Essays on Business Taxation and Development

Anne Brockmeyer

A thesis submitted to the Department of Economics,

for the degree of Doctor of Philosophy, June 2013.

London School of Economics and Political Science

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others.

I confirm that the work for chapter 2 was carried out jointly with equal share by Michael Best, Henrik Kleven, Johannes Spinnewijn, Mazhar Waseem and myself. I contributed 20% of all the analysis in the paper. I declare that chapter 3 is the result of my previous study for the degree of MRes, which I undertook at the London School of Economics and Political Science. That degree was awarded in 2010.

For chapter 1, the following disclaimer applies: "This work contains statistical data from Her Majesty's Revenue and Customs (HMRC) which is Crown Copyright. The research datasets used may not exactly reproduce HMRC aggregates. The use of HMRC statistical data in this work does not imply the endorsement of HMRC in relation to the interpretation or analysis of the information."

I declare that my thesis consists of 43,917 words in total.

For my family.

لاشرف.

Acknowledgment

This work has been supported by the exciting intellectual environment at the LSE and the invaluable advice I received from a number of people. I am very grateful to Tim Besley, Robin Burgess and especially Henrik Kleven, for their guidance, support and collaboration at different stages of my research. I am particularly thankful to Henrik Kleven and Johannes Spinnewijn, and to Michael Best and Mazhar Waseem, for the collaboration which lead to chapter 2 of this thesis. I have learned enormously from this project, and the inspiration I drew from it has served as a catalyst for my other projects. I also thank all the public economics and development faculty at the LSE for helpful feedback over the years.

I am deeply indebted to the staff at the HMRC datalab, especially Lucy Nicholson and Daniele Bega, for providing the data on which chapter 1 is based, and dealing with countless enquiries in an efficient and supportive way. I also owe big thanks to the datalab IT team including Chioma Anaba, Mark Brewin, Lin Edwards, John Haynes, Karen Jackson and Miriam Stratton, for ensuring the smooth functioning of the datalab.

My work has benefited from presentations at the EBRD, HMRC, the Warwick PhD students conference, the Oxford Centre for Business Taxation Annual Symposium, from helpful discussions with Michael Devereux, and from the comments of numerous colleagues, including Michael Best, Shawn Chen, Jon de Quidt, Francisco Costa, Anders Jensen, Li Liu, Simon Loretz, Sam Marden, Claudia Steinwender, Mohammad Vesal, Mazhar Waseem and Tuuli Ylinen. For financial support, I thank the Economic and Social Research Council, the Adeline and Karl Goeltz Foundation and the German National Academic Foundation.

I have been incredibly fortunate to have forged great friendships with my classmates and officemates at the LSE, and my wonderful housemates at Goodenough College, and I thank all of them for their company. Last but not least, I am deeply grateful for the generosity, love and support of my family.

Abstract

This thesis addresses a number of questions on the optimal taxation of firms, with particular emphasis on the challenges to taxation in developing economies. Chapter 1 exploits bunching of firms at a tax kink to identify the effect of a tax rate change on investment. Building on the standard bunching framework, I estimate the frequency distribution of firms around the kink, and the share of bunching firms with excess investment. I apply this approach to administrative tax returns for firms in the United Kingdom and find that excess investment explains up to 20% of bunching. Chapter 2 examines the trade-off between production efficiency and revenue efficiency in taxation under imperfect enforcement. We exploit quasi-experimental variation created by a minimum tax scheme, a production inefficient policy used in many developing countries, which consists of taxing firms on turnover as their profit rate falls below a certain threshold. Using administrative tax records of corporations in Pakistan, we find large bunching around the profit rate kink createded by the minimum tax scheme and estimate that the turnover tax reduces evasion by up to 60-70% of corporate income. Chapter 3 analyzes the impact of interventions by the International Monetary Fund (IMF) on countries' likelihood of adopting the value-added tax (VAT). I discuss how the IMF has promoted VAT adoption by making lending conditional on adoption, providing administrative and technical assistance, and reducing the political costs of adoption. Applying a Cox proportional hazard model to a cross-country panel for the period 1975-2000, I find that countries that are under a lending agreement with the IMF are three times as likely to adopt the VAT than are countries not under a lending agreement.

Contents

List of Figures

List of Tables

List of Appendices

Introduction

This thesis addresses a number of questions on the optimal taxation of firms, with particular emphasis on the challenges to taxation in developing economies. A key motivation for this study is the fact that developing economies raise less tax revenue as a share of GDP than developed economies (17.6% vs. 25%) and are much more reliant on corporate taxation as a source of revenue (42.3% vs. 17.85% of tax revenue, Gordon & Li 2009). I study how investment responds to corporate taxation, discuss the trade-off between production and revenue efficiency in taxation under imperfect enforcement, and examine the determinants of the adoption of the value-added tax (VAT), a consumption tax levied at firm level. The first two chapters exploit quasi-experimental variation from kinks in the corporate tax schedule and use corporate tax return data for the universe of firms in the United Kingdom (UK) and Pakistan. The third chapter estimates a Cox hazard model on cross-country data. Overall, this work contributes to a growing literature on public finance and development reviewed in Besley & Persson (2013).

Chapter 1 exploits bunching of firms at a tax kink as quasi-experimental variation to identify the effect of a tax rate change on investment, and explore how this effect interacts with variation in capital depreciation rates. The idea is that firms with a taxable income slightly above the kink have an incentive to reduce their income to bunch at the kink, and increasing investment is one possible strategy for that. This means that bunching of firms should be accompanied by a spike in investment at the kink. Building on the standard bunching framework, I estimate the frequency distribution of firms around the kink, and the share of bunchers with excess investments at the extensive and intensive margin. I apply this approach to administrative tax return data for the universe of UK firms from 2001-2007, and show that investment by small firms significantly responds to a tax rate change. I find large and significant spikes in the share of capital investors and median capital costs at the 10k kink. The spikes are larger in 2002-2005 when the kink is larger, and for quickly depreciating capital items, which yield larger tax reductions. I estimate that extensive margin investments explain 7.7-19.2% of bunching and intensive margin investments explain 4.3-16.8% of bunching. Evidence from subsample analysis supports the interpretation of the observed behaviour as real investment rather than evasion or avoidance.

Chapter 2 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.

Chapter 3 analyzes the impact of interventions by the International Monetary Fund (IMF) on the likelihood of countries to adopt the VAT. I discuss how the IMF has promoted VAT adoption by making lending conditional on adoption, providing administrative and technical assistance, and reducing the political costs of adoption. I use the Cox proportional hazard model to estimate the impact of IMF lending on VAT adoption in a panel of 125 countries from 1975-2000. I find that countries that are under a lending agreement with the IMF are three times as likely to adopt the VAT than are countries not under a lending agreement. I show that this effect is not driven by omitted variables such as debt crises and provide suggestive evidence against reverse causality. I also find that political factors play a less prominent role in VAT adoption than theory predicts. Political instability does not affect VAT adoption and autocracy reduces the likelihood of adoption only in the absence of IMF lending.

Chapter 1

The Investment Effect of Taxation: Evidence from a Corporate Tax Kink

1.1 Introduction

The impact of tax policy on corporate investment continues to be debated among economists and policymakers. As the corporate income tax is not a tax on pure profits,[1] theory predicts that the tax rate and depreciation allowance schedule should affect firms' investment decisions. A reduction in the tax rate would increase the after-tax return to investment, while an increase in capital depreciation rates would reduce the cost of investment. As investment is much more volatile than other macroeconomic variables and has important multiplier effects on the economy, policymakers have frequently tried to stimulate investment through tax incentives. The 1981 introduction of the investment tax credit by Reagan, the 2002 US bonus depreciation policy,[2] and the zero starting rate for firms with a taxable income below £10,000[3] in the UK, in place during 2002-2005, are just a few examples of such stimulus policies.

While the use of investment tax incentives abounds, it is empirically difficult to demonstrate a positive causal effect of tax policy on investment. The line of causality between investment and tax policy runs both ways, as tax reforms are often motivated by sluggish investment, and a host of other macroeconomic variables affects both tax policy and investment. The literature so far has shown that the Hall-

[1]For the multiple reasons why the corporate income tax is not a tax on pure profits, see Auerbach *et al.* (2010) and Hassett & Hubbard (2002).

[2]Temporary increase in capital depreciation rates.

[3]Henceforth 10k for simplicity.

Jorgenson *user cost of capital* has a large impact on investment[4], but this estimate confounds the effects of different tax policy parameters, namely the tax rate and capital depreciation allowance. Work attempting to disentangle these effects by examining specific tax parameter changes, such as the bonus depreciation policy found only weak evidence for tax incentive effects (e.g. Cohen & Cummins 2006, House & Shapiro 2008).

This paper exploits bunching of firms at a kink point, a taxable income threshold at which the marginal tax rate changes discontinuously, as quasi-experimental variation. This allows me to identify the effect of a tax rate change on investment, and to explore how this effect interacts with variation in capital depreciation rates. The idea is that firms with a taxable income slightly above the kink have an incentive to reduce their income to bunch at the kink, and increasing investment is one possible strategy for that. Investment translates into capital depreciation allowances, which are deducted from taxable income and can measure up to 100% of the underlying expenditure in the year of purchase, depending on the capital type. If firms move to the kink by increasing investment, bunching should be accompanied by a spike in investment at the kink. My estimation builds on the bunching approach of estimating the elasticity of taxable income (ETI), as developed by Saez (2010) and applied to corporations by Bach (2012) and Devereux *et al.* (2013). In addition to the frequency distribution of firms, I estimate the size of the *investment spike*, i.e. the number of firms with higher than predicted investment levels around the kink. I consider the size of the investment spike at both the extensive margin (share of capital investors) and the intensive margin (median capital costs for investors). This allows me to derive the number of *investment bunchers*, bunchers that moved to the kink inter alia by investing.

I apply this approach to study the investment effect of taxation in the UK, a setting which presents two methodological advantages. First, the UK tax schedule provides a compelling source of variation for my study, featuring a large and salient kink at 10k of taxable income during 2001-2005. The size of the tax rate jump at this kink varies over time from 12.5 to 23.75 percentage points, creating stronger incentives to bunch at the kink in years with a larger tax rate jump. In addition, variation in capital depreciation rates across capital types creates variation in the suitability of different capital investments as instruments for bunching, with investments yielding larger allowances in the year of purchase more suitable for bunching. Combining these two types of variation allows me to distinguish bunching-induced variation in investment around the 10k kink from any other changes in investment that occur around the kink. The second advantage of the UK setting is the availability of administrative tax return data, which covers the universe of UK

[4]See Hassett & Hubbard (2002) for a review.

firms for 2001-2007, and contains precise measures of investment-induced taxable income reductions (capital depreciation allowances).

My empirical findings show that investment by small firms significantly responds to a tax rate change. First, I find large and statistically significant spikes in the share of capital investors and median capital costs at the 10k kink. Consistent with the idea of bunching through investment, the spikes are larger in 2002-2005 when the kink is larger, and also larger for quickly depreciating capital items, which yield larger tax reductions. Second, I quantify the contribution of this investment behaviour to bunching, and find that extensive margin investments explain 7.7-19.2% of bunching and intensive margin investments explain 4.3-16.8% of bunching. Third, I provide evidence supporting the interpretation of the observed behaviour as real investment rather than evasion (over-reporting) or avoidance (transfer pricing, income shifting). The bunching response to the kink is stronger in subsamples characterized by high investment propensity (growing firms, manufacturing and retail sector firms, high cost-margin firms), and weaker in subsamples characterized by high ease of evasion (firms with below median turnover and number of employees). The scope to engage in avoidance schemes like cost manipulation through transfer pricing or shifting income to international tax havens is extremely limited for firms with a taxable income around 10k, as less than 3% of them are members of a group or register any overseas income.

This paper contributes to several strands of literature. First, my paper relates to the large literature estimating Hall-Jorgenson style user cost of capital models. This literature either relies on tax reforms as instruments for user cost of capital changes in a generalized methods of moments estimation (Cummins *et al.* 1994; Cummins *et al.* 1996), or compares forecast errors in investment and the user cost of capital (Auerbach & Hassett 1991). My work echoes this literature's findings of a significant effect of tax policy on investment, but presents two methodological advantages. Contrary to previous studies, which confound the effect of different tax policy parameters, my quasi-experimental design allows me to isolate the effect of a tax rate change, and to explore its interaction with capital allowance rates. In addition, my study uses administrative tax return data, which is more accurately measured than the accounting data used in previous studies, is available for a larger sample of firms, and provides direct measures of capital depreciation allowances claimed.

Second, my work is related to the more recent literature examining the effect of the bonus depreciation policy, using difference-in-difference-type strategies (Cohen & Cummins 2006) or a structural approach (House & Shapiro 2008; Auerbach *et al.* 2008). Contrary to this literature, which remains divided on the stimulating effect of the policy, I provide clear evidence for a positive effect of a tax rate decrease, and show that capital depreciation rates matter for the composition of capital investment.

Third, this paper builds on the use of bunching at kink points to estimate the ETI, an approach developed by Saez (2010), Chetty *et al.* (2011) and Kleven & Waseem (2013), and applied to firms by Bach (2012), Devereux *et al.* (2013) and Best *et al.* (2013). I show how estimating the distribution of taxable income *components* such as investment provides information on the underlying drivers of bunching. This empirical approach is not confined to estimating investment responses only, but can also be used more generally to decompose the response to taxation in contexts with kinked tax schedules.

Finally, my paper contributes to research on the decomposition of taxable income changes, providing the first empirical evidence of real responses by firms. Gordon & Slemrod (2000), Saez (2004) and Devereux *et al.* (2013) find evidence that changes in corporate profit margins in the UK and US are partly due to tax-rate induced income shifting from the corporate to the personal tax base. Almunia & Lopez-Rodriguez (2012) and Seim (2012) argue that responses of Spanish firms to an enforcement notch, and of Swedish personal taxpayers to a wealth tax kink are entirely driven by reporting effects. Similarly, Best *et al.* (2013) show that bunching at the minimum tax kink in Pakistan must be largely driven by evasion. This paper is the first to provide well-identified evidence of a real investment response to taxation.

The paper is organized as follows. Section 1.2 develops the empirical strategy. Section 1.3 introduces the context and data. Section 1.4 presents the empirical results. Section 1.5 concludes.

1.2 Empirical Strategy

This section presents a novel empirical strategy for investigating firms' response to taxation. Section 1.2.1 reviews how bunching at kink points is used to estimate the ETI. Section 1.2.2 shows that examining the distribution of investment around tax kinks helps to identify the response of investment to a tax rate change.

1.2.1 Estimating Bunching

As proposed by Saez (2010), I use bunching in the distribution of taxable income around a kink point in the tax schedule to estimate the ETI. Panel A of Figure 1.I illustrates this idea. With a constant marginal tax rate τ_1 and no kink, taxable income π follows a smooth frequency distribution (red dashed line). Now consider the introduction of a higher marginal tax rate τ_2 for profits above some threshold $\pi^\star$ (marked by a black solid line). Assume that firms have a convex cost function (or a concave production function) and maximize after-tax profits. With the new tax regime, all firms with $\pi > \pi^\star$ reoptimize and move to a lower profit level. Firms in some interval $[\pi^\star, \pi^\star + \Delta\pi^\star]$ will no longer find it profitable to produce $\pi > \pi^\star$ and thus move to the kink (black solid line). The new distribution of profits (blue solid line) has a spike at $\pi^\star$ and a lower frequency for $\pi > \pi^\star$.

As the excess mass at the kink must equal the (shaded) area below the distribution from which bunchers move to the kink, the excess mass indicates the income change $\Delta\pi^\star$ of the marginal buncher. To see how this translates into an estimate of the ETI, abstract from income effects by assuming that the jump in the marginal tax rate is small, and consider that the compensated elasticity of taxable income π with respect the tax rate τ is

$$\epsilon = \frac{d\pi^\star}{\pi^\star}\frac{(1-\tau)}{d\tau}. \tag{1.1}$$

Assuming furthermore that the frequency distribution $f(\pi)$ is uniform around the kink, the number of bunchers is

$$B = d\pi^\star f(\pi^\star). \tag{1.2}$$

Transforming B into an estimate of the excess mass b through $b(\tau_1, \tau_2) = B/f(\pi^\star)$, we can thus approximate[5] ϵ as

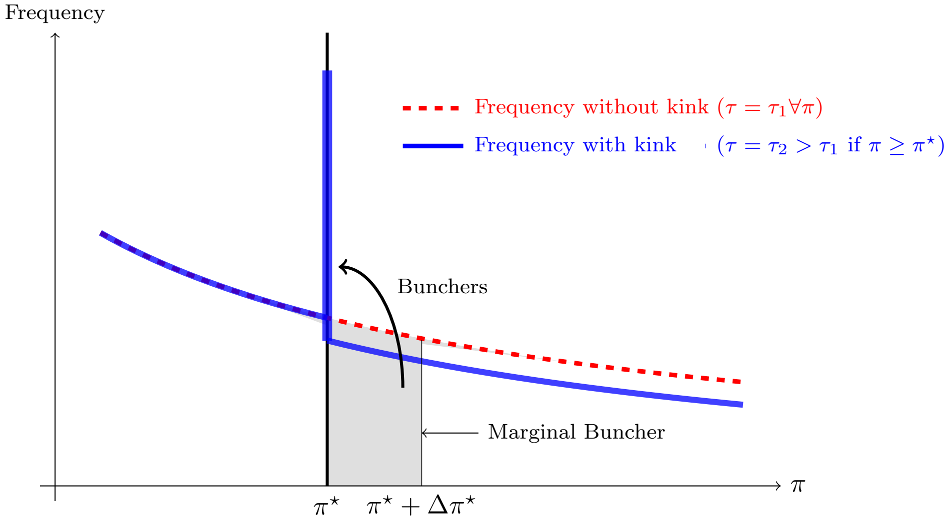
$$\epsilon \simeq \frac{b(\tau_1, \tau_2)}{\pi^\star \cdot log\frac{1-\tau_1}{1-\tau_2}}. \tag{1.3}$$

To implement the estimation empirically, I follow the strategy developed by Chetty *et al.* (2011) and Kleven & Waseem (2013), as illustrated in Panel B of Figure 1.I. The figure shows the empirical frequency (blue curly line) which, due to optimization frictions, does not feature a precise spike but rather diffuse bunching around the kink. I estimate a counterfactual frequency (red solid line) by fitting a

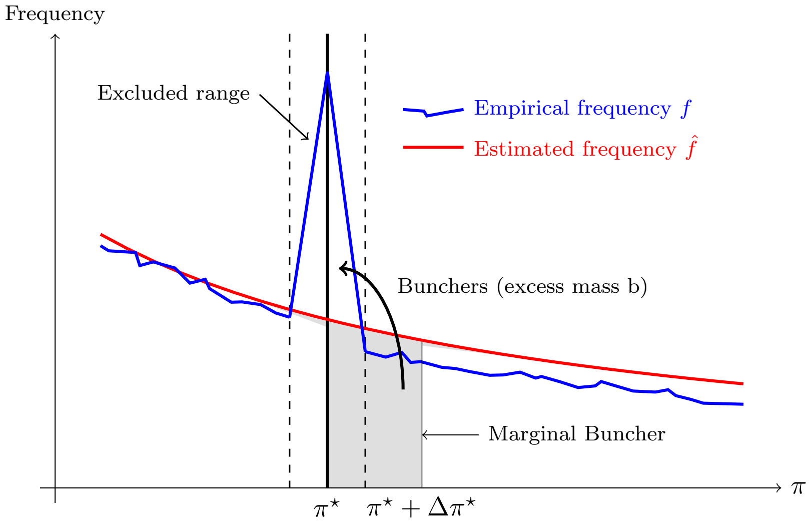
[5] Saez (2010) shows that this approximation also applies to the general case in which the tax rate change at the kink is not small, if one assumes a quasi-linear utility (production) function, so that the compensated and uncompensated elasticities are the same.

Figure 1.I: **Estimating Bunching**

A: Theory



B: Empirical Implementation



Notes: The figure illustrates the theoretical underpinning (Panel A) and empirical implementation (Panel B) of the bunching strategy to estimate the elasticity of taxable income. As Panel A shows, when there is no kink and all firms pay marginal tax rate $\tau = \tau_1$, independently of their taxable income, the frequency distribution of firms is smooth (red dotted line). Now consider that a kink is introduced at income level $\pi^\star$ (marked by a black vertical line), so that firms pay tax rate $\tau_2 > \tau_1$ for all units of income $\pi > \pi^\star$. Assuming that firms have a convex cost function, firms with income in some interval $[\pi^\star, \pi^\star + \Delta\pi^\star]$ move to the kink. Firms at higher income levels also re-optimize and decrease their income, but do not move all the way to the kink. The new frequency distribution (blue solid line) features a spike at the kink and a lower frequency above the kink. The size of the excess mass at the kink is equal to the interval (shaded area) from which firms move to the kink, and thus indicates the income change $\Delta\pi^\star$ of the marginal buncher. Panel B shows the empirical frequency distribution (blue curly line) when there is a kink at $\tau^\star$. In practice, optimization frictions prevent firms from moving straight to the kink, so that bunching takes the form of a diffuse excess mass around the kink, rather than a spike. The empirical implementation consists in fitting a counterfactual frequency $\hat{f}$ (red solid line) as flexible polynomial to the observed frequency f, excluding the range around the kink in which bunching takes place (marked by black dashed lines). The details of the estimation are shown in equation (1.4). Estimating the excess mass b between the empirical and the estimated frequency in the excluded range allows us to approximate the elasticity of taxable income as $\epsilon = b/\left[\pi^\star \cdot log(1-\tau_1, 1-\tau_2)\right]$.

q-order polynomial to the frequency counts f_j of firms with a taxble income in bin π_j, excluding an interval $[\pi_L, \pi_U]$ around the kink point (marked by black dashed lines):

$$f_j = \sum_{l=0}^{q} \beta_l (\pi_j)^l + \sum_{k=\pi_L}^{\pi_U} \gamma_k \cdot \mathbf{1}[\pi_j = k] + \sum_{r \in R} [\rho_r + \mu_r \cdot \pi_j] \cdot \mathbf{1}[\frac{\pi_j}{r} \in N] + v_j. \quad (1.4)$$

The γ_k coefficients allow for different frequencies in the excluded range. As the empirical frequency displays strong round number bunching at low income levels, I control for round number bunching at multiples of $R = \{5k, 10k\}$, and allow these round number effects to change linearly with income. The last term in the equation, v_j, is the error. I account for the fact that bunchers move to the kink from the right by implementing an integration constraint, shifting the counterfactual distribution to the right of the kink upwards, until the area under the counterfactual integrates to the area under the empirical distribution. My empirical specification follows Devereux *et al.* (2013) in that it relies on taxable income bins of £100, a fifth-order polynomial, and an excluded range of $[8k; 12k]$.[6]

The number of bunchers is calculated as the difference between the observed frequency and the predicted frequency $\hat{f}_k$ of equation (1.4), ignoring the contribution of the γ_k coefficients for the excluded range:

$$B = \sum_{k=\pi_L}^{\pi_U} (f_k - \hat{f}_k). \quad (1.5)$$

This can be translated into the excess mass b and the elasticity ϵ, by scaling B by the average predicted frequency in the N_k excluded bins:

$$b = \frac{\sum_{k=\pi_L}^{\pi_U} (f_k - \hat{f}_k)}{\sum_{k=\pi_L}^{\pi_U} \hat{f}_k / N_k}. \quad (1.6)$$

All standard errors are derived from a bootstrap procedure, which samples from the estimated residuals with replacement.

1.2.2 Estimating Investment Bunching

This section supplements the bunching framework with an analysis of investment changes, to quantify their contribution to bunching. Investment in capital items is measured by capital depreciation allowances, which capture the extent to which

[6]See Devereux *et al.* (2013) for robustness checks with different specifications. All estimations rely on a $[3k, 40k]$ estimation range, but the figures in section 1.4 zoom in on the $[3k, 30k]$ range around the kink.

investment reduces taxable income. If firms reduce their taxable income by increasing investment, bunching should be accompanied by an *investment spike* around the kink. The size of the observed spike, combined with the underlying distribution of investment, is used to estimate the number of *investment bunchers*, firms which moved to the kink by increasing their level of investment. The approach is applied to extensive and intensive margin changes in investment in sections 1.2.2.1 and 1.2.2.2 respectively.

1.2.2.1 Extensive Investment Bunching

To detect extensive investment bunching, I investigate whether firms that would not otherwise invest make capital investments to move to the kink. I plot the share s_j of investors, i.e. firms in income bin π_j that report capital depreciation allowance $c_g > 0$ for some capital item g. This plot is illustrated in Panel A of Figure 1.II (blue curly line). In the absence of the kink, the share of investors evolves smoothly as a function of taxable income. Once a kink is introduced, firms intending to bunch start investing in new capital items, move to the kink, and the share of investors spikes at the kink.

To estimate the counterfactual share of investors (red solid line), I fit a q-order polynomial to the share of investors s_j in income bin π_j, excluding bins in an interval $[\pi_L, \pi_U]$ around the kink:

$$s_j = \delta_{01}^E \cdot \mathbf{1}[\pi_j < \pi^\star] + \delta_{02}^E \cdot \mathbf{1}[\pi_j \geqslant \pi^\star] + \sum_{l=1}^{q} \delta_l^E (\pi_j)^l + \sum_{k=\pi_L}^{\pi_U} \theta_k^E \cdot \mathbf{1}[\pi_j = k] + u_j^E. \quad (1.7)$$

The superscript E denotes the extensive margin, u_j is the error term. The empirical specification (bin size, order of polynomial, excluded range) is as in section 1.2.1. The only distinction is that this specification allows for different constants above and below the kink. Indeed, in the UK tax system, firms with a gross income below the 10k kink face a 0% tax rate in 2002-2005 and thus have no incentive to claim capital depreciation allowances. The observed share of firms claiming capital allowances is lower below the kink and discontinuously jumps up at the kink.[7]

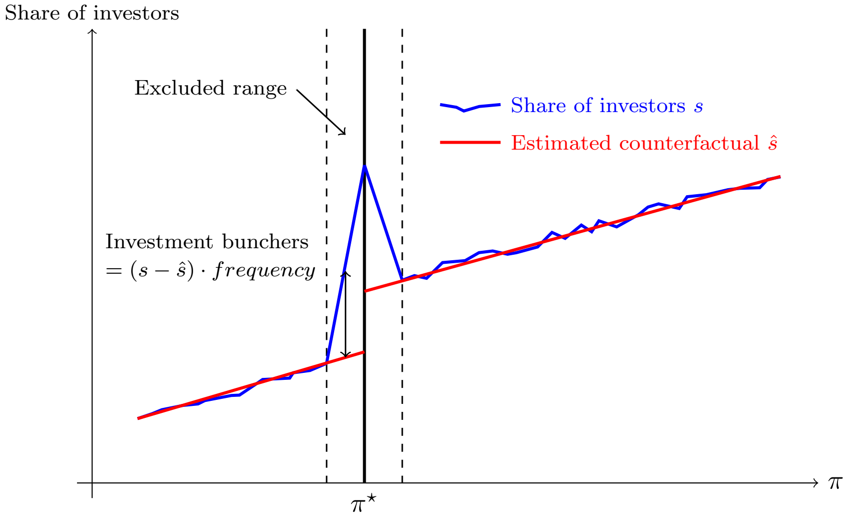
The number of *extensive investment bunchers* is estimated as the excess number of investors in the bunching range:

$$M^E = \sum_{k=\pi_L}^{\pi_U} (s_k - \hat{s}_k) \cdot f_k, \quad (1.8)$$

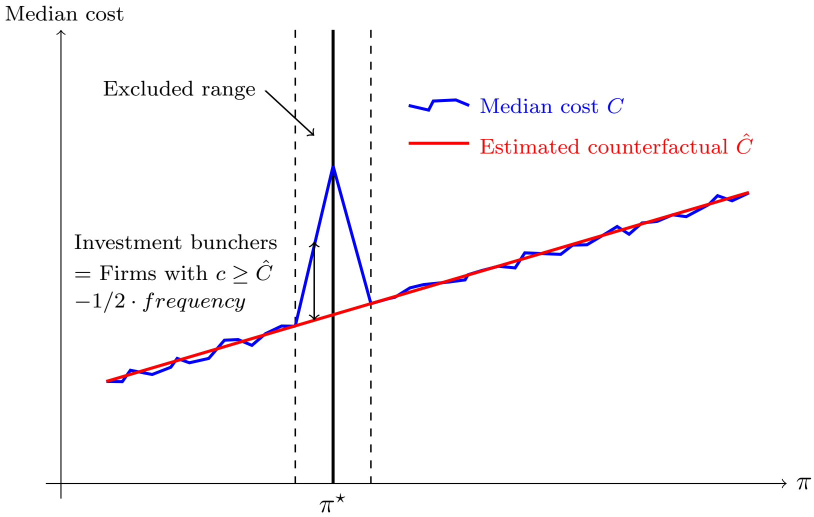
[7] Round number investment spikes are controlled for by the control procedure presented in section 1.4.2.

Figure 1.II: **Estimating Investment Bunching**

A: Extensive Margin



B: Intensive Margin



Notes: The figure illustrates the empirical estimation of investment bunching, for extensive and intensive margin investment changes. Panel A considers the extensive margin. It plots the share of investors s (firms claiming positive capital depreciation allowances, blue curly line) evolves smoothly as a function of taxable income outside of the excluded range (marked by black dashed lines) but features a spike at the kink $\pi^\star$ (marked by a black solid line). Due to optimization frictions, this spike is located not only at the kink, but also covers income levels just below and above the kink, within the excluded range. The fact that s is lower to the left of the excluded range is due to a particularity of the UK tax system and is shown here to facilitate an understanding of the empirical results. During 2002-2005, firms below the kink face a 0% tax rate, and thus have no incentive to claim capital allowances for their investments, unless their gross income would place them above the kink. The empirical strategy consists of estimating the counterfactual share of investors $\hat{s}$ (red solid line) by fitting a flexible polynomial to the observed share s, excluding a range around the kink, as shown in equation (1.7). The number of extensive investment bunchers is estimated as suggested in equation (1.8), as the excess number of investors, i.e. the difference between the observed and counterfactual share of investors in the excluded range. Panel B is similar to Panel A, but considers the intensive margin of investment changes. It shows the observed median cost C (capital depreciation allowance) for the sample of investors (blue curly line) and the counterfactual median cost $\hat{C}$ (red solid line), estimated according to equation (1.11). See section (1.2.2.2) for a discussion of why I consider the median cost to examine intensive investment bunching. The number of intensive investment bunchers is estimated as the excess number of firms with above median capital costs in the excluded range, as show in equation (1.12).

where the counterfactual share of investors is the predicted value from equation (1.7), ignoring the $\hat{\theta}_k^E$coefficients for the excluded range,

$$\hat{s}_j = \hat{\delta}_{01}^E \cdot \mathbf{1}[\pi_j < \pi^\star] + \hat{\delta}_{02}^E \cdot \mathbf{1}[\pi_j \geqslant \pi^\star] + \sum_{l=1}^{q} \hat{\delta}_l^E (\pi_j)^l. \tag{1.9}$$

The contribution of extensive investment changes to bunching, and thus to the ETI, is defined as the share of extensive investment bunchers in the total number of bunchers:

$$Cont^E = \frac{M^E}{B}. \tag{1.10}$$

1.2.2.2 Intensive Investment Bunching

To detect intensive investment bunching, I investigate whether firms move to the kink by increasing the size of their investment. I consider only investors and base my analysis on the median cost (capital depreciation allowance) C_j of firms in income bin π_j. This choices warrents some explanation. First, I examine the median rather than the mean cost, as the latter is much more noisy and influenced by outliers. Second, I consider only investors, as the share of non-investors is above 90% for most capital items. Examining intensive investments in the full sample would thus require using not the median but different (endogenously chosen) percentiles for different capital items.[8]

As illustrated in Panel B of Figure 1.II, the median cost (blue curly line) evolves smoothly as a function of taxable income in the absence of a kink. Once the kink is introduced, firms intending to bunch increase their investment levels, move to the kink, and the median cost spikes at the kink. Analogously to the previous section, I fit a q-degree polynomial to estimate the counterfactual median cost (red solid line), excluding the bins around the kink:

$$C_j = \sum_{l=0}^{q} \delta_l^I (\pi_j)^l + \sum_{k=\pi_L}^{\pi_U} \theta_k^I \cdot \mathbf{1}[\pi_j = k] + u_j^I. \tag{1.11}$$

The superscript I denotes the intensive margin. The number of *intensive investment bunchers* is estimated as the excess number of firms with a cost above the counterfactual median:

$$M^I = \sum_{k=\pi_L}^{\pi_U} \sum_{i \in j} \mathbf{1}[\pi_j = k] \cdot \mathbf{1}[c_{ij} \geq \hat{C}_j] - \frac{1}{2} \sum_{k=\pi_L}^{\pi_U} f_k, \tag{1.12}$$

where c_{ij} is the cost of firm i in income bin j, and the counterfactual median is the

[8]The median, for instance, is constant at 0 for all capital items except short-life machinery, but the 10th percentile is uninformative for short-life machinery, in which almost all firms invest.

predicted value from equation (1.11), ignoring the $\hat{\theta}_k^I$ coefficients for the excluded range,

$$\hat{C}_j = \sum_{l=0}^{q} \hat{\delta}_l^I (\pi_j)^l. \tag{1.13}$$

The contribution of intensive investment bunching is then defined as the share of intensive investment bunchers in the total number of bunchers:

$$Cont^I = \frac{M^I}{B}. \tag{1.14}$$

It should be noted that the group of investors obviously confounds firms that invest regardless of the kink and the extensive investment bunchers that invest to move to the kink. The presence of extensive investment bunchers, which most likely have below-median cost levels, would bias the median cost downward for investors in the excluded range. I therefore consider the intensive margin estimates a lower bound on the investment bunching contribution.

1.3 Context and Data

The UK corporate tax system provides a unique context for applying the new empirical strategy to study the investment response to taxation. Section 1.3.1 presents the tax system and discusses the sources of identifying variation it provides. Section 1.3.2 presents the administrative tax return data used in this study.

1.3.1 UK Corporate Tax System

The corporate income tax in the UK is contributed by approximately 1.5 million registered corporations each fiscal year (1 April in year t until 31 March in $t+1$).[9] In 2011, the corporate income tax in the UK raised £43,763 million, i.e. 10% of total tax revenue and 4% of GDP.[10] Although corporate tax collections are important, the gap between predicted and actual collected tax liabilities is large. The UK tax authority Her Majesty's Revenue and Customs (HMRC) estimates that in 2009 (the most recent tax gap estimates), 9.6% of true corporate tax liabilities remained unpaid (HMRC (2012)).

The UK corporate income tax offers two compelling sources of variation to study the response of investment to tax policy changes. The first source of variation

[9] Author's estimates.
[10] See http://www.hmrc.gov.uk/statistics/receipts/receipts-stats.pdf.

Table 1.I: **UK Corporate Tax Schedule**

Taxable income	2000-2001	2002-2005	2006	2007	2008-2010
0-10k	10	0	19	20	21
10k-50k	22.5	23.75	19	20	21
50k-300k	20	19	19	20	21
300k-1,500k	32.5	32.75	32.75	32.5	29.75
>1,500k	30	30	30	30	28

Notes: The table shows how the marginal corporate tax rate (in percent) has varied across tax brackets (rows) and across fiscal years (columns). The tax is applied to taxable corporate profits, defined as turnover from different income sources, net of recurrent investments, capital depreciation allowances for different of capital items, deductions (e.g. corporate venture scheme relief) and losses carried forward from previous years. For more information on the tax base, see http://www.hmrc.gov.uk/agents/ct/. For more information on the tax rate schedule, see http://www.hmrc.gov.uk/statistics/ct-receipts/table-a6.pdf.

is the tax rate jump (kink) between the first and second brackets in the corporate tax schedule. As Table 1.I shows, the UK corporate tax schedule features five tax brackets, with two large convex kinks, at 10k and 300k of taxable income, and two smaller concave kinks. The kink at 10k is particularly suitable for studying the investment response to taxation. First, in 2002-2005, the 10k kink is larger than any other kink, thus generating strong incentives for bunching.[11] Second, the size of the kink changes over time. In 2001, the kink implies a 12.5 percentage points tax rate change. For 2002-2005, a 0% starting rate is introduced for all firms with a taxable income below 10k, which increases the size of the tax rate jump to 23.75 percentage points and makes the kink more salient. Finally, in 2006, the kink is abolished and the first three tax brackets are merged at a marginal tax rate of 19%.

The second source of variation is the difference in depreciation allowance rates across different capital items, which leads to a differential suitability of capital items for investment bunching. To see why depreciation allowance rates are relevant, consider that the corporate tax base is comprised of turnover net of recurrent investments (salaries, purchases of tradable goods), capital depreciation allowances, deductions (e.g. community investment schemes) and losses carried forward.[12] An

[11] In the presence of optimization frictions, bunching is relatively stronger at larger kinks, as shown in Chetty et al (2011).

[12] Losses thus affect taxable income but are realized in prior periods and cannot be manipulated ex-post to reduce taxable income. Deduction can be used to change taxable income and move to the kink, but my empirical analysis find no evidence of this. Spikes in median deductions or the share of firms claiming deductions at the kink appear only in some subsamples and time periods, and the excess number of firms claiming deductions in the excluded range (deduction bunchers) is not significantly different from 0 when selection effects are controlled for. For more information on how the tax base is derived, see http://www.hmrc.gov.uk/agents/ct/.

Table 1.II: **UK Capital Depreciation Schedule**

	Short-life Machinery	Cars	Long-life Machinery	Industrial Buildings
First Year Allowance	40%	25%/100%	40%	24%
Writing Down Allowance	25%	25%	6%	4%

Notes: The table shows how the capital depreciation schedule varies across capital items, for fiscal years 2001-2007. All capital purchases give rise a to a First Year Allowance (FYA) in the year of purchase, and a Writing Down Allowance (WDA) in the years following the year of purchase. All items are depreciated according to a reducing balance allowance, except for industrial buildings, which are depreciated on a straight line basis. A reducing balance allowance is applied every year to the remaining balance of unclaimed expenditure. Once the balance of unclaimed expenditure is less than £1000, it can be depreciated immediately, under the small pool allowance rule. A straight line allowance means that the same amount of allowance for an expenditure is claimed every year until the total expenditure has been claimed. For more information on capital allowance rates, see http://www.hmrc.gov.uk/statistics/ct-receipts/table-a5.pdf. For more information on capital allowance rules, see http://www.hmrc.gov.uk/manuals/pimmanual/pim3005.htm and http://www.hmrc.gov.uk/capital-allowances/getstarted.htm.

increase in recurrent investments translates one-for-one into a reduction of the tax base. To what extent an increase in capital investments translates into a reduction of the tax base depends on the capital depreciation schedule, summarized in Table 1.II. For each capital item purchased, firms deduct a First Year Allowance (FYA) in the year of purchase, and a Writing Down Allowance (WDA) in all following years. The WDA is applied on a reducing balance basis (straight line basis for buildings), until the cost that remains to be claimed is less than £1000, at which time it is written off.

When considering the use of capital investments for bunching, firms take into account both the FYA and the total depreciation speed of the capital item. The weight given to each of these two determinants depends on the firms' production function and discount factor. Independently of the weights, purchases of short-life machinery (e.g. computers) are most suitable for bunching, as they yield the highest FYA (40%) and highest depreciation speed. As for the other capital items, their ranking in terms of suitability for bunching is ambiguous. For firms valuing the FYA over depreciation speed, purchases of long-life machinery (e.g. electrical systems), which yield a 40% FYA, are most suitable for bunching. For firms valuing depreciation speed over the FYA, investments in cars are most suitable for bunching.

A number of exceptions to these rules are worth mentioning. Investments by small and medium firms in short-life machinery received a 50% FYA in 2004, 2006 and 2007. Furthermore, investments by small firms[13] in information and com-

[13]Small firms are distinct from SMEs (cf. footnote 18) and satisfy at least two of the following

munication technologies (considered short-life machinery) received a 100% FYA in 2000-2003. In addition, a 100% FYA has been in place for specifically designed energy efficient short-life machinery since 1 April 2001 and for low carbon emission cars since 1 April 2002. These exceptions make investments in short-life machinery and cars even more suitable for bunching purposes, compared to investments in other capital assets.

1.3.2 Data

This study is based on the HMRC CT600 panel including the universe of UK corporate tax returns filed for the years 2001-2007.[14] The dataset provides information on taxable income under different headings (foreign, trading, capital etc.), capital allowances, deductions, losses, tax liability and after-tax deductions. The accuracy of the data is extremely high, given its administrative nature and the fact that all returns are electronically checked for consistency. The dataset contains approximately 1.5 million obervations per year. An average of 800,000 firms per year have a taxable income between 3k and 40k, the interval relevant for the investment bunching estimations. Only firms that do not report turnover (10.7% of the sample) are dropped.

The investment variables of interest are recurrent investment and capital allowances for short-life machinery, cars, long-life machinery and buildings. Recurrent investment is constructed as the difference between trading turnover and trading income. To measure capital investments, I use the capital allowance variable rather than the underlying expenditure, as the former precisely captures the investment-induced taxable income reduction.[15] The data does not distinguish between the 100% and 40%/50% FYA capital types within the short-life machinery and cars categories. Furthermore, although capital allowances are intended to cover capital purchases for business use only, it cannot be excluded that some (especially small) firms also claim allowances for personal use items (e.g. laptops).

To examine which types of firms are most likely to engange in investment bunching behavior, I combine the CT600 panel with FAME accounting data, which

requirements: turnover of not more than £2.8 (5.6) million, balance sheet total of not more than £1.4 (2.8) million, number of employees of not more than 50 for FYs ending before (after) 30 January 2004. See http://www.hmrc.gov.uk/manuals/camanual/CA23170.htm.

[14] As the 10k kink disappeared in 2006, and capital depreciation rules changed significantly in 2008, the study does not exploit data for post-2007.

[15] Unfortunately, the expenditure data does not provide information on the timing of capital purchases. Such information would help support the hypothesis of bunching through investment, if bunchers are found to implement a disproportionate share of investment at the end of the accounting period, when they can predict that their gross profit is close to, but above, the kink.

is available for approximately 90% of the CT600 observations. This data is compiled by Bureau van Dijk, based on annual accounts submitted by all tax-registered firms to UK Companies House. Among other things, this data contains information on the company's debt levels and 5-digit SIC industry codes. I aggregate the latter to sectors at the 2-digit level, to achieve subsamples big enough (at least 10% of the full sample) for frequency plots around the 10k kink. I then merge the panel with sector-level data on the share of sales to the final consumer and the median number of employees per firm. These data are compiled by the Office for National Statistics in the Supply and Use Tables and the SME Statistics.[16]

1.4 Empirical Results

This section builds on the finding of a large behavioral response to the 10k kink, as established by Devereux *et al.* (2013) and shown in Figure 1.III. The excess bunching mass of 12.3 translates into a large elasticity of 0.45 with a standard error of 0.016.[17] An examination of the investment level of firms around the kink sheds light on the anatomy of this behavioral response. Section 1.4.1 presents evidence that firms move to the kink by increasing investment. Section 1.4.2, provides estimates of the contribution of this behavior to bunching. Section 1.4.3 discusses evidence supporting the interpretation of the observed behavior as real investment rather than avoidance or evasion through over-reporting.

1.4.1 Evidence of Bunching Through Investment

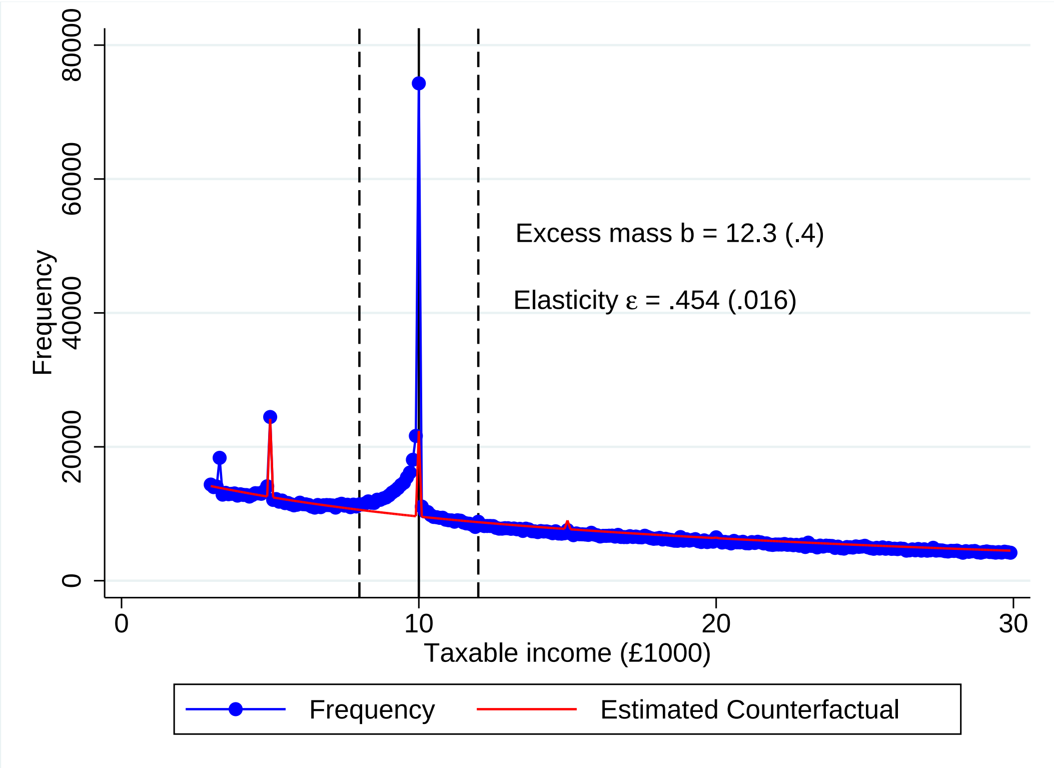
If firms use investments to reduce their taxable income and move to the 10k kink, this should be reflected in a higher share of investors and a higher median capital cost around the kink. Variation in the kink size across time allows us to distinguish the bunching-induced investment changes from any other difference in investment levels between firms around the 10k kink and firms at higher or lower income levels.

---

[16] To avoid selection of subsamples on outcomes, I searched for the earliest available data. However, the median number of employees per firm is only available starting from 2007. The share of sales to the final consumer is for 2000.

[17] In fact, bunching is strongly asymmetric, suggesting that firms misperceive the kink as a notch, i.e. a jump in the average tax rate. A notch generates a strictly dominated area above the tax bracket cutoff, and should thus be accompanied by bunching below the cutoff and a missing mass above the cutoff (see Kleven & Waseem (2013) for evidence from Pakistan). In the UK data, however, there is no missing mass above the cutoff, and the estimates of bunching and investment bunching with an asymmetric excluded range at 10k are almost identical to the results presented here and thus omitted. The asymmetry of bunching does not appear in the figures shown in Devereux *et al.* (2013), which plot the distribution of taxable income in £1000 bins and rely on the subsample of firms with one-year accounting periods.

Figure 1.III: **Bunching at 10k**



Notes: The figure presents the empirical frequency (blue dotted line) and the estimated counterfactual (red solid line) of firms around the 10k kink, for fiscal years 2002-2005. During this period, the marginal tax rate is 0% for incomes below 10k and 23.75% for incomes above 10k. The counterfactual is estimated from equation (1.4), by fitting a flexible polynomial to the observed frequency, excluding a range around the kink. The kink is marked by a vertial solid line; the exluded range is marked by vertical dashed lines. The bin size (£100) and degree of polynomial (5) are chosen to optimize the fit, as in Devereux *et al.* (2013). The estimation allows for round number effects at multiples of 10k up to 20k, and at multiples of 5k, interacted with taxable income. Excess mass b is the excess frequency in the excluded range around the kink, in proportion to the average counterfactual frequency in the excluded range, as estimated following equation (1.6). The elasticity of taxable income ϵ is estimated using equation (1.3). Bootstrapped standard errors are shown in parentheses. Source: HMRC.

Consider first recurrent investments, the largest investment item (on average 97.1% of a firm's costs), and the only one for which an increase translates one-to-one into a reduction in taxable income. Panel A of Figure 1.IV plots the median recurrent cost by taxable income, for 2001, 2002-2005 and 2006-2007. The plots also show the counterfactual median cost, estimated according to equation (1.13), and the size of the spike at 10k. Note that I consider median costs rather than the share of investors, as the latter is not applicable for recurrent investments, for which all firms register a positive amount (unless one of the underlying variables is missing). The spike size is the difference between the observed and estimated median cost in the 10k bin, scaled by the estimated cost. This proxy of investment bunching is distinct from the number of investment bunchers, estimated in section 1.4.2, which focuses on the excess number of investors across the entire excluded range, rather than only at the kink.

As Figure 1.IV shows, there is already a statistically significant spike in recurrent investments at 10k in 2001, when the kink at 10k is present but small. The spike size more than doubles in 2002, when the introduction of the 0-% starting rate increases the size of the tax rate jump and the salience of the kink. In addition to the sharp spike at 10k, the 2002-2005 figure also shows higher cost levels for firms with a taxable income $\in [8k, 10k]$, the income interval below the kink which registers a significant excess mass of firms, as shown in Figure 1.III.[18] Finally, the spike size is reduced to a tenth of its previous size in 2006-2007, when the kink disappears. The spike is still statistically significant, however, which could be explained by misperceptions or adjustment costs that prevent bunchers from lowering their cost immediately after the kink is abolished. Also note that overall investments decrease over time.

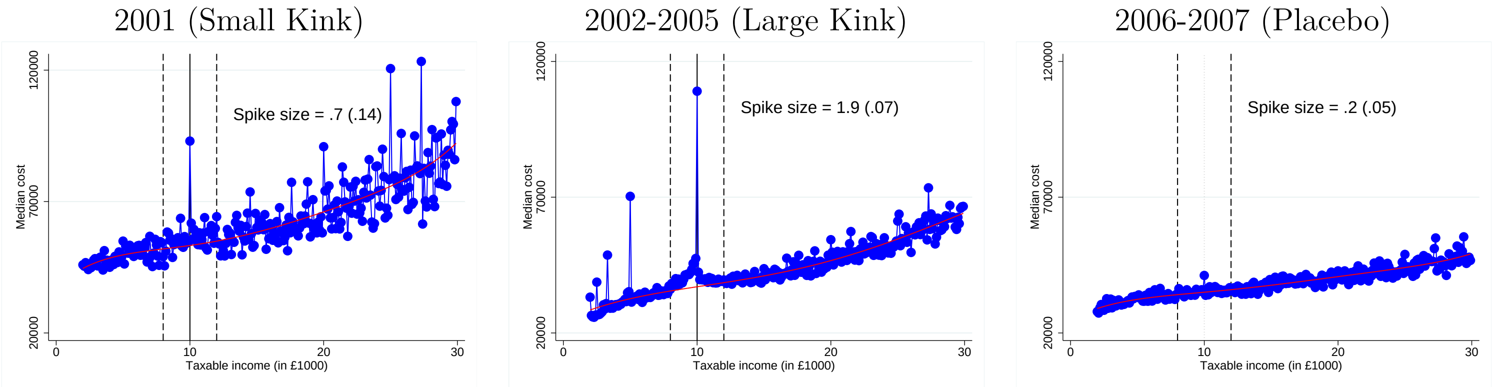
The time pattern for investments in short-life machinery (Panel B), the most quickly depreciating capital item, is very similar to the pattern for recurrent investments. In 2001, the median capital cost, at 10k, is about twice as high as the estimated counterfactual. The spike increases to quadruple the estimated cost in 2002-2005, and falls back to a level only 30% higher than the estimated cost in 2006-2007. Note that, during 2002-2005, the median cost for most income bins below the kink is 0. This is because firms with a taxable income below 10k pay no tax, and thus have no incentive to claim capital allowances, unless their income gross of capital allowances places them above 10k.[19]

[18]See the previous footnote for a discussion of the asymmetric nature of bunching

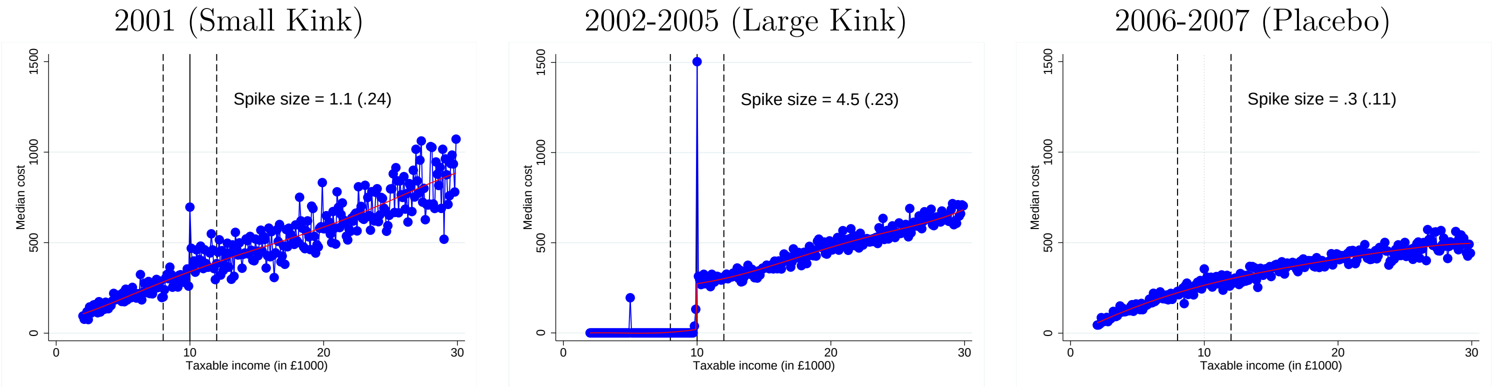
[19]Just as in the frequency plot in Figure 1.III, the cost plots for 2002-2005 display a spike at 5k, although there is no kink at this point. This spike is present even when dropping firms with a taxable income of exactly 5k, suggesting that the excess mass at 5k is at least partly driven by firms that misperceived 5k as a kink and intentially increased their cost to move there.

Figure 1.IV: **Investment Spikes Across Time**

A: Recurrent Investment

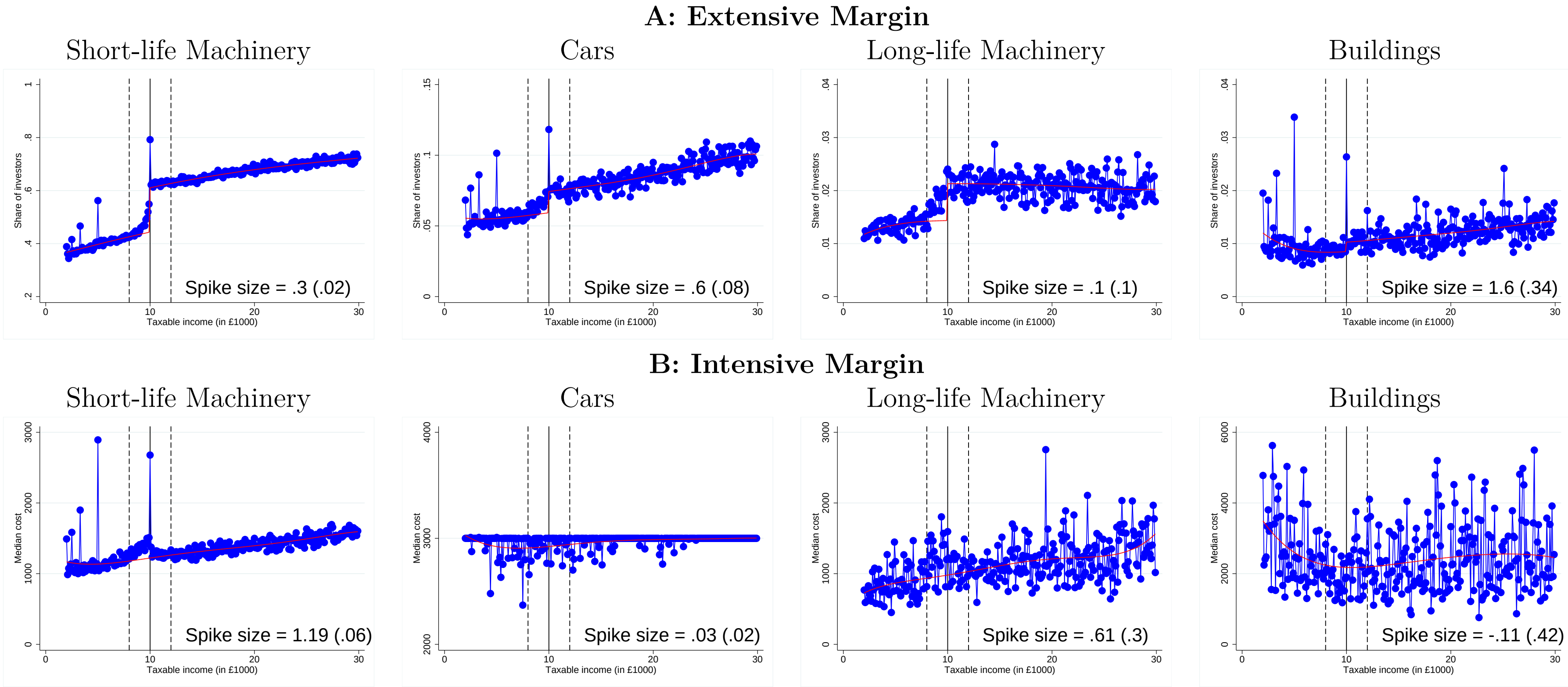


B: Short-life Machinery



Notes: The figure shows plots of the median cost (blue dotted line) by taxable income bins (bin size £100) and the estimated counterfactual median cost (red solid line). Panel A shows median recurrent costs and Panel B shows median costs (capital allowance claims) for short-life machinery, for the years 2001, 2002-2005 and 2006-2007 respectively. In 2001, the 10k kink represented a jump in the marginal tax rate from 10% to 22.5%. In 2002-2005, the jump was from 0% to 23.75%. In 2006-07, there was no kink and the marginal tax rate was 19% for all firms with a taxable income below 50k. The counterfactual median cost is estimated from equation (1.11), by fitting a 5th-order polynomial to the observed median cost, excluding a range around around the kink. For short-life machinery in 2002-2005, the specification allows for different constants above and below the kink. The kink is marked by a vertial solid line (vertical dotted line in 2006-2007, when there is no kink); the excluded range is marked by vertical dashed lines. The spike size is the difference between the observed and counterfactual median cost in the income bin centered at 10k, minus a randomly drawn estimated residual, and scaled by the counterfactual median cost at the kink. Bootstrapped standard errors are shown in parentheses. Source: HMRC.

Figure 1.V: **Investment Spikes Across Capital Items**



Notes: The figure examines investment changes around the 10k kink at the extensive margin (Panel A) and the intensive margin (Panel B). Panel A plots the share of investors (firms with positive capital allowance claims) and Panel B plots the median cost (capital allowance claims for investors, blue dotted line), by taxable income (bin size £100). The plots also show an estimated counterfactual (red solid line). For Panel A, the counterfactual median cost is estimated from equation (1.7), by fitting a 5th-order polynomial to the observed median cost, excluding a range around the kink. The estimation allows for different constants for income bins above and below 10k. The counterfactual for Panel B is estimated in a similar way, using equation (1.11). All figures are for 2002-2005, when the kink represents a tax rate jump from 0% to 23.75%. The kink is marked by a vertial solid line; the excluded range is marked by vertical dashed lines. The spike size is the difference between the observed and counterfactual share of investors (median cost) in the income bin at 10k, minus a randomly drawn estimated residual, and scaled by the counterfactual share of investors (median cost) at the kink. Bootstrapped standard errors are shown in parentheses. Source: HMRC.

As a second way of distinguishing bunching-induced investment spikes from other changes in investment levels around the kink, I exploit variation in the rate of capital depreciation across different capital types. If firms invest in order to bunch, they should primarily invest in quickly depreciating capital items like short-life machinery and cars. Figure 1.V provides evidence for differential investment spikes across capital items, distinguishing between the extensive and the intensive margin (Panels A and B respectively). For short-life machinery, both the share of investors as well as the median cost for investors display a large and statistically significant spike at the kink. For cars, there is a significant spike at the extensive margin only. Intensive margin investment bunching is inhibited by the cap on capital allowance for cars at £3000 per year, which means that the median cost is almost constant at £3000. For long-life machinery and buildings, both the share of investing firms as well as the median cost for investors are much more noisy and the pattern of statistical significance of investment spikes is inconsistent. Although the intensive margin spike for long-life machinery and the extensive margin spike for buildings are statistically significant, they are smaller than other spikes at random levels of taxable income. This might suggest that the empirical methodology used in these distributions is not robust enough to accurately determine their statistical significance. Overall, the pattern of significance of investment spikes over time and across capital types supports the hypothesis that firms bunch by investing.

1.4.2 Estimating the Contribution of Investment Bunching

To estimate the contribution of investment bunching, I calculate the number of investment bunchers, as defined in section 1.2.2, and their contribution $Cont$ to the total number of bunchers. Figures 1.VI and 1.VII show these estimations for short-life machinery, at the extensive and intensive margins respectively. The results for all other capital items are presented in Table 1.III. For recurrent investments, only the intensive margin is applicable. For cars, long-life machinery and buildings, I conduct the estimation for the extensive margin only. This is because the sample of investors for these capital items is too small (less than 15% of the full sample) and the median cost plots are too noisy to credibly estimate intensive margin investment bunching. Note that all contribution estimates are for the period 2002-2005, when the kink is most salient and represents a tax rate jump of 23.75 percentage points.

In the first place, consider the baseline results for extensive investment bunching as shown in Panel A1 of Figure 1.VI. The estimation suggests that investment bunchers represent 19.2% of all bunchers. This result relies on the assumption that the share of investors would evolve smoothly as a function of taxable income if there were no kink, i.e. that the spike in investment at the kink is entirely due to investment bunching behavior. To test this assumption, let $s_{j,t}$ denote the share of investors in income bin j in year t and $s_{j,t-2}$ the share of investors among the *same* group of firms in year $t-2$.[20] Unless bunchers stay at the kink for three consecutive years, there is no reason to expect $s_{j,t-2}$ to spike at the kink. The year $t-2$ is chosen because it is close enough to serve as a reasonable control, but further in the past than year $t-1$, in which the share of investors might be affected by auto-correlation in bunching. Indeed, approximately one-third of all bunchers stay at the kink for more than one year, but only one-sixth of them stay for more than two years. [21]

The joint plot of $s_{j,t}$ and $s_{j,t-2}$ in Panel A2 shows that the share of investors, two years prior to bunching at the kink, also spikes at the kink. This suggests that bunchers might be a selected sample among the firms above the kink. Mechanically, firms with a high investment level (or low profit margin) will require a smaller percentage change in investment to achieve a certain percentage change in profits, to move to the kink. Another possible mechanism to explain the selection effect is adjustment costs. Consider that only firms with sufficiently low adjustment costs move to the kink (e.g. those with good tax advisors), and that this feature is positively correlated with investment propensity. In this case, we would observe an investment spike at the kink, even if bunchers did not change their investment behavior to move to the kink.

Independently of the mechanism driving it, the selection effect should be controlled for in the estimation. For Panel A2, I estimate the number of extensive investment bunchers by subtracting from the previous estimate in equation (1.8) the control difference between the observed and counterfactual share of investors in $t-2$:

$$M_C^E = \sum_{k=\pi_L}^{\pi_U} [(s_{j,t} - \hat{s}_{j,t}) - (s_{j,t-2} - \hat{s}_{j,t-2})] \cdot f_{k,t}. \tag{1.15}$$

Subscript C marks all estimates that control for selection effects. The contribution of extensive investment bunching is defined as

$$Cont_C^E = \frac{M_C^E}{B}. \tag{1.16}$$

Controlling for the selection effect makes the contribution estimate drop to 7.7%. However, this estimate is possibly downward biased because some firms stay at the kink for more than two consecutive years, so that $s_{j,t-2}$ may be affected by investment bunching in $t-2$. I therefore consider the baseline estimate an upper bound and the control estimate a lower bound on the true contribution of investment bunching.

[20] The "full" sample thus considers all firms that report at least three consecutive observations.

[21] The share of investors in $t+1$ is not a good counterfactual, since investments are measured by capital depreciation allowances, so that investment bunchers in t will also be observed as investors in $t+1$, when they continue to depreciate their capital purchase from year t.

To verify the robustness of these results, I implement a second estimation for the sample of one-time bunchers only (Panel B of Figure 1.VI). This allows me to use as counterfactual the share of investors in $t-1$, which should be unaffected by past investment bunching behavior, because I eliminate firms with a taxable income in $[8k, 12k]$ for at least two consecutive years (approximately 13% of the sample). The baseline estimation for this sample (Panel B1) suggests that investment-bunchers represent 37% of all bunchers. This estimate drops to 24.9% only when the selection effect is controlled for (Panel B2). Obviously, these results apply only to one-time bunchers, which might rely on "crude" strategies like investment bunching to a larger extent than more sophsticated repeat bunchers.[22] However, the results for this sample confirm that the investment bunching contribution remains significant even when the selection effect is controlled for.

As Figure 1.VII shows, the estimates for the contribution of intensive investment bunching display a qualitatively similar variation across estimation strategies, but are overall smaller (though statistically significant).[23] In the full sample, intensive investment bunchers represent between 16.8% (baseline estimation) and 4.3% (control estimation) of the total number of bunchers. In the sample of one-time bunchers, the estimates are again slightly higher, ranging from 20.3% (baseline) to 10.2% (control).

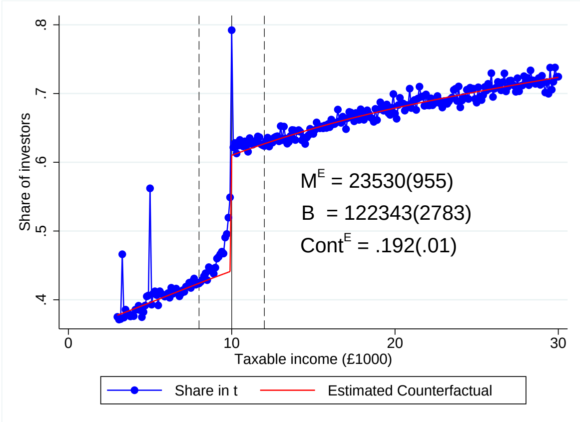
[22] Also notice that the estimates are potentially upward biased because $s_{j,t-1}$ changes discontinuously at $12k$, the upper bound of the excluded range. This is because restricting the sample to one-time bunchers means eliminating a disproportionately large number of observations in the $[8k, 12k]$ interval, compared to bins outside the bunching interval, and the remaining observations have a different investment level than the full sample.

[23] To see how I control for the selection effect in the intensive margin estimation, note that the number of firms in the bunching range is different for period t and the control $t-x$, $x \in \{1, 2\}$. Therefore, contrary to section 1.2.2.2, I estimate not the number of investment bunchers but rather the *share* of investment bunchers among the firms in the bunching range, m_t^I and m_{t-x}^I: $m_t^I = 1/F_{C,t} \cdot \left\{ \sum_{k=\pi_L}^{\pi_U} \sum_{i \in j} \mathbf{1}[\pi_j = k] \cdot \mathbf{1}[c_{ij,t} \geq \hat{C_{j,t}}] - \frac{1}{2} F_{C,t} \right\}$, and accordingly for m_{t-x}^I, with $F_{C,t} = \sum_{k=\pi_L}^{\pi_U} f_{k,t}$, the total number observations within the bunching range in year t, in the relevant sample. The contribution of intensive investment bunching is defined as $Cont_C^I = \frac{(m_t^I - m_{t-x}^I) \cdot F_{C,t}}{B_{C,t}}$.

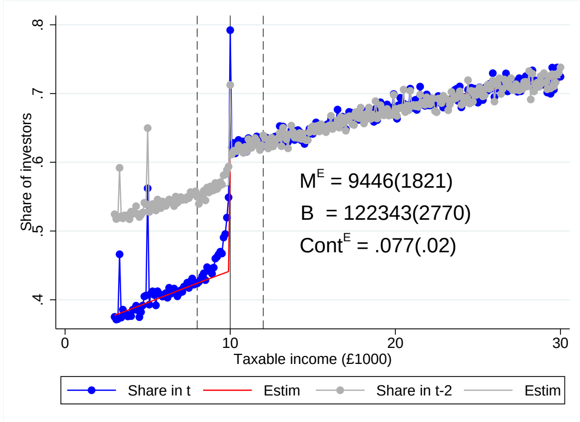
Figure 1.VI: **Extensive Investment Bunching**

A: Full Sample

A1: Baseline

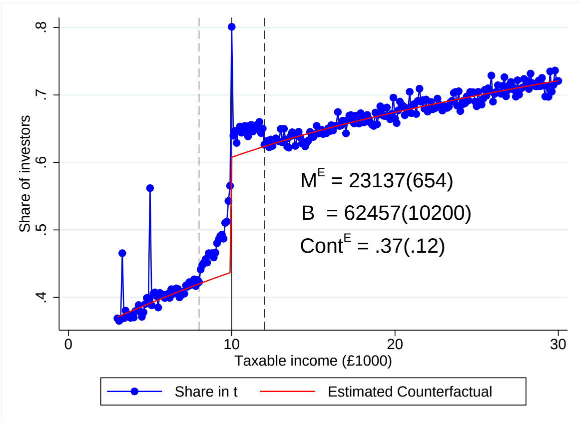


A2: With Control

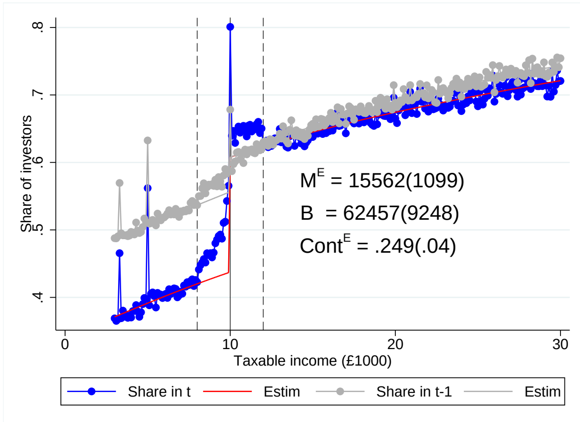


B: One-time Bunchers

B1: Baseline

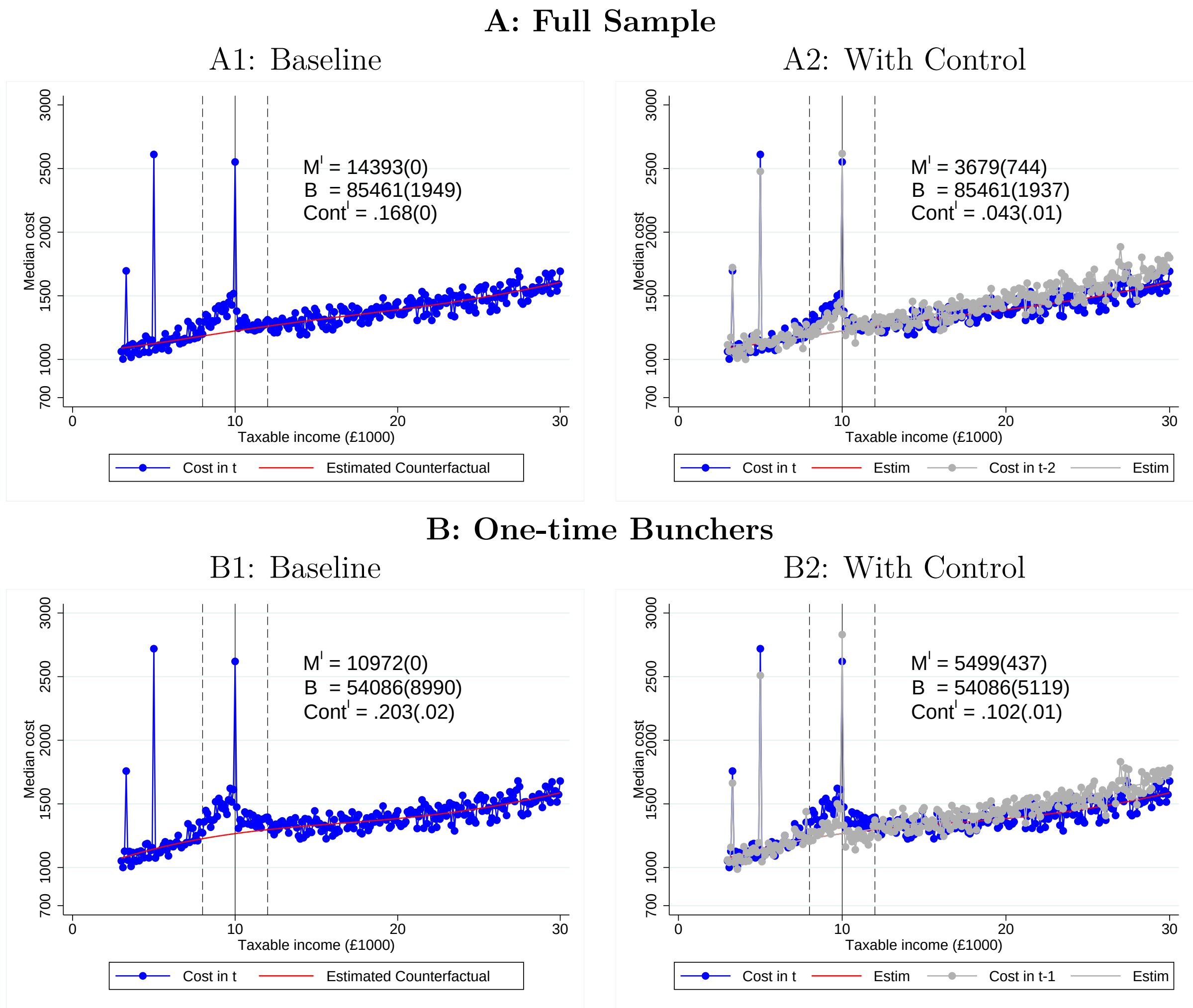


B2: With Control



Notes: The figure displays results for estimating the contribution of investment in short-life machinery to bunching, focusing on extensive margin changes in investment. The data are for 2002-2005, when the 10k kink represents a tax rate jump from 0% to 23.75%. The kink is marked by a vertial solid line; the excluded range is marked by vertical dashed lines. In the full sample (Panel A), I use the share of investors in $t-2$ to control for selection in the group of bunchers. Panel A1 plots the share of investors (firms with positive capital allowance claims for short-life machinery) in year t by income bins in year t (blue dotted line, bin size £100), and its estimated counterfactual (red solid line). The counterfactual is estimated by fitting a flexible polynomial to the observed share, excluding a range around the kink, as shown in equation (1.7). Panel A2 is identical to Panel A1, and also plots the control share of investors in $t-2$ by income bins in t (grey dotted line), and its estimated counterfactual (grey solid line). The number of extensive margin *investment bunchers* (excess investors) M^E is the difference between the observed and counterfactual number of investors in the excluded range for Panel A1 (equation (1.8)), and the difference between the observed and the control counterfactual number of investors in the excluded range for Panel A2 (equation (1.15)). The number of bunchers B is estimated as shown in Figure 1.III, by fitting a flexible polynomial to the observed frequency distribution of firms. B is the difference between the observed and estimated frequency, in the excluded range, as in equation (1.5). The contribution of extensive margin investment to bunching is $Cont^E = M^E/B$. Bootstrapped standard errors are shown in parentheses. In the sample of one-time bunchers (Panel B), in which I eliminate all firms with a taxable income $\in [8k, 12k]$ for two consecutive years, I use the share of investors in $t-1$, by income bins in t, to control for selection in the group of bunchers. Everything else is as in Panel A. Source: HMRC.

Figure 1.VII: **Intensive Investment Bunching**



Notes: The figure displays results for estimating the contribution of investment in short-life machinery to bunching, focusing on intensive margin changes in investment. The data are for 2002-2005, when the 10k kink represents a tax rate jump from 0% to 23.75%. The kink is marked by a vertical solid line; the excluded range is marked by vertical dashed lines. In the full sample (Panel A), I use the median cost in $t-2$ to control for selection in the group of bunchers. Panel A1 plots the median cost for investors (firms with positive capital allowance claims for short-life machinery) in year t, by income bins in year t (blue dotted line, bin size £100), and its estimated counterfactual (red solid line). The counterfactual is estimated by a fitting a flexible polynomial to the observed median cost, excluding a range around the kink, as shown in equation (1.11). Panel A2 is identical to Panel A1, and also plots the control median cost in $t-2$ by income bins in t (grey dotted line), and its estimated counterfactual (grey solid line). The number of intensive margin *investment bunchers* (excess investors) M^I is excess number of firms with a cost (capital allowance claim) above the counterfactual median in the excluded range for Panel A1 (equation (1.12)), and the excess number of firms with a cost above the control counterfactual median in the excluded range for Panel A2 (see footnote 21). The number of bunchers B is estimated as shown in Figure 1.III, by fitting a flexible polynomial to the observed frequency distribution of firms. B is the difference between the observed and estimated frequency, in the excluded range, as in equation (1.5). The contribution of intensive margin investment to bunching is $Cont^I = M^I/B$. Bootstrapped standard errors are shown in parentheses. In the sample of one-time bunchers (Panel B), in which I eliminate all firms with a taxable income $\in [8k, 12k]$ for two consecutive years, I use the share of investors in $t-1$, by income bins in t, to control for selection in the group of bunchers. Everything else is as in Panel A. Source: HMRC.

Table 1.III: **Estimating Investment Bunching**

	Full Sample				One-time Bunchers			
			Baseline	With Control			Baseline	With Control
Investment item	N	Bunchers B	*Cont*	*Cont*	N	Bunchers B	*Cont*	*Cont*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Recurrent Investments	2507423	120689	.238	.018	2182426	55127	.458	-.062
		(2714)	(.01)	(.01)		(11606)	(.06)	(.01)
Short-life machinery I	1479976	85461	.168	.043	1295667	54086	.203	.102
		(1949)	(.00)	(.01)		(8990)	(.02)	(.01)
Short-life machinery E	2558590	122343	.192	.077	2226960	62457	.37	.249
		(2783)	(.01)	(.02)		(10200)	(.12)	(.04)
Cars	2558590	122343	.046	-.002	2226960	62457	.08	-.009
		(3023)	(.0049)	(.009)		(7237)	(.012)	(.012)
Long-life machinery	2558590	122343	.01	.0026	2226624	63823	.013	.007
		(2824)	(.0021)	(.003)		(7965)	(.005)	(.004)
Buildings	2558590	122343	.013	.0001	2226624	63823	.019	-.001
		(2695)	(.0026)	(.005)		(14665)	(.005)	(.006)

Notes: The table presents the results of estimating the contribution of investment changes to bunching at the 10k kink, for different investment items (rows). The estimations are for 2002-2005, when the 10k kink represents a tax rate jump from 0% to 23.75%. The results for recurrent investments and for short-life machinery (I) are for intensive margin investment changes. All other results are for the extensive margin. Columns (1)-(4) present the results for the full sample of firms for which at least three consecutive observations exist. The control strategy for this sample uses the share of investors/median cost in $t-2$ to control for selection in the group of bunchers. Columns (5)-(8) present the results for the sample of one-time bunchers, eliminating all firms with taxable income $\in [8k, 12k]$ for two consecutive years. The control strategy for this sample uses the share of investors/median cost in $t-1$ to control for selection in the group of bunchers. Columns (1) and (5) present the sample size, columns (2) and (6) present the number of bunchers B, and columns (3)-(4) and (7)-(8) present the contribution *Cont* of investment to bunching, with and without controlling for selection effects. The number of bunchers B is estimated as shown in Figure 1.III, by fitting a flexible polynomial to the observed frequency distribution of firms. B is the difference between the observed and estimated frequency, in an excluded range around the kink, as in equation (1.5). The estimation of the contribution of investment bunching *Cont* is as shown in Figures 1.VI and 1.VII. I calculate the investment bunchers M^E (M^I) as the excess number of investors, i.e. the difference between the observed and counterfactual share of investors (median cost) in the excluded range, according to equation (1.8) (equation (1.12)). For the specification with control, I compare the observed share/median cost to the control counterfactual, according to equation (1.15) (footnote 21). The contribution is estimated as $Cont = M/B$. Bootstrapped standard errors are shown in parentheses. Source: HMRC.

Table 1.III shows the estimation results for all investment items. Most contribution estimates are statistically significant only without the selection effect control. The contribution estimates are largest for recurrent investments (45%) and lowest for long-life machinery and buildings (below 2%). With the control, however, all estimates lose significance in at least one of the samples. The control estimates are significant for recurrent investments in the full sample, and for long-life machinery in the sample of one-time bunchers. However, these estimates are statistically significant only at the 10% level, and economically insignificant, as they indicate an investment bunching contribution of below 2%.

Emphasizing the more conservative results for the full sample, I conclude that investments in short-life machinery are an instrument for bunching for a moderate fraction of up to 19.2% of bunchers, and that extensive investments play a larger role than intensive investments. The results do not provide evidence for investment bunching through items other than short-life machinery. It is possible, however, that firms use large-scale changes in capital investments to approach the kink, and fine-tune their bunching behavior with recurrent investments. The latter may be too small to be picked up by the estimation strategy used.

1.4.3 Distinguishing Avoidance, Evasion and Investment

The empirical analysis has so far been agnostic about whether the observed investment spikes at the kink constitute avoidance (e.g. transfer pricing manipulation), evasion (e.g. over-reporting), or real investment. While it is not possible to cleanly decompose the investment spikes into these components, this section discusses preliminary evidence suggesting that the spikes mainly represent a real investment increase. The argument is threefold.

First, it is unlikely that the firms considered in this study engage in tax avoidance strategies such as cost manipulation through transfer pricing or shifting income to international tax havens. The firms in the sample are small and/or relatively unprofitable. Only 0.12% of them register any overseas income, 0.20% register double taxation relief and 6.1% are member of a group.[24] This means that the scope for engaging in tax avoidance schemes is very limited, as most of these schemes rely on international business activity or group membership.

Second, an examination of the context suggests that evasion is unlikely to explain a large part of bunching. Tax compliance in the UK is relatively high compared to many other countries, and tax-registered corporations need to publish

[24] Author's calculation.

accounts and emit VAT receipts to their clients. This means that over-reporting investment would require sophisticated disclosure strategies. Moreover, although the audit probability for small firms is negligible, the cost of an audit is high, and may include the cost of cooperating with auditors, potential penalties, and reputational damage.

To test empirically for the presence of evasion, I examine bunching behavior in subsamples of firms with high *ease of evasion*. If a large part of investment and thus bunching is driven by evasion (over-reporting), we should expect firms with better evasion opportunities to respond more strongly to the kink. Based on previous findings in the tax evasion literature, I consider firm size, the need for financial intermediation and the share of sales to the final consumer as key determinants of the ease of evasion. Kleven *et al.* (2009) develop an agency model in which firms with a large number of employees find it more difficult to sustain a collusion agreement on evasion.[25] Turnover is another measure of firm size and often used by tax inspectors to determine the level of enforcement (accounting requirements, audit frequency), as discussed in Almunia & Lopez-Rodriguez (2012). Gordon & Li (2009) provide a model in which firms that rely on formal credit are more tax compliant, as financial sector transactions are observable to the government.[26] Finally, firms find it easier to evade sales to the final consumer, which are not covered by the VAT paper trail, rather than to evade sales to other (VAT compliant) firms.[27]

Panel A of Table 1.IV displays excess mass and elasticity estimates for the following subsamples: firms with low (below median) turnover, low number of employees, no debt (dummy)[28], and high (above median) share of sales to the final consumer, as compared to the full sample. Although theory predicts that firms in these subsamples can evade taxes with relative ease, the bunching estimates for firms with low turnover and number of employees are significantly lower than those for the full sample. The estimates for firms without debt are not significantly different from the full sample. This provides evidence against the hypothesis that most bunchers move to the kink through over-reporting of costs.[29] However, the fact that firms with a high share of sales to the final consumer bunch significantly more than the full sample suggests that part of the response to the kink might be driven by output evasion, which this group of firms can practice with relative ease.

---

[25]Kumler *et al.* (2013) find evidence from Mexico that supports this theory.

[26]This argument is consistent with cross-country evidence by Bachas & Jensen (2013).

[27] Pomeranz (2013) supports this argument with experimental evidence from Chile.

[28]Firms for which the FAME data does not report any long-term or short-term loans or debt are coded as having no debt.

[29]The estimates for the contribution *Cont* of investment to bunching are more noisy due to the smaller sample size and thus omitted. Qualitatively, however, the size of these estimates evolves in similar ways as the excess mass estimates when comparing subsamples to the full sample.

If avoidance and evasion are not the key drivers of bunching, the investment spikes must constitute real investment. To support this argument, Panel B of Table 1.IV displays bunching estimates for subsamples of firms with different degrees of *investment propensity*, as compared to the full sample. The evidence is consistent with the hypothesis of bunching through investment. The smallest excess mass is registered by the financial sector, which works with financial rather than physical capital and thus has less opportunity for bunching through real investment. The capital intensive manufacturing sector and the stock-intensive retail sector, on the other hand, register a significantly higher excess mass at the kink than the full sample does. The excess mass is also significantly higher for firms with a high (above median) cost margin, high short-life machinery capital investments, high recurrent investments and for growing firms (though the latter difference is significant only at the 10% level). It should also be mentioned that there is no evidence for inter-temporal shifting in investments, so that the investment spikes can be interpreted as an overall increase in investment, rather than a reallocation of investment across years.[30]

Although care was taken to avoid selection on outcomes (e.g. conditioning subsamples on observables in $t-1$), the results of this heterogeneity study obviously rely on correlations and cannot be interpreted as causal evidence. Besides, even though heterogeneity across subsamples is consistent with bunching through investment, it is not clear whether these investments are for business purposes or personal use items. For instance, a plumbing business with a taxable income of 12k may move to the kink by purchasing a computer, which could be used both for bookkeeping purposes and for leisure activities. This interpretation of bunching-induced investments as private consumption is supported by the fact that bunching-induced investments do not seem to stimulate growth. Investment bunchers (and bunchers in general) do not grow faster in the years after they bunch, compared to firms below or above the bunching range.[31]

1.5 Conclusion

This paper has presented a novel empirical strategy for examining the investment effect of tax policy changes. I exploit bunching of firms at a tax kink to examine how a tax rate change affects firms' investment decisions, and how this effect interacts

[30] The $t+1$ share of investors (median cost), by taxable income bins in year t, does not exhibit a downward spike, as would be expected if firms shift investments from year $t+1$ to t. Besides, Devereux *et al.* (2013) also fail to find evidence of inter-temporal profit shifting.

[31] I test for this by plotting the growth rate in $t+1/2/3$ by taxable income bins in year t. I find that firms located at or around the kink do not register higher growth rates than firms above or below the bunching range.

Table 1.IV: **Heterogeneity in Bunching**

Panel A: Ease of Evasion Indicators

Sample	Share	b	ϵ
	(1)	(2)	(3)
Full	1	12.3	.45
		(.4)	(.016)
Low turnover	.50	8.3	.30
		(.2)	(.009)
Low number of employees	.46	10.2	.38
		(.4)	(.013)
No debt	.92	12.1	.45
		(.4)	(.014)
High share of sales to final consumer	.45	14.5	.53
		(.5)	(.019)

Panel B: Investment Propensity Indicators

Sample	Share	b	ϵ
	(1)	(2)	(3)
Manufacturing sector	.12	14.4	.53
		(.5)	(.018)
Retail sector	.12	14.4	.53
		(.7)	(.025)
Service sector	.17	13.1	.48
		(.5)	(.019)
Financial sector	.48	10.2	.38
		(.4)	(.014)
Growing	.43	12.8	.48
		(.5)	(.018)
Low profit margin in $t-1$	.50	16	.59
		(.7)	(.027)
High capital investments in $t-1$	.47	17.7	.65
		(.7)	(.026)
High recurrent investments in $t-1$	.44	15.7	.58
		(.7)	(.028)

Notes: The table shows estimates of the excess mass b at the 10k kink (column (2)) and the elasticity of taxable income ϵ (3), for different subsamples of firms. Column (1) shows the share of the subsample within the sample of firms with a taxable income $\in [3k, 40k]$. The estimations are for 2002-2005, when the 10k kink represents a tax rate jump from 0% to 23.75%. "Low" and "high" refers to firms below and above the sample median on the specified observable characteristics. Firms for which the FAME data does not report any long-term or short-term loans or debt are coded as having no debt (dummy). The sector categorization is based on 2-digit SIC industry codes. Growing firms are those with an above-median growth rate in turnover between $t-1$ and t. Recurrent and capital investments (for short-life machinery) are scaled by turnover. The excess mass b is the excess frequency in the excluded range around the kink, in proportion to the average counterfactual frequency in the excluded range, as estimated following equation (1.6). The details of this estimation are explained in the footnotes to Figure 1.III. The elasticity of taxable income ϵ is estimated using equation (1.3). The estimation of the contribution of investment bunching $Cont$ is calculated as shown in Panel B2 of Figure 1.VI, for the preferred specification (extensive margin, control specification II). The contribution is $Cont = M_c^E/B$, where B is the number of bunchers and $M_C^{\mathrm{E}}$the number of investment bunchers, i.e. the difference between the observed and control counterfactual share of investors in the excluded range, according to equation (1.15). Bootstrapped standard errors are shown in parentheses. Source: HMRC.

with variation in capital depreciation rates across capital types. Building on the bunching approach proposed by Saez (2010), I estimate the frequency distribution of firms around the kink, and the number of firms with excess investments at the intensive and extensive margin. This allows me to calculate the share bunchers that changed their investment levels to move to the kink. This empirical approach is widely applicable to the study of the underlying drivers of taxable income responses to kinked choice sets.

Using tax return data for UK corporations, I provide the first micro-evidence that investment by small firms significantly responds to a tax rate change. Bunching at the 10k kink is accompanied by spikes in the share of capital investors and the median capital cost at the kink. These spikes are larger in 2002-2005 when the kink is larger, and for quickly depreciating capital items, which yield larger tax reductions. I estimate that extensive margin investments explain 7.7-19.2% of bunching and intensive margin investments explain 4.3-16.8% of bunching. The fact that the response to the kink is stronger among firms with a high investment propensity and weaker among firms with a high ease of evasion supports the interpretation of the investment spikes as real investment response rather than evasion through over-reporting.

The findings imply that the introduction of the zero starting rate for UK firms with a taxable income below 10k not only encouraged incorporation by firms below the kink, as examined by Devereux & Liu (2013), but also investment by firms just above the kink. This evidence of locally important investment effects suggests that tax rate changes can be used to stimulate investment. The analysis also confirms that capital depreciation rates matter for the composition of investment, as firms only use investment in short-lived capital items in response to tax rate changes. However, bunching through investment appears to be a small firms phenomenon. The 300k kink in the UK tax schedule also features a significant bunching mass but no investment spikes.

More work remains to be done to understand whether tax-induced investments are productive, or whether small firms exploit the low levels of tax enforcement they face to claim capital allowances for the purchase of private consumption goods. Work in this area would help shed light on the long-run growth effects of tax-induced investments. I also aim to extend the empirical framework to derive an estimate of the elasticity of investment with respect to the tax rate. Moreover, while the proposed empirical approach takes a step towards decomposing taxable income responses, it leaves a large part of the bunching mass at the kink unexplained. The approach should be further developed to identify evasion and avoidance, and explain the share of bunching that is not driven by investment.[32]

[32] I aim to provide some evidence on evasion behavior in the future, by combining the tax return data with administrative audit and penalty reports. HMRC is in the process of preparing these data for disclosure.

Chapter 2

Production vs. Revenue Efficiency With Limited Tax Capacity: Theory and Evidence from Pakistan

2.1 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 2005a; 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 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

[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.

[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

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 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.

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 & Sorensen 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.

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 2005a; 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 the credibility and precision of these estimates are questionable. Our paper contributes a novel methodology for the estimation of evasion using quasi-experimental

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

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 2.2 presents our conceptual framework, which is used in section 2.3 to develop an empirical methodology based on minimum tax schemes. Section 2.4 describes the context and data and section 2.5 presents our results. Section 2.6 numerically analyzes optimal policy, and section 2.7 concludes.

2.2 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.

[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.

2.2.1 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

$$\Pi(y) = (1-\tau)y - c(y) + \tau\mu c(y). \tag{2.1}$$

The profit-maximizing output level solves

$$c'(y) = 1 - \tau\frac{1-\mu}{1-\tau\mu} \equiv 1 - \omega, \tag{2.2}$$

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

$$W = \Pi(y) + \lambda\{T[y, c(y)] - R\}, \tag{2.3}$$

where the firm's output choice satisfies (2.2) 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 (2.1) and (2.3), 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

$$\frac{\partial W}{\partial \tau}/\lambda = M_\tau + \omega \frac{\partial y}{\partial \tau}, \quad (2.4)$$

$$\frac{\partial W}{\partial \mu}/\lambda = M_\mu + \omega \frac{\partial y}{\partial \mu}. \quad (2.5)$$

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. □

2.2.2 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,

$$\Pi(y,\hat{c}) = \hat{\Pi}(y,\hat{c}) - g(\hat{c} - c(y)) = (1-\tau)\,y - c(y) + \tau\mu\hat{c} - g(\hat{c} - c(y)), \quad (2.6)$$

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

$$c'(y) = 1 - \omega, \quad (2.7)$$

$$g'(\hat{c} - c(y)) = \tau\mu. \quad (2.8)$$

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

[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).

[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.

$\Pi(\cdot) = \hat{\Pi}(\cdot) - g(\cdot)$ 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

$$\frac{\partial W}{\partial \tau}/\lambda = M_\tau + \omega\frac{\partial y}{\partial \tau} - \tau\mu\frac{\partial(\hat{c}-c)}{\partial \tau}, \tag{2.9}$$

$$\frac{\partial W}{\partial \mu}/\lambda = M_\mu + \omega\frac{\partial y}{\partial \mu} - \tau\mu\frac{\partial(\hat{c}-c)}{\partial \mu}. \tag{2.10}$$

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*

$$\frac{\tau}{1-\tau} \cdot \frac{\partial \omega}{\partial \tau}(\mu) = G(\mu) \cdot \frac{\varepsilon_{\hat{c}-c}}{\varepsilon_y}, \tag{2.11}$$

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 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

[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.

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,

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

Rewriting this in terms of elasticities, we find

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

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 (2.11), 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 $G\left(\mu\right)$. 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 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 $G(\mu)$

[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.

and increasing the production wedge until formula (2.11) 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 (2.11) also identifies sufficient statistics for determining the optimal tax base and rate in our stylized framework, which we will study empirically.[12]

2.3 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.

2.3.1 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,

$$T[y, c(y)] = \max\{\tau_\pi [y - c(y)], \tau_y y\}. \tag{2.12}$$

[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 (2.11) 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.

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

$$\tau_\pi [y - c(y)] = \tau_y y \quad \Leftrightarrow \quad p \equiv \frac{y - c(y)}{y} = \frac{\tau_y}{\tau_\pi}. \tag{2.13}$$

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 (2.12) 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.

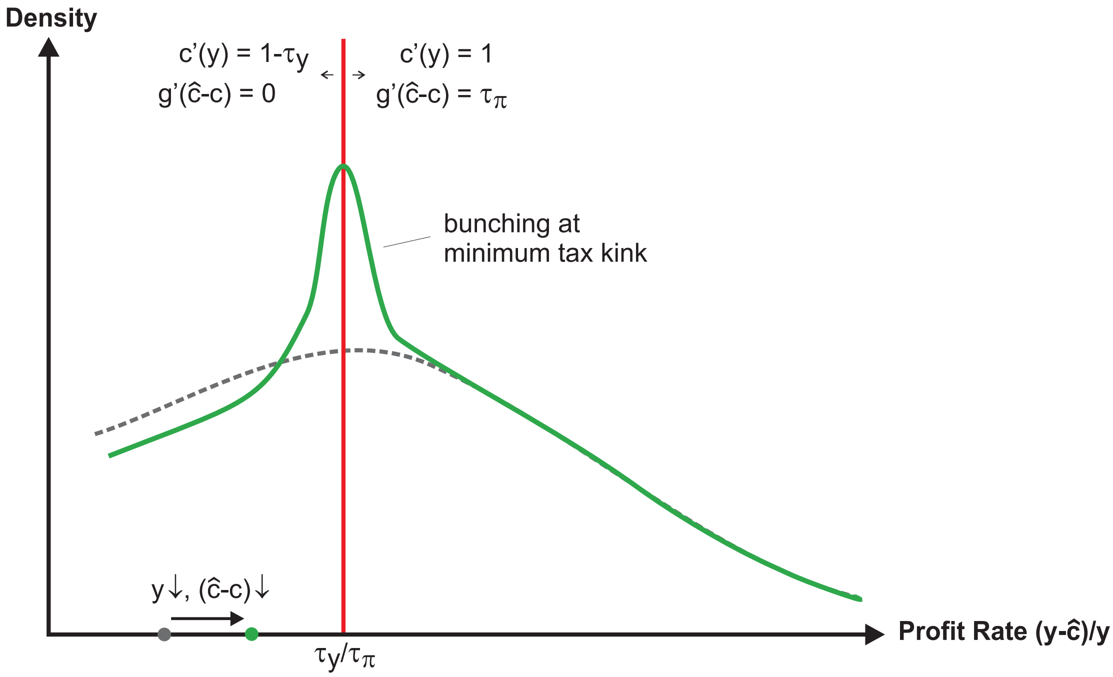
Figure 2.I illustrates how the minimum tax kink at τ_y/τ_π creates bunching in the distribution of profit rates. The figure is based on the general model with both real responses and evasion responses, but we begin by abstracting from evasion responses. The dashed line represents the distribution of 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 2.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

[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.

Figure 2.I: **Empirical Methodology**



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 2.2.2, 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.

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

$$\Delta p = \left[\frac{c}{y} - c'(y)\right]\frac{dy}{y} \simeq \frac{\tau_y^2}{\tau_\pi}\varepsilon_y, \tag{2.14}$$

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 (2.14), 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.

2.3.2 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

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

[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.

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 2.I.

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

$$\Delta\hat{p} = \left[\frac{\hat{c}}{y} - c'(y)\right]\frac{dy}{y} - \frac{d(\hat{c} - c)}{y} \simeq \frac{\tau_y^2}{\tau_\pi}\varepsilon_y - \frac{d(\hat{c} - c)}{y}, \tag{2.15}$$

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 (2.15) 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 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]

2.3.3 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

[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.

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 (2.15) as follows

$$\Delta \hat{p} = \Delta p - d\left(\frac{\hat{c} - c}{y}\right), \tag{2.16}$$

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 effective marginal tax rate under the profit tax were positive (rather than zero), our estimate of the evasion response based on the decomposition in (2.15) 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 (2.15) becomes

$$\begin{aligned} \Delta\hat{p} &= \left[\frac{\hat{c}}{\hat{y}} - c'\left(y\right)\right]\frac{dy}{\hat{y}} - \frac{d\left(\hat{c} - c\right)}{\hat{y}} - \frac{\hat{c}}{\hat{y}}\frac{d\left(y - \hat{y}\right)}{\hat{y}} \\ &\simeq \frac{\tau_y^2}{\tau_\pi}\varepsilon_y\frac{y}{\hat{y}} - \frac{d\left(\hat{c} - c\right)}{\hat{y}} - \left(1 - \frac{\tau_y}{\tau_\pi}\right)\frac{d(y - \hat{y})}{\hat{y}} \end{aligned} \tag{2.17}$$

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

[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.

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 (2.15) is

$$\begin{aligned}\Delta\hat{p} &= \left[\frac{\hat{c}}{y}(1-\sigma) - c'(y)\right]\frac{dy}{\rho(y)\,y} - \frac{d(\hat{c}-c)}{\rho(y)\,y} \\ &\simeq (1-\sigma)\frac{\tau_y^2}{\tau_\pi}\varepsilon_y - \frac{d(\hat{c}-c)}{\rho(y)\,y} \qquad (2.18)\end{aligned}$$

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 (2.15) a lower bound.

[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.

2.4 Context and data

2.4.1 Corporate Taxation: Minimum Tax Regime

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 20,000 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 2.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

[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.

Table 2.I: **Pakistan Corporate 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.

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.

2.4.2 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 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

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

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.

2.5 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 2.3, 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.

2.5.1 Bunching at Minimum Tax Kinks

As shown in section 2.3, 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 2.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.

[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.

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 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]

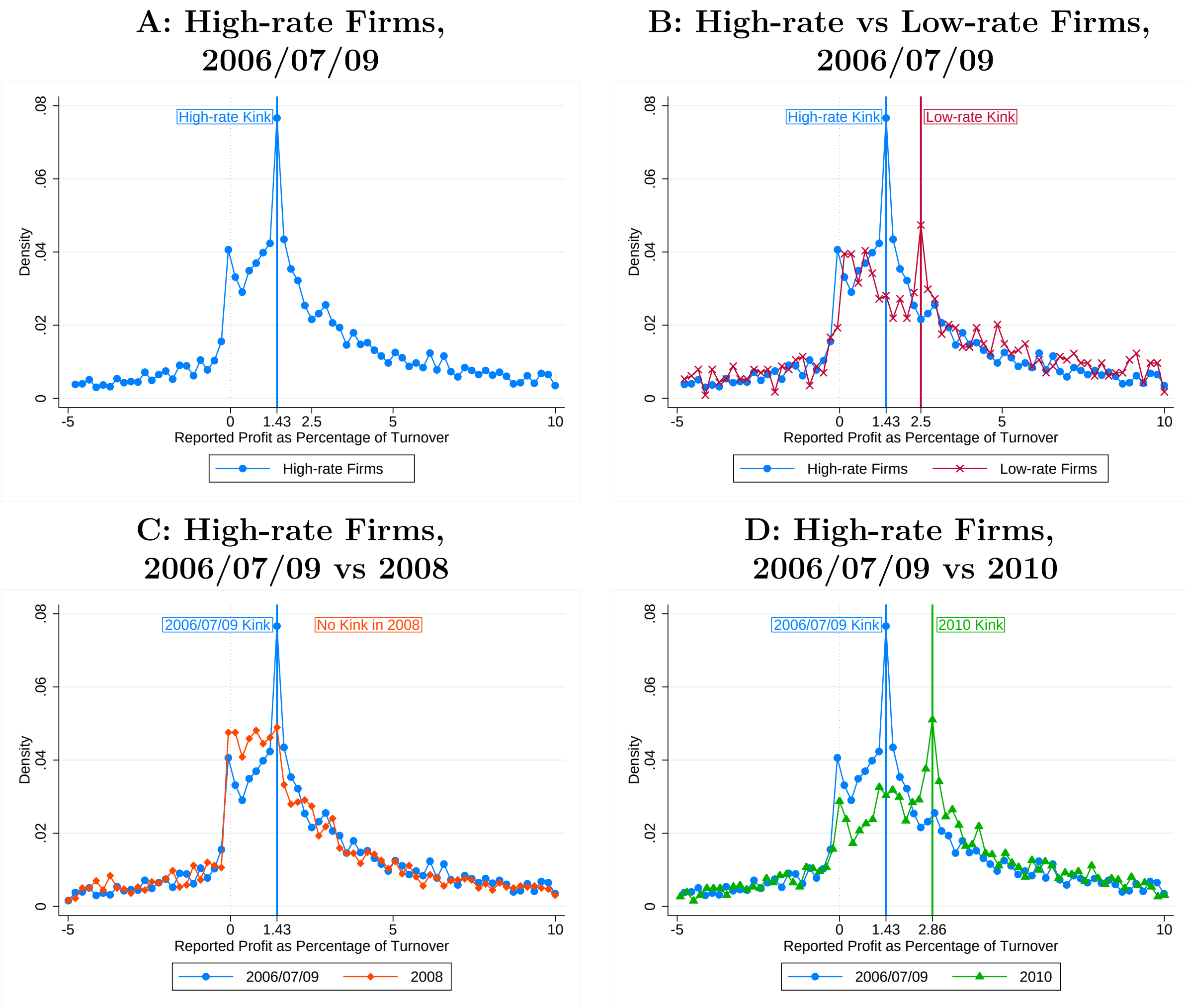
Panels C and D of Figure 2.II exploit time variation in the kink, focusing on the sample of high-rate firms. Panel C shows that excess bunching at 1.43% completely disappears in 2008 when the minimum tax regime (i.e., turnover tax

[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.

[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 2.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 2.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.

Figure 2.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.

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 2.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.

2.5.2 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 2.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

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

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.

2.5.3 Estimating Evasion Responses Using Bunching

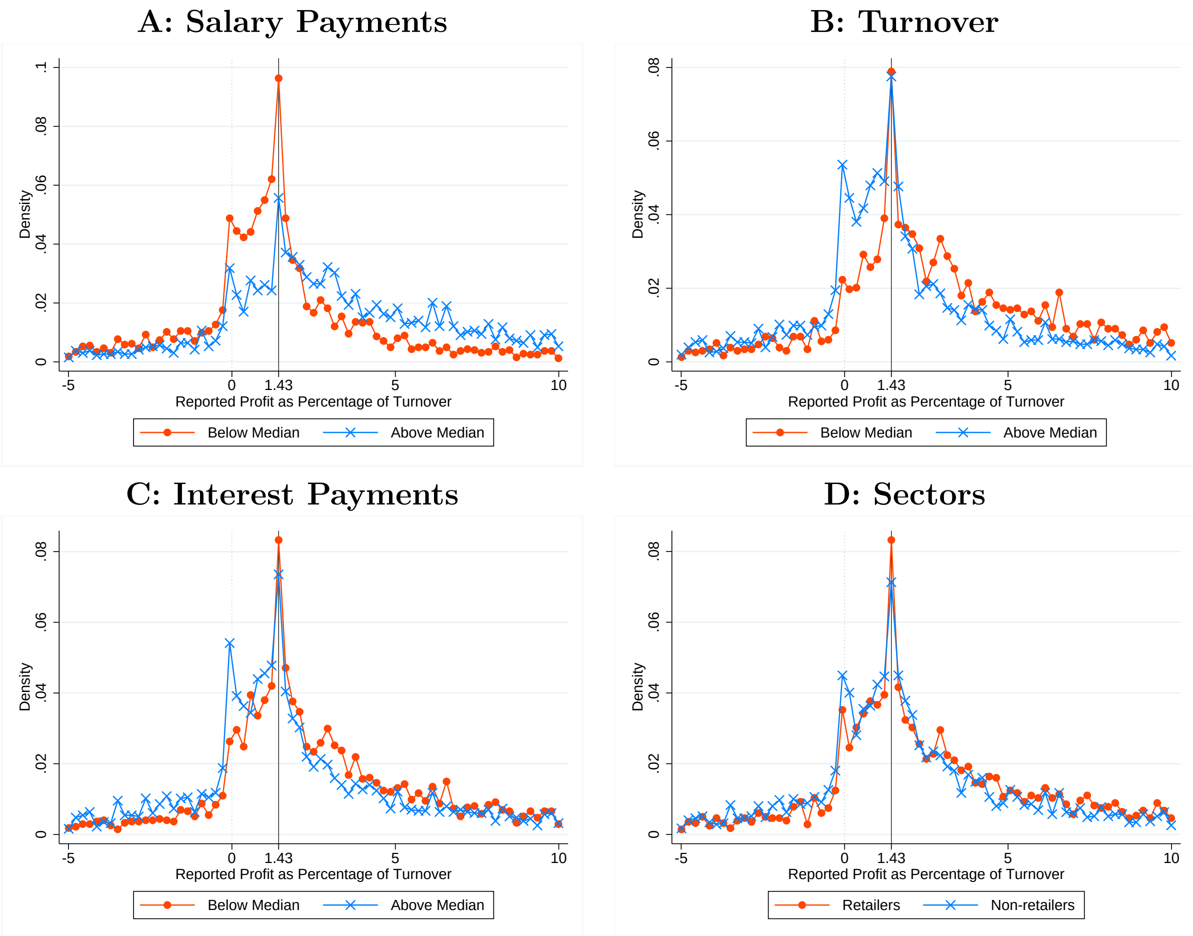
This section presents estimates of excess bunching and uses our model to translate them into estimates of evasion responses. Following Chetty *et al.* (2011), we estimate a counterfactual density—what the distribution would have looked like absent the kink—by fitting a flexible polynomial to the observed 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) profit rate in bin j, the counterfactual density is obtained from a regression of the following form

$$d_j = \sum_{i=0}^{q} \beta_i (p_j)^i + \sum_{i=p_L}^{p_U} \gamma_i \cdot \mathbf{1}[p_j = i] + \nu_j, \tag{2.19}$$

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 (2.19) 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 2.IV compares the empirical density distributions to estimated counterfactual distributions (smooth solid lines) for the four samples examined in Figure 2.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

Figure 2.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 2.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.

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 2.II converts our bunching estimates into evasion responses using the methodology developed in section 2.3. 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 (2.15),

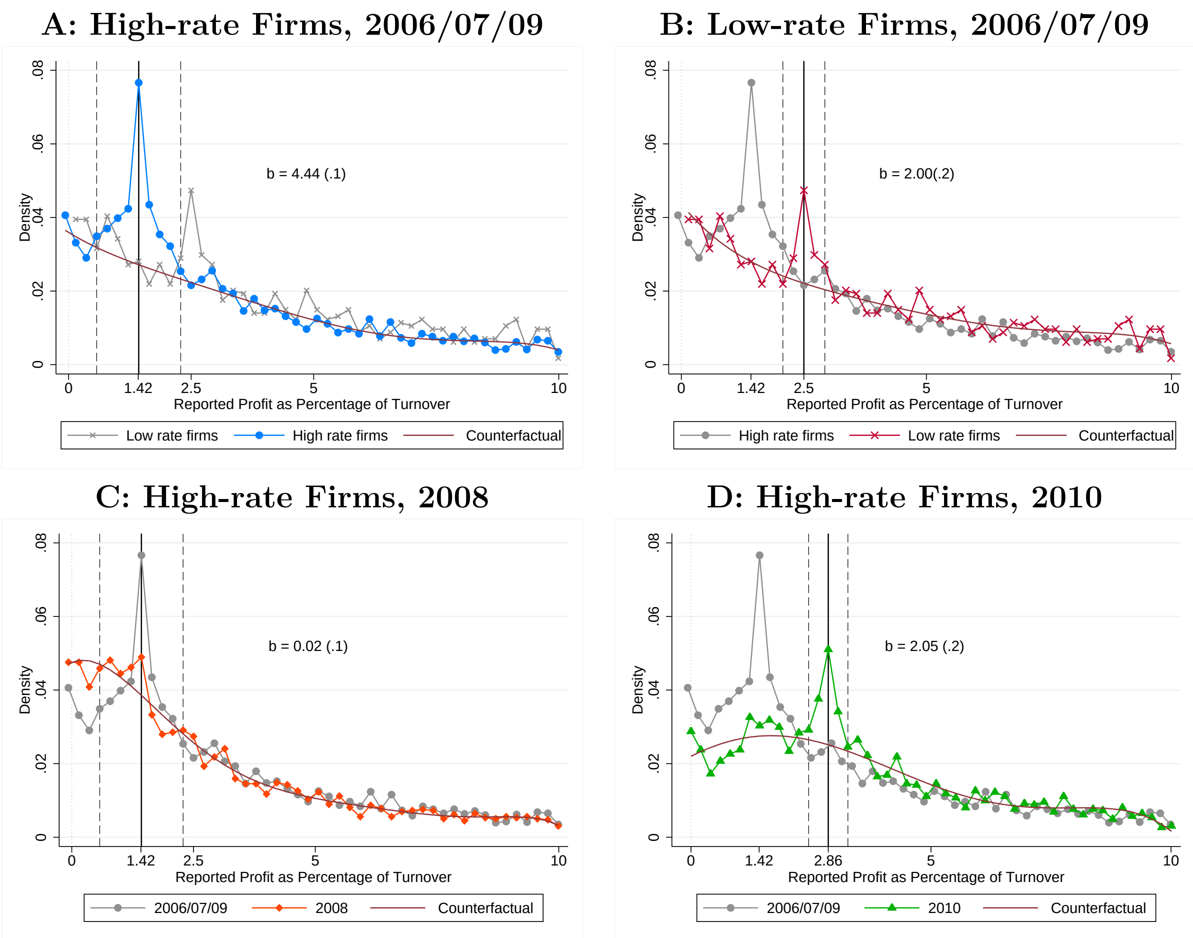
$$\Delta\hat{p} \simeq \frac{\tau_y^2}{\tau_\pi}\varepsilon_y - \frac{d\,(\hat{c} - c)}{y}.$$

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. (2.15). 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.

[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 $binwidth = 0.00214$.

Figure 2.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 2.II. The tax rate schedules and kink locations are explained in the notes to figure 2.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 (2.19). 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.

Table 2.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 2.IV. Column (1) reproduces the bunching estimate b, based on estimating equation (2.19). 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 $\Delta\hat{p} = B/f_0(\tau_y/\tau_\pi) \simeq b \times binwidth$. 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 $\Delta p = [c/y - c'(y)]\, dy/y \simeq \left(\tau_y^2/\tau_\pi\right) \varepsilon_y$. 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 $\Delta\hat{p} = [\hat{c}/y - c'(y)]\, dy/y - [d(\hat{c} - c)/y] \simeq \left(\tau_y^2/\tau_\pi\right) \varepsilon_y - (\tau_y/\tau_\pi)[d(\hat{c} - c)/(y - \hat{c})]$, assuming different real output elasticities $\varepsilon_y \in \{0, 0.5, 1, 5\}$. Bootstrapped standard errors are shown in parentheses.

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 (2.17). 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 2.4. Those tax gaps were measured as a fraction of *true* tax liability, but can easily be converted into tax gaps as a fraction of *actual* taxes paid (corresponding to our estimates in Table 2.II), in which case they would close to 100%.

2.6 Numerical Analysis of Optimal Tax Policy

This section links our empirical results to the stylized model introduced in section 2.2. 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,

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

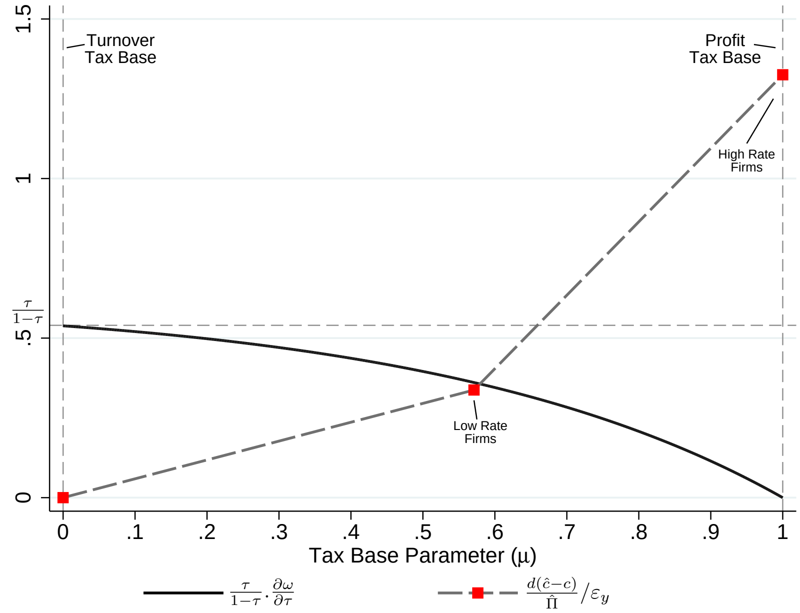
For tax rates below 50 percent, the effective marginal tax rate is decreasing in μ, starting at a maximum of $\frac{\tau}{1-\tau}$ for $\mu = 0$ and falling to 0 for $\mu = 1$. This is illustrated in panel A of Figure 2.V for $\tau = .35$, which is the standard profit tax rate in Pakistan.

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

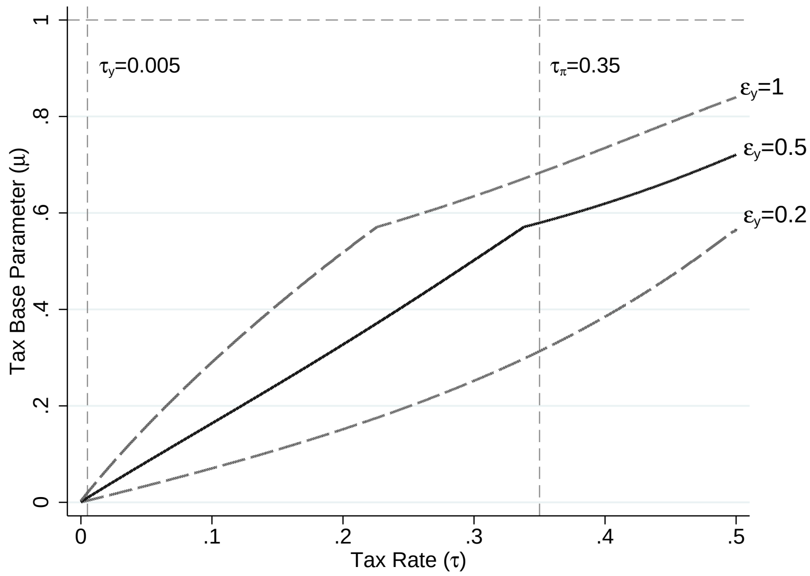
$$G(\mu) \cdot \frac{\varepsilon_{\hat{c}-c}}{\varepsilon_y} = \frac{\hat{c}-c}{\hat{\Pi}} \frac{\varepsilon_{\hat{c}-c}}{\varepsilon_y} \simeq -\frac{d(\hat{c}-c)}{\hat{\Pi}} / \varepsilon_y,$$

Figure 2.V: **Numerical Analysis of Optimal Tax Policy**

A: Optimal Tax Rule



B: Tax Base vs. Tax Rate



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 2.11) 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.

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 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 2.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 2.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,

[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.

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.

2.7 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.

[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.

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.

Chapter 3

Does IMF Conditionality Work? Evidence from VAT Adoption

3.1 Introduction

Adoption of the VAT is an important issue for developing countries' fiscal policy agendas. Developing countries are under pressure to increase public spending on health, education and social services, but their tax systems rely on distortive tax instruments and raise little revenue. Estimates for the period 1996-2001 show that industrialized countries raised on average 25.0% of GDP in tax revenue, whereas developing countries raised only 17.6% of GDP. Developing countries relied on tariffs, corporate income taxes and seignorage for 62.7% of their tax revenue, whereas these taxes represented only 20.2% of tax revenue in industrialized countries (Gordon & Li 2009). The VAT, a broad-based tax levied at all stages of production, is often considered a cure for developing countries' fiscal problems, as it promises to alleviate the distortions of the tax system and increase tax revenue. First introduced in France in 1948, the VAT has spread rapidly and is now used in more than 140 countries, two-thirds of which are low-and middle-income countries (Bird & Gendron 2007). However, its broad tax base and the fact that it often replaces more selective sales and excise taxes also make the VAT an object of political opposition. Besides, adopting the VAT, with its complex recrediting mechanism, is logistically challenging, especially for developing countries that have low administrative capacities. It is therefore important for policymakers to understand what factors determine VAT adoption and how VAT adoption can be facilitated in practice. An institution intimately linked to the spread of the VAT is the International Monetary Fund (IMF). The IMF considers the VAT "a key instrument for securing macroeconomic stability and growth by placing domestic revenue mobilization on a sounder basis." It has promoted the VAT by making lending conditional on VAT adoption and by providing technical assistance for adoption (Ebrill *et al.* 2001; Tanzi 1990). However, IMF lending and technical assistance projects have been highly controversial in the past, and countries have repeatedly failed to implement IMF advocated reforms (Easterly 2005).

This paper therefore analyzes whether IMF interventions in favor of VAT adoption have increased countries' likelihood of adopting the VAT. I discuss how the IMF has promoted VAT adoption by making lending conditional on adoption, providing administrative and technical assistance, and reducing the political costs of adoption. Since only a handful of countries have ever repealed the VAT[1], I consider VAT adoption an absorbing state and estimate the determinants of adoption in a hazard model. The empirical analysis is based on panel data for 26 years (1975-2000) and 125 countries, 60 of which adopted the VAT during the sample period. To proxy IMF interventions, I use a dummy variable that indicates whether or not a country is under a lending agreement with the IMF in a given year.[2] I find that countries under a lending agreement are approximately three times as likely to adopt the VAT as countries not under a lending agreement, and that this effect is not driven by omitted variables such as debt crises. I also show that IMF lending attenuates the negative effect of autocracy on VAT adoption.

My paper contributes to the literature on the VAT as summarized most recently by Keen & Lockwood (2010). This literature has generally estimated the determinants of VAT adoption in probit models which are not fit to accommodate absorbing state events. I improve upon this by modeling VAT adoption in a hazard framework, thus taking into account its absorbing state nature, and by paying particular attention to political factors conditioning VAT adoption. My paper also contributes to the literature on the impact of IMF lending, as represented by Barro & Lee (2005) and Easterly (2005). This literature has focused on distant macroeconomic outcomes and found weak and inconsistent effects of IMF lending. I instead focus on a specific policy reform requested by the IMF and estimate the immediate impact of IMF lending on this reform. The paper also draws inspiration from Besley & Persson (2009) and other contributions to the fiscal capacity literature.

The paper is structured as follows. Section 3.2 discusses the benefits and costs of VAT adoption, the role of the IMF in this process, and summarizes the related literature. Section 3.3 presents the estimation strategy. Sections 3.4 and 3.5

[1] These exceptional cases will be discussed below. In general, countries are unlikely to repeal the VAT because of the high fixed costs of adoption (investments in administrative capacity etc.).

[2] The data do not allow me to distinguish between lending agreements that contain an explicit condition on VAT adoption and/or are accompanied by a technical assistance package, and those that are not.

discuss the data and empirical results. Section 3.6 offers concluding remarks.

3.2 Motivation

This section describes the efficiency and revenue gains from VAT adoption, its administrative and political costs, and the role the IMF has played in the process of VAT adoption. It also discusses related literature.

3.2.1 Benefits of VAT Adoption

According to Bird & Gendron (2007), the VAT is a "broad-based tax, levied at multiple stages of production and distribution, with, crucially, taxes on inputs credited against taxes on output." In fact, the VAT is simply a tax on final consumption that is collected throughout the chain of production. It is usually charged on imports as well as domestic goods, but not on exports. This, together with the recrediting mechanism, ensures that the VAT fulfills the conditions of production efficiency according to Diamond & Mirrlees (1971). The VAT does not create incentives for vertical integration, as do turnover taxes, which are also levied at all stages of production, but which do not allow for the crediting of input taxes against output taxes. The VAT also avoids the excessive rate differentiation that characterizes final sales and excise taxes. These tax instruments impose high tax rates on a narrow range of goods and a zero rate on all other goods, which distorts the relative price between those two types of goods.[3] Finally, the VAT does not distort the relative price of imports and domestically produced goods, as tariffs do (Bird & Gendron 2007), Burgess & Stern 1993). Since the VAT tends to replace tax revenue from turnover, sales and excise taxes, and from tariffs, its adoption is likely to increase the efficiency of the tax system.

The second key reason that the VAT may provide efficiency gains is the difficulty of evading it. In fact, VAT evasion requires the cooperation of a buyer and a seller who have opposing interests. A firm can evade the VAT either by overstating its input purchases or understating its sales. However, the input purchases are another firm's sales, and this second firm is interested in obtaining a correct invoice for its sales. This is why VAT compliance has often been described as self-enforcing

[3] Final sales taxes are intended to be levied on all goods at the stage of final sales. However, these taxes are vulnerable to evasion since it is difficult in practice to distinguish final consumption goods from intermediate goods. This means that final sales taxes also end up being levied on a narrow tax base.

(Bird & Gendron (2007); Ebrill *et al.* 2001).[4]

In summary, theory predicts that the VAT, if correctly implemented, can generate important efficiency gains and reduce the marginal cost of public funds. It is thus an attractive means of taxation, especially for developing countries that have in the past relied on highly distortive tax instruments. Evidence shows that the VAT has indeed become an important source of tax revenue. It now raises 20% of tax revenue worldwide and is the "single most important tax" in many developing countries. In El Salvador, the Republic of Congo, and Tajikistan, for instance, it raised more than 45% of the total tax revenue in 1998-2000 (Bird & Gendron 2007). However, it remains unclear whether VAT adoption has allowed countries to increase total tax revenue, i.e. the tax revenue to GDP ratio. Keen & Lockwood (2006) and (2010) report that the VAT has significantly increased tax revenue to GDP ratios in OECD countries, but has not always done so in developing countries. In particular, their predicted revenue gains from VAT adoption are negative for some African countries. In these countries, the increase in tax revenue from VAT adoption has often been more than offset by decreasing revenue from other sources such as trade taxes, and by the high administrative costs of VAT adoption. These costs of VAT adoption are therefore discussed in the next subsection.

3.2.2 Costs of VAT Adoption

The costs of VAT adoption can be divided broadly into administrative and political costs, although these two are closely linked. The administrative costs of VAT adoption relate to the investments in fiscal capacity a country must make in order to successfully implement the VAT. Since the VAT is levied at all stages of production, it covers a large number of taxpayers and thus requires a large and well-equipped national tax administration. In addition, the VAT recrediting mechanism is based on an extensive system of record-keeping, invoicing and self-assessment by firms. Preparing the business community for this task requires public awareness and education programs (Bird & Gendron 2007). The administrative costs of VAT adoption are thought to differ widely across countries, according to their state of development, their existing fiscal capacity and the design of the VAT (Ebrill *et al.* 2001). Esti-

[4] Since the VAT can only be applied to firms that keep records, there is a concern that it may introduce distortions between formal and informal sector production, especially in developing countries where informal sectors are large. For instance, Piggott & Whalley (2001) and Emran & Stiglitz (2005b) present models in which the VAT can reduce welfare, because it leads to differential taxation of informal and formal sector inputs and thus causes a shift towards inefficient informal production. However, Keen (2008) shows that the VAT can be optimal even in the presence of a large informal sector, if it is combined with a withholding tax. A withholding tax allows for taxes paid on imports to be credited against income tax payments, and thus functions as a tariff for informal sector businesses. A VAT-cum-withholding-tax is therefore preferable to tariffs.

mates by Cnossen (1994) suggest that the annual costs of administering the VAT rise to approximately US$ 100 per registered firm in OECD countries. However, there are, to the best of my knowledge, no similar estimates for developing countries, and no estimates of the fixed costs of VAT adoption.

In addition to its administrative costs, VAT adoption also has political costs. Political costs here are understood as the loss of votes in a democracy and the emergence of violent protest in an autocracy. The VAT is a politically sensitive issue for two reasons. First, the VAT is a broad-based tax and often replaces more selective excise or sales taxes. It is therefore perceived by citizens as a regressive tax that disproportionately hits low income groups.[5] Second, the VAT is unpopular in the business community due to the high administrative costs it imposes on firms. In particular, small-scale businesses in developing countries, which have low administrative and managerial capacity, struggle with the self-assessment mechanism (Bird & Gendron 2007). The administrative and political costs of VAT adoption are closely linked. If the government makes little effort to raise awareness about the VAT and to prepare the business community for it, the political cost of VAT adoption is likely to be higher. The history of VAT adoption in Ghana, discussed in the appendix, illustrates this point.

3.2.3 VAT Adoption and the IMF

A large amount of anecdotal evidence demonstrates how the IMF has promoted VAT adoption. First, the IMF has promoted VAT adoption through conditionality: that is, including VAT adoption in the set of reforms that a country commits itself to implement under IMF lending.[6] For example, the Medium-Term Economic and Fiscal Policy Framework Paper for an loan to Cameroon for the period 1998-2001 states: “Envisaged tax policy changes include (a) the replacement of the sales tax by a value-added tax (VAT), as from January 1, 1999; [...].” VAT adoption was also part of the lending agreement between the IMF and Chad in 1999. In its Letter of Intent[7], the Chadian government promised “the introduction on January 1, 2000 of the value-added tax (VAT) at the single rate of 18 percent, replacing the turnover tax (TCA).” Cameroon and Chad indeed adopted the VAT in 1999 and

[5] In reality, many countries choose to reduce the regressivity of the tax by applying it only to businesses of a certain size and exempting necessity goods. The degree of regressivity thus depends on the context and widely differs across countries (Jenkins *et al.* 2006).

[6] The IMF uses different types of conditionality concepts, all of which broadly relate to a set of policies that the country agrees to implement under IMF guidance. The way compliance with conditionality is evaluated differs across policy prescriptions and lending programs. For more information, see http://www.imf.org/external/np/exr/facts/conditio.htm (accessed April 18th 2010).

[7] The Letter of Intent is the official request for funding from the IMF.

2000 respectively.

Second, the IMF has promoted VAT adoption through technical assistance.[8] The aforementioned agreement between the IMF and Cameroon, for instance, states: "In the area of domestic resource mobilization, Cameroon expects to continue receiving technical assistance from the IMF [...]. The government will receive further IMF assistance in introducing the VAT [...]." The 2001 Article IV consultation report for Chad explains that "In the area of revenues, the valued-added tax (VAT) administration has improved, with technical assistance from the Fund." Finally, Vreeland (2003) and Vreeland (1999) suggest that the IMF has also helped to reduce the political costs of VAT adoption by figuring as a scapegoat. They argue that governments deflect popular anger about VAT adoption by claiming that the IMF forced them to adopt the VAT.

Descriptive data show that the VAT has indeed spread rapidly over the past decades. The VAT was first adopted in France in 1948 and spread to other countries in Europe and Latin America in the 1960s and 1970s. A second wave of adoption brought it to Africa, Asia and the transition economies in the 1980s and 1990s. Figure 3.I traces the number of countries having a VAT and the number of countries adopting a VAT each year for my sample of 125 countries. While only 22 countries had a VAT in 1975, this number increased to 82 by 2000. Today, more than 140 countries have a VAT.

An event study analysis, presented in Figure 3.II, shows that VAT adoption is often accompanied by an IMF intervention. For countries that adopted the VAT during my sample period, the blue bars show the percentage of countries that were under an IMF lending agreement in the years before, during, or after VAT adoption. In each of these three years, approximately 50% of the VAT adopters experienced an IMF intervention. The fourth bar shows that in all other years (i.e. those more than one year before or after VAT adoption), on average only 40.54% of the VAT adopters experienced an IMF intervention. This hints at a bunching of IMF interventions around the VAT adoption date. Further evidence for this is provided by the fifth bar, which displays the average percentage of countries under an IMF lending agreement in any given year for the countries that did not adopt the VAT during the sample period. On average, only 22.13% of these countries experienced an IMF intervention in any given year. IMF interventions were thus more frequent in countries that adopted the VAT during the sample period, and in years preceding the VAT adoption date. The fact that VAT adopters also experienced a higher number of IMF interventions in the year following VAT adoption suggests that the

[8] This can include the training of tax collectors, advising the government on the appropriate tax rate and supporting the design of tax legislation (Ebrill *et al.* 2001).

Figure 3.I: **VAT Adoption Over Time**

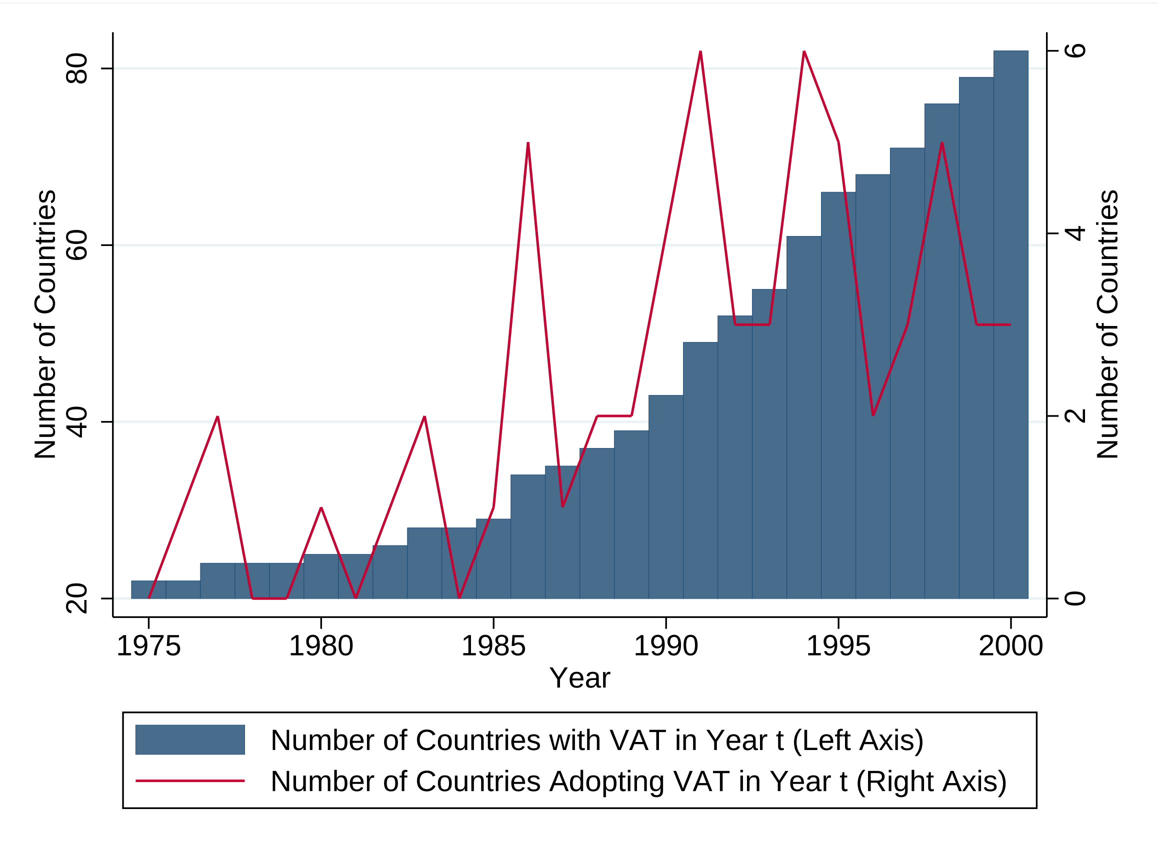
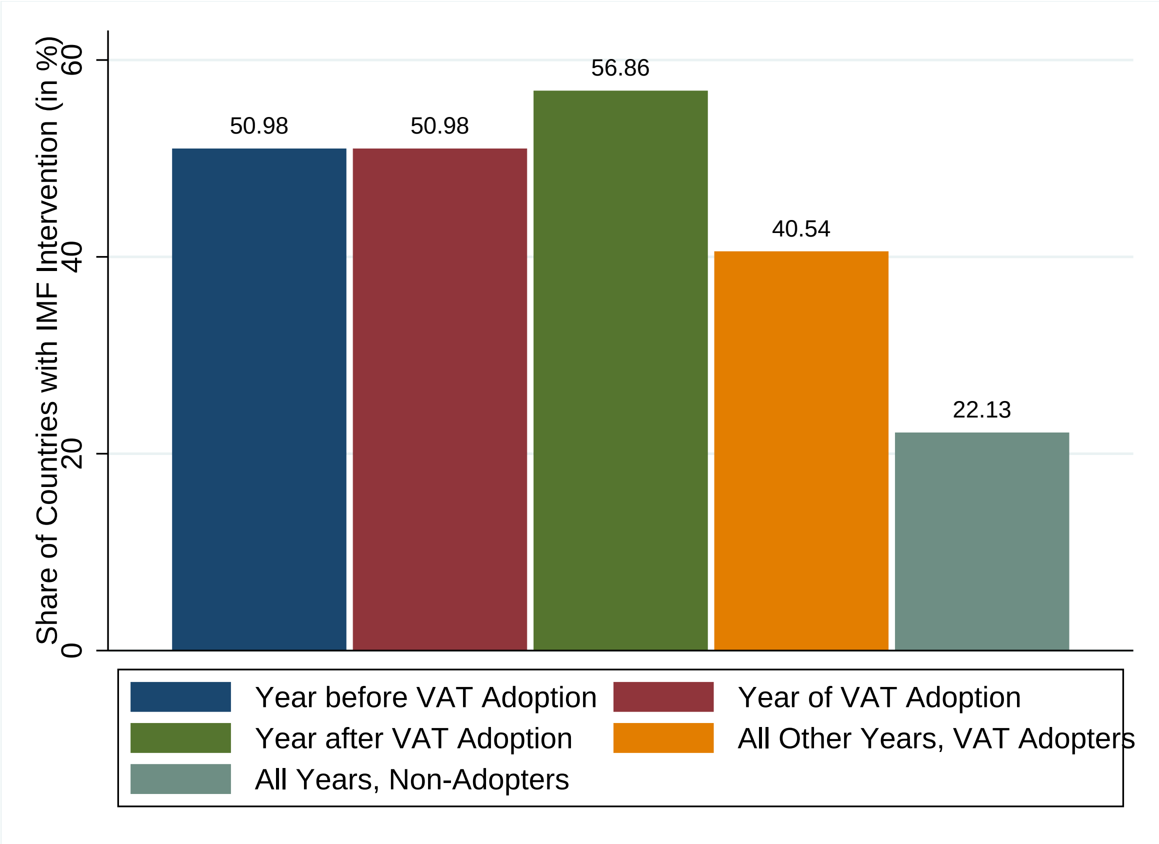


Figure 3.II: **VAT Adoption and IMF Intervention**



line of causality may not only run from IMF lending to VAT adoption, but also the other way around, an issue which will be discussed in more detail in section 3.3.

3.2.4 Related Literature

This paper is related to three different strands of literature, the first of which is the literature on the VAT. Key papers are Keen & Lockwood (2010) and (2006), who estimate the revenue impact of VAT adoption. Other contributions are Desai & Hines (2005), who show that VAT tax revenue is negatively related to trade openness and export performance, and Agha & Haughton (1996), who analyze the determinants of VAT compliance, showing that compliance is higher when rates are lower and less numerous. The only paper that analyzes the determinants of VAT adoption is Keen & Lockwood (2010). However, this paper uses a probit model to explain whether or not a country has a VAT in a given year. This disregards the fact that VAT adoption is in practice an absorbing state, so that the dependent variable, a VAT dummy, is determined to take value one in all periods after the VAT adoption date. After adoption, there is no variation in the dependent variable left to be explained by the independent variables. I improve upon this by modeling VAT adoption in a survival framework, which takes account of the absorbing state nature of this event, as will be discussed in more detail below. My paper also pays particular attention to the political economy of IMF intervention and VAT adoption.

While research on the VAT is scarce, there is a large empirical literature on the impact of IMF lending on various macroeconomic outcomes. Barro & Lee (2005) use an instrumental variable strategy to take into account the endogeneity of IMF lending and find that IMF lending reduces a country's growth rate. Easterly (2005) uses a similar estimation strategy and finds that IMF lending has no impact on a macroeconomic imbalance indicator. Przeworski & Vreeland (2000) rely on a dynamic Heckman selection model to demonstrate that IMF lending leads to increased growth once a country graduates from the lending program. Other contributions are Wei & Zhang (2010), Marchesi (2003), Dicks-Mireaux *et al.* (2000) and Conway (1994). Overall, the effect of IMF lending that these papers identify is at best ambiguous. This might be due to the literature's focus on macroeconomic outcomes rather than on specific policy changes. Understanding whether IMF interventions have actually led to the desired *policy reforms* may help to better predict the impact of these reforms on macroeconomic outcomes. This paper therefore analyzes the IMF's impact on the adoption of the VAT, which is one specific reform the IMF has promoted.

Finally, this paper also draws inspiration from a nascent literature on fiscal investments and state capacity. For instance, Besley & Persson (2009) model the determinants of investments in fiscal and legal capacity, and show that the occurrence of wars, political stability and the existence of inclusive political institutions are among the factors that promote investments in state capacity. Besley & Persson (2010) further develop this framework and relate investments in state capacity to development and growth dynamics. Another contribution is Acemoglu *et al.* (2006), who pay particular attention to the political economy of inequality and show that states with low tax capacity are more likely to emerge when inequality is greater.

3.3 Empirical Strategy

The foundation of my empirical strategy is considering VAT adoption an absorbing state. In theory, countries should be unlikely to repeal the VAT once they have adopted it, because of the large fixed costs of adoption. This is also true in practice. Even though we observe five countries repealing and later reintroducing the VAT,[9] the first VAT adoption in these cases was only an unsuccessful attempt. The

[9] Belize introduced the VAT in 1996, repealed it in 1999, and reintroduced it in 2006. Ghana introduced the VAT in 1995, repealed it three months later, and reintroduced it in 1998.

VAT had to be withdrawn because of lack of administrative capacity and political support, as illustrated by the history of VAT adoption in Ghana, discussed in the appendix. The VAT thus becomes an absorbing state if *successfully* adopted.[10] The next subsection explains why an absorptive dependent variable requires the use of a hazard model, and specifically presents the advantages of the Cox proportional hazard model.[11] I then discuss how the coefficients in this type of model are estimated and identified.

3.3.1 The Cox Proportional Hazard Model

As discussed in Cleves *et al.* (2004), the distribution of the time t until the occurrence of an event Z is most often asymmetric and possibly bimodal. Estimating the model $t_i = \beta_0 + \beta_1 x_i + \epsilon_i$ via an ordinary least squares or probit method, which assumes the normality of the error term, may therefore lead to biased and inconsistent estimates. If the event Z is an absorbing state, the errors in this model would also be strongly auto-correlated. These problems are overcome in a hazard model, which does not require any assumption on the error term. Let T be the time of VAT adoption. A hazard model then estimates the probability of adoption, also called "hazard rate", at time t as a function of time and a vector of independent variables $\mathbf{x}(t)$, which can themselves be time-varying:

$$h(t; \mathbf{x}(t)) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t, \mathbf{x}(t))}{\Delta t}. \tag{3.1}$$

The independent variables $\mathbf{x}(t)$ in my model are a dummy for IMF lending, and proxies for the efficiency and revenue gains, and for the administrative and political costs of VAT adoption. In order to model the relationship between the probability of VAT adoption and the matrix $\mathbf{x}(t)$, one needs to impose a functional form on the hazard rate $h(t; \mathbf{x}(t))$. This paper relies on the semi-parametric proportional hazard specification due to Cox (1972), which is most frequently used in the aforementioned literature. The Cox model describes the hazard rate as the product of an exponential function of the independent variables and a baseline hazard h_0,

Grenada introduced the VAT in 1986, faced it out gradually, and reintroduced it in 2010. Malta introduced the VAT in 1995, repealed it in 1997, and reintroduced it in 1999. Vietnam introduced and repealed the VAT in several steps in the 1970s and reintroduced it in 1999.

[10]For the countries that adopted, repealed and readopted the VAT, I code the VAT dummy as switching to one with the second VAT adoption. In my sample, this concerns only Ghana, since Malta and Vietnam are not in my sample, and Belize and Grenada re-adopted the VAT after 2000. The results are robust to dropping Ghana, Belize and Grenada from the sample, or coding the first introduction as the VAT adoption date.

[11]The class of hazard models is also known as survival or duration models.

which only depends on time:

$$h(t; \mathbf{x}(t)) = exp(\mathbf{x}(t)'\beta)h_0(t). \tag{3.2}$$

This specification has two main advantages. First, it ensures that the logarithm of the hazard rate is a linear function of the independent variables $\mathbf{x}(t)$, which is convenient for interpretation.[12] Second, the model does not require any assumptions on the structure of the error term because it is semi-parametric. It is parameterized with respect to $\mathbf{x}(t)$, but the baseline hazard $h_0(t)$ is left unspecified. The cost of not making a distributional assumption is reduced precision of the estimates. In fact, the coefficient vector β has to be estimated via *partial* rather than ordinary maximum likelihood, as explained in the next subsection.[13]

3.3.2 Estimation Via Partial Maximum Likelihood

The partial maximum likelihood method removes the distributional assumption by only taking into account the ordering of VAT adoption times, but not the time intervals between two adoption events. In a dataset consisting of n countries, $k \leq n$ of which adopt the VAT during the period under observation, and $n - k$ of which are right-censored (i.e. adopted the VAT after the period under observation), let the countries be ordered according to their adoption time. The probability that the ith country adopts the VAT at time t_i is then

$$Pr(t_i = T_i | R(t_i)) = \frac{exp(\mathbf{x}_i(t)'\beta)h_0(t)}{\sum_{j \in R(t_i)} exp(\mathbf{x}_j(t)'\beta)h_0(t)}, \tag{3.3}$$

where $R(t_i)$ is the risk set at time t_i, i.e. the set of countries that have not adopted the VAT until t_i. Notice that the baseline hazard, which is assumed to be the same for all countries, nicely cancels out in this formulation. Now let d_i be an indicator that takes value zero for right-censored observations. The product of the adoption probabilities for each country then yields the partial likelihood function

$$L_p = \prod_{i=1}^{n} \left[\frac{exp(\mathbf{x}_i(t)'\beta)}{\sum_{j \in R(t_i)} exp(\mathbf{x}_j(t)'\beta)} \right]^{d_i}. \tag{3.4}$$

[12]Using the exponential function also ensures that the hazard rate cannot be negative because $h_0(t)$ is a probability and thus greater than or equal to 0.

[13]Wooldridge (2002), Chapter 20, and Lancaster (1990) discuss these and other advantages of the Cox model in more detail.

The corresponding partial log-likelihood function is

$$logL_p = \sum_{i=1}^{n} d_i \left[\mathbf{x}_i(t)'\beta - log \sum_{j \in R(t_i)} exp(\mathbf{x}_j(t)'\beta) \right], \quad (3.5)$$

which can be easily maximized. Cox (1972) and Cox (1975) show that the partial maximum likelihood estimates have the same properties as ordinary maximum likelihood estimates. They are asymptotically normal, asymptotically efficient, consistent and invariant.

Left-censored observations (that is countries that adopted the VAT before the period under observation) do not contribute any information to the estimation of the coefficients. Right-censored observations (that is countries that adopted the VAT after the period under observation) contribute information only to the risk set.[14] We only know that these countries adopted the VAT after the year 2000, the end of the period of observation, but do not know in which year they adopted it. Right-censoring does not bias the coefficients as long as, conditional on the covariates, the time until VAT adoption is independent of the censoring time c_i, which means $D(T_i|\mathbf{x}_i, c_i) = D(T_i|\mathbf{x}_i)$, where $D(\cdot|\cdot)$ denotes the conditional distribution (Wooldridge 2002, Chapter 20). This assumption clearly holds here, because c_i is a constant, 2000. All countries are censored in the year 2000, because the dataset does not extend further.

An important challenge for the Cox model is dealing with "tied failures", or years in which several countries adopt the VAT and the order of adoption is not known. In these cases, the exact composition of the risk set is unknown and the partial likelihood function thus needs to be approximated. Here, I use the computationally simple Breslow method, which declares the risk set to be the same for all countries that adopt the VAT in a given year. It contains all countries that adopt the VAT that year, and all countries that adopt it at a later point in time. The loss of precision from this approximation is small, since the number of tied failures in my sample does not exceed six countries in a given year. The results are also robust to using more sophisticated methods of approximation such as the Efron and exact discrete method (see Efron (1977) for more details on these methods).

Another challenge is that the Cox model cannot accommodate country fixed effects, since its estimation relies on a single observation per country (the time until VAT adoption). This is problematic if one expects the baseline propensity to adopt the VAT to vary across countries. I therefore stratify the sample according to four groups of countries that can be expected to have different baseline propensities of

[14] Right-censoring is equivalent to right-truncation in this model.

VAT adoption. I let the baseline hazard $h_0(t)$ vary across these subsamples but restrict the coefficient vector β to be the same for all subsamples. This means that I compare countries only to other countries in the same subsample. The four subsamples are oil-producing countries (which have high oil revenue and thus little need for the VAT), small island states (where most domestic consumption is imported, which makes the VAT similar to a tariff), rich countries (those in the upper quartile of the income distribution in 1975), and poor countries. Finally, since the errors are most likely correlated across time for each country, I cluster the standard errors at the country level.

3.3.3 Identification Assumption and Challenges

Having discussed how the coefficient vector β can be estimated, I will now address the question of whether it can also be correctly identified. The identification assumption I have to make is that the time path of the independent variables $\mathbf{x}(t)$ does not depend on the time until VAT adoption (Lancaster 1990). This means that

$$P(\mathbf{x}(t, t + \Delta t)|T \geq t + \Delta t, \mathbf{x}(t)) = P(\mathbf{x}(t, t + \Delta t)|\mathbf{x}(t)), \tag{3.6}$$

where $\mathbf{x}(t, t + \Delta t)$ is the path of the independent variables between time t and $t + \Delta t$.

This assumption faces two important challenges. The first is omitted variables. Indeed, one can easily imagine that a shock, such as a debt crisis, might push a country to adopt the VAT and simultaneously trigger an IMF intervention. This would cause a spurious correlation between the IMF intervention and VAT adoption. The above exogeneity condition would then hold only when the probabilities are conditioned on the omitted variables. Without accounting for the omitted variables, the partial likelihood maximization would yield upward biased and inconsistent estimates. Therefore, after presenting the baseline results in section 3.5.1, I explore the impact of various omitted variables in section 3.5.2. I show that controlling for the shocks that one most commonly expects to lead to IMF intervention and possibly VAT adoption does not alter the sign and significance of the coefficient on the IMF dummy.

The second challenge to the identification assumption is reverse causality. For instance, countries that want to adopt the VAT might call for an IMF intervention to assist them in carrying out the necessary administrative reforms.[15] Alterna-

[15] Countries address a Letter of Intent to the IMF to request funding or technical assistance. However, these letters are drafted in cooperation with the local IMF representation. It is therefore difficult to distinguish in practice between supply-driven and demand-driven IMF

tively, the IMF might be more willing to lend to countries that have adopted the VAT, because those countries have a stronger fiscal balance and are more likely to pay back the loan. In these cases, the above exogeneity condition does not hold and the estimated coefficient on the IMF variable would be biased. I therefore conduct a falsification test for reverse causality in section 3.5.3. The test shows that there is no significant correlation between VAT adoption and *future* IMF interventions, once autocorrelation in IMF interventions is controlled for.

3.4 Data

The data I use to test the impact of IMF interventions on VAT adoption contains information on 125 countries over a period of 26 years (1975-2000). The set of countries covers all continents, but excludes the transition economies due to their specific development history.[16] Data on the VAT adoption year was obtained from Ebrill et al. (2001). 60 countries, most of them in Africa and Asia, adopted the VAT during the sample period.

The key independent variable in my analysis is an IMF dummy indicating whether or not the country is under a lending agreement for at least five months of a given year. It is derived from Dreher (2006). I use different leads and lags of the IMF dummy, as well as the combined dummy IMF_{-3} (IMF_{+3}) which takes value one if the country was under a lending agreement in one of the last three years (will be under a lending agreement in one of the forthcoming three years). I lose only few observations in this process, because the IMF data covers a slightly longer period than the baseline data (1970-2004). 25.8% of the country-year points in my sample experienced an IMF intervention. Individual lending agreements normally last between one and four years. Many countries experienced only sporadic one-year interventions, while others were under lending for a continuous period of more than a decade (e.g. Panama from 1970 to 1986).

In addition to IMF interventions, the theory of taxation and the discussion in section 3.2 suggest that VAT adoption depends positively on the revenue and efficiency gains and negatively on the administrative and political costs of VAT adoption. As proxies for the revenue gains of VAT adoption, I use the level of GDP per capita, the share of the agricultural sector in GDP and the share of imports in GDP. The larger a country's GDP and the smaller its agricultural sector, the

---

interventions.

[16] The transition economies all adopted the VAT in the early 1990s, during the IMF-led process of liberalization. Including these countries in the sample might therefore upward bias the estimated IMF effect.

larger the monetary value of taxable transactions that take place and the higher the revenue gains from VAT adoption.[17] The share of imports in GDP does not directly proxy revenue gains but rather the need for additional tax revenue. Countries with a lower share of imports in GDP have lower trade tax revenues and should thus be more in need of the VAT as an additional source of revenue. GDP, the size of the agricultural sector and the share of imports in GDP are also standard variables in tax effort models (Tanzi 1992; Burgess & Stern 1993) that attempt to predict the ratio of tax revenue to GDP in a cross-section of countries.

As proxies for the efficiency gains of VAT adoption, I use the share of trade taxes in total tax revenue and the inflation rate. Heavy reliance on trade and inflation taxes suggests that the current tax system is highly distortive. Assuming that VAT adoption raises the tax system's efficiency to a certain level, VAT adoption would yield higher efficiency gains in countries with more distortive tax systems. Data on GDP per capita, the size of the agricultural sector, imports, tax revenue and trade tax revenue as a share of GDP, and the inflation rate are obtained from Baunsgaard & Keen (2010). The administrative costs of VAT adoption are proxied by the adult illiteracy rate[18] obtained from the World Development Indicators.[19]

Finally, in the absence of clear predictions about the political determinants of VAT adoption, I focus on two political factors that have repeatedly been suggested as determinants of fiscal capacity investments - democracy and political stability. Besley & Persson (2009) show that greater political inclusiveness (democracy) lowers the value of redistribution and increases the value of common interest goods. A more representative regime thus has a higher value for fiscal capacity and is more likely to invest in it. Political stability fosters investments in fiscal capacity, because the ruling group's expected value of public funds increases with its ability to stay in power. The higher the group's likelihood of staying in power, the lower the chances that taxes will be used as a redistributive tool against its interests, and the higher the group's incentives to invest in fiscal capacity. To measure democracy, I use the polity score from the Polity IV dataset[20]. This indicator ranges from -10 to 10, with the highest score characterizing perfectly democratic countries. As proxy for regime stability, I use data on the number of executive changes in a given year from the Cross-National Time Series Archive.[21] When specifically analyzing the

[17] Eventually, I would also like to control for the degree of formality of the economy as a measure of VAT coverage, using data such as those collected by Schneider (2005).

[18] This is of course a very crude proxy, and disregards the fact that the administrative costs of VAT adoption depend at least partially on the level of sophistication of the existing tax system. Data on collection costs for other taxes could be a good proxy for the administrative costs of VAT adoption. At the moment, however, such data are, to the best of my knowledge, not available for a large sample of countries.

[19] http://data.worldbank.org/indicator.

[20] http://www.cidcm.umd.edu/polity/.

[21] http://www.databanksinternational.com/32.html.

political economy of VAT adoption, I also use the number of riots, anti-governmental demonstrations, coups d'Etat and cabinet changes, from the same dataset.

In my tests for robustness and omitted variable bias, I use data on the share of aid in GNI and the GDP growth rate from the World Development Indicators, as well as the debt over GNI ratio from the Global Development Finance Indicators[22]. The use of these data will be explained below.

All independent variables I use are potentially endogenous. In the absence of a theoretical model, the direction of causality between these variables and VAT adoption is unclear. I minimize endogeneity concerns by lagging the independent variables.[23] I estimate all models by separately using the first, second and third lag of the independent variables, and their average over the years $t-1$ to $t-3$. All results are robust across these specifications.

Means of all variables in 1975 and 2000 and other summary statistics are displayed in Tables 3.I (economic variables) and 3.II (political variables). The figures indicate that the countries which had adopted a VAT by 2000 had a higher level of GDP per capita at both the beginning and the end of the sample period, although the difference is not significant at the 10% level in 1975. VAT adopters also had a lower share of imports in GDP and a lower share of trade tax revenue in total tax revenue in 1975 and 2000, and these differences are significant at the 1% level. Finally, VAT adopters were significantly more democratic as measured by the polity score in 2000, and slightly so in 1975.[24] Together, these results suggest a strong correlation between VAT adoption and GDP per capita, reliance on trade taxes, and democracy. These effects will reappear in the next section, which presents the results of my empirical analysis.[25]

---

[22]http://data.worldbank.org/data-catalog/global-development-finance.

[23]Using lags avoids endogeneity bias only under the assumption that that economic agents do not anticipate VAT adoption and adjust their economic accordingly, which is unlikely to hold.

[24]It also seems as if VAT adopting countries had a significantly higher number of riots, anti-governmental demonstrations and cabinet changes in 2000, but these results are driven by outliers.

[25]An important aim for future research is the extension of the dataset, because some weaknesses in the empirical results (referred to below) are due to the short sample period. I have already compiled data on VAT adoption dates for the entire period 1948-2010. Data on the economic and political variables used are readily available from the aforementioned datasets and the Penn World Tables. The IMF dummy, however, still needs to be coded for the years before 1970 and after 2004, using information from IMF annual reports and article IV consultation reports. When doing this, I would also like to collect information on whether lending agreements contained an explicit VAT condition, and whether they were accompanied by a technical assistance package. Being able to distinguish between countries that had an IMF agreement *without* VAT conditionality, those that had an agreement *with* VAT conditionality, and those that had an agreement with conditionality *and* a technical assistance package would allow me to test for the different IMF impact channels.

Table 3.I: **Summary Statistics for Economic Variables**

Variable	Min	Max	N	Mean if VAT=1	N	Mean if VAT=0	N	t-value	Mean if VAT=1	N	Mean if VAT=0	N	t-value
		All Years				**1975***					**2000***		
GDP per Capita	85	56372	3059	6234.947	75	3557.23	27	-1.545	9703.316	82	3356.49	39	-2.744
				(1530.248)		(924.960)			(943.256)		(1209.67)		
GDP Growth	-50.2	71.188	2463	3.251	52	2.504	26	-0.443	3.085	59	3.638	41	0.688
				(0.382)		(0.790)			(0.769)		(1.829)		
Imports/GDP	0	251	2889	33.458	71	48. 827	29	3.252	38.761	67	53.714	35	3.010
				(2.54)		(4.857)			(1.853)		(5.633)		
Tax Revenue/GDP	0.1	55.6	2991	23.896	52	18.346	13	-1.754	21.760	75	18.916	42	-1.311
				(2.450)		(4.857)			(1.432)		(2.666)		
Trade Tax Rev./GDP	0	33.3	2983	2.9990	52	6.580	12	3.022	2.077	72	4.417	42	3.782
				(0.246)		(0.692)			(0.381)		(1.881)		
Inflation Rate	-13	11750	2681	21.606	66	9.703	27	-1.040	5.974	79	5.709	31	-0.110
				(1.304)		(1.922)			(7.277)		(1.166)		
Agriculture/GDP	0	74	2911	21.954	71	27.304	23	1.328	17.629	70	20.593	32	0.963
				(1.610)		(2.886)			(1.939)		(3.787)		
Adult Illiteracy Rate	1.6	93.7	2288	46.221	60	47.207	28	0.164	27.351	60	24.750	28	-.544
				(2.859)		(3.359)			(3.580)		(4.269)		
Debt/GNI	0	824.539	2023	28.850	46	19.372	18	-1.936	75.022	55	70.237	31	-0.444
				(2.541)		(4.376)			(6.450)		(8.665)		
Aid/GNI	0	211	3045	0.636	22	6.6	80	3.216	4.085	82	8.275	40	2.255
				(0.250)		(0.966)			(0.746)		(2.184)		

Notes: Standard errors are shown in parentheses, all amounts are in millions of US$, * data on inflation rate is for 1976, ** data on adult illiteracy rate is for 1999.

Table 3.II: **Summary Statistics for Political Variables**

Variable	Min	Max	N	Mean if VAT=1	N	Mean if VAT=0	N	t-value	Mean if VAT=1	N	Mean if VAT=0	N	t-value
		All Years				**1975**					**2000**		
Polity IV Score	-10	10	2693	-0.773	75	-2.961	26	-1.244	5.012	77	-0.427	28	-4.341
				(0.624)		(1.173)			(0.912)		(1.409)		
Num. of Riots	0	26	3140	0.506	77	0.093	32	-1.548	0.280	82	0.046	43	-1.834
				(0.090)		(0.032)			(0.169)		(0.068)		
Num. of Demonstrations	0	26	3142	0.480	77	0.5	32	0.057	0.927	82	0.093	43	-2.994
				(0.199)		(0.044)			(0.186)		(0.283)		
Num. of Coups d'Etat	0	2	3153	0.769	78	0.030	33	-0.804	0.024	82	0.023	43	-0.039
				(0.017)		(0.023)			(0.035)		(0.030)		
Num. of Cabinet Changes	0	5	3150	0.500	78	0.406	32	-0.706	0.512	82	0.279	43	-2.544
				(0.055)		(0.069)			(0.074)		(0.098)		
Numb. of Exec. Changes	0	4	3153	0.282	78	0.181	33	-0.790	0.182	82	0.116	43	-0.961
				(0.182)		(0.116)			(0.070)		(0.101)		

Notes: Standard errors are shown in parentheses.

3.5 Empirical Results

This section presents the results of my empirical analyses. It first discusses the baseline results, then presents the tests for omitted variables and reverse causality, and finally analyzes the impact of political factors on VAT adoption.

3.5.1 Baseline Results

Stylized facts indicate that the time span between the decision to adopt a VAT and the implementation of the VAT law varies across countries and ranges from a few months to several years. This is true even when considering only countries that adopted the VAT under IMF supervision.[26] I therefore conduct my baseline analysis using the combined dummy IMF_{-3} for an IMF intervention in one of the last three years as the key independent variable. The results are displayed in Table 3.III. In column (1), I control for the full set of economic and political determinants of VAT adoption lagged by one period. In column (2), I use the averages over year $t-1$ to $t-3$ of these variables, to ensure that their impact is not driven by short-run fluctuations. In both specifications, the impact of an IMF intervention in one of the last three years is positive and statistically significant at the 5% level. To understand the economic significance of the results, one needs to derive the relative hazard ratio of country i, which is under an IMF lending agreement, to country j, which is not under a lending agreement, all else equal:

$$\frac{h_i(t)}{h_j(t)} = exp((1-0)\beta) = exp(\beta). \tag{3.7}$$

Using the more conservative estimate from column (2), I find that a country which had an IMF intervention in one of the last three years is $2.962 = exp(1.086)$ times more likely to adopt the VAT than a country that had no intervention. The effect is thus not only statistically but also economically significant.[27] Moreover, this effect is much larger than the effect estimated in Keen & Lockwood (2010), which finds that countries that participate in a non-crisis IMF program have a 25% higher chance of

[26]Chad, for instance, adopted the VAT within a year of announcing this reform in 1999 (Letter of Intent 1999, Article IV consultation report 2010). Guyana, on the other hand, had planned VAT adoption since 2004 or earlier and achieved adoption in 2007 only (Article IV consultation reports 2004 and 2008). IMF fiscal experts suggest that a preparatory period of 18 to 24 months is necessary to set up a VAT (Tait 1988).

[27]I also analyzed the individual impacts of different types of IMF lending agreements by decomposing the IMF dummy into separate dummies for the main lending programs (Extended Funds Facility, Poverty Reduction and Growth Facility, Standby Arrangement and Structural Adjustment Facility). The coefficients of these dummies are significant in some specifications, but not in others. This is probably due to the small sample size and small number of VAT adoptions that identify each coefficient. Using the extended dataset may eventually provide a more informative picture on the relative effects of different lending programs.

adopting the VAT. The probit estimation, which allows the control variables to play a role in predicting the VAT dummy even after the adoption date, thus downward biases the true IMF effect.

Among the independent variables in this model, only the coefficients on GDP per capita and the share of imports in GDP are statistically significant, and they have the expected sign. Countries with a higher level of GDP per capita and a lower share of imports in GDP are more likely to adopt the VAT. Note that, since the baseline hazard is not identified in this model, there is no measure for the goodness of fit.

As a robustness test, I estimate the baseline model of different subsamples to examine whether the IMF effect is driven by a specific group of countries. In particular, countries with a low level of GDP per capita, low tax revenue and high aid inflows may be more dependent on IMF lending and thus more susceptible to IMF influence. I therefore estimate the baseline model excluding countries placed in the upper or bottom quartile of the distribution of these variables in 1975. In these estimations, I control only for the independent variables which showed some sign of significance in the baseline results, GDP per capita and imports as a share of GDP in $t-1$. This limits the loss of observations and reduces the standard errors, but does not affect the size of the coefficient on IMF_{-3}. Column (3) of Table 3.III shows the results of this model for the full sample and columns (4)-(9) show the results for the different subsamples.[28] The figures demonstrate that the impact of an IMF intervention in one of the last three years is robust to excluding countries with particularly high or low GDP per capita, tax revenue, and aid as a share of GNI. All coefficients are significant at least at the 5% level and the size of the coefficients does not vary significantly across subsamples. This shows that the size and significance of the coefficient on IMF lending is not limited to specific subsamples.[29] However, my identification strategy does not allow me to conclude that this is evidence of a causal effect of the IMF intervention. To shed more light on the direction of causality, I test for the presence of omitted variables and reverse causality.

[28] Note that the number of observations differs across subsamples, due to missing values and discontinuity in the variables I use to define the subsamples.

[29] I also find that the effect of IMF lending is not limited to particular points in time (results available upon request). First, I show that the effect of an IMF intervention in the last three years is robust to excluding the first five and last five years of the sample period. Second, I estimate a model with two IMF dummies, one for the first three years of lending, and one for all other years (IMF_oth_{-3}) and this yields a coefficient on IMF_oth_{-3}, which is significant at the 5% level. The coefficients on the dummy for the first three years of lending is insignificant due to high standard errors. Results are similar when singling out the first four or five years of lending. This shows that the effect of IMF lending is not limited to the first three to five years of lending.

Table 3.III: **Baseline Results**

Variable	Lagged Controls	Averaged Controls	Reduced Controls	High GDPpc	Low GDPpc	High Taxrev	Low Taxrev	High Aid/GNI	Low Aid/GNI
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IMF_{-3}	1.321***	1.086**	1.340***	1.480***	1.711***	1.155**	1.018**	1.515***	1.202**
	(0.508)	(0.475)	(0.404)	(0.458)	(0.505)	(0.494)	(0.514)	(0.502)	(0.500)
$GDPpc_{t-1}$	0.000*	0.000**	0.000	0.000	0.001**	0.000***	0.000**	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
IMP/GDP_{t-1}	-0.025***	-0.033***	-0.016**	-0.014*	-0.027***	-0.033***	-0.028**	-0.025***	-0.009
	(0.008)	(0.012)	(0.008)	(0.007)	(0.009)	(0.012)	(0.013)	(0.008)	(0.008)
$Tradetaxrev_{t-1}$	-1.086	0.862							
	(1.449)	(1.721)							
$Inflation_{t-1}$	-0.004	0.003							
	(0.008)	(0.010)							
$Agric/GDP_{t-1}$	-0.027	-0.025							
	(0.017)	(0.019)							
$Illiteracy_{t-1}$	0.012	0.010							
	(0.010)	(0.011)							
$Polity_{t-1}$	0.006	-0.001							
	(0.032)	(0.033)							
EX_CHA_{t-1}	0.473	0.467							
	(0.316)	(0.544)							
Observations	850	728	1484	1186	1034	648	560	1066	868

Notes: The table shows the results of a Cox hazard model estimation of VAT adoption. The dependent variable is a dummy that switches on once a country has adopted the VAT. The key independent variable is IMF_{-3}, a dummy indicating whether a country had an IMF intervention in the past three years. The control variables are explained in section 3.4. Columns (1)-(3) present the results for the full sample, with different specifications of the control variables (all controls as simple lags, all controls as averages over $t-1$ to $t-3$, and a reduced set of simple lagged controls). Columns (4)-(9) present results for different subsamples, as explained in section 3.III. Robust standard errors, clustered at the country level, are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

3.5.2 Testing for Omitted Variables

Omitted variables are a key challenge to my identification strategy. Indeed, any shock that causes both VAT adoption and IMF intervention would result in a spurious correlation between these two events. An external debt crisis is most obviously correlated with IMF intervention and possibly VAT adoption. A debt crisis often requires an IMF intervention to provide short-term lending and re-establish investors' confidence. It may also push a country to adopt a VAT to increase tax revenue and service the debt. Shocks to tax revenue, imports or GDP growth may also destabilize public finances in a way that necessitates both an IMF intervention and VAT adoption. I therefore test whether including proxies for these shocks in my regressions reduces the significance of the IMF impact on VAT adoption. If the IMF effect in the baseline results was driven by omitted variables, it should disappear once the omitted variables are included.

As debt crisis proxy, I use the dummy $Debt\_S$ to indicate whether the difference in the external debt to GNI ratio between year t and $t-1$ was in the upper decile of the distribution of differences. The dummy is correlated with the debt crisis dummy used in Reinhart & Rogoff (2009) at the 1% level of significance, which suggests that it is a good proxy. I use this proxy instead of the dummy from Reinhart & Rogoff (2009), because the latter is defined for only one-third of the country-year points in my sample. I also use dummy variables for tax revenue and imports shocks, $Taxrev\_S$ and IMP/GDP_S. These dummies indicate whether the change in the tax to GDP ratio and in the imports to GDP ratio respectively were in the lower decile of the distribution. The choice of the upper and lower deciles as the cut-off levels for these dummies is of course an arbitrary one, but the results are robust to using the upper and lower quantiles and quartiles as cut-off levels. Finally, I also control for growth shocks by using the growth rate, a dummy for negative growth, and a dummy for a growth rate less than -5%. For the debt, tax revenue, imports and growth shocks, I use a three-year combined dummy that indicates whether the respective shock took place in one of the last three years. This ensures that I capture the impact of a shock on VAT adoption, even if it does not materialize immediately. In all regressions, I control for the first lag of GDP per capita and imports as a share of GDP.

Table 3.IV: **Testing for Omitted Variables**

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IMF_{-3}	1.340***	1.097**	1.050***	1.101***	1.341***	1.355***	1.361***
	(0.404)	(0.483)	(0.390)	(0.389)	(0.404)	(0.399)	(0.400)
$Debt\_S_{-3}$		0.501					
		(0.369)					
$Taxrev\_S_{-3}$			0.148				
			(0.331)				
IMP/GDP_S_{-3}				0.088			
				(0.358)			
GDP_growth_{t-1}					0.005		
					(0.018)		
Neg_growth_{-3}						0.176	
						(0.271)	
$Growth\_-5_{-3}$							-0.483
							(0.636)
Observations	1484	1146	1263	1336	1474	1450	1447

Notes: The table shows the results of a Cox hazard model estimation of VAT adoption. The dependent variable is a dummy that switches on once a country has adopted the VAT. The key independent variable is IMF_{-3}, a dummy indicating whether a country had an IMF intervention in the past three years. The coefficients on the control variables $GDPpc_{t-1}$ and IMP/GDP_{t-1} are omitted. Columns (2)-(7) add indicators for various macroeconomic shocks that could trigger an IMF intervention, as discussed in section 3.5.2. Robust standard errors, clustered at the country level, are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The results are displayed in Table 3.IV. They show that none of the shock variables predicts VAT adoption and that the inclusion of these shocks does not alter the size and significance of the IMF dummies. An IMF intervention in the past three years roughly triples the likelihood of VAT adoption even in the presence of these shocks. This would suggest that my results are not driven by omitted variables. There are two reasons, however, why including these shock dummies may not reduce the significance of the IMF coefficients, even if the IMF effect is driven by omitted variables. First, the effect may be driven by omitted variables other than those that I control for. A second concern is that the dummies I use may not be good proxies for the shocks for which I want to control.

In conclusion, although the tests I conduct do not allow me to completely rule out the existence of omitted variable bias, they at least indicate that proxies for the shocks most commonly related to IMF interventions do not alter the observed IMF effect on VAT adoption. This still leaves the possibility that the results are driven by reverse causality, the second challenge to my identification strategy.

3.5.3 Testing for Reverse Causality

In order to investigate the presence of reverse causality, I conduct a falsification test by estimating the baseline model with a measure of future IMF interventions on the right-hand side. If the line of causality runs from the IMF intervention to VAT adoption, and if there are no anticipatory effects of future IMF interventions on VAT adoption, the coefficient on the lead IMF interventions should not be significant.[30] There are, however, two problems with such a test. First, there are reverse causality scenarios in which the IMF intervention and VAT adoption take place simultaneously: for example if a country requests an IMF intervention in order to adopt the VAT. Therefore, insignificance of the coefficient on lead IMF interventions is no *clear* sign for the absence of reverse causality. The test I conduct is thus a very crude falsification test. Second, as the IMF program lasts from one to four years, and contracts are often renewed, there is significant autocorrelation in IMF interventions. This could result in a significant correlation between VAT adoption and future IMF interventions, even in the absence of reverse causality. I therefore

[30] While possible in theory, anticipatory effects are highly unlikely in practice. There is no evidence for ex ante conditionality in the sense that the IMF makes VAT adoption a condition for future intervention (although some programs such as the Flexible Credit Line are available only to countries that have shown good fiscal performance, cf.http://www.imf.org/external/np/exr/facts/fcl.htm). Article IV consultation reports that precede a lending agreement occasionally mention the adoption of a VAT, but do not make it an explicit condition for the conclusion of a lending agreement. Besides, since the process of adopting the VAT can span several years, and countries that apply for an IMF loan need financial support immediately, it is unlikely that they would agree to VAT adoption as an ex ante condition.

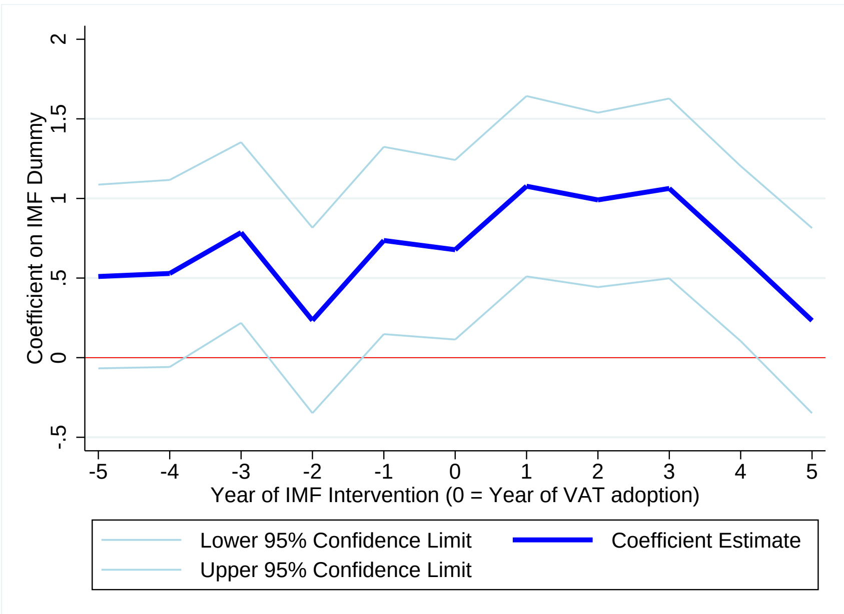
conduct the falsification test with two different measures of future IMF interventions: the lead of the previously employed IMF dummy, and the lead of a proxy for IMF program "onset," which takes into account only the first or first two years of each IMF program.

The results using the lead IMF dummy, IMF_{+3}, are displayed in columns (1) and (2) of Table 3.V. This dummy is constructed analogously to the dummy IMF_{-3} used in the previous section, and indicates whether the country was under an IMF lending agreement in one of the years $t+1$, $t+2$ and $t+3$. In column (1), I control for the full set of economic and political determinants of VAT adoption, and in column (2) only for GDP per capita and imports, lagged by one period. As the figures show, the coefficient of an IMF intervention in one of the forthcoming three years is highly significant, suggesting a strong correlation between VAT adoption and future IMF interventions. In an attempt to further explore the time pattern of this correlation, I also estimate the baseline model with simple rather than combined leads and lags of the IMF dummy. Here I control only for the first lag of GDP per capita and imports. The coefficients of the IMF dummy from the fifth lag to the fifth lead are plotted in Figure 3.III. In this figure, the year of VAT adoption is set to zero and the coefficient on an IMF intervention z years before (after) VAT adoption is plotted at x-coordinate -z (z). The coefficients are significant from the third lag to the third lead, except for the second lag.[31] This confirms that both IMF interventions before and after the VAT adoption date are significantly correlated with VAT adoption.

One possible explanation for this could be the presence of omitted variables. However, the results discussed in the previous section already showed that including potentially omitted variables does not alter the effect of the lagged IMF dummy. The same is true for the lead IMF dummy. The correlation of VAT adoption and an IMF intervention in the following three years is of similar size to the correlation of the lagged IMF dummy, and the size and significance of the correlation changes little when proxies for various types of shocks are controlled for.

[31] The insignificance of the second lag is surprising but can safely be regarded as random due to the small sample size. Indeed, since my sample includes only 60 VAT adopters, the coefficients of the IMF dummies are sensitive to small changes in the number of observations that identify them. For the first three years after VAT adoption, the year of adoption, and the first and third year before adoption, the number of countries that experienced an IMF intervention in that specific year lies between 24 and 26. However, only 19 countries had an IMF intervention two years before VAT adoption. I expect the insignificance of the second lag to disappear once I use data that covers a larger number of VAT adoptions.

Figure 3.III: **Coefficients on IMF Dummies**



Notes: The figure plots the coefficients on the IMF dummy (=1 if country is under an IMF intervention) in a Cox hazard model of VAT adoption, for IMF interventions 5 years before and after the year of VAT adoption. The estimations control for GDP and imports in year $t-1$.

Table 3.V: **Testing for Reverse Causality**

Variable	Full Controls	Reduced Controls	Full controls				Reduced Controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IMF_{+3}	1.520***	1.242***								
	(0.500)	(0.352)								
IMF_onset1_{-3}			0.592				0.539*			
			(0.380)				(0.311)			
IMF_onset1_{+3}				0.511				0.277		
				(0.367)				(0.317)		
IMF_onset2_{-3}					0.965**				0.838**	
					(0.442)				(0.339)	
IMF_onset2_{+3}						0.608				0.386
						(0.414)				(0.317)
Observations	837	1484	837	837	837	837	1484	1484	1484	1484

Notes: The table shows the results of a Cox hazard model estimation of VAT adoption. The dependent variable is a dummy that switches on once a country has adopted the VAT. In columns (1) and (2), the independent variable is IMF_{+3}, a dummy indicating whether a country had an IMF intervention in years $t+1$ to $t+3$. Columns (3)-(10) test the effect of the IMF_onset dummy, which is discussed in Section 3.5.3 and indicates whether the country experienced the onset (first year or first two years) of an IMF intervention in the last three or following three years (indicated by subscript -3 and $+3$ respectively). The coefficients on the control variables are omitted. For columns (1)-(6), the control variables are as in Table 3.III, column (1). For columns (7)-(10), the control variables are as in Table 3.III, column (3). Robust standard errors, clustered at the country level, are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Another possible explanation is autocorrelation in IMF interventions. To examine whether the correlation between VAT adoption and future IMF interventions persist once autocorrelation in IMF interventions is controlled for, I code a measure of IMF program onset. Unfortunately, my data do not allow me to cleanly distinguish among different IMF lending agreements. I can distinguish among four different *types* of lending programs[32], but cannot distinguish between different lending agreements in the same program, unless they are separated by a year without lending. I thus code program-specific dummies that take into account only the first year (first two years) of a continuous period of lending, and then sum up these program dummies to obtain the dummies IMF_onset1, for the first year of each lending period, and IMF_onset2, for the first two years of each lending period.[33] These dummies are an imperfect measure of program "onset" because they ignore lending agreements that immediately follow the previous agreement. Approximately 11% of the country-year points experience the first year of an IMF lending agreement during the sample period, and 14.5% experience the first or second year of an agreement.

The results using the combined three-year lag and lead versions of these IMF onset dummies are presented in columns (3)-(10) of Table 3.V. In columns (3)-(6), I use the full set of controls and in columns (7)-(10), I use only GDP per capita and imports, all lagged by one period. With the full set of controls, countries that experienced the first year of an IMF lending agreement in years $t-1$ to $t-3$ (column (3)) are not more likely to introduce the VAT in year t (probably due to the small number of observations that identify this coefficient), but countries that experienced the first *or second* year of a lending agreement in the past three years (column (5)) are more likely to introduce the VAT at the 5% level of significance. More importantly, countries that will experience the onset of IMF lending in the forthcoming three years (columns (4) and (6)) are *not* more likely to introduce the VAT. The results are similar but stronger when using only the reduced set of controls. Countries that experienced the first or first two years of a lending agreement in the past three years (columns (7) and (9)) are more likely to adopt the VAT at least at the 10% level, but countries that experience program onset in the forthcoming three years (columns (8) and (10)) are not more likely to introduce the VAT. Thus, even an imperfect measure of program onset, which does not take into account all lending agreements, shows that lagged IMF interventions are significantly correlated to VAT adoption, but that future IMF interventions are not. This provides credence to the hypothesis that the line of causality runs from IMF interventions to VAT adoption and not the reverse.

[32]Extended Fund Facility, Poverty Reduction and Growth Project, Standby Agreement and Structural Adjustment Facility

[33]No country was ever under two different programs at the same time, so the sum of the program dummies is still a dummy variable.

However, even if I can show that future IMF interventions are not correlated to VAT adoption, this leaves the possibility that the IMF effect is driven by *simultaneous* reverse causality, in the sense that the IMF is more likely to intervene in countries that *plan* to adopt the VAT. This is a challenge that cannot be addressed in the Cox hazard framework. An instrumental variable strategy would be the natural way to correctly identify the impact of an IMF intervention on VAT adoption. Barro & Lee (2005) and Easterly (2005), for example, use various geopolitical factors as instruments for IMF lending. However, two-stage methods for hazard models are only in a nascent stage of development, mainly due to the non-linear nature of these models (Box-Steffensmeier & Jones 2004).

3.5.4 The Political Economy of VAT Adoption

Given that VAT adoption is a politically sensitive topic, it is interesting to examine how the political context affects the interplay between VAT adoption and IMF interventions. As mentioned before, I focus here on the effect of democracy and regime stability, which have been identified by Besley & Persson (2009) as factors promoting fiscal investments. However, instead of using proxies for democracy and stability, I use proxies for *auto*cracy and *in*stability, and analyze the direct impact of these factors on VAT adoption as well as their interaction with the IMF effect. This allows me to investigate the idea brought forth by Vreeland (2003), according to which an IMF intervention may help countries to implement unpopular reforms even if the political context is not favorable to that implementation. A loose interpretation of this idea would suggest that an IMF intervention might attenuate the negative effect of autocracy and political instability on VAT adoption. One would thus expect the coefficients on the autocracy and instability measures to have a negative sign and the interaction term between each of these factors and the IMF dummy to have a positive sign.[34]

I use three different measures of autocracy, all lagged by one period:[35] the negative of the polity IV score Neg_Polity_{t-1}, an autocracy dummy indicating whether the polity score was weakly negative, $Autoc0_{t-1}$, and a autocracy dummy indicating whether the polity score was below 5, $Autoc5_{t-1}$. These are the cut-off levels traditionally used in the economics and political science literature respectively (Besley & Kudamatsu 2008; Persson & Tabellini 2008; Fearon 2008). Again, I control for GDP per capita and imports as a share of GDP in period $t-1$. Columns (1)-(3) of Table 3.VI display the effect of these autocracy measures on VAT adoption.

[34] This obviously remains a rather intuitive prediction, and a theoretical treatment of the political determinants of VAT adoption is ultimately needed to shed more light on the lines of causality, as I explain below.

[35] The results are similar when using the second or third lag of the autocracy measures.

For each of the three measures, the coefficients of the autocracy indicator and of its interaction with the IMF dummy are both highly significant, of approximately the same size and of opposing signs. The autocracy measure thus cancels out when the country is under an IMF lending agreement. This can be interpreted as follows: in the presence of an IMF intervention, the degree of autocracy has no impact on the likelihood of VAT adoption. Absent an IMF intervention, however, autocratic countries are less likely to adopt the VAT. It thus seems that an IMF intervention attenuates the negative effect of autocracy on VAT adoption.[36]

The effect of different measures of political instability on VAT adoption is displayed in columns (4)-(9) of Table 3.VI. As instability measures, I use the number of riots, anti-governmental demonstrations, coups d'Etat, cabinet changes and changes in the executive in the last three years. The coefficients of all instability measures and their interaction with the IMF dummy are insignificant, and these results are robust to using the simple lag of the instability measures, their five-year averages, or other instability measures such as the variance of the polity score, or the number of changes in the polity score over the last five years (results available upon request). This means that political instability does not affect VAT adoption, and the IMF impact on VAT adoption does not differ across countries according to their degree of political instability.

[36]Whether an IMF intervention has a beneficial effect on VAT adoption in non-autocratic countries is unclear. The first specification, in which the IMF dummy is highly significant, suggests that this is the case. In the second and third specifications, however, in which I use autocracy dummies rather than the continuous autocracy measure, the coefficient of the IMF dummy is insignificant. This would mean that democracies under IMF lending are no more likely to introduce the VAT than democracies without IMF lending. Verification with cross-tabulate revealed that the problem is not one of lack of variation (too few observations in the off-diagonal cells). Still, using a larger sample might allow me to overcome these empirical irregularities.

Table 3.VI: **The Political Economy of VAT Adoption**

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IMF_{-3}	1.531***	0.264	0.099	1.351***	1.339***	1.348***	0.976**	1.459***
	(0.511)	(0.484)	(0.643)	(0.415)	(0.445)	(0.411)	(0.444)	(0.460)
Variable	Neg_Polity_{t-1}	$Autoc0_{t-1}$	$Autoc5_{t-1}$	$RIOTS_{-3}$	$ANTIGOV_{-3}$	$CDETAT_{-3}$	CAB_CHA_{-3}	EX_CHA_{-3}
Coefficient	-0.166***	-2.007**	-1.656**	0.029	0.061	0.218	-0.164	0.227
	(0.056)	(0.8987)	(0.739)	(0.045)	(0.065)	(0.829)	(0.246)	(0.322)
Interaction Effect	0.177***	-2.396**	1.944**	-0.035	-0.032	-0.562	0.250	-0.347
	(0.062)	(0.956)	(0.924)	(0.042)	(0.065)	(0.921)	(0.267)	(0.366)
Observations	1228	1228	1228	1412	1415	1419	1419	1419

Notes: The table shows the results of a Cox hazard model estimation of VAT adoption. The dependent variable is a dummy that switches on once a country has adopted the VAT. The key independent variable is IMF_{-3}, a dummy indicating whether a country had an IMF intervention in the past three years. The coefficients of the control variables, as shown in Table 3.III column (1), are omitted. In each column, a different political dummy variable (specified in row (3) and discussed in section 3.5.4) as well as its interaction with the IMF dummy is added to the estimation. Rows (4) and (5) show the coefficient of the political dummy and of its interaction with the IMF dummy. Robust standard errors, clustered at the country level, are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

In summary, the effects of autocracy and regime instability on VAT adoption are weaker than those predicted by the fiscal investment literature. Autocracy reduces the likelihood of VAT adoption, but this effect is attenuated by an IMF intervention, and regime instability seems to have no impact on adoption. These results are robust to using various different proxies for autocracy and instability. It is thus unlikely that they are driven by measurement errors in the variables used. A more plausible explanation is that VAT adoption is inherently different from other types of fiscal investments and thus not driven by the same determinants. In particular, the IMF seems to play an overarching role in promoting the VAT and possibly crowds out the effects of other political and economic determinants of VAT adoption. Overall, these findings highlight the importance of developing a theoretical framework to study how the political economy context affects VAT adoption and the interplay between VAT adoption and IMF lending.

3.6 Conclusion

This paper has analyzed the impact of IMF lending on VAT adoption in a hazard model framework. It finds that IMF lending approximately triples a country's likelihood of adopting the VAT. I show that this result is not driven by omitted variables such as debt crises and shocks to growth, tax revenue and imports, and I provide suggestive evidence against reverse causality. I also find that IMF interventions reduce the negative effect of autocracy on VAT adoption. My methodological approach faces two shortcomings, which open directions for future research.

First, I do not have clear theoretical predictions about the mechanisms through which the IMF affects VAT adoption. A discussion of the context suggests that the IMF can influence VAT adoption through lending conditionality, the provision of technical assistance, and a reduction of the political costs of adoption. Embedding these ideas into a political agency model would allow for a better understanding of these different impact mechanisms. The model could then yield testable predictions of the relative importance of the different IMF impact channels, and their interplay with the political economy context.

The second shortcoming is that the empirical strategy does not allow me to cleanly identify a causal effect of IMF interventions on VAT adoption. I provide evidence against omitted variable bias and show that future IMF interventions are not correlated with VAT adoption, once autocorrelation in IMF interventions is controlled for. However, this still leaves the possibility that the results are driven by simultaneous reverse causality. Investigating the IMF's influence on other types of fiscal reforms in developing countires might provide better opportunities for identifying a causal effect.

References

Acemoglu, Daron, Ticchi, Davide, & Vindigni, Andrea. 2006 (December). *Emergence and Persistence of Inefficient States.* NBER Working Papers 12748. National Bureau of Economic Research.

Agha, Ali, & Haughton, Jonathan. 1996. Designing Vat Systems: Some Efficiency Considerations. *The Review of Economics and Statistics*, **78**(2), 303 – 308.

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

Almunia, Miguel, & Lopez-Rodriguez, David. 2012. *The Efficiency Cost of Tax Enforcement: Evidence from a Panel of Spanish Firms.* Working Paper.

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

Assibey-Mensah, George O. 1999. The Value-Added Tax in Ghana. *Public Budgeting and Finance*, **19**(2), 76 – 89.

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., & Hassett, Kevin. 1991. Recent U.S. investment behavior and the tax reform act of 1986: A disaggregate view. *Carnegie-Rochester Conference Series on Public Policy*, **35**(1), 185–215.

Auerbach, Alan J., Chaney, Eric, & Hassett, Kevin. 2008. *Dividend Taxes, Partial Expensing and Business Fixed Investment: The Case of the Bush Tax Cuts.* Working Paper.

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 Trainls, Tax Compliance and Development*. Mimeo.

Barro, Robert J., & Lee, Jong-Wha. 2005. IMF programs: Who is chosen and what are the effects? *Journal of Monetary Economics*, **52**(7), 1245 – 1269. Political economy and macroeconomics.

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

Besley, Timothy, & Kudamatsu, Masayuki. 2008. Making autocracy work. *Chap. 11 of: Institutions and economic performance, Elhanan Helpman (Ed.)*. Cambridge, Massachusetts: Harvard University Press.

Besley, Timothy, & Persson, Torsten. 2009. The Origins of State Capacity: Property Rights, Taxation, and Politics. *American Economic Review*, **99**(4), 1218 – 1244.

Besley, Timothy, & Persson, Torsten. 2010. State Capacity, Conflict and Development. *Econometrica*, **78**(1), 1 – 34.

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.

Best, Michael, Brockmeyer, Anne, Kleven, Henrik, Spinnewijn, Johannes, & Waseem, Mazhar. 2013. *Production vs Revenue Efficiency with Low Tax Compliance: Theory and Evidence from Pakistan*. Working Paper.

Bird, Richard M., & Gendron, Pierre-Pascal. 2007. *The VAT in Developing and Transitional Countries*. Cambrdige: Cambridge University Press.

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

1(1), 1–27.

Box-Steffensmeier, Janet M., & Jones, Bradford S. 2004. *Event History Modeling: A guide for Social Scientists*. Cambridge: Cambridge University Press.

Burgess, Robin, & Stern, Nicholas. 1993. Taxation and Development. *Journal of Economic Literature*, **31**(2), 762 – 830.

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.

Cleves, Mario A., Gould, William G., & Gutierrez, Roberto G. 2004. *An Introduction to Survival Analysis Using Stata*. Stata Press.

Cnossen, Sijbren. 1994. Administrative and comlicance costs of the VAT: A review of evidence. *Tax Notes*, **63**, 1609 – 1626.

Cohen, Darrel, & Cummins, Jason. 2006. *A retrospective evaluation of the effects of temporary partial expensing*. Tech. rept.

Conway, Patrick. 1994. IMF lending programs: Participation and impact. *Journal of Development Economics*, **45**(2), 365 – 391.

Cox, David R. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**(2), 187 – 220.

Cox, David R. 1975. Partial Likelihood. *Biometrika*, **62**(2), 269 – 276.

Cummins, Jason G., Hassett, Kevin A., Hubbard, R. Glenn, Hall, Robert E., & Caballero, Ricardo J. 1994. A Reconsideration of Investment Behavior Using Tax Reforms as Natural Experiments. *Brookings Papers on Economic Activity*, **1994**(2), 1–74.

Cummins, Jason G., Hassett, Kevin A., & Hubbard, R. Glenn. 1996. Tax reforms and investment: A cross-country comparison. *Journal of Public Economics*, **62**(1-2), 237–273.

Desai, Mihir A., & Hines, James R. Jr. 2005. *Value-added taxes and International Trade: The Evidence*. mimeo.

Devereux, Michael, & Liu, Li. 2013. *Small Business Incorporation: the Role of Corporation Taxation*. Working Paper.

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., & Sorensen, 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.

Dicks-Mireaux, Louis, Mecagni, Mauro, & Schadler, Susan. 2000. Evaluating the effect of IMF lending to low-income countries. *Journal of Development Economics*, **61**(2), 495 – 526.

Dreher, Axel. 2006. IMF and Economic Growth: The Effects of Programs, Loans, and Compliance with Conditionality. *World Development*, **34**(5), 769 – 788.

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).

Easterly, William. 2005. What did structural adjustment adjust?: The association of policies and growth with repeated IMF and World Bank adjustment loans. *Journal of Development Economics*, **76**(1), 1 – 22.

Ebrill, Liam, Keen, Michael, Bodin, Jean-Paul, & Summers, Victoria. 2001. *The Modern VAT*. Washington, D.C.: International Monetary Fund.

Efron, Bradley. 1977. The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, **72**(359), 557 – 565.

Emran, M. Shahe, & Stiglitz, Joseph E. 2005a. *On Selective Indirect Tax Reform in Developing Countries*. Tech. rept.

Emran, Shahe M., & Stiglitz, Joseph E. 2005b. On selective indirect tax reform in developing countries. *Journal of Public Economics*, **89**(4), 599 – 623. Cornell - ISPE Conference on Public Finance and Development.

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

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

Fearon, James D. 2008. Evidence on paths to and from democracy. *Chap. 7 of: Institutions and economic performance, Elhanan Helpman (Ed.)*. Cambridge, Massachusetts: Harvard University Press.

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.

Gordon, Roger, & Slemrod, Joel. 2000. Are 'Real' Responses to Taxes Simply Income Shifting Between corporate and Personal Tax Bases? *Pages 240–280 of:* Slemrod, Joel (ed), *Does Atlas Shrug? The Economics Consequences of Taxing the Rich*. Russel Sage Foundation and Harvard University Press.

Gruber, Jonathan, & Rauh, Joshua. 2007. How Elastic is the Corporate Income Tax Base? *In:* an Alan Auerbach, James R. Hines, & 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*.

HMRC. 2012. *Measuring Tax Gaps 2012, Tax Gap Estimates for 2010-11*. Statistical Report by Her Majesty Revenue and Customs.

House, Christopher L., & Shapiro, Matthew D. 2008. Temporary Investment Tax Incentives: Theory with Evidence from Bonus Depreciation. *American Economic Review*, **98**(3), 737–68.

Jenkins, Glenn, Jenkins, Hatice, & Kuo, Chun-Yan. 2006. *Is the Value-Added Tax Naturally Progressive?* mimeo.

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.

Keen, Michael, & Lockwood, Ben. 2006. Is the VAT a Money Machine?. *National Tax Journal*, **59**(4), 905 – 928.

Keen, Michael, & Lockwood, Ben. 2010. The value added tax: Its causes and consequences. *Journal of Development Economics*, **92**(2), 138 – 151.

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, Soren, & 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.

Lancaster, Tony. 1990. *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.

Marchesi, Silvia. 2003. Adoption of an IMF programme and debt rescheduling. An empirical analysis. *Journal of Development Economics*, **70**(2), 403 – 423.

Osei, Philip D. 2000. Political Liberalisation and the Implementation of Value Added Tax in Ghana. *The Journal of Modern African Studies*, **38**(2), 255 – 278.

Persson, Torsten, & Tabellini, Guido. 2008. The growth effect of democracy: Is it heterogenous and how can it be estimated. *Chap. 13 of: Institutions and economic performance, Elhanan Helpman (Ed.)*. Cambridge, Massachusetts: Harvard University Press.

Piggott, John, & Whalley, John. 2001. VAT Base Broadening, Self Supply, and the Informal Sector. *The American Economic Review*, **91**(4), 1084 – 1094.

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

Przeworski, Adam, & Vreeland, James R. 2000. The Effect of IMF Programs on Economic Growth. *Journal of Development Economics*, **62**(2), 385 – 421.

Reinhart, Carmen, & Rogoff, Kenneth. 2009. *This Time is Different: Eight Centuries of Financial Folly*. Princeton, New Jersey: Princeton University Press.

Saez, Emmanuel. 2004. Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications. *Pages 117–174 of: Tax Policy and the Economy, Volume 18*. NBER Chapters. National Bureau of Economic Research, Inc.

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.

Schneider, Friedrich. 2005. *Shadow Economies of 145 Countries all over the World: Estimation Results over the Period 1999 to 2003.* mimeo.

Seim, David. 2012. *Real or Evasion Responses to the Wealth Tax? Theory and Evidence from Sweden.* Working Paper.

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.

Tait, Alan A. 1988. *Value-added tax: International practice and problems.* Washington, D.C.: International Monetary Fund.

Tanzi, Vito. 1990. *The IMF and tax reform.* IMF Working Paper 90/39.

Tanzi, Vito. 1992. Structural factors and tax revenue in developing countries: A decade of evidence, Ian Goldin & Alan Winters (Ed.). *Chap. 12 of: Open economies: Structural adjustment and agriculture.* Cambridge University Press.

Vreeland, James R. 1999. *The IMF: Lender of last resort or scapegoat?* mimeo.

Vreeland, James R. 2003. Why do governments and the IMF enter into agreement? Statistically selected cases. *International Political Science Review*, **24**(3), 321 – 343.

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

Wei, Shang-Jin, & Zhang, Zhiwei. 2010. Do external interventions work? The case of trade reform conditions in IMF supported programs. *Journal of Development Economics*, **92**(1), 71 – 81.

Wooldridge, Jeffrey W. 2002. *Econometric Analysis of Cross-Section and Panel Data.* Cambridge, Massachusetts: MIT Press.

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

APPENDICES

Appendix A

Appendix for Chapter 2

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 (2.19), for the subsamples considered in table 2.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 2.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 2.II ($q = 5$). Bootstrapped standard errors are shown in parentheses.

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 2.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 2.IV. Column (1)-(3) repeat the first three columns of table 2.II. Column (1) reproduces the bunching estimate b, based on estimating equation (2.19). 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 2.3.3. 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

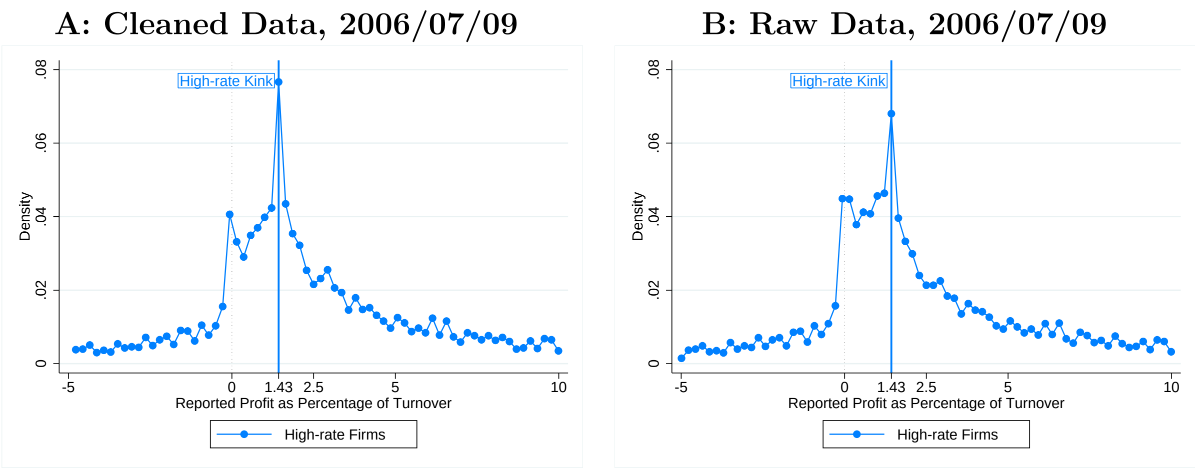
Notes: Panel A of this table explains the consistency checks applied to the data. For all consistency checks, a tolerance threshold of 5% is used. Panel B displays the sample size for different steps in the cleaning process. Capital K is equity plus retained earnings. Note that the implied turnover tax 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).

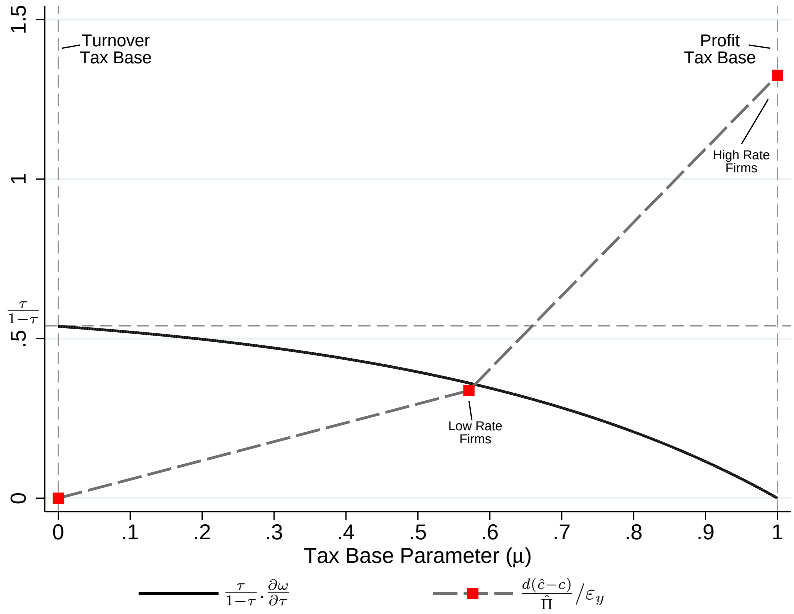
Figure A.I: **Bunching among High-rate Firms**



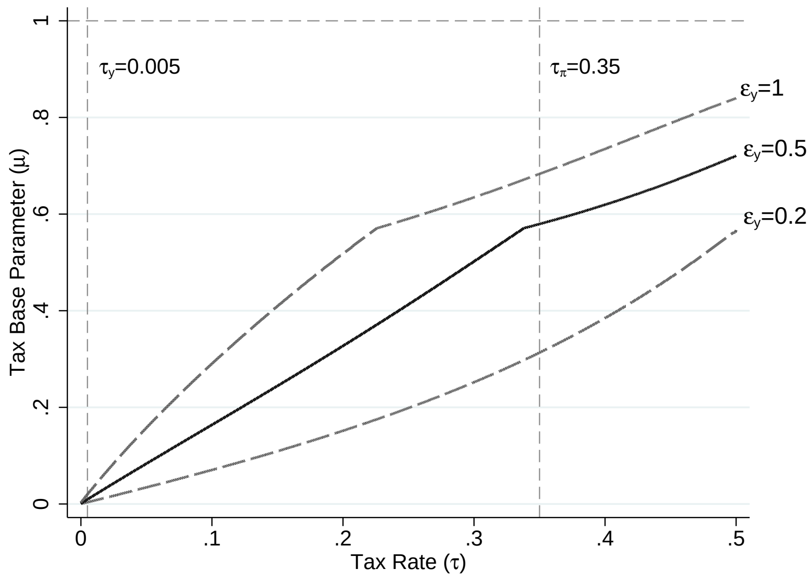
Notes: This figure shows bunching in the sample of high-rate firms in 2006/07/09, in the cleaned data (panel A, same as panel A in figure 2.II), and the raw data (panel B). The raw data contains all observations that report turnover and profits, before consistency checks I and II are applied, and before 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 2.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 Analysis of Optimal Tax Policy with Distortionary Profit Tax**

A: Optimal Tax Rule



B: Tax Base vs. Tax Rate



Notes: This figure analyzes the robustness of the numerical policy analysis shown in Figure 2.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 2.11) 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.

Appendix B

Appendix for Chapter 3

B.1 List of Countries and VAT Adopters

Countries in the sample: Algeria, Antigua and Barbuda, Argentina, Australia, Austria, Bahamas, Bahrain, Bangladesh, Barbados, Belgium, Belize, Benin, Bhutan, Bolivia, Botswana, Brunei, Burkina Faso, Burundi, Cameroon, Canada, Central African Rep., Chad, Chile, China, Colombia, Comoros, Congo, Rep., Costa Rica, Côte d'Ivoire, Denmark, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Ethiopia, Fiji, Finland, France, Gabon, Gambia, Germany, Ghana, Greece, Grenada, Guatemala, Guinea, Guyana, Haiti, Honduras, Iceland, India, Indonesia, Iran, Ireland, Italy, Jamaica, Japan, Jordan, Kenya, Kiribati, Korea, Kuwait, Lesotho, Luxembourg, Madagascar, Malawi, Malaysia, Maldives, Mali, Mauritania, Mauritius, Morocco, Mozambique, Myanmar, Namibia, Nepal, Netherlands, New Zealand, Niger, Nigeria, Norway, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Portugal, Rwanda, Samoa, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Singapore, Solomon Islands, South Africa, Spain, Sri Lanka, St. Kitts-Nevis, St. Lucia, St. Vincent and Gr., Suriname, Swaziland, Sweden, Switzerland, Syria, Tanzania, Thailand, Togo, Tonga, Trinidad and Tobago, Tunisia, Uganda, United Kingdom, United States, Uruguay, Vanuatu, Venezuela, Zambia, Zimbabwe.

VAT adopters and adoption year: Algeria (1992), Australia (2000), Bangladesh (1991), Barbados (1997), Benin (1991), Burkina Faso (1993), Cameroon (1999), Canada (1991), Chad (2000), China (1994), Congo, Rep. (1997), Dominican Republic (1983), Egypt (1991), Fiji (1992), Finland (1994), Gabon (1995), Ghana (1998), Greece (1987), Guatemala (1983), Guinea (1996), Haiti (1982), Honduras (1976), Iceland (1990), Indonesia (1985), Japan (1989), Kenya (1990), Korea (1977), Madagascar (1994), Malawi (1989), Mali (1991), Mauritania (1995), Mauritius (1998), Morocco (1986), Mozambique (1999), Namibia (2000), Nepal (1997), New Zealand (1986), Niger (1986), Nigeria (1994), Pakistan (1990), Panama (1977), Papua New

Guinea (1999), Paraguay (1993), Philippines (1988), Portugal (1986), Samoa (1994), Senegal (1980), Singapore (1994), South Africa (1991), Spain (1986), Sri Lanka (1998), Switzerland (1995), Tanzania (1998), Thailand (1992), Togo (1995), Trinidad and Tobago (1990), Tunisia (1988), Uganda (1996), Vanuatu (1998), Venezuela (1993), Zambia (1995).

B.2 The History of VAT Adoption in Ghana

The Government of Ghana passed a VAT bill in December 1994. Stipulating VAT adoption in March 1995, the bill left no time for awareness campaigns and the necessary investments in administrative capacity. Citizens were left uninformed about the functioning of the VAT, most businesses were unable to carry out the self-assessment, and the state did not have the capacity to process the VAT recrediting applications. In May 1995, popular anger against VAT adoption was taken to the streets in violent protests that left four people dead. Although the government withdrew the VAT shortly after these incidents, the political damage it had created was substantial. In the 1996 parliamentary elections, the government lost 56 of its 189 seats. Ghanaian political scientists clearly relate this to the attempted VAT adoption (Osei (2000), Assibey-Mensah (1999)). Ghana finally adopted the VAT successfully in March 1998 - this time after intense preparations.

B.3 Variable Definition

Variable	Definition
IMF_t	Dummy = 1 if the country was under a lending agreement with the IMF for at least five months in year t.
IMF_{+3}	Dummy = 1 if the country was under a lending agreement with the IMF for at least five months in at least one of the years $t+1$, $t+2$ and $t+3$.
IMF_{-3}	Dummy = 1 if the country was under a lending agreement with the IMF for at least five months in at least one of the years $t-1$, $t-2$ and $t-3$.
$GDPpc_{t-1}$	GDP per capita in $t-1$.
IMP/GDP_{-3}	Imports as a share of GDP in $t-1$.
$Tradetaxrev_{t-1}$	Trade tax revenue as a share of tax revenue in $t-1$.
$Inflation_{t-1}$	Annual inflation rate in $t-1$.
$Agric/GDP_{t-1}$	Share of agricultural GDP in total GDP in $t-1$.
$Illiteracy_{t-1}$	Adult illiteracy rate in $t-1$.
$Polity_{t-1}$	PolityIV score in $t-1$.
EX_CHA_{t-1}	Number of changes in the executive in $t-1$.
$Debt\_S_{-3}$	Debt shock dummy = 1 if the increase in debt from year $t-1$ to year t was in the upper decile of the distribution of differences, in at least one of the years $t+1$, $t+2$ and $t+3$.
$Taxrev\_S_{-3}$	Tax revenue shock dummy = 1 if the decrease in tax revenue as a share of GDP from year $t-1$ to year t was in the upper decile of the distribution of differences, in at least one of the years $t-1$, $t-2$ and $t-3$.
IMP/GDP_S_{-3}	Imports shock dummy = 1 if the decrease in imports as a share of GDP from year $t-1$ to year t was in the upper decile of the distribution of differences, in at least one of the years $t-1$, $t-2$ or $t-3$.
GDP_growth_{t-1}	Annual GDP growth rate.
Neg_growth_{-3}	Negative growth dummy = 1 if the GDP growth rate was negative in at least one of the years $t-1$, $t-2$ or $t-3$.
$Growth\_-5_{-3}$	Dummy = 1 if the GDP growth rate was less than -5% in at least one of the years $t+1$, $t+2$ or $t+3$.
IMF_onset1_{-3}	Dummy = 1 if an IMF program (as defined in section 5.3) started in one of the years $t-1$, $t-2$ or $t-3$.
IMF_onset2_{-3}	Dummy = 1 if the first or second year of an IMF program took place in one of the years $t-1$, $t-2$ or $t-3$.
Neg_Polity_{t-1}	Negative of the polityIV score in $t-1$.
$Autoc0_{t-1}$	Autocracy dummy = 1 if the polityIV score was weakly negative in year $t-1$.
$Autoc5_{t-1}$	Autocracy dummy = 1 if the polityIV score was below 5 in year $t-1$.
$RIOTS_{-3}$	Total number of riots in years $t-1$, $t-2$ or $t-3$.
$ANTIGOV_{-3}$	Total number of antigovernment demonstrations in years $t-1$, $t-2$ or $t-3$.
$CDETAT_{-3}$	Total number of coups d'Etat in years $t-1$, $t-2$ or $t-3$.
CAB_CHA_{-3}	Total number of major cabinet changes in years $t-1$, $t-2$ or $t-3$.
EX_CHA_{-3}	Total number of changes in the executive in years $t-1$, $t-2$ or $t-3$.