasion level $\hat{c}-c(y)$, independently of the real output level y. This independence simplifies the analysis without changing the main substance of our results.

[9] The assumption that the private and social costs of evasion are the same is important for efficiency and optimal tax results (Slemrod 1995; Slemrod & Yitzhaki 2002; Chetty 2009). Examples of social evasion costs include productivity losses from operating in cash, not keeping accurate accounting books, and otherwise changing the production process to eliminate verifiable evidence. Including the evasion cost as a social cost is the natural starting point for developing countries where the revenue loss from evasion is a first-order social concern. In fact, the big-picture question on the trade-off between production and revenue efficiency that motivates this paper would be a moot point if evasion were socially costless.

share of reported profits. The evasion rate $G(\mu)$ satisfies $G(0) = 0$ and is monotonically increasing in μ, while $\frac{\partial\omega}{\partial\tau}(\mu) = \frac{1-\mu}{(1-\tau\mu)^2} \geq 0$ satisfies $\frac{\partial\omega}{\partial\tau}(0) = 1$, $\frac{\partial\omega}{\partial\tau}(1) = 0$ and is monotonically decreasing in μ whenever $\tau \in \left[0, \frac{1}{2-\mu}\right]$.

Proof. For $\mu = 1$, an increase in the tax base has a second-order negative impact on production efficiency, but a first-order positive impact on evasion reduction, i.e., $\frac{\partial W}{\partial\mu}/\lambda = M_\mu - \tau\frac{\partial(\hat{c}-c)}{\partial\mu} < 0$. Notice that this result holds for $\lambda = 1$ as well, in which case $M_\mu = 0$.

For $\mu = 0$, a decrease in the tax base has a second-order negative impact on evasion reduction, but a first-order positive impact on production efficiency. Notice that $\frac{\partial W}{\partial\mu}/\lambda = M_\mu + \tau\frac{\partial y}{\partial\mu} > 0$ if M_μ is sufficiently small. However, since the impact on evasion is of second order, we can use the same argument as before to argue that a tax-neutral increase in μ and τ, for a given y, will increase y and thus increase welfare, starting from $\mu = 0$.

To characterize the relation between the tax rate τ and the tax base μ, consider again an increase $d\mu$ in μ and $d\tau = \frac{\tau\hat{c}}{y-\mu\hat{c}}d\mu$ such that the mechanical welfare effects cancel out. The welfare effect through the change in y is like in the proof of Proposition 1. Hence,

$$\begin{aligned} dW/\lambda &= \omega\frac{\partial y}{\partial\omega}\left[\frac{\partial\omega}{\partial\tau}\frac{\tau\hat{c}}{y-\mu\hat{c}} + \frac{\partial\omega}{\partial\mu}\right]d\mu - \tau\mu\frac{\partial(\hat{c}-c)}{\partial\tau\mu}\left[\frac{\partial\tau\mu}{\partial\tau}\frac{\tau\hat{c}}{y-\mu\hat{c}} + \frac{\partial\tau\mu}{\partial\mu}\right]d\mu \\ &= \omega\frac{\partial y}{\partial\omega}\frac{\partial\omega}{\partial\tau}\left[\frac{\tau\hat{c}}{y-\mu\hat{c}} + \frac{\partial\omega/\partial\mu}{\partial\omega/\partial\tau}\right]d\mu - \tau\mu\frac{\partial(\hat{c}-c)}{\partial\tau\mu}\left[\mu\frac{\tau\hat{c}}{y-\mu\hat{c}} + \tau\right]d\mu \end{aligned}$$

Rewriting this in terms of elasticities, we find

$$dW/\lambda = \left\{\frac{\tau}{1-\tau}\frac{\partial\omega}{\partial\tau}\hat{\Pi}(y,\hat{c})\,\varepsilon_y - [\hat{c}-c]\,\varepsilon_{\hat{c}-c}\right\}\frac{\tau y}{y-\mu\hat{c}}d\mu.$$

Notice that $dW/\lambda = 0$ is required for the initial level of τ and μ to be optimal, and so the expression in the proposition follows. □

Hence, in the presence of profit evasion, it is always optimal to introduce at least some production inefficiency by setting $\mu < 1$. To understand the optimal tax rule (11), note that the left-hand side $\frac{\tau}{1-\tau}\cdot\frac{\partial\omega(\mu)}{\partial\tau}$ reflects the effective marginal tax wedge on real production. This production wedge is equal to $\frac{\tau}{1-\tau}$ when $\mu = 0$, equal to zero when $\mu = 1$, and typically monotonically decreasing between those two extremes.[10] At the social optimum, the production wedge must be equal to the ratio between the evasion and output elasticities $\varepsilon_{\hat{c}-c}/\varepsilon_y$ scaled by the evasion rate $G(\mu)$. This rate is equal to zero when $\mu = 0$ and is monotonically increasing in μ. The formula highlights the trade-off between production efficiency (captured by the real output elasticity) and revenue efficiency (captured by the evasion elasticity) when setting the tax base μ. If the evasion elasticity is small relative to the real output elasticity ($\varepsilon_{\hat{c}-c}/\varepsilon_y \approx 0$), the production efficiency concern will be strong relative to the revenue efficiency concern, and so it will be socially optimal to move close to a pure

[10] The cross-derivative $\frac{\partial^2\omega}{\partial\tau\partial\omega}$ may switch signs such that the production wedge may be locally increasing in μ for $\tau < 1/(2-\mu)$. Hence, this can only occur when the tax rate is at least 50 percent.

profit tax by setting $\mu \approx 1$ (such that $\frac{\tau}{1-\tau} \cdot \frac{\partial \omega(\mu)}{\partial \tau} \approx 0$). Conversely, if the evasion elasticity is large relative to the real output elasticity, the revenue efficiency concern will be relatively strong and this makes it optimal to move towards the output tax by lowering μ, thereby simultaneously decreasing the evasion rate $G(\mu)$ and increasing the production wedge until formula (11) is satisfied.[11] The former case is arguably the one that applies to a developed country context, whereas the latter case captures a developing country context. Our stylized framework thus highlights the starkly different policy recommendations in settings with strong vs. weak tax capacity. Finally, note that the optimal tax formula (11) also identifies sufficient statistics for determining the optimal tax base and rate in our stylized framework, which we will study empirically.[12]

3 Empirical Methodology Using Minimum Tax Schemes

Using our conceptual framework, this section develops an empirical methodology that exploits a type of minimum tax scheme common to many developing countries, including Pakistan which we consider in the empirical application below. Under this type of minimum tax scheme, if the profit tax liability of a firm falls below a certain threshold, the firm is taxed on an alternative, much broader tax base than profits. The alternative tax base is typically output/turnover (e.g., in Pakistan), and we focus on this case to be consistent with our empirical application. We show that such minimum tax schemes give rise to (non-standard) kink points in firms' choice sets, and that they produce differential quasi-experimental variation in the incentives for real production and compliance.

3.1 Minimum Tax Kink and Bunching (Without Evasion)

We first consider the baseline model without evasion. Firms report turnover y and costs $c(y)$ and pay the maximum of a profit tax ($\mu = 1, \tau_\pi$) and an output tax ($\mu = 0, \tau_y$) where $\tau_y < \tau_\pi$. That is,

$$T[y, c(y)] = \max\{\tau_\pi [y - c(y)], \tau_y y\}. \tag{12}$$

Firms thus switch between the profit tax and the output tax when

$$\tau_\pi [y - c(y)] = \tau_y y \quad \Leftrightarrow \quad p \equiv \frac{y - c(y)}{y} = \frac{\tau_y}{\tau_\pi}. \tag{13}$$

[11] The optimal tax rate τ changes endogenously as μ changes to satisfy the revenue constraint.

[12] Our decomposition into real output and evasion elasticities is not in contradiction with the sufficiency of taxable income elasticities for welfare analysis (Feldstein 1995, 1999). It is possible to rewrite equation (11) in terms of the elasticities of taxable profits with respect to the tax rate τ and the tax base μ, respectively. If taxable profits are more responsive to an increase in the tax rate than to an increase in the tax base, this implies a relatively low efficiency cost associated with the tax base increase and therefore a low optimal μ. The presence of evasion, however, suggests an explanation for why these taxable profit responses may diverge as evasion is expected to respond in opposite directions to an increase in the tax rate ($\tau \uparrow$) and an increase in the tax base ($\mu \downarrow$). Our empirical methodology builds on this decomposition into real responses and evasion.

This implies a fixed cutoff τ_y/τ_π for the profit rate p (profits as a share of turnover): if the profit rate is higher than this cutoff, firms pay the profit tax; otherwise they pay the output tax. As the profit rate crosses the cutoff, the tax rate and tax base change discontinuously, but the tax liability (12) is continuous. Hence, this is a kink (a discontinuous change in *marginal* tax incentives) as opposed to a notch (a discontinuous change in *total* tax liability), but a conceptually different type of kink to those explored in previous work (Saez *et al.* 2012; Chetty *et al.* 2011) due to the joint change in tax rate and tax base.[13] In the model without evasion, firms choose only real output based on the marginal return $1-\omega$, which changes from 1 (profit tax) to $1-\tau_y$ (output tax) at the kink.

Figure I illustrates how the minimum tax kink at τ_y/τ_π creates bunching in the distribution of profit rates. The figure is based on the general model with both real responses and evasion responses, but we begin by abstracting from evasion responses. The dashed line represents the distribution of profit rates before the introduction of a minimum tax (i.e., under a profit tax). Assuming a smooth distribution of firm productivities (through heterogeneity in cost functions), this baseline distribution of profit rates is smooth and we denote it by $f_0(p)$. The introduction of a minimum tax (i.e., an output tax for $p \leq \tau_y/\tau_\pi$) reduces the marginal return to output from 1 to $1-\tau_y$ for firms initially below the cutoff. Those firms respond by reducing their output levels, which leads to an increase in their profit rates under decreasing returns to scale (such that marginal costs $c'(y)$ larger than average costs $c(y)/y$). This creates a right-shift in the profit rate distribution below the cutoff (with no change above the cutoff) and produces excess bunching exactly at the cutoff. Allowing for optimization error (as in all bunching studies), there will be bunching *around* the cutoff rather than a mass point precisely at the cutoff, as illustrated in Figure I. Finally, it is important to note that real output reductions below the kink produces excess bunching only under decreasing returns to scale. In the case of constant (increasing) returns to scale, the model without evasion predicts zero bunching (hole) at the minimum tax kink.[14]

Bunchers at the kink point τ_y/τ_π come from a continuous segment $[\tau_y/\tau_\pi - \Delta p, \tau_y/\tau_\pi]$ of the baseline distribution $f_0(p)$ absent the kink, where Δp denotes the profit rate response by the marginal bunching firm. Assuming that the kink is small, the total amount of bunching is given by $B = \Delta p \cdot f_0\left(\frac{\tau_y}{\tau_\pi}\right)$. Hence, based on estimates of excess bunching B and a counterfactual density at the kink $f_0\left(\frac{\tau_y}{\tau_\pi}\right)$, it is possible to infer the profit rate change Δp induced by the kink. In the model without evasion, this profit rate response is directly proportional to the real output elasticity. Assuming again that the minimum tax kink is small,[15] total differentiation yields

$$\Delta p = \left[\frac{c}{y} - c'(y)\right]\frac{dy}{y} \simeq \frac{\tau_y^2}{\tau_\pi}\varepsilon_y, \tag{14}$$

[13]See Kleven & Waseem (2013) for further discussion of the conceptual distinction between kinks and notches.

[14]The possibility of non-decreasing returns to scale therefore only strengthens our main conclusion below that, once we allow for both real and evasion responses, bunching at minimum tax kinks tends to be driven mainly by evasion.

[15]The small-kink assumption is common in bunching studies and has the advantage of avoiding parametric specifications of the cost functions $c(.), g(.)$.

where we use that $c'(y) = 1$ and $\frac{c}{y} = 1 - p \simeq 1 - \frac{\tau_y}{\tau_\pi}$ in the vicinity of the cutoff. The output elasticity is defined as $\varepsilon_y \equiv \frac{dy/y}{d(1-\omega)/(1-\omega)}$ and we use that $\frac{d(1-\omega)}{1-\omega} = -\tau_y$ when crossing the kink.

Based on equation (14), we note that large bunching (large Δp) will translate into an extremely large output elasticity. This follows from the observation that τ_y^2/τ_π will in general be a tiny number, because output tax rates are always small due to the fact that output/turnover is a very broad base (for example, τ_y is at most 1% in the case of Pakistan). The intuition for this result is that the combined changes in tax base μ and tax rate τ offset each other to create a very small change in the real return to output $1-\omega$, which makes the minimum tax kink a very small intervention in a model without evasion. Hence, the presence of large bunching around minimum tax kinks (which is what we find empirically) cannot be reconciled with believable real output elasticities in a model without tax evasion and therefore represents *prima facie* evidence of evasion.[16] The next section characterizes bunching responses in the model with evasion.

3.2 Minimum Tax Kink and Bunching (With Evasion)

We now turn to the model with evasion. Firms switch from profit to output taxation when the reported profit rate $\hat{p} = (y - \hat{c})/y$ falls below the cutoff τ_y/τ_π. This kink point is associated with a differential change in the incentives for real output and compliance. The marginal return to real output $1-\omega$ changes from 1 to $1-\tau_y$ when switching from profit to output taxation as before, whereas the marginal return to tax evasion $\tau\mu$ changes from τ_π to 0. Hence, for firms whose reported profit rate falls below the cutoff τ_y/τ_π absent the minimum tax, the introduction of the minimum tax reduces real output (loss of production efficiency) and increases compliance (gain in revenue efficiency). Assuming decreasing returns to scale, both effects increase the reported profit rate below the cutoff and thus produce bunching from below as shown in Figure I.

Using the decomposition $d\hat{c} = d(\hat{c} - c) + dc$ and totally differentiating, we now obtain

$$\Delta \hat{p} = \left[\frac{\hat{c}}{y} - c'(y)\right] \frac{dy}{y} - \frac{d(\hat{c} - c)}{y} \simeq \frac{\tau_y^2}{\tau_\pi} \varepsilon_y - \frac{d(\hat{c} - c)}{y}, \tag{15}$$

where we again use $c'(y) = 1$ and $\frac{\hat{c}}{y} = 1 - \hat{p} \simeq 1 - \frac{\tau_y}{\tau_\pi}$. The bunching response $\Delta \hat{p}$ thus depends on both the real output response and the evasion response, but in very different ways. The real output response will be small under any potentially believable output elasticity (due to the scale factor τ_y^2/τ_π as described above), and so a large bunching response $\Delta \hat{p}$ must imply a large evasion response to the output tax. While we cannot separately estimate real output and evasion responses using only one minimum tax kink, equation (15) allows for a bounding exercise on the evasion response under different assumptions about ε_y. Because of the smallness of the factor τ_y^2/τ_π, the

[16] In theory, the real output elasticity could be very large if the production technology is close to constant returns to scale (the elasticity goes to infinity as we converge to constant returns to scale). However, as explained above near-constant returns to scale implies near-constant profit rates even under large output responses and therefore no output-driven bunching. In other words, real output elasticities are large precisely in situations where output-driven bunching at the minimum tax kink must be small, and so the observation of large bunching cannot be credibly explained by real responses under near-constant returns to scale.

estimated evasion response will be insensitive to ε_y. Furthermore, if in addition to the presence of a minimum tax scheme, there is *random* variation in the output tax rate τ_y applying to this scheme (giving us more than one observation of $\Delta\hat{p}$ for the same values of the output elasticity ε_y and the evasion response $d(\hat{c}-c)$ under the randomness assumption), it would be possible to separately estimate the real and evasion responses.[17] Random variation in the profit tax rate τ_π is not as useful for separately estimating output and evasion responses, because the profit tax rate directly affects the evasion reponse $d(\hat{c}-c)$ to the minimum tax kink (and so does not give us additional observations of $\Delta\hat{p}$ for the same values of $d(\hat{c}-c)$).[18]

3.3 Robustness

The kink implied by the minimum tax scheme changes the incentives for production and evasion differentially. The analysis above shows that when combining a pure profit tax with a small turnover tax, the change in real incentives at the kink is minor, implying that substantial bunching provides evidence for evasion. This section shows that the key insight—that bunching at the minimum tax kink reflects mostly evasion—is robust to a number of generalizations.

Distortionary Profit Tax

The assumption that the profit tax corresponds to a tax on pure economic rent is very strong and stands in sharp contrast to the large body of literature analyzing the effective marginal tax distortion created by corporate income taxation and its real impact on corporations (see Hassett & Hubbard 2002). However, relaxing this simplifying assumption only strengthens our conclusion that observed bunching must be driven overwhelmingly by evasion responses. Other things equal, the introduction of real distortions in the profit tax regime implies that when firms move from profit to turnover taxation real incentives will deteriorate by less or potentially improve. This additional effect *by itself* implies that the minimum tax scheme improves real incentives below the kink, so that firms respond by increasing their output. An increase in output reduces a firm's true profit rate ($\Delta p \leq 0$) under non-increasing returns to scale and thus moves it *away* from the kink. Rewriting equation (15) as follows

$$\Delta\hat{p} = \Delta p - d\left(\frac{\hat{c}-c}{y}\right), \tag{16}$$

we see clearly that also in the case of a distortionary profit tax (implying $\Delta p \leq 0$ other things equal) real responses cannot be responsible for bunching at the minimum tax kink (corresponding to $\Delta\hat{p} > 0$). Hence, our argument does not rely on the small change in the production incentives around the kink, but only on the production incentives not being much smaller for the turnover tax (with a small τ_y) than for the profit tax such that Δp is small or negative. We conclude that if the

[17] We do have time variation in τ_y in our empirical application, but such variation is not plausibly random and so we focus on the bounding approach to separately estimate evasion responses.

[18] We do have variation in τ_π in our empirical application that is arguably exogenous. Even if this does not allow us to separately estimate output and evasion responses, it is still very useful for providing an additional identification check on our bunching strategy.

effective marginal tax rate under the profit tax were positive (rather than zero), our estimate of the evasion response based on the decomposition in (15) would provide a lower bound.

Distortionary Turnover Tax: Cascading and Extensive Responses

As we argued above, a low turnover tax rate generates only small distortions to *firm-level* production incentives at the *intensive* margin. Importantly, this does not imply that the overall distortionary effect of turnover taxes is small. First, even a small turnover tax could cause significant production inefficiencies because of *cascading* through the production chain. Cascading effects of multiple-stage production can imply effective distortions far higher than statutory tax rates (Keen 2013). Crucially, though, such general equilibrium distortions affect the distribution of firms on both sides of the minimum tax kink and therefore cannot generate bunching at the kink. The absence of cascading effects in our bunching estimates is a great advantage for our ability to identify evasion responses to tax base switches, but cascading would still form a potentially important part of an overall welfare evaluation of minimum tax schemes.

Second, due to the breadth of the base, even a small turnover tax can create a large average tax rate on profits for firms with low (actual) profit rates, which could cause firms to respond along the *extensive margin*. Crucially for our analysis, since tax liability is *continuous* around the minimum tax kink, the switch between profit and turnover taxation does not create extensive responses in the *vicinity* of the kink; extensive responses to the turnover tax will only be potentially important further away from the kink. Conceptually, excess bunching is a measure of the intensive response and is therefore not affected by potential extensive responses, but such effects would again be relevant for a broader policy evaluation.[19]

Output Evasion

The model can be extended to include output evasion whereby firms report output $\hat{y}$ which may differ from their true output y and face a convex cost of doing so. In this case, firms' profits are given by

$$\Pi = y - c\left(y\right) - \tau\left(\hat{y} - \mu\hat{c}\right) - g\left(\hat{c} - c\left(y\right), y - \hat{y}\right)$$

and the analog of equation (15) becomes

$$\begin{aligned} \Delta\hat{p} &= \left[\frac{\hat{c}}{\hat{y}} - c'\left(y\right)\right]\frac{dy}{\hat{y}} - \frac{d\left(\hat{c} - c\right)}{\hat{y}} - \frac{\hat{c}}{\hat{y}}\frac{d\left(y - \hat{y}\right)}{\hat{y}} \\ &\simeq \frac{\tau_y^2}{\tau_\pi}\varepsilon_y\frac{y}{\hat{y}} - \frac{d\left(\hat{c} - c\right)}{\hat{y}} - \left(1 - \frac{\tau_y}{\tau_\pi}\right)\frac{d(y - \hat{y})}{\hat{y}} \end{aligned} \tag{17}$$

[19] In particular, since profit rates vary across sectors due to differences in technology, a *uniform* turnover tax rate (as in our empirical application to Pakistan) creates differential average tax rates on profits across different sectors and therefore distorts the sectoral allocation of labour and capital. Such effects may call for sector-specific turnover tax rates depending on, for example, the average profit rate in each sector.

decomposing the bunching response $\Delta\hat{p}$ into a real response in the first term and the two evasion responses. Given the lower tax rate ($\tau_y \ll \tau_\pi$), the incentives to underreport output are arguably smaller under the output tax than under the profit tax, which would further increase the bunching response. The key point to note is that this expression preserves the feature that the real output response in the first term will be small as it is scaled by τ_y^2/τ_π and so large bunching responses must reflect some combination of output and cost evasion responses. While we have done our welfare analysis in the presence of cost evasion only, the insights do not depend on the particular form evasion takes, as long as evasion is easier under a profit tax than under the output tax. Moreover, if it were true that the minimum tax makes it easier to misreport output, then firms would reduce their reported output more under the turnover tax, moving them *away* from the kink, so the presence of bunching also directly supports the notion that evasion is easier under a profit tax regime.[20]

Pricing Power

The model can also be extended to incorporate pricing power by firms. In this case, firm profits are given by

$$\Pi = (1-\tau)\,\rho\,(y)\,y - c\,(y) + \tau\mu\hat{c} - g\,(\hat{c} - c\,(y))$$

where $\rho\,(y)$ is the price the firm receives, which depends negatively on output y. In this model, the analog of equation (15) is

$$\begin{aligned}\Delta\hat{p} &= \left[\frac{\hat{c}}{y}\,(1-\sigma) - c'\,(y)\right]\frac{dy}{\rho\,(y)\,y} - \frac{d\,(\hat{c}-c)}{\rho\,(y)\,y} \\ &\simeq (1-\sigma)\,\frac{\tau_y^2}{\tau_\pi}\varepsilon_y - \frac{d\,(\hat{c}-c)}{\rho\,(y)\,y} \qquad (18)\end{aligned}$$

where $\sigma \equiv -\frac{\partial\rho(y)}{\partial y}\frac{y}{\rho(y)} > 0$ is the price elasticity the firm faces and the second equality follows by using $\hat{c}/\rho\,(y)\,y = 1-\tau_y/\tau_\pi$ and $c'\,(y) = \rho(y)\,(1-\sigma)$ at the kink. Firms now reduce their prices when increasing output. Hence, the more elastic the demand, the less true profits will change in response to real incentives. The term multiplying the output elasticity is smaller than when we assume firms have no pricing power and so we conclude that the presence of pricing power only strengthens our interpretation of observed bunching and makes our estimate of the evasion response based on the decomposition in (15) a lower bound.

4 Context and data

4.1 Corporate Taxation: Minimum Tax Scheme

The corporate income tax is an important source of revenue in Pakistan and currently raises 2.5% of GDP, which comprises about 25% of all federal tax revenues. The tax is contributed by more than

[20] Here we have assumed that it is not possible for firms to misreport their output for the minimum tax without simultaneously misreporting their output for the profit tax. In the Pakistani context this is reasonable as firms report output once and this is used to calculate both the profit tax and the minimum tax liabilities.

20,000 corporations filing their tax returns every year. The scale of non-compliance is expected to be large in Pakistan, but credible evidence on the amount of corporate tax evasion has been lacking due to problems with data and methodology. The Federal Board of Revenue (FBR) reports an estimate of the corporate evasion rate equal to 45%, but does not provide information on the estimation.[21] A study by the World Bank (2009) estimates the evasion rate to be as high as 218% of actual corporate income tax payments, drawing on an input-output model for a selected group of sectors. It is because of the concern about corporate non-compliance that policy makers in Pakistan have devised a tax scheme, which ensures that every operational corporation pays a minimum amount of tax every year. The minimum tax scheme, which has been in place since 1991, combines a tax rate τ_π on annual corporate profits (turnover minus deductible costs) with a smaller tax rate τ_y on annual corporate turnover, requiring each firm to assess both tax liabilities and pay whichever is higher.[22]

As explained above, the minimum tax scheme implies that a firm's tax base depends on whether its profit rate (corporate profits as a share of turnover) is above or below a threshold equal to the tax rate ratio τ_y/τ_π. This profit rate threshold represents a kink point where the tax base and tax rate change discretely. The kink point varies across different groups of firms and across time. First, Pakistan offers a reduced profit tax rate for recently incorporated firms. All companies which register after June 2005, have no more than 250 employees, have annual sales below Rs. 250 million, and paid-up capital below Rs. 25 million are eligible for a lower profit tax rate. Second, both the profit tax rate and the turnover tax rate undergo changes during the time period we study. Table I catalogs these variations across firms and over time, which we exploit in our empirical analysis. Importantly, the definitions of the tax bases to which these rates are applied remain the same for the entire period under consideration. In the years we study, around 50 percent of the taxpayers are liable for turnover taxation as their reported profit rates put them below the minimum tax kink.

4.2 Data

Our study uses administrative data from FBR, covering the universe of corporate income tax returns for the years 2006-2010.[23] Since July 2007, electronic filing has been mandatory for all companies, and over 90% of the returns used in our study were filed electronically. Electronic filing ensures that the data has much less measurement error than what is typically the case for developing countries. As far as we know, this is the first study to exploit corporate tax return data for a developing country. The filed returns are automatically subject to a basic validation check that uncovers any

[21] See FBR (2012-2013).

[22] When the turnover tax is binding, firms are allowed to carry forward the tax paid in excess of the profit tax liability and adjust it against next year's profit tax liability, provided that the resulting net liability does not fall below the turnover tax liability for that year. Such adjustment, if not exhausted, can be carried forward for a period of up to five years (three years in 2008 and 2009). In the data, we observe that only 1.3% of firms claim such carry forward, indicating either that firms are unaware of this option or that their profit tax liability net of carry forward drop below output tax liability, in which case carry forward cannot be claimed. In any case, the potential for carry forward attenuates the size of the minimum tax kink and works against the (strong) bunching that we find.

[23] In Pakistan, tax year t runs from July 1 of year t to June 30 of year $t+1$.

internal inconsistencies like reconciling tax liability with reported profit. Besides this validation check, the tax returns are considered final unless selected for audit.

Two aspects of the data are worth keeping in mind. First, our dataset contains almost all active corporations. As corporations also act as withholding agents, deducting tax at source on their sales and purchases, it is almost impossible for an operational corporation not to file a tax return. FBR takes the view that registered corporations which do not file tax returns are non-operational. Second, besides the corporations in our data, the population of firms in Pakistan include both unincorporated firms subject to personal income taxation and informal firms operating outside the tax net. In general, corporate and personal income taxes may lead to shifting between the corporate and non-corporate sectors as well as between the formal and informal sectors (Waseem 2013). We do not study these (interesting) effects here; our bunching estimates capture intensive-margin responses *conditional on being a corporate tax filer*, and are therefore not affected by incorporation or informality responses.

We limit our analysis to firms that report both profit and turnover, and either the incorporation date or the profit tax liability, which are required for allocating firms to the high and low profit rate groups.[24] We also subject the data to a number of checks for internal consistency detailed in table A.III of the web appendix. Our final dataset contains 24,290 firm-year observations.

5 Empirical Results

This section presents the results of our empirical analysis, examing how firms respond to the minimum tax policy. We first present evidence that there is sharp bunching at the minimum tax kink as predicted by our analysis in section 3, and that it is caused by the presence of the kink. We then analyze heterogeneity in bunching across different subsamples, showing that it is consistent with the predictions of the previous literature on tax evasion, and finally we use the observed bunching to estimate the magnitude of evasion responses.

5.1 Bunching at Minimum Tax Kinks

As shown in section 3, the type of minimum tax scheme observed in Pakistan should lead to excess bunching by firms around a threshold profit rate (profits as a share of turnover) equal to the ratio of the two tax rates, τ_y/τ_π. Figure II shows evidence that firms do indeed bunch around this minimum tax kink. The figure shows bunching evidence for different groups of firms (panels A and B) and different years (panels C and D), exploiting the variation in the kink across these samples. We plot the empirical density distribution of the reported profit rate (profits as percentage of turnover) in bins of approximately 0.2 percentage points.[25] Panel A shows the density for high-rate firms (facing

[24] Table A.IV in the web appendix compares the firms we lose to those we are able to use based on firm characteristics reported in the tax returns.

[25] The exact bin width is chosen to ensure that kink points are always located at bin centres. This requires us to slightly vary the bin width between panels A-C and panel D as the distance between the two kinks in panel D is different due to the higher turnover tax rate.

a profit tax rate of 35%) in the years 2006, 2007 and 2009 pooled together, since for those firms and years the minimum tax kink is at a profit rate threshold of $\frac{\tau_y}{\tau_\pi} = \frac{0.5\%}{35\%} = 1.43\%$ (demarcated by a solid vertical line in the figure). The density exhibits large and sharp bunching around the kink point. Since there is no reason for firms to cluster around a profit rate of 1.43% other than the presence of the minimum tax scheme, this represents compelling evidence of a behavioral response to the scheme. Notice also that there is a modest amount of "natural" excess mass around the zero-profit point (much smaller than at the kink point) as many firms generate very little income.[26]

Panels B-D provide identification checks ensuring that excess bunching at the minimum tax kink is indeed a response to the tax system (as opposed to a spurious property of the profit rate distribution) by exploiting variation in the minimum tax kink across firms and over time. Panel B compares high-rate firms to low-rate firms during the years 2006, 2007 and 2009, when the latter group of firms face a reduced profit tax rate of 20% and therefore a minimum tax kink located at $\frac{\tau_y}{\tau_\pi} = \frac{0.5\%}{20\%} = 2.5\%$. Besides the different location of the kink, our model implies that the size of the kink is smaller for low-rate firms than for high-rate firms in terms of evasion incentives (as the change in the evasion incentive $\mu \cdot \tau$ at the kink equals the profit tax rate τ_π) but not real incentives (as the change in the real incentive $1 - \omega$ at the kink equals the turnover tax rate τ_y). Hence, we expect to see both that low-rate firms bunch in a different place and that the amount of bunching is smaller (if evasion is important), and this is precisely what panel B shows. Even though bunching is smaller for low-rate firms, it is still very clear and sharp. Outside the bunching areas around the two kinks, the low-rate and high-rate distributions are very close and exhibit the same (small) excess mass around zero.[27]

Panels C and D of Figure II exploit time variation in the kink, focusing on the sample of high-rate firms. Panel C shows that excess bunching at 1.43% completely disappears in 2008 when the minimum tax regime (i.e., turnover tax below a profit rate of 1.43%) is removed. The 2008 density instead exhibits a larger mass of firms with profit rates between 0 and 1.43%. The distributions in panel C are consistent with our theoretical prediction that the introduction of an output tax below a profit rate threshold creates bunching coming from below. Finally, panel D shows that bunching moves from 1.43% to 2.86% in 2010, when the doubling of the output tax rate shifts the kink. This change is accompanied by an overall decrease in the mass of firms with profit rates between 0 and 2, again illustrating that bunchers move to the kink from below.

Taken together, the panels of Figure II provide compelling evidence that firms respond to the incentives created by the minimum tax scheme. The substantial amount of bunching observed around kink points (which are associated with weak *real* incentives as explained above) suggests that evasion responses are quantitatively important.

[26] A small number of firms in the data report *precisely* zero profits, which represents a form of "round-number bunching" as analyzed in detail by Kleven & Waseem (2013). To eliminate this effect driven by the salience of zero, the empirical distributions in Figure II exclude observations with $\pi = 0$, so the excess mass around zero is not driven by round-number bunching. Figure A.I in the web appendix reproduced panel A of figure II, using the raw data before the consistency checks and without dropping firms with $\pi = 0$. Our results are robust to including firms reporting $\pi = 0$.

[27] The low-rate distribution is more noisy than the high-rate distribution because the former represents a much smaller fraction (about 16%) of the population of firms.

5.2 Heterogeneity in Evasion Opportunities

Since our theoretical framework predicts that bunching responses will be driven primarily by evasion, we expect that firms for which evasion is easier will display greater bunching. To explore this, we split the data using indicators of evasion opportunities identified by the previous tax evasion literature and compare bunching in those subsamples. First, Kleven *et al.* (2009) develop an agency model in which firms with a large number of employees find it more difficult to sustain collusion on evasion, and Kumler *et al.* (2013) find evidence that supports this theory in Mexico, so we split the sample by firm size as proxied for by the wage bill and turnover.[28] Second, Gordon & Li (2009) provide a model where firms that rely on formal credit are more tax compliant (as this creates a paper trail that governments can observe), an argument that is consistent with the cross-country evidence in Bachas & Jensen (2013), so interest payments as a proportion of turnover is our second indicator of evasion opportunities. Third, firms selling to final consumers have more opportunity to evade than firms selling to other firms (due to the absence of a verifiable paper trail on the former), consistent with the experimental evidence for Chile by Pomeranz (2013), so we split the sample into retailers and non-retailers.

We focus on the group of high-rate firms, which represents most of the data and therefore gives us more power to detect heterogeneity. Panels A and B of Figure III plot the density of the profit rate around the kink at 1.43%, splitting the sample by median salary payments (scaled by turnover) and median turnover, respectively. As predicted by the theory, we find that small firms respond more strongly to the kink. In both panels, the small-firm distribution exhibits a larger spike at the kink than the large-firm distribution. Panel C shows bunching for firms with more or less need for financial intermediation, as proxied for by reported interest payments as a fraction of turnover. While the difference in excess bunching is less pronounced than in the previous panels, the density for firms with below median interest payments does exhibit a larger spike at the kink, again in accordance with the theory. Finally, panel D examines bunching by sector, dividing the firms into "retailers" (firms that sell at least partly to final consumers) and "non-retailers" (firms that sell exclusively to other firms). In line with the theory, there is again a larger spike at the kink in the retailer subsample. We therefore conclude that the patterns of bunching heterogeneity across firms are consistent with evasion being the main mechanism, using the indicators of evasion opportunity identified by the existing literature.

5.3 Estimating Evasion Responses Using Bunching

This section presents estimates of excess bunching and uses our model to translate them into estimates of evasion responses. Following Chetty *et al.* (2011), we estimate a counterfactual density—what the distribution would have looked like absent the kink—by fitting a flexible polynomial to the observed density, excluding observations in a range around the kink that is (visibly) affected by bunching. Denoting by d_j the fraction of the data in profit rate bin j and by p_j the (midpoint)

[28] Unfortunately, the number of employees is not available in the data.

profit rate in bin j, the counterfactual density is obtained from a regression of the following form

$$d_j = \sum_{i=0}^{q} \beta_i (p_j)^i + \sum_{i=p_L}^{p_U} \gamma_i \cdot \mathbf{1}[p_j = i] + \nu_j, \tag{19}$$

where q is the order of the polynomial and $[p_L, p_U]$ is the excluded range. The counterfactual density is estimated as the predicted values from (19) omitting the contribution of the dummies in the excluded range, i.e. $\hat{d}_j = \sum_{i=0}^{q} \hat{\beta}_i (p_j)^i$, and excess bunching is then estimated as the area between the observed and counterfactual densities in the excluded range, $\hat{B} = \sum_{j=p_L}^{p_U} \left(d_j - \hat{d}_j \right)$. Standard errors are bootstrapped by sampling from the estimated errors with replacement.

Figure IV compares the empirical density distributions to estimated counterfactual distributions (smooth solid lines) for the four samples examined in Figure II: high-rate firms in 2006/07/09 in panel A, low-rate firms in 2006/07/09 in panel B, high-rate firms in 2008 (placebo) in panel C, and high-rate firms in 2010 in panel D. In each panel, the solid vertical line represents the kink point while the dashed vertical lines demarcate the excluded range around the kink used in the estimation of the counterfactual.[29] To better evaluate the estimated counterfactuals, each panel also shows the empirical distribution for a comparison sample in light grey (low-rate firms in panel A, high-rate firms in panel B, 2006/07/09 in panels C and D). The observation that in all cases the empirical distribution for our comparison sample lines up well with the estimated counterfactual, particularly around the kink, provides a further validation of our estimates.

The figure also displays estimates of excess bunching scaled by the average counterfactual density around the kink, i.e. $b = \hat{B}/E(\hat{d}_j \mid j \in [p_L, p_U])$. In general, these bunching estimates are large and strongly statistically significant, except in the placebo analysis of panel C where bunching is close to zero and insignificant. Excess bunching is larger for high-rate firms in 2006/07/09 ($b = 4.44\,(0.1)$) than for low-rate firms in the same period ($b = 2.00\,(0.2)$), consistent with the fact that a lower profit tax rate implies a smaller change in the evasion incentive at the kink. Futhermore, excess bunching by high-rate firms is larger during the years 2006/07/09 than in year 2010 ($b = 2.05\,(0.2)$), possibly because optimization frictions prevent some firms from responding to the change in the location of the kink in the short run.

Table II converts our bunching estimates into evasion responses using the methodology developed in section 3. As shown earlier, the amount of bunching translates to a reported profit rate response via the relationship $\Delta\hat{p} = B/f_0 \left(\frac{\tau_y}{\tau_\pi}\right) \simeq b \times binwidth$,[30] and this profit rate response is in turn linked to the combination of real output and evasion responses via equation (15),

[29] The excluded range $[p_L, p_U]$ is set to match the area around the kink in which the empirical density diverges from its smooth trend; four bins on either side of the kink in panels A and C, two bins on either side in panels B and D. The order of the polynomial q is five (seven for 2008), chosen such as to optimize the fit. Table A.I shows that the estimates are fairly sensitive to the choice of excluded range and polynomial degree in panel A, but less so in the other panels.

[30] Since b equals bunching divided by the counterfactual density in discrete bins, we have to multiply b by binwidth to obtain the profit rate response. The binwidth underlying b is 0.214 percentage points for most estimates and so $binwidth = 0.00214$.

$$\Delta\hat{p} \simeq \frac{\tau_y^2}{\tau_\pi}\varepsilon_y - \frac{d\left(\hat{c}-c\right)}{y}.$$

The table shows estimates of excess bunching b in column (1), the profit rate response $\Delta\hat{p}$ in column (2), the real output elasticity ε_y assuming zero evasion in column (3), and the evasion response assuming different real output elasticities $\varepsilon_y \in \{0, 0.5, 1, 5\}$ in columns (4)-(7). Evasion responses are reported as percentages of taxable profits (evasion *rate* responses) instead of percentages of output in eq. (15). Evasion rates in terms of output are easily converted to evasion rates in terms of profits, using the fact that $\frac{y-\hat{c}}{y} = \frac{\tau_y}{\tau_\pi}$ at the kink. The different rows of the table show results for the main subsamples considered in the bunching figures.

The following main findings emerge from the table. First, in a model without evasion, the bunching we observe implies phenomenally large real output elasticities, ranging from 15 to 133 across the different samples. These elasticities are all far above the upper bound of the range of values that can be considered realistic and so we can comfortably reject that model.[31] The reason for the large elasticities in this model is the combination of large observed bunching and the tiny variation in real incentives at the kink. Second, when we allow for tax evasion in the model, it becomes possible to reconcile observed bunching with reasonable values of the real output elasticity combined with large (but not implausible) evasion responses. Column (3) provides an upper bound on the evasion response, assuming a zero real output response. In this case, the evasion response ranges from 14.7% to 66.7% of profits across the different populations, with high-rate firms in 2006/07/09 featuring the largest response. Third, the evasion estimates are very robust to the real output elasticity even though we allow for elasticities up to 5, much higher than the empirical literature suggests is justified. The reason for this robustness is again that real incentives at the kink are extremely small. Hence, while we cannot separately identify both real and evasion responses using the minimum tax kink, we can provide very tight bounds on the evasion response due to the particular set of incentives provided by the minimum tax kink.[32]

Finally, it should be noted that the evasion responses which we estimate using bunching at minimum tax kinks are responses to the *differential* evasion opportunities offered by profit and output taxation. The fact that we see such large bunching is *prima facie* evidence that it is much harder to evade output taxes than profit taxes, consistent with the motivation of the policy and our stylized conceptual framework. In the extreme case where output taxes offer zero evasion opportunity (as in our stylized model), our estimates would capture *total* tax evasion by firms in Pakistan. More realistically, if output taxes offer some scope for evasion as well, our estimates of evasion responses are lower bounds on the total evasion level by firms in Pakistan.[33]

[31] For example, Gruber & Rauh (2007) estimate that the elasticity of corporate taxable income with respect to the effective marginal tax rate in the United States is 0.2. Taking that estimate at face value in the Pakistani context, with all the caveats that entails, would imply that $0.2 = \frac{dCTI}{d\omega}\frac{\omega}{dCTI} = \frac{\partial CTI/\partial y}{CTI/y}\varepsilon_y$ and so even a real output elasticity of 15 would require marginal taxable profits to be 1.33% of average taxable profits to be reconcilable with their estimate.

[32] We find that the estimates of the evasion rate response are very similar when assuming that only output can be evaded, using equation (17). The results are reported in table A.II.

[33] This is consistent with the fact that our evasion estimates are lower than the existing tax gap estimates for Pakistan discussed in section 4. Those tax gaps were measured as a fraction of *true* tax liability, but can easily be

6 Numerical Analysis of Tax Policy Implications

This section links our empirical results to the stylized model introduced in section 2. The model characterizes the trade-off between production and revenue efficiency when setting a uniform tax rate and tax base, while our empirical analysis of Pakistan's minimum tax scheme identifies sufficient statistics allowing us to evaluate this trade-off. As discussed earlier, the conceptual framework ignores general equilibrium cascading effects that will make turnover taxes less desirable, other things equal. Hence, this section should not be seen as a complete analysis of the optimality of turnover taxes, but as an analysis of how far the partial equilibrium evasion argument by itself can move the optimal policy away from production efficiency. If the evasion mechanism cannot justify a large move away from profit taxes towards turnover tax in the context of our model, then turnover taxes are unlikely to be an attractive policy even in countries with limited tax capacity.

At the optimum, as shown in Proposition 2, the tax policy is characterized by

$$\frac{\tau}{1-\tau} \cdot \frac{\partial \omega}{\partial \tau}(\mu) = G(\mu) \cdot \frac{\varepsilon_{\hat{c}-c}}{\varepsilon_y}.$$

While this optimal tax rule needs to be considered jointly with a revenue requirement to determine the optimal tax base and rate, our analysis can shed light on the optimality of the tax base μ for a given tax rate τ. This crucially depends on how both sides of the optimal tax rule change as the tax system moves from profit taxation ($\mu = 1$) towards turnover taxation ($\mu = 0$), which we analyze below.

The left-hand side of the optimal tax rule reflects the effective marginal tax rate. Our theoretical model fully determines how this depends on the tax base parameter,

$$\frac{\tau}{1-\tau} \cdot \frac{\partial \omega}{\partial \tau}(\mu) = \frac{\tau}{1-\tau} \cdot \frac{1-\mu}{(1-\tau\mu)^2}.$$

For tax rates below 50 percent, the effective marginal tax rate is decreasing in μ, starting at a maximum of $\frac{\tau}{1-\tau}$ for $\mu = 0$ and falling to 0 for $\mu = 1$. This is illustrated in panel A of Figure V for $\tau = .35$, which is the standard profit tax rate in Pakistan.

The right-hand side of the optimal tax rule equals the inverse of the output elasticity scaled by the evasion rate response,

$$G(\mu) \cdot \frac{\varepsilon_{\hat{c}-c}}{\varepsilon_y} = \frac{\hat{c}-c}{\hat{\Pi}} \frac{\varepsilon_{\hat{c}-c}}{\varepsilon_y} \simeq -\frac{d(\hat{c}-c)}{\hat{\Pi}} / \varepsilon_y,$$

using $d(\tau\mu)/(\tau\mu) = -1$ at the minimum tax kink. This evasion rate response is exactly what we have estimated in our empirical analysis. For the firms facing a profit tax rate of $\tau = .35$, our empirical analysis suggests an estimate of the right-hand side of 1.22 for $\varepsilon_y = .5$. This exceeds the left-hand side, which is bounded from above by $\frac{\tau}{1-\tau} = \frac{.35}{.65} = .54$. The evasion rate by high-rate firms

converted into tax gaps as a fraction of *actual* taxes paid (corresponding to our estimates in Table II), in which case they would close to 100%.

in response to the current profit tax is thus *too high* relative to the effective tax wedge, indicating that welfare could be increased by broadening the tax base. This would reduce the evasion rate, but increase the effective tax wedge.

To reveal how quickly the right-hand side of the optimal tax rule decreases when moving towards a turnover tax (i.e., reducing μ), we can use our estimate of the evasion rate response at the kink for firms facing a low profit tax rate of $\tau = .20$.[34] Due to the lower tax rate, the evasion incentive $\tau\mu$ for the low-rate firms is equal to $\frac{.20}{.35} = 57\%$ of the evasion incentive for the high-rate firms. This reduction in the tax evasion incentive corresponds to a decrease in the tax base parameter from $\mu = 1$ to $\mu = .57$ (for a fixed tax rate) using that the evasion incentive $\tau\mu$ is symmetric in the rate and the base. Our estimate of the evasion rate response for the low-rate firms would imply an estimate for the right-hand side of the optimal tax rule, when evaluated at $\mu = .57$, equal to .34. The substantially smaller evasion rate response for the low-rate firms implies that the right-hand side is marginally below the left-hand side for $\mu = .57$. This is illustrated in panel A of Figure V. Extrapolating linearly between these two points, we find that the two sides of the optimal tax rule are equal for $\mu = .578$, which suggests that only about half of the costs should be deductible from the corporate tax base at a tax rate of 35 percent. While this exact number relies on the specific assumptions of our model and calibration, the result indicates that the full cost deductibility granted by profit taxation are far from optimal when accounting for evasion.

Our conclusion is robust to different tax rates and output elasticities. Panel B of figure V shows all combinations of the tax rate and tax base for which the trade-off between production and revenue efficiency is optimized as captured by the optimal tax rule.[35] The optimal tax base moves further towards the turnover tax base when decreasing the tax rate, with an optimal μ close to zero for a tax rate of .005, which is the minimum tax rate in Pakistan. The figure also illustrates that a higher output elasticity would move the optimal tax system closer to profit taxation for any tax rate, but still far from a pure profit tax base with full cost deductibility, while a lower output elasticity would move the system closer to turnover taxation.[36] The caveat mentioned at the beginning notwithstanding, the numerical analysis clearly illustrates how production-inefficient tax instruments could become desirable when tax capacity is limited.

[34]Here we make the (strong) assumption that the variation in the profit tax rate across firms can be viewed as exogenous.

[35]Here we use that the right-hand side decreases to 0 when moving all the way to turnover taxation, as implied by our model. We thus have three points for the right-hand side as a function of μ (i.e., for $\mu \in \{0, .57, 1\}$ and use a linear extrapolation in between.

[36]The numerical analysis assumes that the profit tax in Pakistan is non-distortionary (i.e.,$\mu = 1$). Figure xx in the web appendix shows the same graphs assuming that the the effective marginal tax rate under profit taxation in Pakistan equals $\omega = .2$, implied by $\mu = .53$. The estimates of the evasion rate response for the high-rate and low-rate firms now provide an estimate of the right-hand side evaluated at$\mu = .53$ and $\mu = \frac{.20}{.35} \times .53$ respectively. The optimal tax base parameter is lower in absolute terms, but somewhat closer to the (current) profit base parameter in relative terms.

7 Conclusion

In this paper we have studied the trade-off between preserving production efficiency and preventing the corrosion of revenues by evasion faced by governments with limited tax enforcement capacity. In contrast to models without evasion in which the optimal tax base is as close as possible to pure profits (preserving production efficiency), we have shown theoretically that in the presence of evasion, the optimal tax base sacrifices some production efficiency in order to curtail evasion levels. Our optimality conditions relate the optimal marginal tax wedge on real production to the elasticities of real production and tax evasion with respect to the tax wedge, sufficient statistics that we study empirically. We have also developed a novel empirical approach showing that minimum taxation schemes of a type that is common throughout the developing world can be used to estimate tight bounds on the evasion response to switching from profits to a broader turnover base. Exploiting the small change in real incentives and the large change in evasion incentives presented by the minimum tax scheme on corporations in Pakistan, we estimate that the switch from a profit tax to a turnover tax reduces evasion levels by up to 60-70% of corporate income. Linking these estimates back to our conceptual framework, our tax rule implies that the optimal tax system has a base that is far broader than profits—in Pakistan, as little as 50% of costs should be deductible when taxed at 35%.

It should be noted that our policy conclusions are based on a setting in which tax enforcement capacity is exogenously given, which is a reasonable approximation of the short-medium term environment of a developing country. In the longer run where tax capacity is endogenous to economic development (Kleven *et al.* 2009) and investments in state capacity (Besley & Persson 2011, 2013), the policy recommendations would obviously change. In fact, the large compliance gains to production-inefficient policies that we estimate shows the potentially large social returns to greater tax capacity.

Finally, since the kind of minimum tax scheme required for our methodology is ubiquitous throughout the world, the approach could be replicated in many different countries. Applying the approach to countries with different tax enforcement capacities could shed further light on how tax policy is shaped by tax enforcement capacity and on the returns to greater tax capacity.

Table I: Tax Schedule

Year	Tax Rates (%)		
	Profit (high)	Profit (low)	Turnover
2006	35	20	0.5
2007	35	20	0.5
2008	35	20	0
2009	35	20	0.5
2010	35	25	1

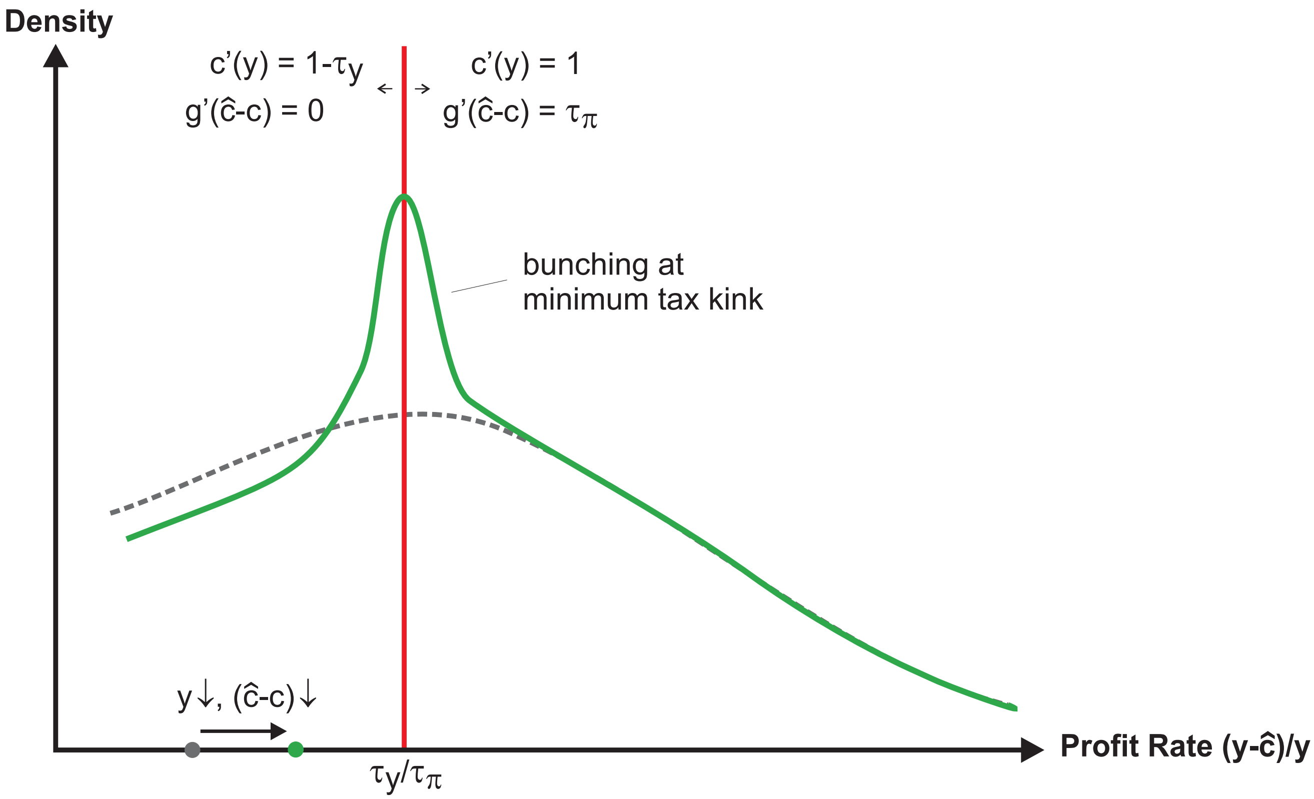
Notes: The table presents Pakistan's corporate income tax schedule for fiscal years 2006 to 2010. Fiscal year t runs from July 1 of year t to June 30 of year $t+1$. Tax rates are given in percentages. The low tax rate applies to firms which registered after June 2005, have no more than 250 employees, have annual sales of not more than Rs. 250 million, and paid-up capital of not more than Rs. 25 million. Our empirial analysis takes into account that the thresholds for some of these requirements change over time. All firms that do not meet these criteria are liable for the high tax rate. Firms calculate their net profit and turnover tax liablity (the tax code allows for specific deductions under each type of tax) and pay whichever liability is higher. Firms are allowed to carry forward the tax paid in excess of the profit tax liability and can adjust it against next year's liability to the extent that the net liability does not fall below the output tax liability for that year. Such adjustment, if not exhausted, can be carried forward for a further period of up to five years (three years in 2008 and 2009). In the data, we observe that only 1.3% of firms claim such carry forward, which indicates that firms are either unaware of this option or observe their profit tax liability net of carry forward drop below output tax liability, in which case carry forward cannot be claimed. We also exclude banks and financial firms, which face a standard tax rate of 38% in 2006 and 154 firms in sectors that were selectively given a lower turnover tax rate in 2010.

Table II: Estimating Evasion Responses

	Observed Responses		Without Evasion	With Evasion			
	Bunching (b)	**Profit Rate ($\Delta\hat{p}$)**	**Output Elasticity (ε_y)**	**Evasion Rate Response**			
				$\varepsilon_y = 0$	$\varepsilon_y = 0.5$	$\varepsilon_y = 1$	$\varepsilon_y = 5$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High-rate Firms, 2006/07/09	4.44	1.0	133.3	66.7	66.4	66.2	64.2
	(0.1)	(0.03)	(3.8)	(2.0)	(2.0)	(2.0)	(2.0)
Low-rate Firms, 2006/07/09	2.00	0.4	34.3	17.1	16.9	16.6	14.6
	(0.2)	(0.04)	(3.0)	(1.5)	(1.5)	(1.5)	(1.5)
High-rate Firms, 2010	2.05	0.4	14.7	14.7	14.2	13.7	9.7
	(0.2)	(0.03)	(1.2)	(1.2)	(1.2)	(1.2)	(1.2)

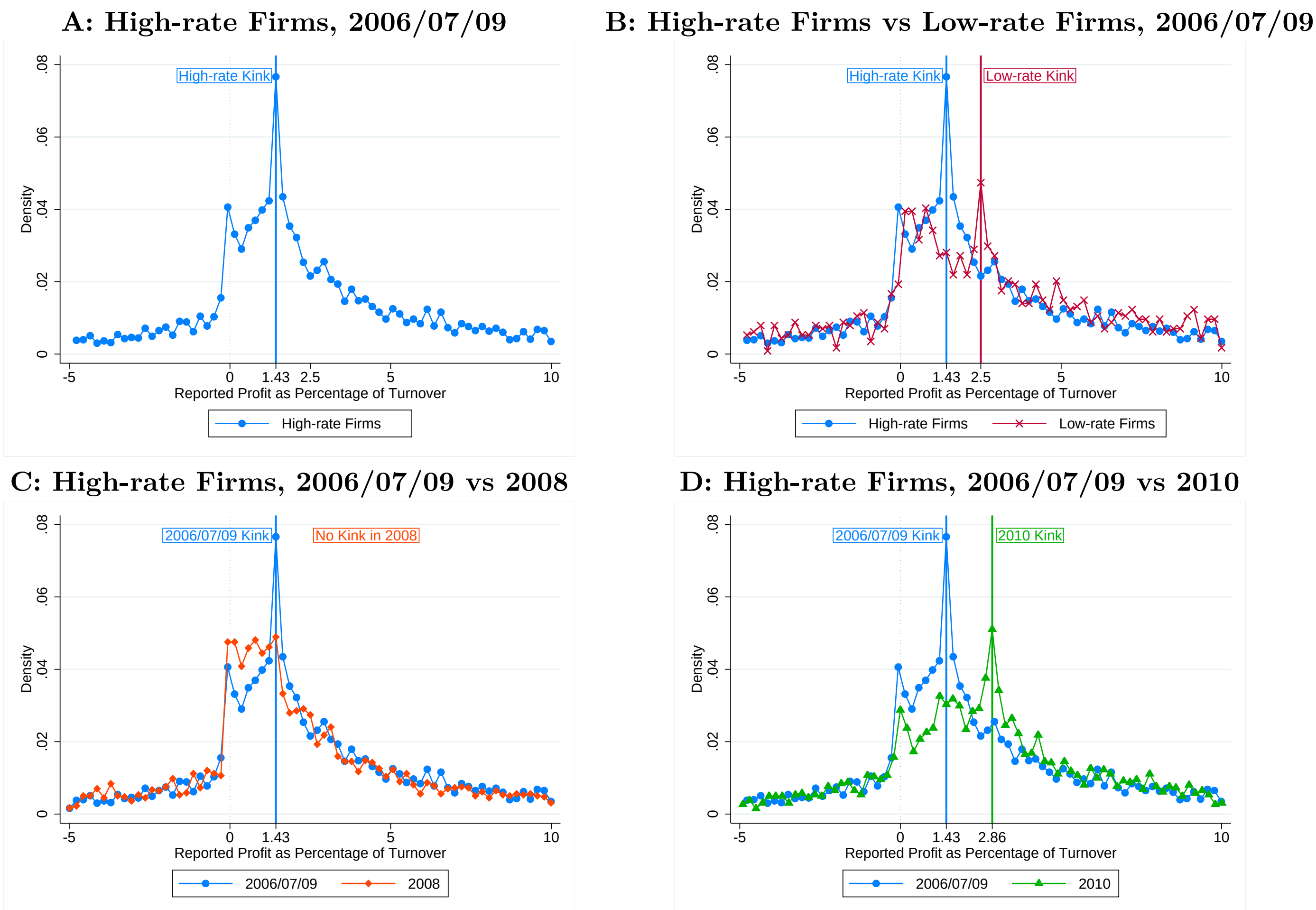
Notes: This table presents bunching, elasticity and evasion estimates for the subsamples considered in panels A, B and D of figure IV. Column (1) reproduces the bunching estimate b, based on estimating equation (19). Bunching b is the excess mass in the excluded range around the kink, in proportion to the average counterfactual density in the excluded range. Column (2) presents an estimate of the profit rate response associated with b, based on the relationship $\Delta\hat{p} = B/f_0\left(\tau_y/\tau_\pi\right) \simeq b \times binwidth$. Column (3) presents estimates of the real output elasticity ε_y for the model without evasion. This model is based on the assumption that bunching is purely due to a real output response. ε_y is estimated using the relationship $\Delta p = \left[c/y - c'\left(y\right)\right] dy/y \simeq \left(\tau_y^2/\tau_\pi\right)\varepsilon_y$. Columns (4)-(7) present estimates of the evasion response as percentage of taxable profits (evasion *rate* responses), for the model with evasion. This model allows for bunching to be driven by both evasion and real output response. The evasion response estimates are based on $\Delta\hat{p} = \left[\hat{c}/y - c'\left(y\right)\right] dy/y - \left[d\left(\hat{c} - c\right)/y\right] \simeq \left(\tau_y^2/\tau_\pi\right)\varepsilon_y - \left(\tau_y/\tau_\pi\right)\left[d\left(\hat{c} - c\right)/\left(y - \hat{c}\right)\right]$, assuming different real output elasticities $\varepsilon_y \in \{0, 0.5, 1, 5\}$. Bootstrapped standard errors are shown in parentheses.

Figure I: Bunching Methodology Using Minimum Tax Schemes



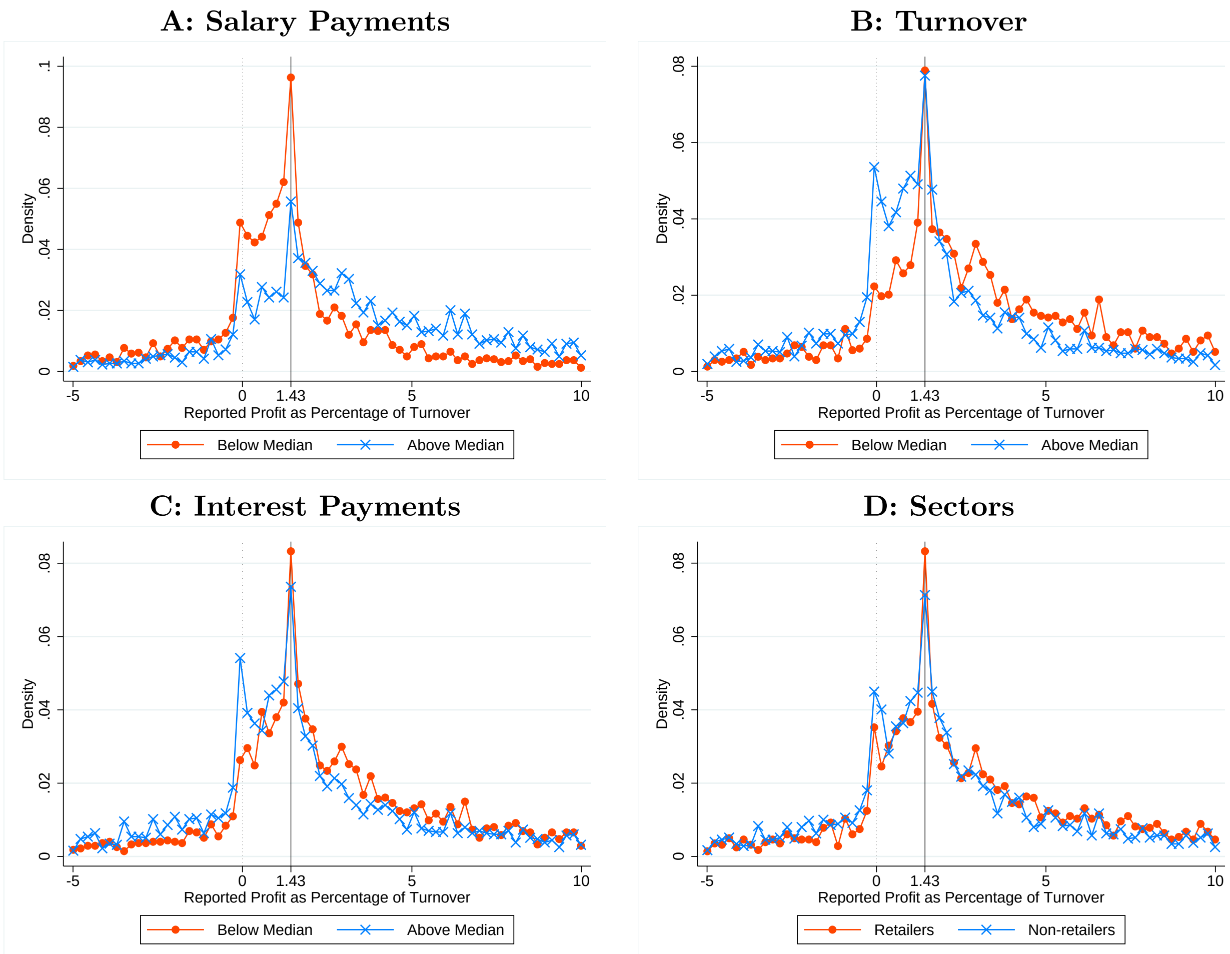
Notes: The figure illustrates the implications of the introduction of a minimum tax on the observed density distribution of reported profit rates $\hat{p} = (y - \hat{c})/y$. The grey dashed line shows the smooth distribution of profit rates that would be observed in the absence of the minimum tax, while the green, solid line shows the distribution of profit rates that is observed in the presence of the minimum tax. As discussed in section 2.2, under the profit tax, firms' optimality conditions are given by $c'(y) = 1$ and $g'(\hat{c} - c) = \tau_\pi$. Firms whose optimal reported profit rate under the profit tax is smaller than τ_y/τ_π will adjust their production and reporting decisions in response to the introduction of the minimum tax to satisfy $c'(y) = 1 - \tau_y$ and $g'(\hat{c} - c) = 0$, causing them to decrease both output y and cost evasion $\hat{c} - c$. Both responses move their reported profit rate up towards the kink. Firms whose profit rate was close to the kink before introduction of the minimum tax pile up at the kink, which gives rise to an observed excess mass around the kink when accounting for optimization errors.

Figure II: Bunching Evidence



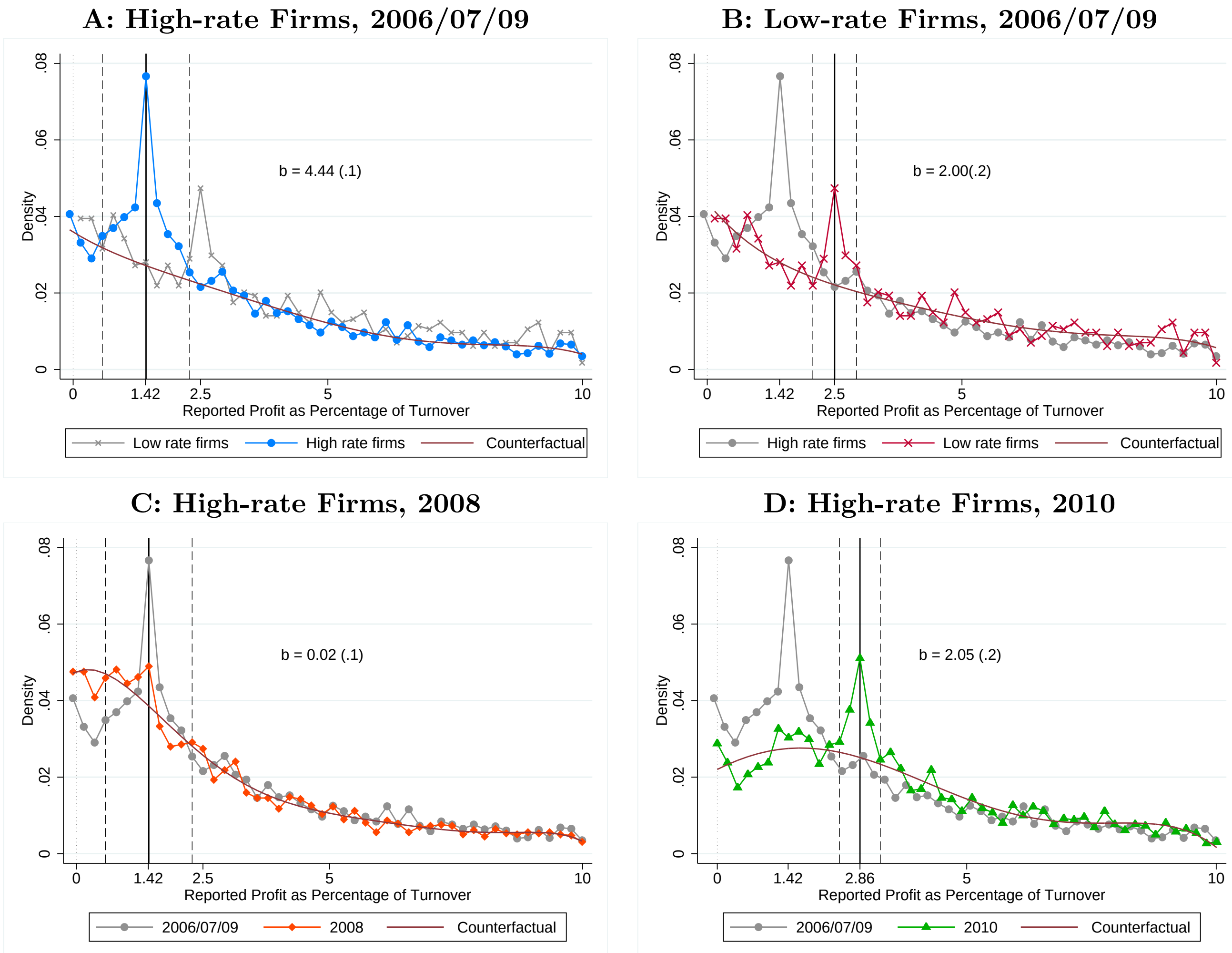
Notes: The figure shows the empirical density distribution of the profit rate (reported profit as percentage of turnover), for different groups of firms and time periods. Firms calculate their profit tax liability (taxed at rate τ_π) and turnover tax liablity (taxed at rate τ_y) and pay whichever liability is higher. This minimum tax regime introduces a kink at a profit rate equal to the tax rate ratio τ_y/τ_π; firms move from the profit tax scheme for profit rates above the kink to the turnover tax scheme for profit rates below the kink. For high-rate firms in 2006/07/09 (panel A), $\tau_\pi = 0.35$ and $\tau_y = 0.005$, placing the kink at a profit rate of 1.43%. For low-rate firms in 2006/07/09 (panel B), $\tau_\pi = 0.25$ and $\tau_y = 0.005$, placing the kink at a profit rate of 2.5%. For high rate firms in 2008 (panel C), the minimum tax scheme is abolished and all firms' profits are taxed at rate$\tau_\pi = 0.35$, so that there is no kink. For high-rates firms in 2010 (panel D), $\tau_\pi = 0.35$ and $\tau_y = 0.01$, placing the kink at a profit rate of 2.86%. Kink points are marked by vertical solid lines, and the colour of the kink line matches the colour of the corresponding density. The zero profit point is marked by a vertical dotted line. The bin size is 0.214 (0.204 for 2010), chosen so that all kink points are bin centres.

Figure III: Heterogeneity in Bunching



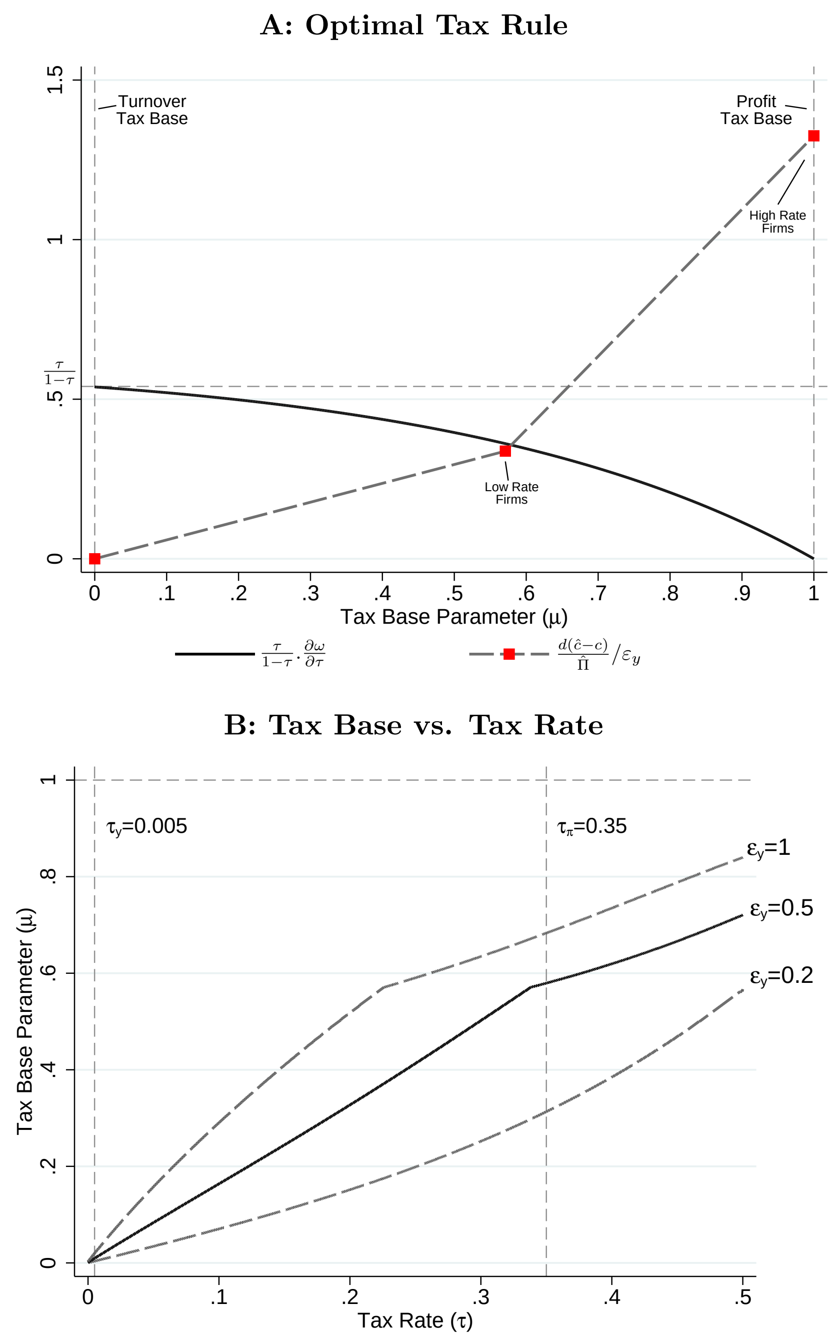
Notes: The figure shows the empirical density distribution of the profit rate (reported profit as percentage of turnover), for different subsamples within the high-rate firms group in 2006/07/09. The tax rate schedule is explained in the notes to figure II. For high-rates firms in 2006/07/09, the minimum tax regime places the kink at a profit rate of 1.43% (firms with a profit rate above this threshold fall under the profit tax scheme, while firms with a profit rate below fall under the turnover tax scheme). In panels A-C, the high-rate firms sample is split by median salary payments, turnover and interest payements respectively. Salary payments and interest payments are scaled by turnover. The red (light) density is for firms below the median, and the blue (dark) is for firms above the median. Panel D splits the sample by sector, into retailers (red density) and non-retailers (blue density). The kink point is marked by a vertical solid line. The zero profit point is marked by a dotted line. The bin size in all panels is 0.214, chosen so that the kink point is a bin centre.

Figure IV: Bunching Estimation



Notes: The figure shows the empirical density distribution of the profit rate (reported profit as percentage of turnover, dotted dark graph), an empirical counterfactual density (dotted light graph), and the estimated counterfactual density (solid graph), for the different groups of firms and time periods considered in figure II. The tax rate schedules and kink locations are explained in the notes to figure II. The empirical counterfactual is the high-rate firms density in 2006/07/09 for panels B-D, and the low-rate firms density in 2006/07/09 for panel A. The counterfactual density is estimated from the empirical density, by fitting a fifth-order polynomial (seventh-order for 2008), excluding data around the kink, as specified in equation (19). The excluded range is chosen as the area around the kink that is visibly affected by bunching. Kink points are marked by vertical solid lines; lower and upper bounds of excluded ranges are marked by vertical dashed lines. The zero profit point is marked by a dotted line. The bin size for the empirical densities is 0.214 (0.204 for 2010), so that the kink points are bin centres. Bunching b is the excess mass in the excluded range around the kink, in proportion to the average counterfactual density in the excluded range. Bootstrapped standard errors are shown in parentheses.

Figure V: Numerical Analysis of Optimal Tax Policy



Notes: The figure presents a numerical analysis of optimal tax policy, integrating our stylized theoretical framework and our empirical results. The solid black curve in panel A plots the left-hand side of the optimal tax rule (equation 11) as a function of μ for $\tau_\pi = 0.35$. The three red markers on the dashed gray curve show respectively the right-hand side of the optimal tax rule for $\tau_\pi = 0.35$, evaluated at $\mu = 0$, at $\mu = {}^{0.2}/_{0.35}$and $\mu = 1$. The value for $\mu = 0$ is directly implied by our theoretical model, while the values for$\mu = {}^{0.2}/_{0.35}$and $\mu = 1$ are based on the estimates of evasion rate response for the low-rate firms and for the high-rate firms respectively, assuming an output elasticity of $\varepsilon_y = .5$. By extrapolating between these three estimates, we find that the optimal tax base implied by the tax rule equals $\mu = 0.578$. In panel B we replicate this exercise to find the optimal tax base as a function of the tax rate for three different levels of the output elasticity.

References

Allingham, Michael G., & Sandmo, Agnar. 1972. Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics*, **1**, 323–338.

Andreoni, James, Erard, Brian, & Feinstein, Jonathan. 1998. Tax complicance. *Journal of Economic Literature*, **36**, 818–860.

Auerbach, Alan J. 2002. Taxation and Corporate Financial Policy. *Chap. 8, pages 1251–1292 of:* Auerbach, Alan J., & Feldstein, Martin (eds), *Handbook of Public Economics, Volume 3*.

Auerbach, Alan J., Devereux, Michael P., & Simpson, Helen. 2010. Taxing Corporate Income. *Chap. 9 of: Dimensions of Tax Design: the Mirrlees Review*. Oxford University Press.

Bach, Laurent. 2012 (January). *Tax Collection and Private Governance: Evidence from French Businesses*. Working Paper.

Bachas, Pierre, & Jensen, Anders. 2013. *Information Trails, Tax Compliance and Development*. Mimeo.

Baunsgaard, Thomas, & Keen, Michael. 2010. Tax revenue and (or?) trade liberalization. *Journal of Public Economics*, **94**(9-10), 563–577.

Besley, Timothy, & Persson, Torsten. 2011. *Pillars of Prosperity: The Political Economy of Development Clusters*. Princeton University Press.

Besley, Timothy, & Persson, Torsten. 2013. *Taxation and Development*. Chapter for Handbook of Public Economics, Vol. 5. forthcoming.

Boadway, Robin, & Sato, Motohiro. 2009. Optimal Tax Design and Enforcement with an Informal Sector. *American Economic Journal: Economic Policy*, **1**(1), 1–27.

Cage, Julia, & Gadenne, Lucie. 2012. *The Fiscal Cost of Trade Liberalization*. Working paper.

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

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

Devereux, Michael, Liu, Li, & Loretz, Simon. 2013. *The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records*. Oxford University Centre for Business Taxation Working Paper 12/23.

Devereux, Michael P., & Sørensen, Peter Birch. 2006. *The Corporate Income Tax: International Trends and Options for Fundamental Reform*. Economic Papers. European Commission.

Dharmapala, Dhammika, Slemrod, Joel, & Wilson, John Douglas. 2011. Tax Policy and the Missing Middle: Optimal Tax Remittance with Firm-Level Administrative Costs. *Journal of Public Economics*, **72**, 1036–1047.

Diamond, Peter A, & Mirrlees, James A. 1971. Optimal Taxation and Public Production I: Production Efficiency. *American Economic Review*, **61**(1), 8–27.

Dwenger, Nadja, & Steiner, Viktor. 2012. Profit Taxation and the Elasticity of the Corporate Income Tax Base: Evidence from German Corporate Tax Return Data. *National Tax Journal*, **65**(March), 117–150.

Emran, M. Shahe, & Stiglitz, Joseph E. 2005. On Selective Indirect Tax Reform in Developing Countries. *Journal of Public Economics*, **89**, 599–623.

Ernst & Young. 2013. *Worldwide Corporate Tax Guide*. Tech. rept. Ernst & Young.

FBR. 2012-2013 (July-September). *Quarterly Review*. Tech. rept. Federal Board of Revenue.

Feldstein, Martin. 1995. Behavioral Responses to Tax Rates: Evidence from the Tax Reform Act of 1986. *Ameican Economic Review*, **81**, 674–680.

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

Fuest, Clemens, & Riedel, Nadine. 2009. Tax evasion, tax avoidance and tax expenditures in developing countries: A review of the literature. *Report prepared for the UK Department for International Development (DFID)*.

Gordon, Roger, & Li, Wei. 2009. Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics*, **93**(7-8), 855–866.

Gruber, Jonathan, & Rauh, Joshua. 2007. How Elastic is the Corporate Income Tax Base? *In:* Hines, James R., Auerbach, Alan, & Slemrod, Joel (eds), *Taxing corporate income in the 21st century*. Cambridge University Press.

Hassett, Kevin A., & Hubbard, R. Glenn. 2002. Tax Policy and Business Investment. *Chap. 9, pages 1293–1343 of:* Auerbach, Alan J., & Feldstein, Martin (eds), *Handbook of Public Economics, Volume 3*.

Kawano, Laura, & Slemrod, Joel. 2012. *The Effect of Tax Rates and Tax Bases on Corporate Tax Revenues: Estimates With New Measures of the Corporate Tax Base*. NBER Working Paper 18440.

KEEN, MICHAEL. 2008. VAT, Tariffs, and Withholding: Border Taxes and Informality in Developing Countries. *Journal of Public Economics*, **92**, 1892–1906.

KEEN, MICHAEL. 2013. *Targeting, Cascading and Indirect Tax Design.* IMF Working Paper 13/57.

KLEVEN, HENRIK J., & WASEEM, MAZHAR. 2013. Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *Quarterly Journal of Economics*, **128**, 669–723.

KLEVEN, HENRIK JACOBSEN, KREINER, CLAUS THUSTRUP, & SAEZ, EMMANUEL. 2009. *Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries.* NBER Working paper 15218.

KLEVEN, HENRIK JACOBSEN, KNUDSEN, MARTIN B., KREINER, CLAUS THUSTRUP, PEDERSEN, SØREN, & SAEZ, EMMANUEL. 2011. Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, **79**(3), 651–692.

KOPCZUK, WOJCIECH. 2005. Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics*, **89**(11-12), 2093–2119.

KUMLER, TODD, VERHOOGEN, ERIC, & FRÍAS, JUDITH. 2013. *Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico.* Working Paper, Columbia University.

POMERANZ, DINA. 2013. *No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax.* Working Paper, Harvard Business School.

SAEZ, EMMANUEL. 2010. Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, **2**(3).

SAEZ, EMMANUEL, SLEMROD, JOEL, & GIERTZ, SETH. 2012. The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature*, **50**, 3–50.

SLEMROD, JOEL. 1995. Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986. *American Econonom Review: Papers and Proceedings*, **85**, 175–180.

SLEMROD, JOEL. 2001. A General Model of the Behavioral Response to Taxation. *International Tax and Public Finance*, **8**, 119–128.

SLEMROD, JOEL. 2007. Cheating Ourselves: The Economics of Tax Evasion. *Journal of Economic Perspectives*, **21**, 25–48.

SLEMROD, JOEL, & KOPCZUK, WOJCIECH. 2002. The optimal elasticity of taxable income. *Journal of Public Economics*, **84**(1), 91–112.

Slemrod, Joel, & Weber, Caroline. 2012. Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, **19**(1), 25–53.

Slemrod, Joel, & Yitzhaki, Shlomo. 2002. Tax avoidance, evasion, and administration. *Chap. 22, pages 1423–1470 of:* Auerbach, A. J., & Feldstein, M. (eds), *Handbook of Public Economics*. Handbook of Public Economics, vol. 3. Elsevier.

Slemrod, Joel, Blumenthal, Marsha, & Christian, Charles. 2001. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, **79**(3), 455–483.

Waseem, Mazhar. 2013. *Taxes, Informality and Income Shifting: Evidence from a Recent Pakistani Tax Reform*. Working Paper, London School of Economics.

Yitzhaki, Shlomo. 1974. Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics*, **3**, 201–202.

A Appendix

Table A.I: Robustness of Bunching Estimates

Panel A: Varying the Order of Polynomial

Order of Polynomial	3	4	5	6	7
High-rate Firms, 2006/07/09	4.28	3.92	4.44	6.05	5.53
	(.1)	(.1)	(.1)	(.1)	(.1)
Low-rate Firms, 2006/07/09	1.93	2.04	2	2.47	2.5
	(.2)	(.2)	(.2)	(.2)	(.2)
High-rate Firms, 2010	2.55	2.25	2.05	1.48	1.41
	(.2)	(.2)	(.2)	(.2)	(.1)

Panel B: Varying the Number of Excluded Bins

Number of Excluded Bins	1	2	3	4	5
High-rate Firms, 2006/07/09	1.83	2.57	3.65	4.44	2.23
	(.1)	(.1)	(.1)	(.1)	(.1)
Low-rate Firms, 2006/07/09	1.7	2	2.01	1.48	1.45
	(.1)	(.2)	(.2)	(.3)	(.3)
High-rate Firms, 2010	1.82	2.05	2.6	2.43	2.34
	(.1)	(.2)	(.2)	(.2)	(.3)

Notes: The table presents estimates of the excess mass b, for different specifications of the estimating equation (19), for the subsamples considered in table II. Bunching b is the excess mass in the excluded range around the kink, in proportion to the average counterfactual density in the excluded range. Panel A presents estimates for different choices of the order of polynomial $q \in \{3, 4, 5, 6, 7\}$, for the excluded range chosen as in table II (4 bins on either side of the kink for high-rate firms in 2006/07/09, 2 bins otherwise). Panel B presents estimates for different choices of the excluded range ($1 - 5$ bins on either side of the kink), for the order of polynomial chosen as in table II ($q = 5$). Bootstrapped standard errors are shown in parantheses.

Table A.II: Estimating Output Evasion Responses

	Observed Responses		Without Evasion	With Evasion			
	Bunching (b)	Profit Rate ($\Delta\hat{p}$)	Output Elasticity (ε_y)	Evasion Rate Response			
				$\varepsilon_y = 0$	$\varepsilon_y = 0.5$	$\varepsilon_y = 1$	$\varepsilon_y = 5$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High-rate Firms, 2006/07/09	4.44	1.0	133.3	67.6	67.4	67.1	67.5
	(0.1)	(0.03)	(3.8)	(2.0)	(2.0)	(2.0)	(2.0)
Low-rate Firms, 2006/07/09	2.00	0.4	34.3	17.6	17.3	17.1	15
	(0.2)	(0.04)	(3.0)	(1.7)	(1.7)	(1.7)	(1.7)
High-rate Firms, 2010	2.05	0.4	14.7	15.1	14.6	14.1	10.0
	(0.2)	(0.03)	(1.2)	(1.3)	(1.3)	(1.3)	(1.3)

Notes: This table analyzes the robustness of the estimates reported in table II when allowing for output evasion rather than cost evasion. The table presents bunching, elasticity and evasion estimates for the subsamples considered in panels A, B and D of figure IV. Column (1)-(3) repeat the first three columns of table II. Column (1) reproduces the bunching estimate b, based on estimating equation (19). Bunching b is the excess mass in the excluded range around the kink, in proportion to the average counterfactual density in the excluded range. Column (2) presents an estimate of the profit rate response associated with b, based on the relationship $\Delta\hat{p} = B/f_0\left(\tau_y/\tau_\pi\right) \simeq b \times binwidth$. Column (3) presents estimates of the real output elasticity ε_y for the model without evasion. This model is based on the assumption that bunching is purely due to a real output response. ε_y is estimated using the relationship $\Delta p = \left[c/y - c'\left(y\right)\right] dy/y \simeq \left(\tau_y^2/\tau_\pi\right)\varepsilon_y$. Columns (4)-(7) present estimates of the output evasion response as percentage of taxable profits (evasion *rate* responses), for the model with output evasion but no cost evasion, as presented in section 3.3. This model allows for bunching to be driven by both output evasion and real output response. The evasion response estimates are based on $\Delta\hat{p} = \epsilon_y\left(y/\hat{y}\right)\left(\tau_y/\tau_\pi\right) - \left(1-\tau_y/\tau_\pi\right)\left(\tau_y/\tau_\pi\right)\left[d(y-\hat{y})/(\hat{y}-\hat{c})\right]$, assuming different real output elasticities $\varepsilon_y \in \{0, 0.5, 1, 5\}$. Bootstrapped standard errors are shown in parentheses.

Table A.III: Data Cleaning Steps

Panel A: Sample Definition

Sample	**Definition**
Firms Reporting Profits & Turnover	Firms reporting profits Π, turnover y and incorporation date D. Based on Π and y, derive implied tax liablities $\tilde{T}^y$, $\tilde{T}^{\Pi}_H$ and $\tilde{T}^{\Pi}_L$ (high and low profit rate).
Consistency Check I	Drop firm if reported and implied tax liability inconsistent i.e. $T^y \neq \tilde{T}^y$ or $T^{\Pi} \neq \tilde{T}^{\Pi}_H$ and $T^{\Pi} \neq \tilde{T}^{\Pi}_L$. If $T^{\Pi} = \tilde{T}^{\Pi}_H$ or $\tilde{T}^{\Pi}_L$, assign {H,L}. If T^{Π} missing, assign {H,L} based on y, D and capital K.
Consistency Check II	Drop firm if reported and implied taxpayer status inconsistent, i.e. if $T^y > T^{\Pi}$ and $\tilde{T}^y < \tilde{T}^{\Pi}$; $T^y < T^{\Pi}$ and $\tilde{T}^y > \tilde{T}^{\Pi}$; $\tilde{T}^y > \tilde{T}^{\Pi}$ and T^y missing ; $\tilde{T}^y < \tilde{T}^{\Pi}$ and T^{Π} missing.

Panel B: Sample Size

Step	**Year**	**High-rate Firms**	**Low-rate Firms**
Raw data	2006/07/09	45,284	
	2008	21,445	
	2010	21,584	
Firms Reporting Profits & Turnover	2006/07/09	10,228	2,899
	2008	4,515	1,546
	2010	4,862	1,867
After Consistency Check I	2006/07/09	10,265	2,198
	2008	4,706	1,126
	2010	5,212	1,418
After Consistency Check II	2006/07/09	9,472	1,966
	2008	4,706	1,115
	2010	4,678	1,239

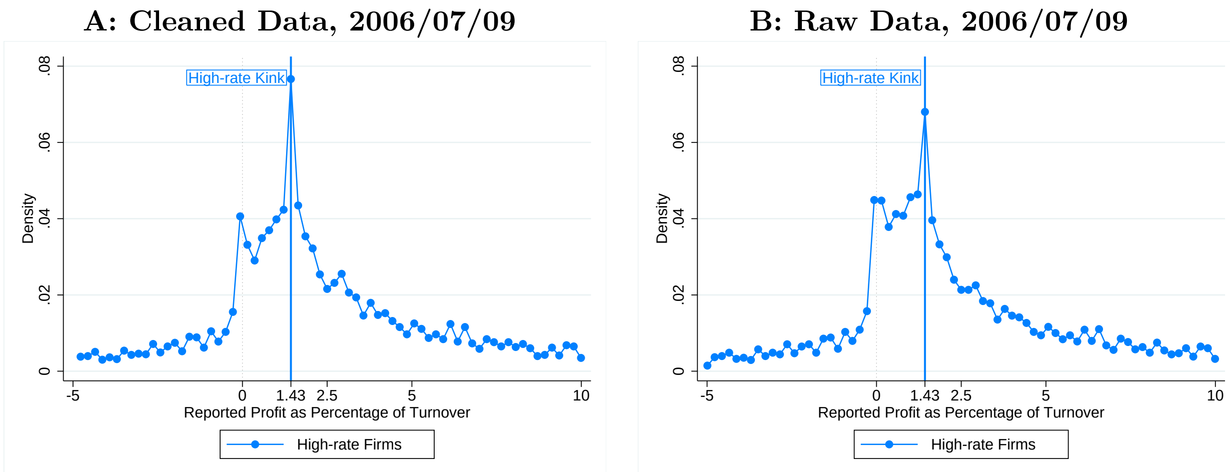
Notes: Panel A of this table explains the consistency checks applied to the data. For all consistency checks, a tolerance threshold of 5% is used. Panel B displays the sample size for different steps in the cleaning process. Capital K is equity plus retained earnings. Note that the implied turnover tax liability used for consistency check I is gross implied turnover tax liability minus net deductions (which are deduced from the tax liablity before the taxpayer status - turnover or profit taxpayer - is determined). For the same reason, the profits to turnover ratio used for consistency check II and for the bunching graphs is (profits-net reductions)/turnover, for firms that report positive net reductions.

Table A.IV: Comparison of Missing and Non-missing Observations

Panel A: Firms Reporting Profits and Turnover				
	N (1)	Median (2)	Mean (3)	SD (4)
Profits	17,358	0.1	-35.3	1624
Turnover	17,358	25.1	711.7	5579.6
Salary	7,865	8.8	63	265.6
Interest	9,726	0.9	84.8	901.2
Share of Low-rate Firms	17358		0.18	
Panel B: Firms Reporting Profits Only				
Profits	13,155	0	-69.8	2483.8
Salary	3,500	15	105.5	514
Interest	5,176	0.7	157.9	1472.3
Share of Low-rate Firms	11814		0.14	
Panel C: Firms Reporting Turnover Only				
Turnover	8,551	9.2	454.5	7399.1
Salary	3,078	5	37.8	271.8
Interest	3,767	0.7	40.2	259.9
Share of Low-rate Firms	8,551		0.27	

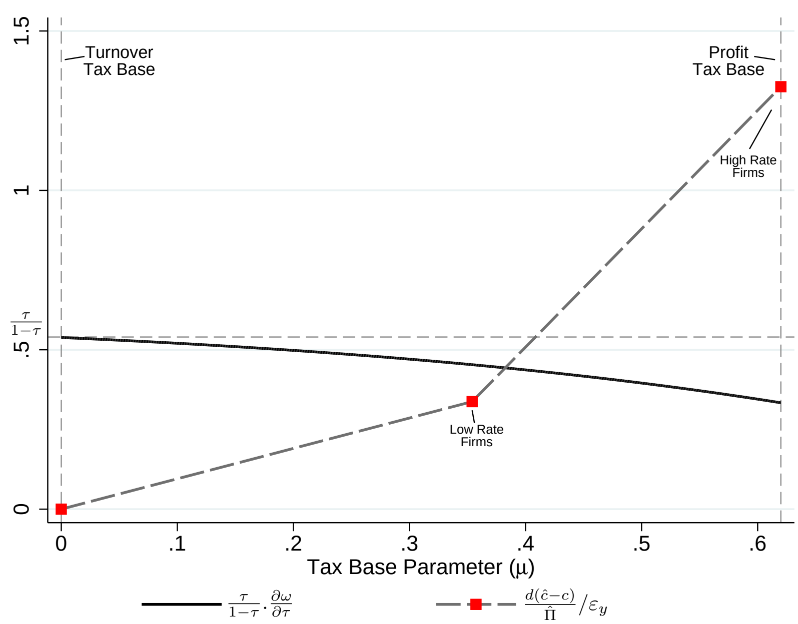
Notes: The table compares different samples of firms depending on wether or not they report profits and turnover. Panel A considers firms that report both profits and turnover. Panel B considers firms that report profits only. Panel C considers firms that report turnover only. Columns (1)-(4) report the number of observations, median, mean and standard deviation for different observable characteristics (turnover, profits, salary payments, interest payments, share of small firms). All statistics are in million Pakistani Rupees (PKR).

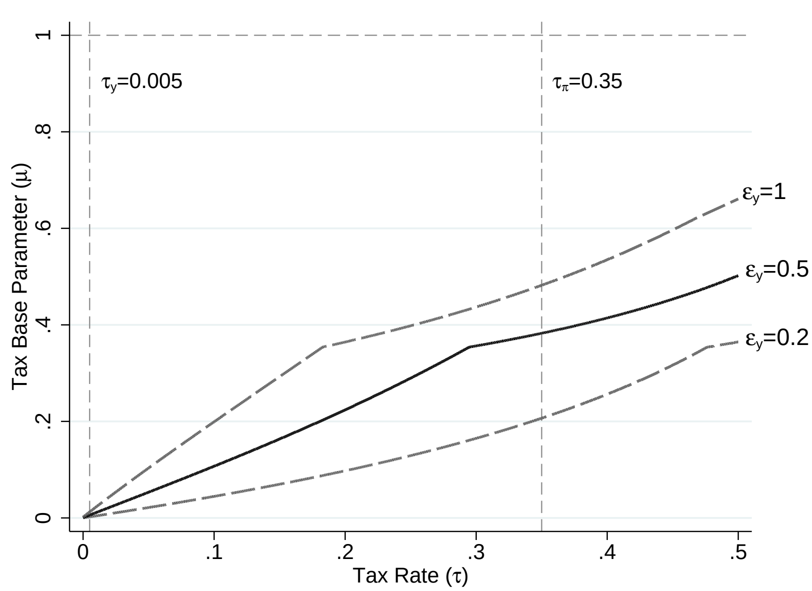
Figure A.I: Comparison of Bunching for Cleaned and Raw Data



Notes: This figure shows bunching in the sample of high-rate firms in 2006/07/09, in the cleaned data (panel A, same as panel A in figure II), and the raw data (panel B). The raw data contains all observations that report turnover and profits, before consistency checks I and II are applied, and before firms reporting $\pi = 0$ are dropped. The graphs show the empirical density distribution of the profit rate (reported profit as percentage of turnover). The tax rate schedule is explained in the notes to figure II. For high-rates firms in 2006/07/09, the minimum tax regime places the kink at a profit rate of 1.43% (firms with a profit rate above this threshold fall under the profit tax scheme, while firms with a profit rate below fall under the turnover tax scheme). The kink point is marked by a vertical solid line. The zero profit point is marked by a vertical dotted line. The bin size is 0.214, chosen so that the kink point is a bin centre.

Figure A.II: Numerical Policy Analysis with Distortionary Profit Tax





Notes: This figure analyzes the robustness of the numerical policy analysis shown in Figure V when using a distortionary profit tax as a benchmark. We rescale our tax base parameter μ so that the implied tax wedge ω equals 17% – the mean effective tax rate on profits reported in Gruber & Rauh (2007) – implying $\mu = 0.62$ when $\tau_\pi = 0.35$. The solid black curve in panel A again plots the left-hand side of the optimal tax rule (equation 11) as a function of the rescaled μ for $\tau_\pi = 0.35$. The three red markers on the dashed gray curve show respectively the right-hand side of the optimal tax rule at $\mu = 0$, at $\mu = (0.2/0.35) \times 0.62$ based on the evasion rate response estimated for the low-rate firms, and at $\mu = 0.62$ based on the evasion rate response estimated for the high-rate firms. By extrapolating between these three estimates, we find that the optimal tax base implied by the tax rule equals $\mu = 0.383$ as compared to $\mu = 0.578$ for a completely nondistortionary profit tax. In panel B we replicate the exercise to find the optimal tax base as a function of the tax rate for three different levels of the output elasticity.