Using Movement of Exemption Cutoff to Estimate Tax Evasion: Evidence from Pakistan*

Mazhar Waseem†

February 2017

Abstract

I contribute a simple, new approach to estimate tax evasion directly from taxpayers' own declarations. The approach relies on the fact that the costs of evasion are extremely low up to a given threshold and rise sharply after that. Facing such costs, an optimizing taxpayer would report true income at zero tax rate but would evade the component of income that entails near-zero cost as the rate increases marginally above zero. Using quasi-experimental variation and administrative data from Pakistan, I confirm the predictions of the model and show that at least 70% of self-employment and 1% of wage income is evaded in the country.

Keywords: Efficiency, Income Tax, Tax evasion
JEL Classification: H21, H24, H26

*I thank seminar participants at the University of Manchester, Institute of Fiscal Studies, Oxford University Centre for Business Taxation, VATT Institute for Economic Research, the IMF, the World Bank, and National Tax Association Annual Conference for helpful suggestions and comments.

†Department of Economics, University of Manchester, Arthur Lewis Building, Oxford Road, Manchester. (mazhar.waseem@manchester.ac.uk)

I Introduction

Tax evasion is pervasive through out the world, particularly in developing economies. And though it affects the incidence, efficiency, and fairness of a tax system in a fundamental way (Andreoni *et al.*, 1998; Slemrod & Yitzhaki, 2002), no unified and satisfactory approach exists to estimate its extent or distribution. The existing approaches to estimate it are either too costly or rely on arbitrary, ad hoc assumptions.[1] This paper contributes a simple methodology that can be used to measure tax evasion directly from taxpayers' own declarations. The data and tax variation needed for the methodology are largely available so that it can be replicated in a variety of contexts.

The natural starting point to identify evasion from taxpayers' own filings is to exploit variation created by tax reforms. Reported income responses to tax reforms, however, conflate real and evasion margins, and generally there is no intuitive way to separate the two. Equally important, even if we are somehow able to separate the two margins, we would only be estimating the *changes* in and not the *level* of tax evasion, which is problematic as changes are rarely informative on levels. Any credible approach that seeks to measure evasion from agents' behavioral responses to taxation, therefore, must address the twin challenges of separating real margin from evasion and changes from levels.

A distinguishing feature of modern tax systems is that they make extensive use of third-party reporting to secure tax compliance. Third-party reports, which are filed by institutions such as employers and banks, provide independent information on taxable income earned by individuals directly to the government. This reduces the costs of detecting tax evasion drastically, as a simple matching of these reports with income tax returns uncovers any underreporting. By contrast, detecting underreporting of income not covered by third-party reports is extremely difficult, as gathering independent information on such income is costly and the governments have limited resources. This particular feature of the enforcement environment when introduced into the standard Allingham and Sandmo model of tax evasion implies that the costs of evasion faced by a taxpayer are roughly of a reverse-L shape: they are extremely low up to a given threshold but turn sharply after that. Intuitively, the threshold represents the portion of income that leaves no verifiable information trail and thus entails near-zero detection probability and evasion cost.

One consequence of such evasion costs, which has not been explored in the literature so far, is the discontinuity they create in the earnings supply function: an optimizing tax-

[1] The most credible approach to estimate tax evasion involves auditing a random sample of tax returns intensively to determine the gap between reported and actual tax liability. This approach has been used by the Internal Revenue Service in its Tax Compliance Measurement Program (TCMP), results summarized in (Slemrod, 2007), and by academic researchers, more recently by (Kleven *et al.*, 2011). Besides needing a sophisticated audit capacity—which unfortunately would be lacking amongst countries where evasion needs to be measured the most—the approach entails daunting administrative and political costs. Owing to these costs, the IRS had to discontinue its program of random audits in 1995.

payer would report true earnings at zero tax rate but would evade the component of income that can be evaded at a trivial cost whenever tax rate increases marginally above zero. This discontinuity is helpful in isolating a lower bound on tax evasion from earnings responses produced by tax changes. To see this consider two small, similar-sized, and discrete tax reductions, one of which cuts tax rate all the way to zero while the other does not. Earnings response to the *to-zero* change would consist of three components: (1) a labor supply component resulting from an increase in effort; (2) an evasion component resulting from a continuous response of tax evasion to the decrease in tax rate; and (3) an evasion component resulting from a discrete change in tax evasion as the rate is reduced to zero.[2] Compared to this, earnings response to the *not-to-zero* change would consist of the first two components only. To the extent that the two tax changes are similar in size and scope, the difference between the two responses would thus identify the third component only. Since at any positive tax rate it is optimal for a taxpayer to evade this portion of income, the difference provides a lower bound on the *level* of tax evasion. The closer the evasion costs are to a reverse-L shape, the closer the lower bound would be to the actual evasion level in the economy.

The empirical signature of the discontinuity in the earnings supply function would be that earnings responses induced by to-zero tax changes would characteristically be different from those induced by not-to-zero changes. If we have access to such rate changes, we can not only test the existence of the discontinuity but can also estimate the minimum evasion rates for different groups of taxpayers in the economy. I implement the approach in Pakistan, where the desired tax variation is created by a series of sharp changes in the personal income tax schedules of the country. Pakistan has two income tax schedules, one each for self-employed and wage earners. A taxpayer is classified self-employed (wage earner) if her wage income does not exceed (exceeds) 50% of taxable income and is then taxed according to the assigned schedule on entire taxable income. The schedules are not indexed to inflation, and bracket boundaries, in particular the exemption cutoff—the threshold below which income can be taken without any payment of tax, need to be moved every few years to avoid bracket creep. This study focuses on the period 2006–2011 during which the two schedules were comprehensively revised once, but the exemption cutoff was increased twice for self-employed and four times for wage earners. The tax changes produced by these reforms are particularly suited to the approach developed here in that they (1) are small; (2) are targeted to a similar area of the income distribution; and (3) a few of them reduce the rate to zero (movements of the exemption cutoff), while the others do not.

I use administrative data from the Federal Board of Revenue (FBR) in Pakistan, which comprises the population of income tax returns filed in 2006–2011. The data contain vari-

[2] Earnings response to taxation could potentially contain a number of real and avoidance margins. For simplicity, I lump all real margins together in a labor supply response and all avoidance margins together in an evasion response. It would become clear later that any finer distinction than this is not needed in this setting.

ables corresponding to line items on the return form including (1) the decomposition of taxable income by source, such as self-employment, wages, and capital; (2) a brief profit and loss account; and (3) tax liability computations. The Pakistani tax code requires all *registered* taxpayers to file a return even if their income falls below the exemption cutoff. This means that earnings reported at zero tax rate are also observed, which is a key requirement of the approach.

In the first set of empirical results, I present nonparametric evidence confirming that earnings responses produced by to-zero tax changes are orders of magnitude larger than ones produced by similar-sized not-to-zero changes. While the elasticities underlying the former are larger than fifteen, those underlying the latter are close to zero. Comparing the evolution of individual line items on the return form, I further show that the difference between the two sets of responses indeed reflects tax evasion. And finally I document that tax evasion is large (more than 70% of reported earnings) even when tax rate is extremely low (just half a percent). In combination, these facts establish that the earnings supply function has a large discontinuity at the bottom but is not particularly elastic otherwise: reported earnings go down by a lot whenever tax rate increases marginally above zero but do not change significantly on any further increase in the rate.

Having established that the model passes the empirical test, I then estimate the lower bound on tax evasion for the two main groups of taxpayers—self employed and wage earners—in the country. To estimate earnings responses produced by the rate changes, I exploit the fact that very similar taxpayers are assigned different tax schedules depending upon the share of wages in their total income. Differential changes in these schedules over time and across taxpayers allow me to isolate tax-driven responses from any macro-driven changes in difference-in-differences and triple-difference research designs. The claim here is not that the assignment to the two schedules is random; it is rather that absent any tax changes reported income in the two groups would have evolved similarly.

I offer three pieces of evidences supporting this assumption. First, I present nonparametric evidence confirming that the pre-reform trends were parallel. Second, I show that the three double-difference coefficients in the triple difference regressions, which capture any preexisting differences across comparison groups and the difference over time for groups not affected by tax changes, are statistically indistinguishable from zero. Placebo experiments, where I pretend that the rate changes occurred one year earlier than they actually did, further reinforce the finding. Finally, I demonstrate that the results stay virtually unaffected when alternative parametric time trends, including industry-specific and region-specific trends, are introduced into the regressions. The robustness of the results derives primarily from a lack of macroeconomic trend in the studied outcomes. In fact, the pre-reform evolution of outcomes is so flat and stable that even the time-series evidence is convincing in this setting.

Comparing earnings responses produced by the to-zero and not-to-zero changes, I estimate that at least 70% of self-employment and 1% of wage income is evaded in the country. Because responses produced by the not-to-zero changes are quantitatively small and statistically insignificant, these lower bounds are fairly close to actual evasion levels in the economy. I run a number of auxiliary tests to show that the estimates are remarkably robust to alternative specifications. To explore heterogeneity, I estimate evasion for different groups of taxpayers within self-employed separately, finding that within sub-groups stratified by major firm characteristics, such as size, industry, and region, tax evasion does not vary lot once the position of the taxpayer in the income distribution is controlled for nonparametrically.

In addition to raising the obvious fairness concerns, the evasion pattern I uncover has important efficiency implications. That evasion jumps up immediately after tax rate increases from zero and stays almost constant thereafter implies that the excess burden of raising the rate marginally above zero is orders of magnitude larger than that of making the change elsewhere. The issue is not important in rich countries, as self-employed, for whom the effect is particularly strong, are only a small fraction of the total tax base. But in developing countries, where self-employed could comprise more than half of the tax base, the inefficiency arising from it could easily dwarf the inefficiency from other distortions created by the tax system. Note the finding that the private costs of evasion in this setting are small does not imply that the efficiency loss created by it is also small. Tax evasion in developing countries is typically achieved by using inferior production technologies, such as operating on cash-only basis, or by keeping the firm-size below its optimal level. The negative externalities flowing from such choices increase the efficiency cost of evasion far beyond the one implied by private costs only.[3] Even if we ignore these externalities, reducing tax evasion through improved enforcement would still be desirable in developing economies, as it is likely to be a more efficient way to generate revenue than raising distortionary taxes (Kaplow 1990; Mayshar 1991; Chetty 2009).

This paper relates to two different strands of literature. First, it contributes a simple, new methodology to estimate evasion to the existing approaches in the tax compliance literature (see Slemrod & Weber, 2012 for a survey). The methodology is less costly than the audit-based approach (Slemrod *et al.*, 2001; Kleven *et al.*, 2011) and requires fewer assumptions than the consumption-based approach (Pissarides & Weber, 1989; Lyssiotou *et al.*, 2004;

[3] Chetty (2009) has recently formalized the argument made in the literature earlier, notably by Slemrod (1995) and Slemrod & Yitzhaki (2002), that with externalities tax evasion/avoidance may not create the same excess burden as pure labor supply responses induced by taxation. Chetty (2009), however, considers a case more attuned to rich country settings, where tax evasion/avoidance creates a positive externality only—some part of the avoided tax is recovered later in the form of fines. For reasons noted above, tax evasion is likely to generate an overall negative externality in developing country settings, and thus excess burden it creates would be much larger than one implied by private costs only.

Feldman & Slemrod, 2007). The tax variation needed for it is not uncommon, as the exemption cutoff is moved regularly in a number of countries. For instance, Piketty & Qian (2009) document that between 1986 and 2008 the exemption cutoff was moved nine times in India and three times in China. Similarly, a number of countries, particularly in the developing world, make taxpayers report their earnings at zero tax rate.[4] The approach can, therefore, be replicated in other jurisdictions.

Second, this paper contributes to the new tax responsiveness literature (see Saez *et al.*, 2012 for the survey). Though this literature is rich, no existing study, to my knowledge, has focused on small tax reforms around the zero tax rate. As a result, the discontinuity at the bottom of the earnings supply function and its implications for tax policy have not been explored so far. In fact, a common assumption in this literature that the earnings supply function is log-linear signifies the settled opinion that the elasticity does not vary a lot along the curve. This paper shows that in a low-tax-capacity setting the earnings supply function is extremely elastic at the bottom and almost inelastic higher up the tax scale. Taking this nonlinearity into account is essential for important tax policy debates in developing countries. For instance, a seemingly appealing policy to broaden the tax base by reducing rates may not remain optimal under plausible social preferences once this nonlinearity is taken into account.

The rest of this paper is organized as follows. Section II develops the conceptual framework underlying the approach; section III describes the data, institutional background, and research design; section IV presents the empirical results; and section V concludes.

II Conceptual Framework

In this section, I develop the methodology that can be used to estimate tax evasion from taxpayers' own declarations directly. I begin with a stylized setting, making assumptions on taxpayers' preferences and government policy. But show later that the strategy is robust to generalization and remains valid under a broad set of conditions.

II.A Baseline Model

Consider an agent who decides how many hours to work (l) at a fixed wage rate w. The agent is required to pay income tax at a rate $\tau \in [0,\, 1]$ on its labor income wl but can reduce the tax liability by hiding e units of income on paying a resource cost $\Gamma(e)$.[5] Denoting con-

[4]For example, Bangladesh requires all individuals who have acquired a Tax Identification Number to file a tax return. Similarly, India makes all firms and companies and the UK all self employed to file a tax return regardless of their income during the year.

[5]Modeling the costs of evasion in this reduced-form way through a deterministic function was pioneered by Mayshar (1990) and has become standard since then (see for example Slemrod 2001; Chetty 2009; Best *et al.*

sumption by c and disutility of labor by $\psi(l)$, the agent's utility maximization problem can be written as

$$\begin{aligned} \text{(1)} \qquad & max_{l,e} \quad u(c,l,e) = c - \psi(l) - \Gamma(e) \\ & \text{s.t. } c = (1-\tau)(wl - e) + e. \end{aligned}$$

Optimizing behavior in this setup implies that the agent will evade income up to the point that the marginal cost of evasion $\Gamma'(e)$ equals the tax rate τ. The standard way to think about the evasion costs is that they are expected tax and penalty payments that would be recovered in case the evasion is detected. Assuming that the agent is risk-neutral, faces an endogenous probability of detection $p(e)$, and that the government applies a penalty θ on the evaded tax, these costs are given by

$$\text{(2)} \qquad \Gamma(e) = p(e)[(1+\theta)\tau e].$$

Since both tax rate and penalty are proportional here, the shape of the evasion costs is determined solely by the detection probability $p(e)$. A consistent theme in the recent compliance literature is that in addition to being an increasing, convex function of evasion, this probability depends a lot on whether the income is subject to third-party reporting. In particular, the probability is close to one if the income is covered by third-party reports and close to zero if it is not. To build this interaction into the model, I assume that $\underline{e}$ units of income of the agent are not covered by third-party information and therefore face a near-zero detection probability.

The marginal evasion cost faced by the agent in this setting

$$\text{(3)} \qquad \Gamma'(e) = p(e)[(1+\theta)\tau] + p'(e)[(1+\theta)\tau e]$$

depends upon both the first- and second-order properties of the detection probability and is shown graphically in Figure IA. The cost is fairly low at the bottom where income is self-reported but increases sharply at the cutoff $\underline{e}$ where third-party reported units of income begin. Note that such an S-shaped marginal evasion cost function arises naturally in models where third-party reporting is embedded into the standard Allingham and Sandmo model of tax evasion (see for example Figure 1 in Kleven *et al.* 2011).

To save on notation, I modify the above model slightly and assume that the marginal cost of evading up to $\underline{e}$ units of income is zero. The modified marginal cost function is shown in Figure IB. The assumption does not make any material difference to the results but simplifies the exposition considerably. It essentially implies that the costs of evasion

2015).

$\Gamma(e)$ are of the following form (see Figure IC)

$$\Gamma(e) = \begin{cases} \underline{g} & \text{if } e \leq \underline{e} \\ g(e) = p(e)[(1+\theta)\tau e] & \text{if } e > \underline{e}. \end{cases} \tag{4}$$

The agent can evade up to $\underline{e}$ units of income on paying a small fixed cost $\underline{g}$ but faces regular costs $g(e)$ with $g'(e) > 0$ and $g''(e) > 0$ on any evasion beyond that.

The optimum in this model is described by the following conditions

$$\psi'(l^*) = (1-\tau)w \tag{5}$$

$$[g'(e^*) - \tau]e^* = 0 \tag{6}$$

$$\tau e^* \geq \underline{g} \qquad \text{if } e^* > 0. \tag{7}$$

The first two of these are the standard first-order conditions pinning down optimal hours and evasion, and the third is the participation constraint providing that the agent will enter into evasion only if the benefit of doing so τe^* exceeds the fixed cost $\underline{g}$. I assume that the fixed cost $\underline{g}$ is small relative to the threshold $\underline{e}$ so that the following condition holds at any positive tax rate

$$\underline{g} \; \leq \; \tau\underline{e} \qquad \forall \tau > 0. \tag{8}$$

The inequality is a simplifying assumption which ensures that evading up to $\underline{e}$ units of income is optimal at any positive tax rate. Without this assumption, evasion would jump from 0 to $\underline{e}$ at the rate $\underline{\tau} \equiv \frac{\underline{g}}{\underline{e}} > 0$. It is because the marginal cost of evading up to $\underline{e}$ units of income in this setup is zero and the participation constraint gets satisfied at the rate $\underline{\tau}$. Inequality (8) implies that that the jump takes place immediately as the rate increases marginally above zero rather than at $\tau = \underline{\tau} > 0$. To the extent that the minimum positive tax rate in the empirical application is greater than $\underline{\tau}$, assuming $\underline{\tau} \approx 0$ has no substantive implication but simplifies the exposition.

Conditions (5) to (8) map tax rate τ to optimal hours $l(\tau)$, evasion $e(\tau)$, and reported earnings $z(\tau) = wl(\tau) - e(\tau)$. I assume that regularity conditions $\psi'(.) > 0$; $\psi''(.) > 0$ and $\psi(0) = \psi'(0) = 0$ are satisfied so that the choice of hours is interior. In addition, I assume that $g(\underline{e}) = g'(\underline{e}) = 0$ and $g'^{-1}(.)$ exists for all τ such that it can be represented by a normalized inverse function $\kappa(\tau) \equiv g'^{-1}(\tau) - \underline{e}$ having the property that $\kappa(0) = \kappa'(0) = 0$.

These assumptions generate an evasion function

$$e(\tau) = \begin{cases} 0 & \text{if } \tau = 0 \\ \underline{e} + \kappa(\tau) & \text{if } \tau > 0, \end{cases} \tag{9}$$

which has three properties: (1) there is no evasion at zero tax rate $e(0) = 0$, (2) evasion jumps to $\underline{e}$ whenever tax rate increases above zero, and (3) evasion increases smoothly with tax rate from that point onwards $\kappa'(\tau) > 0$.

This simple model generates two important predictions which can be tested if the relevant data are available.

PREDICTION P1: *Reported earnings will be a discontinuous function of tax rate*

$$z(\tau) = \begin{cases} wl_0 & \text{if } \tau = 0 \\ wl(\tau) - \underline{e} - \kappa(\tau) & \text{if } \tau > 0. \end{cases} \tag{10}$$

PREDICTION P2: *Evasion will be high even when tax rate is fairly close to zero* $e(\tau = \epsilon) = \underline{e} \quad \epsilon > 0$.

Intuitively, a taxpayer weighs the benefits of evasion $\tau.e$ against its costs $\underline{g} + g(e)$ to choose between $e^* = 0$ and $e^* \geq \underline{e}$. True earnings are reported at the zero rate because there is no benefit of evasion but it entails a small fixed cost $\underline{g}$. As the rate increases above zero, the benefits of evasion outweigh the costs and it jumps to the level $e^* \geq \underline{e}$. The discrete jump in evasion in turn reflects itself as a discrete drop in reported earnings, resulting in a discontinuity in the earnings supply function as illustrated in Figure ID.

Empirically, the most important signature of the discontinuity would be that earnings response to a to-zero tax reform would characteristically be different from that to a similar-sized not-to-zero reform. To see this formally, consider two discrete reforms $\Delta\tau_A(\tau_A \rightarrow 0)$ and $\Delta\tau_B(\tau_B \rightarrow \tau_B^{'})$ of the same size ($\Delta\tau_A = \Delta\tau_B = \Delta\tau$). Earnings responses induced by these reforms are given by

$$\begin{aligned} \Delta z_A(\tau_A \rightarrow 0) &= w\Delta l(\tau_A \rightarrow 0) - \Delta e(\tau_A \rightarrow 0) \\ \Delta z_B(\tau_B \rightarrow \tau_B^{'}) &= w\Delta l(\tau_B \rightarrow \tau_B^{'}) - \Delta e(\tau_B \rightarrow \tau_B^{'}) \end{aligned} \tag{11}$$

Suppose that the two reforms satisfy the following conditions

CONDITION C1: *The tax changes are small* $\Delta\tau \approx 0$.

CONDITION C2: *The tax changes are similar-sized and fairly adjacent to each other* $\tau_A^{'} \approx \tau_B^{'}$.

Then the fact that $l(\tau)$ and $\kappa(\tau)$ are smooth, continuous functions implies that

$$\begin{aligned} \Delta l(\tau_A \to 0) &\approx \Delta l(\tau_B \to \tau_B^{'}) \\ \Delta\kappa(\tau_A \to 0) &\approx \Delta\kappa(\tau_B \to \tau_B^{'}) \end{aligned} \tag{12}$$

so that the difference between the two earnings responses captures $\underline{e}$

$$\Delta z_A(\tau_A \to 0) - \Delta z_B(\tau_B \to \tau_B^{'}) \approx \underline{e}. \tag{13}$$

The intuition for the result is provided in Figure ID. Taxable income response to the to-zero change $\Delta z_A(\tau_A \to 0)$ consists of the movement along the earnings supply curve $z(\tau)$ and the movement along the horizontal axis from $wl_0 - \underline{e}$ to wl_0. By contrast, the response to the other change $\Delta z_B(\tau_B \to \tau_B^{'})$ consists of the movement along the supply curve only. As long as conditions C1 and C2 are satisfied, along the supply curve movements would nearly be equal so that the difference would identify $\underline{e}$. Thus, by simply comparing earnings responses to to-zero and not-to-zero rate changes we can not only test the two predictions of the model but can also estimate $\underline{e}$ from agents' own declarations. Since at any positive tax rate it is optimal to evade at least $\underline{e}$ units of income, it provides a lower bound on tax evasion.

It is important to emphasize that the threshold $\underline{e}$ would vary across taxpayers depending mainly upon the composition of their income. For a taxpayer whose income comprises both third-party and self-reported components, $\underline{e}$ should be seen as the self-reported component. For a taxpayer whose entire income is third-party reported, such as a wage earner, the threshold should be seen as compensation that can be taken in a non-taxed form such as cash-in-hand. And finally note that it is rarely feasible to evade completely even if all income is self-reported, as a few categories of consumption, such as credit card purchases, and changes in wealth, such as assets held with financial institutions, leave verifiable information trails that can easily be traced by the government at the time of audit.[6] Thus, for a taxpayer whose entire income is self-reported, such as a self-employed individual, the threshold should be seen as the portion of income over and above the verifiable components of consumption and saving. Respecting this heterogeneity, I would estimate $\underline{e}$ separately for the groups of taxpayers—self-employed individuals and wage earners—for whom it is likely to be similar.

[6] There is some evidence that tax authorities in developing countries use the fact that a few categories of consumption can easily be verified through information from external sources to enhance tax compliance. For example, the Pakistani authorities ask taxpayers to report a breakdown of consumption in the tax return, requiring especially to separate the expenditure on eight categories they feel can easily be verified. These categories include electricity, gas, and telephone usage; residential rent, insurance, and water bills; foreign and domestic travel; running and maintenance expenses on motor vehicles; expenses on the education of children; club membership fees.

II.B Robustness to Generalizations

I have described the above methodology assuming *(i)* risk-neutrality, *(ii)* proportional taxation, *(iii)* linear penalty applied to the evaded tax, and *(iv)* additive separability between effort and evasion costs. Under these assumptions, tax evasion increases with tax rate and does not depend on the labor supply decision of the agent. It is however easy to see from equations (11) to (13) that the strategy remains valid even if one or all of these assumptions are relaxed. In the general setting, the relationship between tax evasion and marginal tax rate becomes ambiguous and evasion could potentially increase after a decrease in the rate (Slemrod & Yitzhaki 2002). But since the methodology is based on comparing earnings responses to the to-zero and not-to-zero tax changes, as long as conditions C1 and C2 are met the difference between the two set of responses would identify $\underline{e}$ even if the evasion and earnings supply functions are backward bending or if evasion depends upon the labor supply choice of the agent. Intuitively, all components of the response that change smoothly with tax rate are netted out here, leaving behind the portion of income which is reported at the zero rate but not at any positive rate. In terms of Figure ID, it implies that the exact shape of the supply curve in the range $\tau \in (0, 1]$ plays no part in the methodology because along the curve movements cancel each other out.

Relatedly, the strategy is based on a static framework but taxpayers report income every year, and therefore their year t declarations could potentially be influenced by intertemporal considerations. For example, taxpayers might not report their true income in period t when tax rate is reduced to zero because of the worry that their $t - s$ declarations would expose them to audit or that any large income growth in $t + s$ periods would take them into the positive tax rate territory. Note, however, that these dynamic incentives provide one additional reason taxpayers might not report their true income at the zero tax rate, which given the lower-bound interpretation of the results is not a concern in this setting. Similarly, tax-induced intertemporal income shifting—taxpayers experiencing the reduction of tax rate bring income forward to reduce their future tax liability—would not matter here, as such responses would be smooth, continuous function of tax rate and therefore would not differ significantly across similar-sized to-zero and not-to-zero tax reductions.

III Context, Data and Research Design

III.A Context

Like other developing countries, personal income tax is an important and growing source of revenue for Pakistan. Its share in federal tax receipts has been rising steadily in recent years, accounting for roughly 13% of the receipts in 2013 (FBR, 2014). The tax is collected through two distinct schedules, one each for self-employed and wage earners. A taxpayer is classified self-employed (wage earner) if her wage income does not exceed (exceeds) 50% of taxable income and is then taxed according to the assigned schedule on entire taxable income. The two schedules, shown in Appendix Figure A.I, specify *average* tax rate as a function of taxable income. To calculate tax liability, a taxpayer simply multiplies her taxable income with the rate applicable in the corresponding bracket. The schedules are individual-based, there is no universal deduction other than that earnings below the exemption cutoff are not taxed, itemized deductions such as charitable donations are applied only after the tax liability has been calculated, and there is no system of tax credits or transfers interacting with the schedules.[7] The tax system, thus, is relatively simple, making the marginal tax rate on an additional rupee easy to work out, and which, therefore, must be salient in taxpayers' earnings and reporting decisions.

The most important feature of the tax system from the perspective of this paper, however, is that the two schedules are not indexed to inflation and therefore need to be revised every few years to avoid bracket creep.[8] During the period considered in this study, 2006–2011, the schedule for self-employed was comprehensively revised in 2010, but the exemption cutoff was moved twice, in 2010 and 2011. Similarly, the schedule for wage earners was comprehensively revised in 2008, but the exemption cutoff was moved four times, in 2008, 2009, 2010, and 2011.[9] These tax reforms are particularly suited to the requirements of this paper, as they create a series of plausibly exogenous to-zero and not-to-zero tax changes targeted to a very similar area of the income distribution. Figure II plots these changes, distinguishing the former from the other using dark color.

The Pakistani institutional environment offers three primary benefits to the methodology developed here. First, tax changes created by the reforms are *small*. Pre-reform rates, in particular those at the bottom of the income distribution, were extremely low, ranging between 0.25% to 5% (Figure A.I). The movement of the exemption cutoff essentially replaced these low rates with the zero rate, resulting in small, discrete tax changes. For example, the percent change in net-of-tax rate implied by the four movements of the exemption cutoff for wage earners is always less than one percent (Figure IIC-F). Second, the to-zero and not-to-zero changes created by the reforms broadly satisfy conditions C1 and C2, as these changes (1) are applied to a comparable area of the income distribution and (2) are almost of the same size. For example, the 2008 not-to-zero change and 2011 to-zero change for wage

---

[7] Pakistan has a small, means-tested income transfer program targeted to extremely poor households. Given, however, that the income tax exemption cutoff is set around the 80th percentile of the income distribution, the sets of taxpayers and transfer recipients do not overlap.

[8] Inflation is generally high in Pakistan and hovered around 10% during the periods considered in this study 2006-11.

[9] All these movements were upward movements. In fact, the exemption cutoff has never been revised downward in the history of the country. This creates strong, legitimate expectation that tax rate once reduced to zero would not be raised back to the positive territory.

earners are exactly similar other than that the latter reduces the rate to zero whereas the former does not. In addition to the tax reforms, the notches in the baseline tax system act as an additional source of not-to-zero tax changes. Earnings responses generated by the notches have already been estimated in Kleven & Waseem (2013), providing an additional reference against which the to-zero responses could be compared. Finally, as the main motivation behind the tax reforms was to avoid bracket creep, they are essentially narrow in focus and do not make significant changes to the tax code other than adjusting the bracket boundaries. This is useful as tax changes that are part of a comprehensive package of reforms conflate behavioral responses to price changes with other contemporaneous changes in the tax code.

One additional advantage of the Pakistani context is that earnings reported at zero tax rate are also observed. Two provisions in the tax code make it possible. First, a provision introduced in 2009 mandates all *registered* taxpayers to file a return even if no tax is payable. Before 2009, another provision in the code required taxpayers to file for period t if income in any of the two previous periods, $t-1$ and $t-2$, was above the exemption cutoff. Compliance with these filing requirements is vigorous, meaning that post-reform earnings of taxpayers experiencing the reduction of the rate to zero are largely observed.

The Pakistani tax system is based on the principle of self-assessment, and a return filed by a taxpayer is considered final unless it is selected for audit. Tax audits, however, are infrequent and not very effective, and compliance is largely secured through a combination of third-party reporting and tax withholding. All wage-income is third-party reported, and the tax is deducted at source by the employer. Evading wage income is, therefore, generally not feasible unless the employer agrees to collude. Self-employment income though self-reported is subject to a series of tax withholding schemes. These schemes withhold tax on specific transactions at source, which is adjusted later against the tax liability at the time of filing. For example, tax is withheld at the time of import of raw materials, the payment for contracts, cash withdrawal from banks, and utility bills payment. This type of withholding though not fully informative on the real size of the tax base, nevertheless restricts the ability of a taxpayer to evade completely.

III.B Data

I use administrative data from the Federal Board of Revenue in Pakistan that include income tax returns filed by self-employed individuals and wage earners in 2006-2011 and a set of taxpayer characteristics. The tax-return dataset contains variables corresponding to line items on the return form, including a brief profit and loss account, the decomposition of taxable income by source, and tax computations. The taxpayer characteristics dataset contains information captured at the time of registration such as the date of registration, gender, and location of a taxpayer. Appendix A, provides a detailed description of variables used in the empirical analysis.

Pakistan introduced a system of electronic return filing from July 2009, making it mandatory on all wage earners and a few categories of self-employed to file electronic returns.[10] Almost all returns of wage earners and around 20% returns of self-employed corresponding to tax years 2008-2011 have been filed electronically. The rest of the returns were filed at designated bank branches and were digitized by an IT firm for FBR. Throughout the period covered by this study, the FBR has been using the data for automated processing and payment of VAT and income tax refunds, which has ensured that the data were kept updated and relatively free from errors.

Appendix Table A.I reports descriptive statistics of the data. The analysis sample (columns 3-4) differs from the full sample (columns 1-2) on three dimensions. First, the research design used in this paper is based on panel analysis, comparing within-taxpayer changes in earnings ($\log \frac{z_{it+1}}{z_{it}}$) over time. Consequently, the analysis sample for period t gets restricted to taxpayers for whom $\log \frac{z_{it+1}}{z_{it}}$ is defined. Second, as the main focus of this paper are earnings responses to the movement of the exemption cutoff, I do not include taxpayers who have base period earnings (z_{it}) too far away from the exemption cutoff.[11] For self-employed, the analysis sample, accordingly, includes taxpayers with $z_{it} \in (80\text{K } 500\text{K}]$, which constitutes around 94% of the population. The wage income distribution has more spread, and the analysis sample, therefore, includes all taxpayers with $z_{it} \in (140\text{K } 700\text{K}]$, which constitutes around 62% of the population. In one of the robustness checks, I show that the results are not affected if this sample restriction is relaxed. Third, I drop taxpayers from the analysis sample for whom log change in earnings ($\log \frac{z_{it+1}}{z_{it}}$) is less than the 1st percentile or exceeds the 99th percentile of the corresponding pooled distribution. Such winsorizing is common in the literature to deal with the extreme outliers (see for example Gruber & Saez 2002).

All empirical results in this paper, unless otherwise specified, are based on the analysis sample with the following three categories of taxpayers dropped: (1) female taxpayers because the exemption cutoffs for them are slightly higher than male taxpayers in 2006–09,[12] (2) partners in partnership firms as their earnings are subject to a different tax regime (Waseem, 2016), (3) taxpayers who switch from self-employed to wage earners and vice versa from concerns that such switching may be endogenous to tax changes. These taxpayers are only a small fraction of the population (rows 4, 5, and 11 of the table), and the empirical results, therefore, are based on more than 96% of the potential analysis sample.

[10] Self-employed individuals required to file electronically include those *(i)* who are registered to remit VAT on their sales and *(ii)* who intend to file a tax refund claim.

[11] Of course, I do not impose any restriction on z_{it+1}.

[12] Doing the analysis separately for the two genders is difficult because female taxpayers are less than 3% of the analysis sample (row 11 of the table).

III.C Research Design

The ideal experiment to implement the methodology would require giving two randomly selected groups of taxpayers small, equal-sized tax changes, one of which brings the rate to zero and the other does not. While such an experiment is hard to come by, the Pakistani tax reforms provide a close alternative. They create two sources of tax variation: (1) between-schedules variation which results from the two schedules undergoing differential changes over time; and (2) within-schedule variation which results from differential rate changes across brackets of the same schedule. To implement the methodology, I exploit both sources of variation in event study research designs detailed below.

III.C.1 Self-Employment Income

Between-schedules Research Design.—The research design is motivated by the fact that self-employment income of taxpayers classified wage-earners is taxed through a different tax schedule. Table A.I (row 2) shows that there are a significant number of such taxpayers (6.4%) who can be used to control for any macroeconomic changes conflating the tax-driven responses. The research design, accordingly, compares the evolution of self-employment income across taxpayers categorized self-employed and wage earners to estimate earnings responses to the rate changes shown in Figure IIA–B. The identification assumption here, as discussed below, is not that the self-employed or wage-earner status is *randomly* assigned; it is rather that the self-employment income would have evolved similarly across the two groups in the absence of tax changes.

Specifically, I estimate the following model

$$\Delta\log\, z_{it}^{S} = \alpha + \beta\, SE_i + \mathbf{year}_t\, \boldsymbol{\gamma} + \mathbf{SE}_i \times \mathbf{post}_t\, \boldsymbol{\delta} + \boldsymbol{X}_{it}\, \boldsymbol{\mu} + u_{it}, \tag{14}$$

where $\Delta\log\, z_{it}^{S}$ is log change in self-employment income for taxpayer i from period t to $t+1$, SE_i indicates that i is classified self-employed, $\mathbf{year}_t$ is a vector of year fixed effects, $\boldsymbol{X}_{it}$ are a set of controls, and $\mathbf{SE}_i \times \mathbf{post}_t$ is a vector of four dummies indicating if the observation relates to (1) a self-employed in the post-reform periods (*SE* $\times$ *post*); (2)-(3) a self-employed experiencing a to-zero change, (*SE* $\times$ *to-zero* $\times$ 2009) and (*SE* $\times$ *to-zero* $\times$ 2010); and (4) a self-employed experiencing the not-to-zero change (*SE* $\times$ *not-to-zero* $\times$ 2009).[13] Note that the partition of $\mathbf{SE}_i \times \mathbf{post}_t$ into four areas requires the additional assumption that income growth within a given area of the income distribution remains stable from one year to the other. Specification (14) contains a built-in test of this assumption: the coefficient on the first dummy variable *SE* $\times$ *post* captures income growth in the area of the income dis-

[13] Given that the dependent variable is defined as log change in income from period t to $t+1$, the effect of a tax change introduced in period $t+1$ would be reflected in the coefficient for the period t.

tribution that does not experience any tax change. A statistically insignificant coefficient on the variable would be a direct evidence in support of the assumption. I would also present nonparametric evidence supporting the assumption in section IV.A.1 of the paper.

There are three potential threats to identification in this setup: (1) self-employment income may not be on a common trend across the two groups, (2) the composition of the sample may be changing around the time of the reform in a way creating a correlation between the error term and the interaction dummies, and (3) the control group itself may be affected by tax changes. I present graphical evidence establishing that the self-employment income was trending similarly in the two groups in the pre-reform periods. In fact, the preexisting evolution of the outcome is so flat and stable that even the time-series evidence is convincing in this setting. I take two precautions to address concerns from a potential change in the composition of the sample. First, as noted in section III.B, I drop few taxpayers who switch from one group to the other. This ensures that there is no endogenous selection into treatment and that the classification of a taxpayer remains time-invariant. Second, I demonstrate that all results stay virtually unaffected if estimated on a constant-composition balanced panel of taxpayers.

Self-employment earnings of the control group are taxed through wage-earners' tax schedule. As this schedule undergoes a few changes of its own between 2006 and 2011, one may worry that the difference-in-differences estimator might be contaminated by these influences. Note, however, that all rate changes experienced by control taxpayers in 2010-11 are tax *decreases* (Figure II). Any bias, therefore, even if it exists would only push the self-employment income response downwards. To further address this concern, I also present estimates from regressions that restrict the control group to taxpayers for whom the mechanical change in tax rate from period t to $t+1$ is always zero.[14] Practically, however, this turns out to be a careful precaution only and the estimates with and without this exclusion are indistinguishable.

Triple-difference Research Design.—Between-schedules research design is inefficient as it does not exploit within-self-employed variation created by the reforms. To exploit this variation, I also estimate the following triple-difference model

$$\begin{aligned}\Delta \log z_{it}^{S} = {} & \alpha_0 + \alpha_1\, \textit{to-zero}_{it} + \alpha_2\, SE_i + \mathbf{year}_t\, \boldsymbol{\gamma} \\ & + \beta_1\, \textit{to-zero}_{it} \times \textit{post}_t + \beta_2\, SE_i \times \textit{post}_t + \beta_3\, \textit{to-zero}_{it} \times SE_i \\ & + \textbf{to-zero}_{it} \times \mathbf{SE}_i \times \mathbf{post}_t\, \boldsymbol{\delta} + \boldsymbol{X}_{it}\, \boldsymbol{\mu} + u_{it}\end{aligned} \tag{15}$$

where $\textit{to-zero}_{it}$ is an indicator that a taxpayer i has self-employment income less than PKR 350K in period t, $\textbf{to-zero}_{it} \times \mathbf{SE}_i \times \mathbf{post}_t$ is a vector of two triple-interaction dummies one

[14]Mechanical change in tax rate is defined as the statutory tax rate change from period t to $t+1$ on the base period income i.e. $\Delta\tau_{it} = \tau_{t+1}(z_{it}) - \tau_t(z_{it})$.

each for 2009 and 2010, and all other variables have the same definitions as in (14). The triple-difference coefficients in this specification signify tax-induced, additional earnings growth reported by self-employed who experience the reduction of rate to zero relative to the other taxpayers.

One key advantage of the specification over (14) is its transparency. The coefficients on the first two double-interaction terms capture if self-employment income in the post-reform periods evolves differently from the preexisting trend for below-cutoff wage earners and above-cutoff self-employed. The coefficient on the last double-interaction term captures any preexisting differences across below-cutoff wage earners and below-cutoff self-employed. The insignificance of these coefficients, thus, provides a direct test of the identifying assumption underlying the model. This transparency, however, comes at a cost for I am unable to separately estimate the impact of the not-to-zero change. But this turns out to be a minor issue as the estimates from (14) and the nonparametric evidence convincingly show this impact to be statistically indistinguishable from zero.

III.C.2 Wage Income

The between-schedules variation for wage income is quite limited, as less than one percent of self-employed have positive wage income (Table A.I, row 3). Compared to this within-schedule variation for wage income is quite rich, as the wage income distribution is more spread out. This motivates the following research design that pools both sources of variation

$$\Delta\log z_{it}^{W} = \alpha + \mathbf{year}_t\,\boldsymbol{\gamma} + \mathbf{treat}_{it}\,\boldsymbol{\delta} + \boldsymbol{X}_{it}\,\boldsymbol{\mu} + u_{it}, \tag{16}$$

where $\Delta\log z_{it}^{W}$ is log change in wage income for taxpayer i from period t to $t+1$, $\mathbf{year}_t$ is a vector of year fixed effects, $\boldsymbol{X}_{it}$ are a set of controls, and $\mathbf{treat}_{it}$ is a vector of two dummies [*to-zero not-to-zero*] indicating if the observation relates to a wage earner affected by a to-zero or a not-to-zero change.

Given that the identification here comes largely from the within-schedule variation, the major threat to identification is mean-reversion. More specific, earnings of a taxpayer in the bottom of the distribution in period t are expected to go up in period $t+1$ for nontax reasons if the transitory component of earnings is mean-reverting. Considering that the to-zero rate reductions are applied at the bottom of the distribution, mean-reversion can reinforce the tax-driven response, causing an upward-bias in the estimates. To deal with this concern, I follow the literature (Gruber & Saez 2002; Kleven & Schultz 2014) and add base-period income controls, including a ten-piece spline of log base-period income, into the regressions. To test if this adequately controls for mean-reversion, I conduct placebo analysis pretending that each reform took place one year prior to its actual implementation.

Considering that the reforms were unanticipated, the placebo regressions would indicate if taxpayers in areas of the income distribution affected by the reforms experience significantly different wage-growth for reasons orthogonal to tax changes. In addition to mean-reversion, the three general concerns noted for specification (14) are also relevant here. To address these concerns, I take the precautions and report the robustness checks noted above for wage income as well.[15]

IV Empirical Results

In this section, I first show that consistent with the predictions of the model (1) earnings responses generated by the to-zero rate changes are fundamentally different from those generated by the similar-sized not-to-zero changes; (2) the difference between the two reflects tax evasion; and (3) tax evasion is large even at extremely low rates. I then use the difference between the two sets of responses to estimate the lower bound on tax evasion in Pakistan.

IV.A Self-employment Income

IV.A.1 Nonparametric Evidence

Figure III plots the evolution of self-employment income from 2006 to 2011 for taxpayers classified self-employed by the tax code. To construct the diagram, I group taxpayers into PKR 20,000 bins on the basis of their base period income (z_{it}^S) and then plot mean log change in income from year t to $t+1$, $\mathbb{E}\,[\log \frac{z_{it+1}^S}{z_{it}^S} | z_{it}^S \in b]$, in each bin b. These plots provide nonparametric evidence on how self-employment income growth in various areas of the income distribution responds to the 2010-11 tax changes. Two features of the evidence are noteworthy here. First, the growth rate is remarkably stable over time and homogeneous across the income distribution in periods of no tax change. Second, there is a striking difference between responses to the two types of rate changes: while income of taxpayers experiencing the reduction of the rate to zero jumps dramatically, that of taxpayers experiencing a similar-sized not-to-zero rate reduction does not change at all.

Was the dramatic income growth at the bottom of the distribution in 2010-11 *caused* by the reduction of the rate to zero? That it (1) is concentrated precisely in the region between the old and new cutoffs, (2) is very strong at the bottom and then tapers off monotonically as we move towards the new cutoffs, and (3) is indistinguishable from the pre-reform level just above the new cutoffs strongly suggests that it was. To further reinforce the causal link,

[15] One of such precautions is that I drop control group taxpayers (self-employed with positive wage income) for whom the mechanical change in tax rate from period t to $t+1$ is not zero. This affects only a few observations in 2010–11.

Panel B of the figure looks at the evolution of self-employment income of taxpayers classified wage earners by the tax code. To the extent that self-employment income is subject to common macro shocks, any nontax factors affecting it in 2010–11 could easily be detected here. However, in sharp contrast to Panel A all curves in Panel B are tight to each other, suggesting that the 2010–11 responses are indeed driven by the tax changes. Figure IV, which is a difference-in-differences analog of Figure III, formalizes this conclusion by demonstrating: (1) there are no significant preexisting differences across the treatment and control groups (Panel A), (2) the reduction of the rate to zero causes a sharp surge in income reported by the treatment group (Panels B and C), and (3) the additional reported income is as large as 70% of the base period income at the bottom (Panel D).

One important requirement of the agenda in this paper is to establish that the earnings responses shown above reflect tax evasion and not adjustment along real margins such as effort. Before looking at the evidence, note that the rate changes behind the 2010–11 responses are extremely small, ranging between 0.5 and 8 percent (Figure II). Such small price changes, under realistic assumptions on the elasticity of effort, cannot explain the magnitude of the observed responses other than through a large and discontinuous drop in tax evasion.[16] To probe the question further, I look at the responses of individual line items reported on the tax return. The six items considered here form the profit and loss account of a taxpayer, and while all of them are expected to increase with effort, some can be misreported easily than the other. Differential evolution of these items can help us uncover the nature of the observed responses.

Figure V illustrates this analysis. Each panel of the figure plots mean log change in the line item from period t to $t+1$ as a function of the self-employment income in period t. Since the sets of taxpayers in various bins here are the same as in Figure IIIA, the analysis should be seen as the decomposition of the response depicted there. The plots show that the line items do not respond uniformly: annual sales and costs respond aggressively, profit and loss expenses respond moderately,[17] and imports do not respond at all. This pattern is consistent with a tax evasion explanation of the observed responses. Had the growth of self-employment income been a result of an increase in effort, all line items would have responded uniformly. Instead, easy-to-misreport items respond more aggressively than the other. Of all the items, imports is perhaps the hardest to misreport, since such misreporting can easily be detected through Customs records. Its non-responsiveness, therefore, provides the cleanest evidence that the observed jump in reported earnings is driven by a large drop in tax evasion. Panels E–F reinforce this conclusion. A surge in real activity triggered by an

[16]The real elasticity has to be larger than 40 to reconcile the observed responses at the bottom of the distribution to have arisen solely from changes in effort.

[17]Profit and loss expenses are input costs such as wages, rents, accounting and legal fees, electricity, and interest paid on loans. Although these costs can be over-reported, it is difficult to do so considering that these can potentially be verified at the time of audit.

unanticipated decrease in taxes is likely to result in the running down of inventory. Contrary to this, inventories at the end of 2010 and 2011 register stronger growth.

IV.A.2 Regression-based Estimates

Table I reports the results from equation (14). The first column contains the estimates from the baseline specification, while the rest of the columns add additional control variables or experiment with the replacement of the year fixed effects with parametric time trends. Three findings emerge from the analysis. First, consistent with the visual evidence the to-zero changes generate extremely strong responses. The corresponding estimates are always large, statistically significant, and remarkable robust to alternative specifications. Column (1) of the table, for example, shows that the reduction of the rate to zero caused a 28 log-points additional income growth in the treatment group in the first year after the reform. Relative to the pre-reform average of 5 log-points, this represents a roughly six-fold increase. Considering that the average net-of-tax rate change behind the response was only around 1.7 log-points, the estimate translates into a taxable income elasticity larger than 15. This elasticity is orders of magnitude larger than the one estimated in Kleven & Waseem (2013) for the the same set of taxpayers using tax variation that does not involve moving to or away from the zero rate. Second, the similar-sized, not-to-zero tax decrease generates no response at all. The corresponding estimates are always of opposite sign, small, and statistically insignificant in four of the eight specifications. Third, the coefficient on the double-interaction term ($SE \times post$) is small and statistically insignificant in most cases. This further confirms the nonparametric evidence that absent tax changes income growth remains stable from one year to the other in all areas of the income distribution, providing a strong support to the assumption underlying model (14).

Table II reports the estimates from the triple-difference specification (15). Unsurprisingly, the results are in line with those in Table I. The coefficients on the two triple-interaction dummies, which capture earnings responses to the to-zero changes, are very similar to the corresponding estimates in Table I.[18] The coefficients on the three double-interaction dummies are always small and statistically insignificant. This shows that the estimates of interest are driven by tax changes and not by any underlying differences in earnings trends, providing direct evidence on the validity of the research design. In section IV.A.4, I conduct a number of additional robustness checks to further cement this point.

[18]It is important to note that the magnitude of the to-zero coefficients from the triple- and double-difference specifications (Table II vs. Table I) are not directly comparable because of the slight difference in the definition of the to-zero dummy in both specifications. Please see section III.C.1 for details.

IV.A.3 Evasion Estimates

Having established that the model passes the empirical test, I use formula (13) to estimate the lower bound on income underreported by self-employed in Pakistan. These estimates are shown in Table III. I divide the region below the new exemption cutoff into six segments and present estimates separately for each segment. The columns of the table correspond to the terms in formula (13): column (1) to the income segment; column (2) to the earnings response produced by the to-zero change; column (3) to the earnings response produced by the equal-sized not-to-zero change; column (4) to the difference between the two responses; and column (5) to the average evasion rate in the segment.

To compute the estimates in column (2), I follow three steps. I first replace the two $SE \times cutoff \times post$ dummies in (14) with twelve $SE \times segment \times post$ dummies to estimate self-employment income response separately in each segment in 2010 and 2011. I then add the two yearly estimates for each segment to compute a medium-run estimate of the response in the segment. And finally, I multiply the estimate from the second step with average income in the segment to convert it into rupees. The estimates in column (3) are computed using taxable income elasticities reported in Kleven & Waseem (2013). These elasticities relate to the same set of taxpayers and were estimated using variation created by the notches in the baseline Pakistani tax system (2006-09). I do not employ the not-to-zero estimates from specification (14) here, because (1) they are statistically not different from zero and (2) there could be a concern the tax change was applied higher up the income distribution. The tax changes implied by the notches satisfy conditions C1 and C2 perfectly, as they are (1) small, (2) applied to the same area of income distribution, and (3) do not involve the zero rate.[19]

The evasion rates reported here have two notable features. First, tax evasion is large even for taxpayers facing a tax rate of merely 0.5%. This behavior is strictly consistent with the prediction P2 of the model, illustrating that evasion jumps to $\underline{e} \gg 0$ whenever the rate increases marginally above zero. Second, the evasion rate is roughly constant up to the income of PKR 150K and then declines monotonically. It is important to note that earnings responses of taxpayers closer to the new cutoff are not unconstrained, as reporting true income would take them into the positive rate territory. Formally, it is not optimal for a taxpayer experiencing the reduction of tax rate from $\tau > 0$ to $\tau = 0$ to report true income as long as

$$z(\tau) + \underline{e} \geq \overline{z},$$

where $\overline{z}$ is the new cutoff. There would, thus, exist an interval $[\overline{z} - \underline{e},\ \overline{z}]$ below the new

[19] In computing the estimates for column (3), I keep the magnitude of the tax rate change the same as for column (2). More specific, for a segment k I convert the elasticity ε_k into earnings response Δz_k using the formula $\Delta z_k = \varepsilon_k.\bar{z}_k.\hat{\Delta}(1-\tau_k)$, where $\bar{z}_k$ is average income in the segment, and $\hat{\Delta}(1-\tau_k)$ is the proportional net-of-tax rate change experienced by taxpayers in the segment because of the reform.

cutoff where evasion would not approach zero even when the rate drops to zero. This interval would be larger if taxpayers have dynamic considerations, trying to keep not only their current but also future income below the cutoff. A monotonically declining response, therefore, arises naturally in this model and means that the unconstrained evasion rate is observed only at the bottom where taxpayers are too far away from the new cutoff. Considering this, I conclude that the lower-bound on the evasion rate of self-employment income in Pakistan, as implied by the first four rows of the table, is 70%.

IV.A.4 Robustness

The insignificance of the *SE* $\times$ *post* term in specification (14) and all the double-difference terms in the triple-difference specification (15) demonstrates that the identification assumptions underlying these models are satisfied. Figure A.II provides visual underpinning to the result, illustrating that the preexisting earnings trends were flat, stable, and parallel in the two groups. One formal manifestation of the lack of macro trend in outcomes is that the time-series estimates, reported in Appendix Table A.II, are indistinguishable from the corresponding difference-in-differences estimates. Tables I, II, and A.II also show that the results are insensitive to (1) including additional control variables; (2) replacing the year fixed effects with parametric time trends; (3) adding a full set of industry, year, and industry $\times$ year fixed effects; and (4) restricting the control samples to taxpayers who experience no tax change. In this section, I report results from additional robustness checks.

I begin by showing that the results are not affected if the composition of the sample is held fixed. Appendix Figures A.III and A.IV replicate the analysis in Figures III and IV, restricting the sample to a balanced panel of taxpayers who file in all six years 2006-2011. Appendix Tables A.III and A.IV carry the corresponding sets of estimates, showing no qualitative change from the earlier results. Appendix Table A.V addresses three additional concerns. Column (2) drops taxpayers who *bunch* at the notches in the baseline tax schedule from concerns that their reported income might be affected by the strong, local incentives created by the notches or that these taxpayers might be special. Columns (3)–(4) increase the range of the data from $z_{it} \in (80\text{K}\ \ 500\text{K}]$ in the baseline results to $z_{it} \in (0\ \ 500\text{K}]$ in column (3) and $z_{it} > 0$ in column (4). Columns (5)–(8) add additional control variables into specification (14). Reassuringly, the results from the alternative specifications, columns (2)–(8), are identical to the baseline results.

Another common concern in the tax responsiveness studies is mean-reversion. Given that the tax variation exploited here is uncorrelated with the base period income, mean-reversion is unlikely to be a problem in this setting. To test this formally, I conduct a placebo analysis, estimating (14) on the pre-reform periods only. Appendix Table A.VI shows the results. The coefficient on the *to-zero* $\times$ *SE* $\times$ 2008 dummy is extremely small (generally less than one log-point), which demonstrates that reported income at the bottom of the distribution does not change significantly from one year to the other for any nontax reason. The coefficient on the *not-to-zero* $\times$ *SE* $\times$ 2008 dummy, however, is similar to one in the original regression. This suggests that the wrong-signed coefficient on the not-to-zero change in the original regression likely reflects extremely mild mean-reversion higher up the income distribution.

IV.A.5 Heterogeneity

To explore heterogeneity, I estimate the following triple-difference analog of equation (14)

$$\begin{aligned}(17)\qquad \Delta\log z_{it}^{S} &= \alpha_0 + \alpha_1\, \textit{to-zero}_{it} + \alpha_2\, \textit{trait}_i + \mathbf{year}_t\, \boldsymbol{\gamma} \\ &+ \textbf{to-zero}_{it} \times \mathbf{post}_t\, \boldsymbol{\delta} + \beta_1\, \textit{to-zero}_{it} \times \textit{trait}_i + \beta_2\, \textit{trait}_i \times post_t \\ &+ \textbf{to-zero}_{it} \times \mathbf{trait}_i \times \mathbf{post}_t\, \boldsymbol{\eta} + \boldsymbol{X}_{it}\, \boldsymbol{\mu} + u_{it},\end{aligned}$$

where $\textit{trait}_i$ is a taxpayer characteristic indicator and $\mathbf{post}_t$ is a vector of two dummies one each for 2009 and 2010. To avoid making strong functional-form assumptions, all characteristics are introduced into the equation nonparametrically. The coefficients on the triple-interaction dummies capture differential response of taxpayers with a given characteristics j. Since characteristics are not randomly assigned, one problem with this specification is the estimates might simply reflect that taxpayers with different characteristics are located in different areas of the income distribution. This is especially problematic in the current setup because the response, as shown in Figure IIIA, declines monotonically along the income distribution when we move towards the new cutoff. To make the comparison more meaningful, I also show results from an alternative approach, where I group taxpayers into bins of PKR 20K on the basis of their base period income and run the above regression separately in each bin. I then generate aggregate estimates as the weighted average of the bin-level estimates, with the weights provided by the distribution of characteristic j in the binned income distribution. This approach is similar to matching and compares taxpayers of different characteristics by matching them on the basis of their base period income. Appendix Figure A.V provides graphical justification for the approach.

The five characteristics explored here are commonly regarded in the compliance literature to be negatively correlated with tax evasion. The most important of these is firm size. It is widely believed that larger firms have lower ability to hide their output and thus are more tax compliant (see for example Gordon & Li 2009; Kopczuk & Slemrod 2006; Kleven *et al.* 2016). Appendix Table A.VII investigate this hypothesis. The results from the simple comparison in odd-numbered columns are consistent with the hypothesis, but they no longer hold when the position of a taxpayer in the income distribution is controlled for nonparametrically (even-numbered columns). Appendix Tables A.VIII and A.IX extend this analysis to five other characteristics, comparing manufacturers to non-manufacturers, regular tax filers to irregular tax filers, VAT-registered taxpayers to the other taxpayer, electronic return filers to manual return filers, and young taxpayers to old taxpayers. The results are generally consistent with those in Table A.VII, suggesting that the response does not vary a lot when taxpayers with different characteristics but in the same area of the income distribution are compared. The analysis, thus, shows that the correlates of compliance commonly discussed in the literature may not have a causal relationship.

IV.B Wage Income

IV.B.1 Nonparametric Evidence

Figure VI shows the evolution of wage income from 2006 to 2011 for taxpayers classified wage earners by the tax code. The diagram is constructed analogously to Figure III and plots the growth of wage income from period t to $t+1$ as a function of the base period income $\mathbb{E}\left[\log \frac{z^W_{it+1}}{z^W_{it}} | z^W_{it} \in b\right]$, where b are bins of PKR 20K. One important feature of the evolution here is that the growth rate is not homogeneous across years like the one for self-employment income shown in Figure IIIA. Because of this, it is hard to distinguish tax-driven response from secular wage trends in the simple raw data plot.

To obtain a first-pass evidence on the tax-driven response, I therefore follow a simple strategy and regress log changes in wage income from period t to $t+1$ on a full set of year fixed effects. The residuals from the regression are then regressed on four yearly dummies, one each for 2007 to 2010. These later regressions are run separately in the PKR 20K bins, and the estimated coefficients and 95% confidence intervals are plotted in Figure VII. Clearly, once a common year effect is partialled out the residual income growth is homogeneous across years and over the income distribution other than that it spikes whenever tax rate is brought to zero. Though this spike is not as prominent as that of self-employment income, the overall pattern is consistent with the model in that bringing the rate to zero gives rise to a behavior fundamentally different in character from that by other rate changes. In the next section, I formalize this analysis by presenting the regression-based estimates.

IV.B.2 Regression-based Estimates

Table IV reports the results from equation (16). I begin with the baseline specification in column (1) and then add successively more control variables, permuting among the combinations of controls for mean-reversion—log base period income and a ten-piece spline of log base period income—and other controls in the rest of the columns. To test the adequacy of the mean-reversion controls, panel B reports the estimates from placebo regressions, where

I pretend that all rate changes took place one year earlier than they actually did. Appendix table A.X runs additional robustness checks, which include inter alia relaxing the sample restrictions.

The main findings are the following. First, the estimates corresponding to the to-zero changes are always economically meaningful, statistically significant, and considerably larger than the corresponding placebo estimates. Given the rate changes underlying these responses are extremely small (always less than one percent, Figure II), these estimates translate into huge elasticities. For instance, the elasticity implied by the baseline estimate in column (1) of Table IV is more than three, which is orders of magnitude larger than the largest elasticity, 0.04, estimated by Kleven & Waseem (2013) for wage earners in Pakistan. Second, the estimates corresponding to the not-to-zero changes are much smaller and generally indistinguishable from the corresponding placebo estimates. The result attains even more significance when one considers that the average size of the not-to-zero rate changes is more than twice that of the to-zero changes. Third, the placebo coefficients though small are generally distinguishable from zero, meaning that the base-period income controls do not fully account for mean-reversion. In calculating the evasion rate, I, therefore, adjust the to-zero coefficient by the corresponding placebo coefficient, assuming that the extent of mean-reversion does not vary from one year to the other.

The overriding conclusion that emerges from the analysis is that wage income also behaves according to the prediction of the model, though its response is much weaker than the self-employment income response. The weaker response here should not be surprising, since wage income is third-party reported and therefore cannot be evaded easily. Insofar as the model passes the empirical test here as well, the differences between the two set of responses reveals the evasion rate of wage income in the country. The extent of such tax evasion, however, is small, roughly around 1% of the reported income.

IV.B.3 Discussion

How do evasion rates I estimate compare against similar estimates from other countries? Before making such a comparison, it is important to emphasize that tax evasion uncovered here relates to the population of tax filers only. The to-zero changes created by the Pakistani reforms affect a large area of the income distribution,[20] and therefore evasion rates I estimate are broadly representative of the population of tax filers. But they do not capture individuals and firms who operate outside the formal tax regime yet have incomes above the taxable limit. Comparing within-filers evasion in Pakistan with other countries, includ-

[20] For example, the new exemption cutoff for self-employed in 2010 was at more than the 50th percentile of the 2009 distribution, meaning that the to-zero changes created by it impacted more than half of all self-employed. Similarly the new exemption cutoff for wage earners in 2010 was at around 15th percentile of the corresponding 2009 distribution.

ing the UK (Pissarides & Weber, 1989), the US (Slemrod, 2007), Denmark (Kleven *et al.*, 2011), and Greece (Artavanis *et al.*, 2016), reveals similar pattern and magnitudes: evasion of self-reported income is quite high and evasion of wage-income is almost negligible.[21] This similarity is consistent with the hypothesis that noncompliance amongst tax filers is driven by the information constraints the government faces (Kleven *et al.*, 2011) and not by any differences in social and political institutions of the country or intrinsic motivation of taxpayers (Andreoni *et al.*, 1998).

The cross-country comparison above does not account for the fact that a greater proportion of income reported in Pakistan is not backed by information reports, as self-employed comprise a much larger fraction of the tax base (more than 50% in Pakistan vs. less than 5% in rich countries). Nor does it take into account that the country has a much larger informal sector relative to one in advanced economies. Considering the two factors would take aggregate noncompliance in Pakistan much above that in the compared countries. It is however difficult to work out the exact level of aggregate noncompliance in the country, as the existing estimates of the size of the informal sector do not take into account that some of the taxpayers operating in the informal regime may have incomes below the exemption cutoff and therefore may not be in breach of the tax laws of the country.

V Conclusions

Like all illegal activities, tax evasion is notoriously difficult to measure. This paper develops a simple methodology that can be used to measure evasion directly from taxpayers' own declarations. The approach relies on the fact that the costs of evasion are extremely low up to a given threshold but turn sharply after that. Facing such costs an optimizing taxpayer would report true income at zero tax rate but would evade the component of income that entails near-zero cost as the rate increases marginally above zero. This creates a discontinuity in the earnings supply function which can be used to isolate a lower bound on tax evasion. For this purpose, we need to compare taxable income response when tax rate goes to or away from zero with the response to a similar tax change in the positive-rate range. Intuitively, such comparison nets out all components of the response that change smoothly with tax rate, leaving behind the portion of income which is reported at the zero rate but not

[21] I find that self-employed in Pakistan evade at least 70% of their *reported* income, which implies a rate of 42% in terms of *true* income. The corresponding rates in terms of true income for the other countries reported in the above studies are: 35% for the UK, 50% for the US, 41.6% for Denmark, and 43-45% for Greece. Of the four studies only Kleven *et al.* (2011) and Tax Compliance Measurement Program (Slemrod, 2007) estimate tax evasion by wage earners, finding it to be less than one percent (in both Pissarides & Weber, 1989 and Artavanis *et al.*, 2016 tax evasion of wage income is zero by assumption). The only exception to these generally similar estimates is a recent study (Best, 2014), which finds that around 3.6% of wage income goes unreported in Pakistan. Though the number is closer to the top-end estimates in Table IV, the difference could also reflect that the study includes top-income employees for whom evasion could be higher.

at any positive rate.

I implement the methodology in Pakistan using variation created by a series of sharp changes in the two income tax schedules of the country. I first show that the behavior of taxpayers conforms strictly with the predictions of the model. I then estimate that at least 70% of self-employment and 1% of wage income goes unreported in the country. Given that reported income is not particularly sensitive to rate changes in the positive-tax range, the lower bounds are not far from actual evasion levels in the country.

The reporting behavior I uncover has both fairness and efficiency implications. First, tax evasion lowers the effective tax rate faced by self-employed to just one-half the statutory rate. In addition to violating horizontal equity, the dispersion in effective taxation across occupations would distort human capital acquisition and occupational choices of agents. Although no reliable estimates exist, the persistence of large self-employment sectors in developing economies (Jensen, 2015) suggests that this distortion could be large. Second, the step-shaped evasion function discovered in this paper means that the deadweight loss of raising the rate marginally above zero is orders of magnitude larger than increasing it elsewhere. Existing tax policy in developing countries does not seem to have adjusted to this reality. For example, with this pattern of evasion the very low tax rates employed by Pakistan in 2006-08 would not be optimal under plausible social preferences. Finally, large-scale and wide-spread evasion means that the gains from investment into enforcement capacity could be large, though the optimal level of such investment would depend upon the distortionary impact of the already employed tax instruments.

References

ANDREONI, JAMES, ERARD, BRIAN, & FEINSTEIN, JONATHAN. 1998. Tax complicance. *Journal of Economic Literature*, **36**, 818–860.

ARTAVANIS, NIKOLAOS, MORSE, ADAIR, & TSOUTSOURA, MARGARITA. 2016. Measuring Income Tax Evasion using Bank Credit: Evidence from Greece*. *The Quarterly Journal of Economics*.

BEST, MICHAEL CARLOS. 2014. *The Role of Firms in Workers' Earnings Responses to Taxes: Evidence From Pakistan*. London School of Economics Working Paper.

BEST, MICHAEL CARLOS, BROCKMEYER, ANNE, KLEVEN, HENRIK JACOBSEN, SPINNEWIJN, JOHANNES, & WASEEM, MAZHAR. 2015. Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan. *Journal of Political Economy*, **123**(6), 1311–1355.

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.

FBR. 2014. Biannual Review, January-June 2013-14. **13**(4). Federal Board of Revenue, Pakistan, http://goo.gl/gYHPoA.

FELDMAN, NAOMI E, & SLEMROD, JOEL. 2007. Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns. *The Economic Journal*, **117**(518), 327–352.

GORDON, ROGER, & LI, WEI. 2009. Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics*, **93**(7-8), 855–866.

GRUBER, JON, & SAEZ, EMMANUEL. 2002. The elasticity of taxable income: evidence and implications. *Journal of Public Economics*, **84**(1), 1 – 32.

JENSEN, ANDERS. 2015. *Employment Structure and the Rise of the Modern Tax System*. London School of Economis Working Paper.

KAPLOW, LOUIS. 1990. Optimal taxation with costly enforcement and evasion. *Journal of Public Economics*, **43**(2), 221–236.

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, & SCHULTZ, ESBEN ANTON. 2014. Estimating Taxable Income Responses Using Danish Tax Reforms. *American Economic Journal: Economic Policy*, **6**(4), 271–301.

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

KLEVEN, HENRIK JACOBSEN, KREINER, CLAUS THUSTRUP, & SAEZ, EMMANUEL. 2016. Why can modern governments tax so much? An agency model of firms as fiscal intermediaries. *Economica*, **83**(330), 219–246.

KOPCZUK, WOJCIECH, & SLEMROD, JOEL. 2006. Putting Firms into Optimal Tax Theory. *American Economic Review Papers and Proceedings*, **96**(2), 130–134.

LYSSIOTOU, PANAYIOTA, PASHARDES, PANOS, & STENGOS, THANASIS. 2004. Estimates of the Black Economy Based on Consumer Demand Approaches. *The Economic Journal*, **114**(497), 622–640.

MAYSHAR, JORAM. 1990. On measures of excess burden and their application. *Journal of Public Economics*, **43**(3), 263–289.

MAYSHAR, JORAM. 1991. Taxation with Costly Administration. *The Scandinavian Journal of Economics*, **93**(1), 75–88.

PIKETTY, THOMAS, & QIAN, NANCY. 2009. Income Inequality and Progressive Income Taxation in China and India, 19862015. *American Economic Journal: Applied Economics*, **1**(2), 53–63.

PISSARIDES, CHRISTOPHER A., & WEBER, GUGLIELMO. 1989. An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, **39**(1), 17 – 32.

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

SLEMROD, JOEL. 1995. Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986. *The American Economic Review*, **85**(2), 175–180.

SLEMROD, JOEL. 2001. A General Model of the Behavioral Response to Taxation. *International Tax and Public Finance*, **8**(2), 119–128.

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

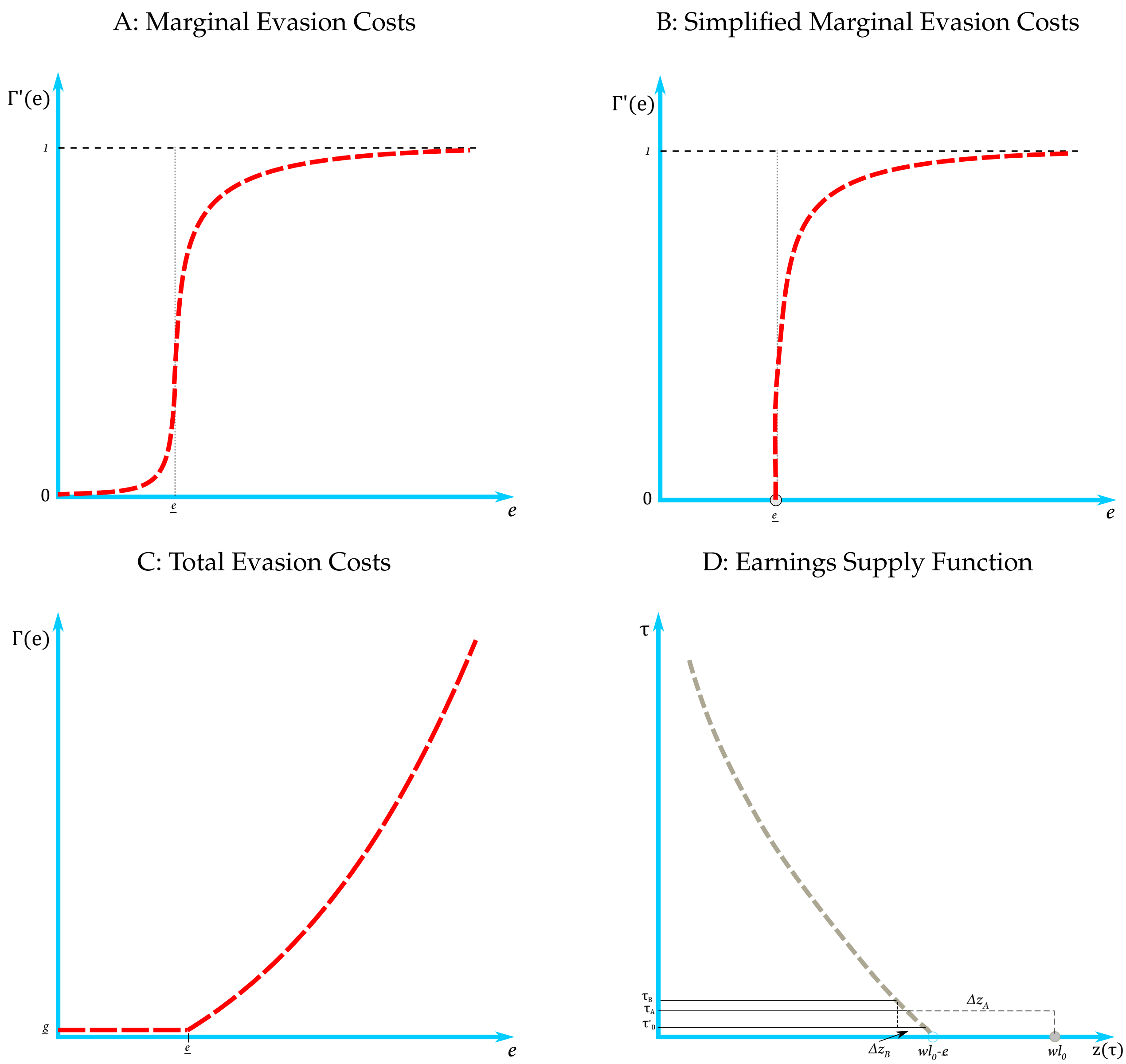
SLEMROD, JOEL, & WEBER, CAROLINE. 2012. Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, **19**(1), 25–53.

SLEMROD, JOEL, & YITZHAKI, SHLOMO. 2002. Tax avoidance, evasion, and administration. *In:* AUERBACH, A. J., & FELDSTEIN, M. (eds), *Handbook of Public Economics*, vol. 3.

SLEMROD, JOEL, BLUMENTHAL, MARSHA, & CHRISTIAN, CHARLES. 2001. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, **79**(3), 455 – 483.

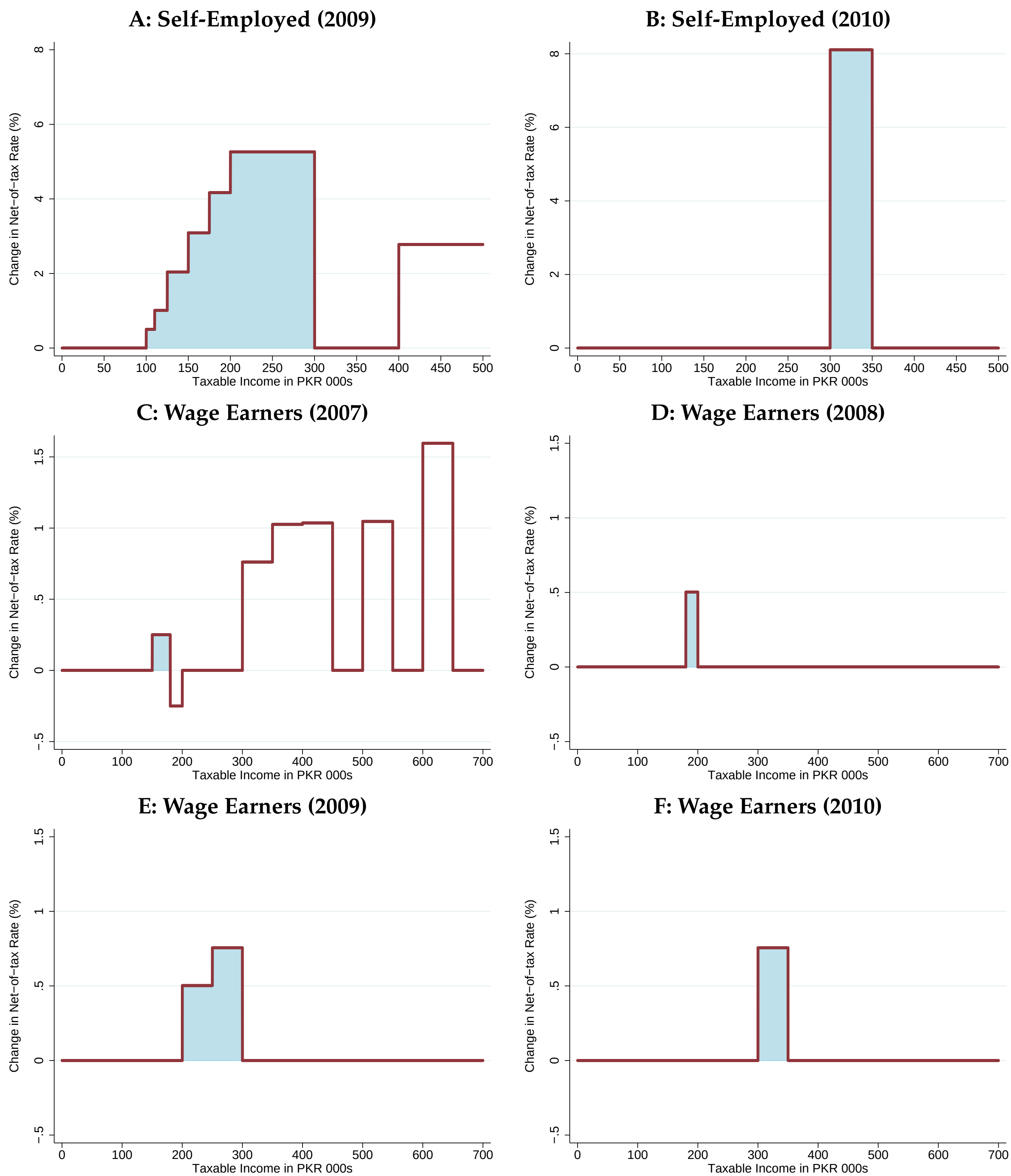
WASEEM, MAZHAR. 2016. Taxes, Informality and Income Shifting: Evidence from a Recent Pakistani Tax Reform. Working Paper, University of Manchester.

Figure I: Conceptual Framework



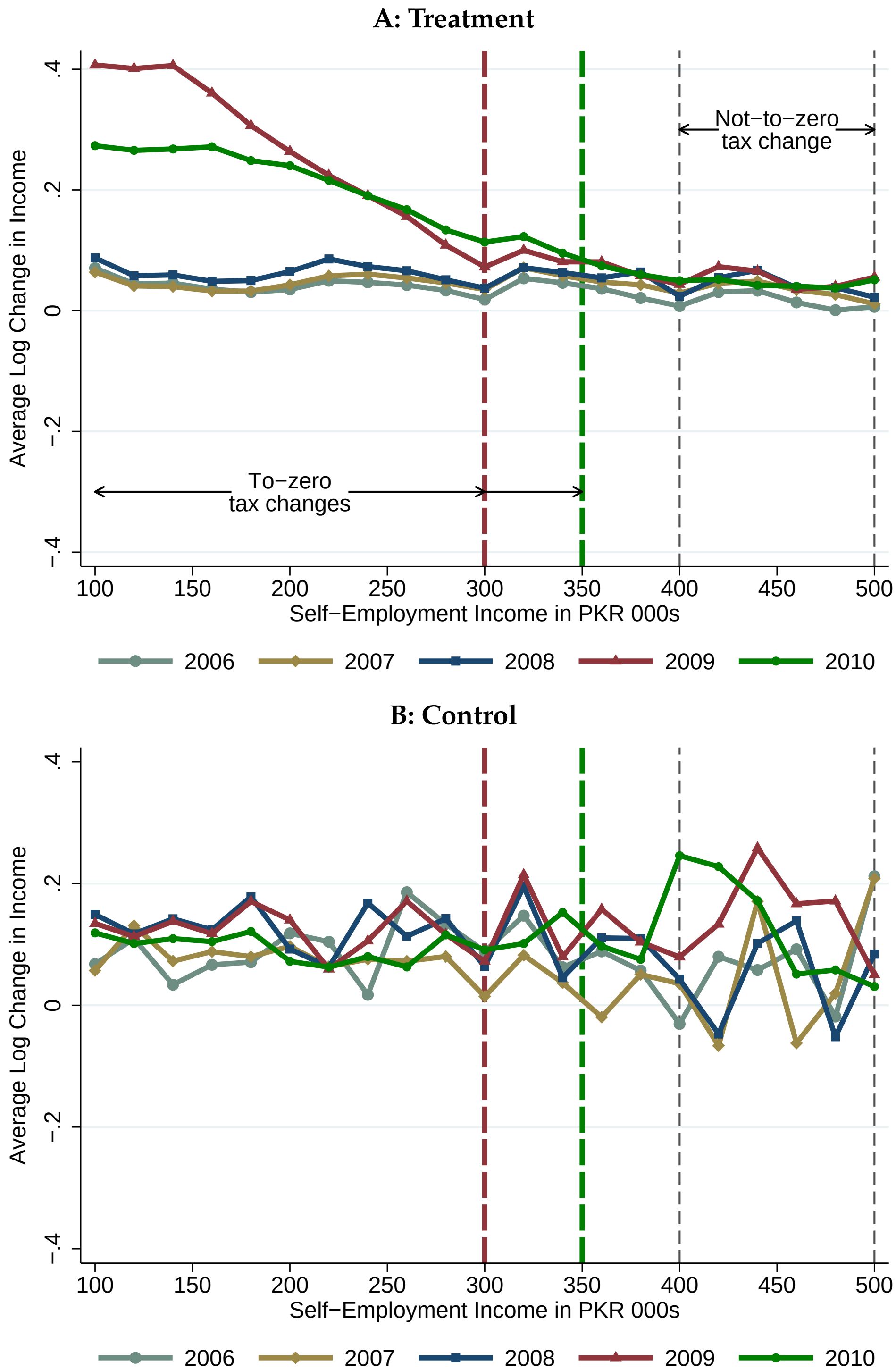
Notes: The figure illustrates the methodology to estimate the lower bound on tax evasion. Panel A plots the marginal evasion costs function (3), depicting the sharp uptick in evasion costs once the units of income covered by third-party information begin. Panel B is the simplified version of the function, which assumes the marginal cost of evading units of income not covered by third-party information to be zero. The total evasion costs function (4) implied by the simplified version is portrayed in panel C. It captures the idea that a taxpayer can evade up to $\underline{e}$ units of income on paying a small fixed cost $\underline{g}$ but faces the standard—increasing and convex—costs after that. Panel D displays the earnings supply function (10), illustrating how an optimizing taxpayer facing costs of evasion of the shape in Panel C would behave at various tax rates: the taxpayer would report true income wl_0 at the zero rate but discretely lower income $wl_0 - \underline{e}$ at a rate marginally above zero. The discontinuity means that the difference between the taxable income response to a to-zero reform $\Delta z_A(\tau_A \rightarrow 0)$ and a not-to-zero reform $\Delta z_B(\tau_B \rightarrow \tau_B^{'})$ identifies $\underline{e}$, the lower bound on tax evasion. Intuitively, any smooth change in reported income caused by a change in rate (movement along the curve) is netted out, leaving behind the discrete change in income from $wl_0 - \underline{e}$ to wl_0 (movement along the horizontal axis).

Figure II: Tax Variation



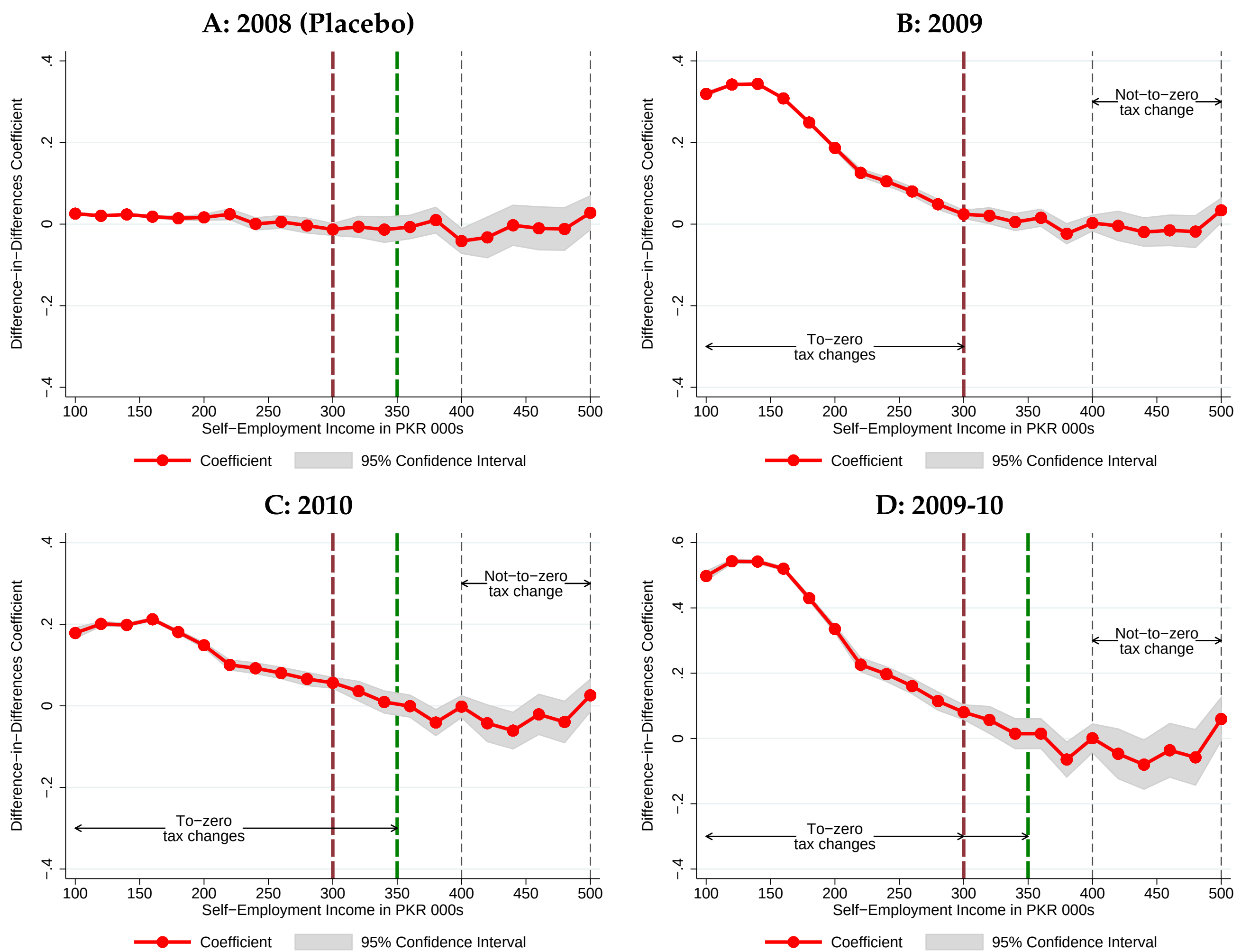
Notes: The figure displays tax variation created by the Pakistani reforms from 2006 to 2011. All curves plot percentage change in net-of-tax rate from period t to $t+1$ as a function of the base period income. Other than a small, narrow tax increase in 2007-08 for wage earners (Panel C), all reforms—both rate changes and movement of the bracket boundaries—result in a reduction of the tax rate, meaning percent changes in the net-of-tax rate created by them are invariably positive. The to-zero changes, which reduce the rate to zero, have been shaded with the light-blue color to distinguish them from the not-to-zero changes.

Figure III: Self-Employment Income Response



Notes: The figure compares how self-employment income responds to the to-zero and not-to-zero changes shown in Figure IIA–B, testing the main prediction of the model that responses produced by the to-zero rate changes would be uncharacteristically large. To construct the plots, taxpayers are grouped into bins of PKR 20,000 on the basis of their base period income. Then, average log change in income from year t to $t+1$ is plotted as a function of the base period income in the bin. Self-employed with base period income up to PKR 300K (350K) experience the reduction of the rate to zero going from 2009 to 2010 (from 2010 to 2011), whereas self-employed with base period income more than PKR 400K but not more than PKR 500K experience a roughly similar sized, not-to-zero rate reduction going from 2009 to 2010 (Panel A). Panel B displays the corresponding curves for the self-employment income of taxpayers classified wage earners by the tax code. This group of taxpayers does not experience the rate changes.

Figure IV: Self-Employment Income Response

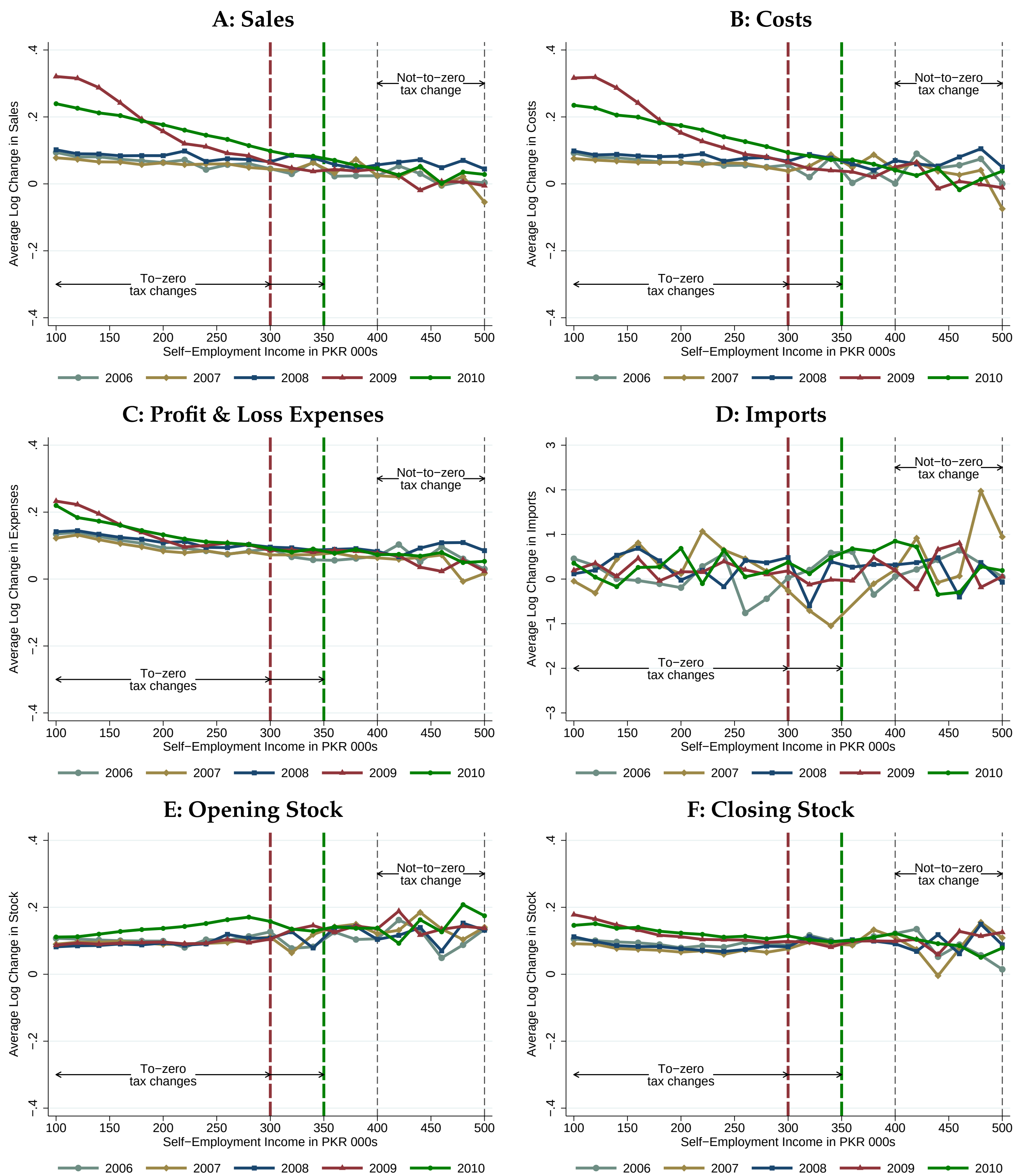


Notes: The figure displays a difference-in-differences analog of Figure III, illustrating self-employment income response to the rate changes. The panel for period t plots the corresponding coefficient δ_t from the following regressions

$$\Delta \log z_{it}^{S} = \alpha + \beta\, SE_i + \gamma\, \text{year}_t + \mathbf{SE_i} \times \mathbf{T_t}\, \boldsymbol{\delta} + u_{it},$$

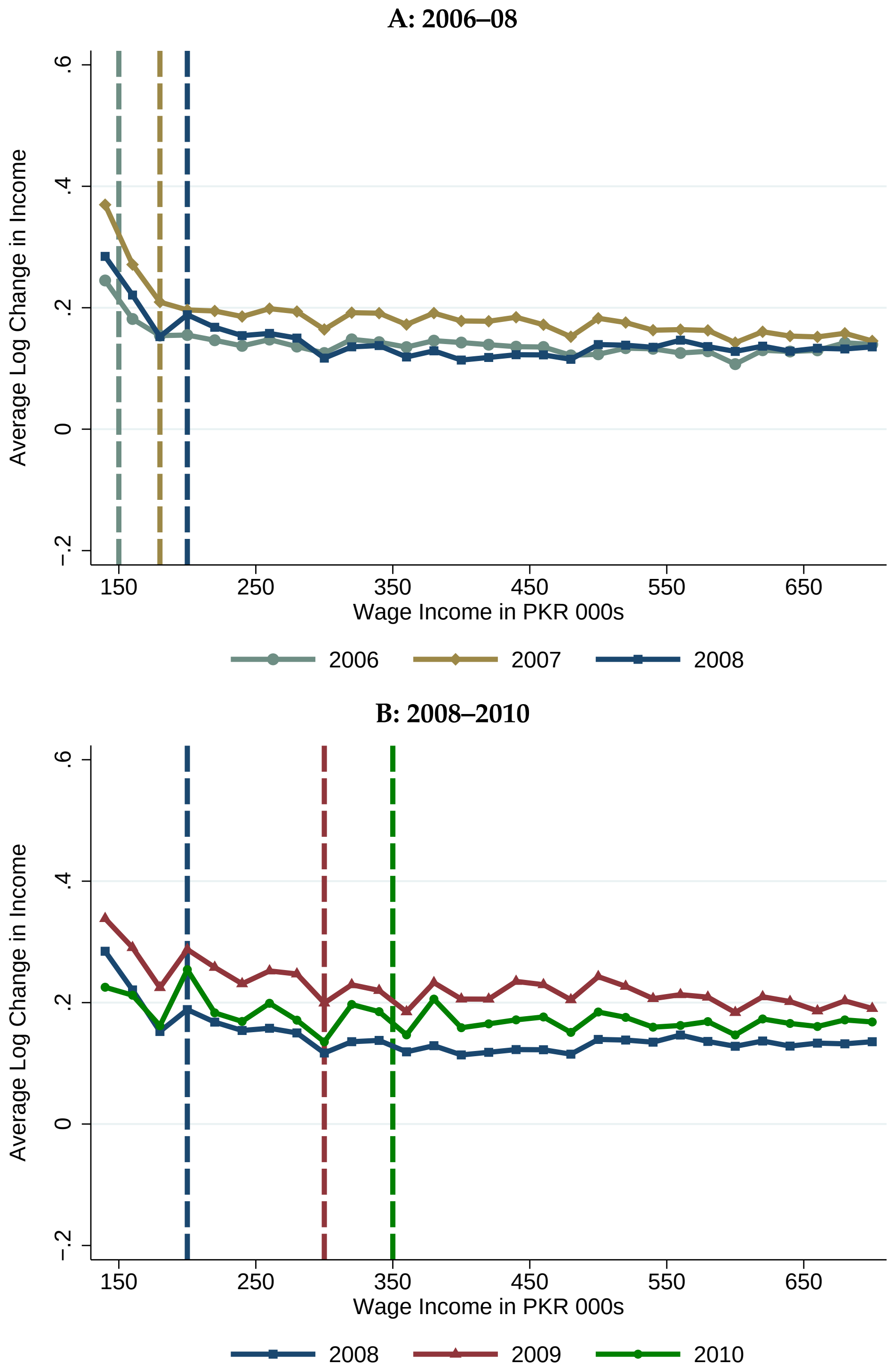
where year_t is a linear time trend and $\mathbf{SE_i} \times \mathbf{T_t}$ is a vector of interaction dummies. For the placebo regressions (Panel A), these interactions comprise three treatment $\times$ year dummies, one each for 2008–2010. For the regressions in Panels B and C, the 2008 interaction is dropped. Panel D illustrate the sum of 2009 and 2010 coefficients, reflecting aggregate response to the tax changes. The regressions are run separately in each bin so that the displayed coefficients reflect average additional earnings growth from period t to period $t+1$ experienced by the treated taxpayers with the base period income in the bin relative to the control taxpayers in the same bin. The 95% confidence interval around the coefficient is represented by the gray area in the plots. The standard errors have been clustered at the individual level. Vertical lines demarcate the areas where the to-zero and not-to-zero changes shown in Figure IIA–B were applied to.

Figure V: Line Items Response



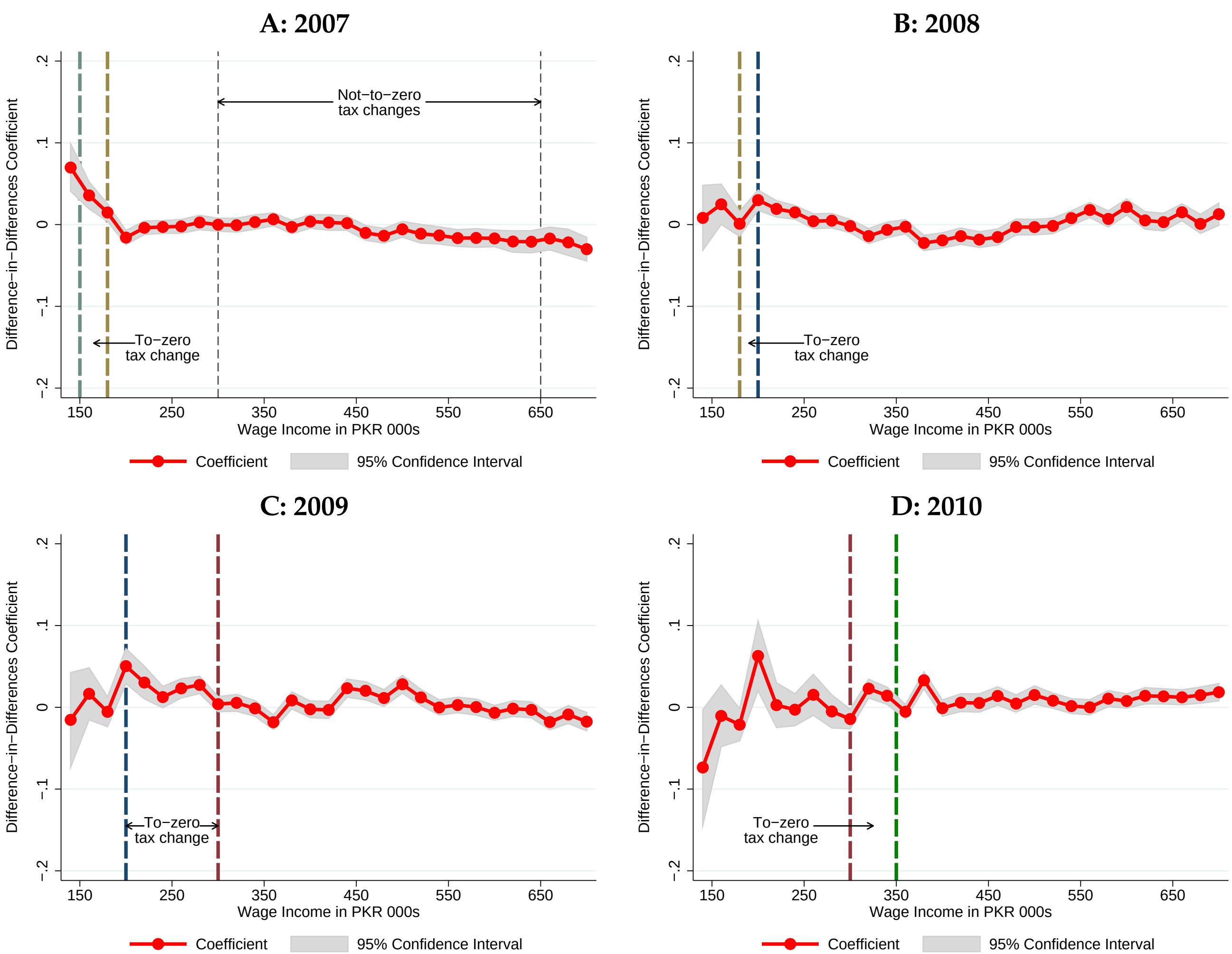
Notes: The figure looks at the anatomy of the self-employment income response depicted in Figure III, investigating if easy-to-misreport components of income—line items—respond differently from the other components. The detailed description of the line items used here are provided in Appendix A. To construct the plot, taxpayers are grouped into bins of PKR 20,000 on the basis of their base period income. Then, log change from period t to $t+1$ averaged across taxpayers in the bin is plotted as a function of the base period income in the bin. Vertical lines demarcate the areas where the to-zero and not-to-zero changes shown in Figure IIA–B were applied to.

Figure VI: Wage Income Response



Notes: The figure shows wage income growth from 2006–2011. Taxpayers are grouped into bins of PKR 20,000 on the basis of their base period income. Then, log change in income from period t to $t+1$ averaged across taxpayers in the bin is plotted as a function of the base period income in the bin. Vertical lines demarcate the exemption cutoff in the corresponding year. Taxpayer located in the area below the exemption cutoff for period t but above the exemption cutoff for period $t-1$ experience the reduction of the rate to zero. The magnitude of these to-zero changes are shown in Figure IIC–F. Taxpayers having wage income above PKR 180K experience not-to-zero changes shown in Figure IIC going from 2008 to 2009. The effects produced by these changes would be reflected in the curve corresponding to 2008.

Figure VII: Wage Income Response



Notes: The figure compares how wage income responds to the to-zero and not-to-zero rate changes shown in Figure IIC–F, investigating if its behavior is also consistent with the predictions of the model. I regress log changes in wage income from period t to $t+1$ on a full set of year fixed effects. The residuals from the regression are then regressed on a treatment group dummy and four yearly dummies one each for 2007 to 2010. These later regressions are run separately in the PKR 20K bins. The figure plots the estimated coefficients and 95% confidence intervals on the four yearly dummies from these regression. The coefficients represent the extent to which the detrended wage growth of taxpayers classified wage-earners in a given area of the income distribution is different in year t from the other years. The standard errors have been clustered at the individual level. The control group here comprises taxpayers whose wage income is subject to the schedule for self-employed. Vertical lines demarcate the areas where the to-zero and not-to-zero tax changes were applied to.

Table I: Difference-in-differences Estimates of Self-Employment Income Response

	Dependent variable: Log change in self-employment income from period t to $t+1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
to-zero × *SE* × 2009	0.281	0.281	0.297	0.297	0.297	0.296	0.261	0.260
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
to-zero × *SE* × 2010	0.156	0.156	0.145	0.144	0.145	0.144	0.142	0.142
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
not-to-zero × *SE* × 2009	-0.017	-0.017	-0.001	-0.001	-0.001	-0.001	-0.021	-0.021
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)
SE × *post*	-0.012	-0.010	-0.005	-0.005	-0.005	-0.005	-0.007	-0.006
	(0.006)	(0.007)	(0.002)	(0.002)	(0.002)	(0.002)	(0.008)	(0.008)
Fixed effects:								
Year	Yes	Yes	No	No	No	No	Yes	Yes
Industry	No	No	No	No	No	No	Yes	Yes
Industry × year	No	No	No	No	No	No	Yes	Yes
Time trend:								
Linear	No	No	Yes	Yes	No	No	No	No
Separate linear	No	No	No	No	Yes	Yes	No	No
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Pre-reform mean of the dependent variable	0.049	0.049	0.049	0.049	0.049	0.049	0.068	0.068
Observations	1,221,167	1,220,219	1,221,167	1,220,219	1,221,167	1,220,219	260,722	260,084

Notes: The table reports the estimates from equation (14). The first two columns correspond to the baseline specification; columns (3)-(6) replace year fixed effects with parametric time trends—linear in columns (3)-(4) and separate linear in columns (5)-(6); columns (7)-(8) include year, industry and industry × year fixed effects, allowing taxpayers in each industry their own earnings growth trend. The details of the industry classification are provided in Appendix A. I do not observe industry classification for all taxpayers, owing to which the numbers of observations in the last two columns are lower. Reassuringly, however, even for this restricted sample the results are similar to the baseline results. The control group here (*SE*=0) comprises taxpayers whose self-employment income is subject to the schedule for wage earners. All even-numbered columns drop control group taxpayers for whom the mechanical change in tax rate from period t to $t+1$ is not zero. Standard errors are in parenthesis, which have been clustered at the individual level.

Table II: Triple-difference Estimates of Self-Employment Income Response

	Dependent variable: Log change in self-employment income from period t to $t+1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
to-zero $\times$ *SE* $\times$ 2009	0.313	0.314	0.279	0.280	0.297	0.304	0.270	0.270
	(0.025)	(0.026)	(0.007)	(0.008)	(0.012)	(0.014)	(0.026)	(0.027)
to-zero $\times$ *SE* $\times$ 2010	0.176	0.177	0.130	0.131	0.148	0.155	0.144	0.145
	(0.025)	(0.026)	(0.007)	(0.008)	(0.012)	(0.014)	(0.026)	(0.027)
to-zero $\times$ *post*	-0.037	-0.038	0.005	0.004	-0.013	-0.020	-0.040	-0.041
	(0.025)	(0.026)	(0.006)	(0.007)	(0.012)	(0.013)	(0.026)	(0.026)
to-zero $\times$ *SE*	-0.009	-0.016	0.021	0.014	0.007	-0.004	0.008	0.003
	(0.020)	(0.020)	(0.013)	(0.014)	(0.015)	(0.015)	(0.020)	(0.021)
SE $\times$ *post*	-0.037	-0.037	0.003	0.003	0.004	0.003	-0.015	-0.015
	(0.024)	(0.025)	(0.003)	(0.003)	(0.003)	(0.003)	(0.025)	(0.026)
Fixed effects:								
Year	Yes	Yes	No	No	No	No	Yes	Yes
Industry	No	No	No	No	No	No	Yes	Yes
Industry $\times$ year	No	No	No	No	No	No	Yes	Yes
Time trend:								
Linear	No	No	Yes	Yes	No	No	No	No
Separate linear	No	No	No	No	Yes	Yes	No	No
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Pre-reform mean of the dependent variable	0.049	0.049	0.049	0.049	0.049	0.049	0.068	0.068
Observations	1,221,167	1,220,219	1,221,167	1,220,219	1,221,167	1,220,219	260,722	260,084

Notes: The table reports the estimates from equation (15). The first two columns correspond to the baseline specification; columns (3)-(6) replace year fixed effects with parametric time trends—linear in columns (3)-(4) and separate linear in columns (5)-(6); columns (7)-(8) include year, industry and industry $\times$ year fixed effects, allowing taxpayers in each industry their own earnings growth trend. The definition of the industry variable is provided in Appendix A. I do not observe industry classification for all taxpayers, owing to which the numbers of observations in the last two columns are lower. Reassuringly, however, even for this restricted sample the results are similar to the baseline results. The control group here (*SE*=0) comprises taxpayers whose self-employment income is subject to the schedule for wage earners. All even-numbered columns drop control group taxpayers for whom the mechanical change in tax rate from period t to $t+1$ is not zero. Standard errors are in parenthesis, which have been clustered at the individual level.

Table III: Estimates of the Rate of Evasion of Self-Employment Income

Income	Earnings Response ($\tau \to 0$)	Earnings Response ($\tau \not\to 0$)	Difference	Evasion Rate (%)
(1)	(2)	(3)	(4)	(5)
80-100K	72,187	334	71,853	74.6
	(1,256)	(189)	(1,271)	(1.3)
100-150K	92,880	604	92,276	72.4
	(792)	(342)	(862)	(0.7)
150-200K	93,003	573	92,430	52.5
	(948)	(209)	(970)	(0.6)
200-250K	71,203	762	70,441	31.0
	(1,126)	(278)	(1,160)	(0.5)
250-300K	29,469	689	28,780	10.2
	(1,183)	(251)	(1,209)	(0.4)
300-350K	20,404	429	19,975	6.1
	(912)	(156)	(925)	(0.3)

Notes*:* The table presents the estimates of the rates of evasion of self-employment income from equation (13). Column (1) shows the income segment; column (2) the earnings response produced by the to-zero change; column (3) the earnings response produced by the equal-sized, not-to-zero change; column (4) the difference between the two responses; and column (5) the average evasion rate in the segment. The estimates in column (2) are computed from the self-employment income response to the movements of the exemption cutoff in 2010–11 shown in Table I. The estimates in column (3) are computed using elasticities estimated for the same group of taxpayers in Kleven & Waseem (2013). The details on how the estimates in columns (2)–(3) have been computed are in section IV.A.3. The difference between the two set of estimates represents average $\underline{e}$ for self-employed in the segment. The evasion rate implied by the difference in column(4) is shown in column (5). The standard errors are in parenthesis.

Table IV: Difference-in-differences Estimates of Wage Income Response

	Dependent variable: Log change in wage income from period t to $t+1$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A: Tax-Driven Response									
to-zero	0.019	0.019	0.016	0.022	0.021	0.018	0.022	0.021	0.018
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
not-to-zero	0.004	0.003	0.004	0.007	0.007	0.008	0.007	0.007	0.008
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
B: Placebo									
to-zero	0.009	0.007	0.006	0.010	0.007	0.006	0.010	0.007	0.006
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
not-to-zero	0.008	0.007	0.007	0.014	0.013	0.014	0.014	0.013	0.014
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Fixed effects:									
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Region × year	No	No	Yes	No	No	Yes	No	No	Yes
Base-period income controls:									
Log base-period income	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Spline of log base-period income	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Pre-reform mean of the dependent variable	0.162	0.162	0.162	0.162	0.162	0.162	0.162	0.162	0.162
Observations	240,804	240,798	240,798	240,804	240,798	240,798	240,804	240,798	240,798

Notes: This table reports the estimates from equation (16). Column (1) corresponds to the baseline specification, and the subsequent columns add more control variables, permuting among the combinations of controls for mean reversion—log base period income and a ten-piece spline in log base period income—and other controls. The definition of the region variable is provided in Appendix A. Panel B carries the results from placebo regressions corresponding to each specification, assuming that all tax changes took place one year earlier than they actually did. Standard errors are in parenthesis, which have been clustered at the individual level.

A Online Appendix

A.A Details of Variables

(*i*) **Taxable Income.** Income derived by a person in a year from all sources such as employment, business, ownership of capital, minus three minor deductions Zakat, Workers' Welfare Fund, and Workers' Participation Fund.[22] The deductions, however, are extremely rare for wage earners and self-employed, and practically taxable income is the aggregate income of a taxpayer in a year.

(*ii*) **Wage Income.** Income received by an employee in a tax year from employment including wages, leave pay, bonuses, commission, fees, and gratuity.

(*iii*) **Self-Employment Income.** Profits from any business carried out by a person other than those from partnerships or corporations.

(*iv*) **Self-Employed.** The tax code classifies a person self-employed if her wage income does not exceed 50% of taxable income.

(*v*) **Wage Income.** The tax code classifies a person wage earner if her wage income exceeds 50% of taxable income.

(*vi*) **Sales.** The variable is from the brief profit and loss account included in the tax return form, and denotes the full revenue for the year from selling goods or services.

(*vii*) **Costs.** The variable is also from the profit and loss account, and denotes what in accounting is referred to as the cost of sales. The cost is calculated by adding opening stock, net purchases, and manufacturing and trading expenses, and then taking away the closing stock.

(*viii*) **Profit and Loss Expenses.** The variable is also from the profit and loss account, and denotes input costs such as wages, rents, accounting and legal fees, electricity, and interest paid on loans.

(*ix*) **Imports.** Imports are costs net of import duties and taxes that are incurred on imported inputs / investment goods.

(*x*) **Opening Stock.** Inventory at the start of the year.

(*xi*) **Closing Stock.** Inventory at the end of the year.

[22] Zakat is a charitable payment made annually by Muslims as a religious obligation. Workers Welfare Fund is paid by industrial establishment having taxable income exceeding Rs. 500,000 in a year. Workers' Participation Fund is paid by industrial establishments having more than 50 workers.

(*xii*) **Electronic Return Filer.** I categorize a taxpayer electronic return filer if any of the six returns from 2006–2011 was filed electronically.

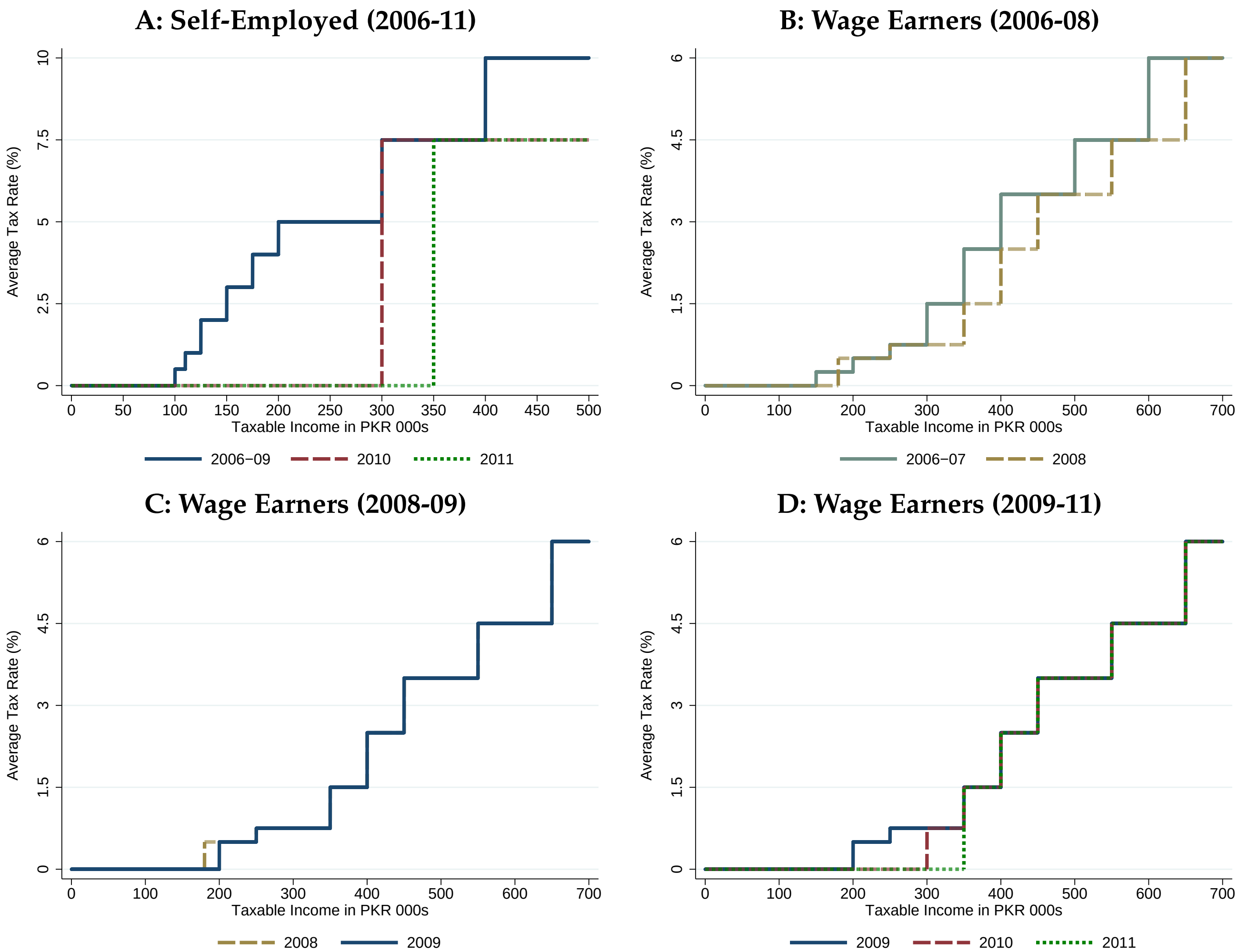
(*xiii*) **VAT-registered.** The variable indicates if a self-employed taxpayer is registered with the FBR to remit VAT on its sales.

(*xiv*) **Region.** The variable codes the tax district a taxpayer's registered office is located in.

(*xv*) **Tax office.** The variable represents the tax office that is responsible for audit and enforcement of an individual's tax liability. Pakistan has two types of tax offices: (1) Large Taxpayer Units, which are located in Karachi, Lahore and Islamabad and cater to the top tax contributors; (2) Regional Tax Offices, which are located in twelve cities and administer the rest of taxpayers.

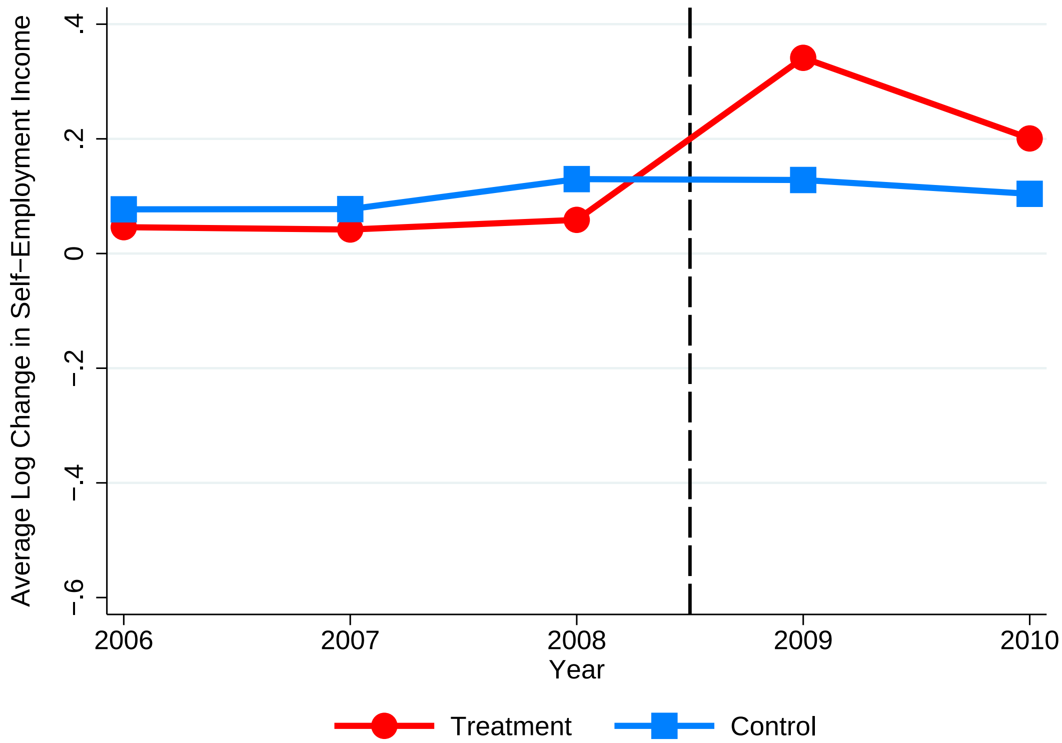
(*xvi*) **Industry.** The variable represents the 2-digit industry code a self-employed taxpayer operates in.

Figure A.I: Tax Schedules



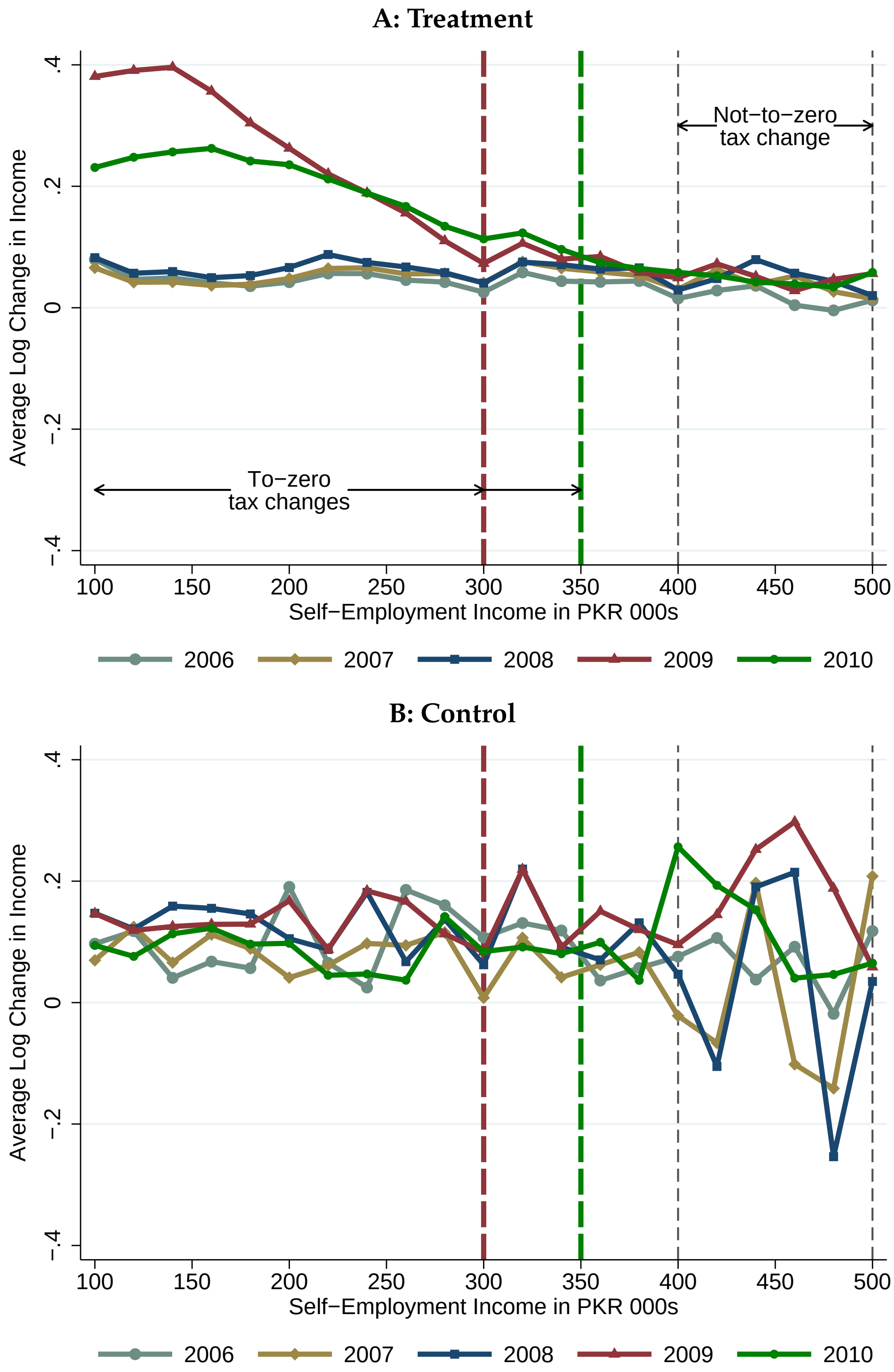
Notes: The figure shows the two personal income tax schedules of Pakistan from 2006 to 2011. A period t plot illustrates the statutory tax rate in the period as a function of taxable income. The Pakistani schedules specify average rather than marginal tax rate in a bracket. The schedule for self-employed underwent three changes during the period. First, the exemption cutoff was increased from PKR 100K to PKR 300K in 2010 and to PKR 350K in 2011. Second, the average tax rate on taxable income more than PKR 400K but not more than PKR 500K was reduced from 10% to 7.5% in 2010. There were more changes to the schedule higher up the income distribution in 2010, which I do not show here. The schedule for wage earners was comprehensively revised in 2008 (Panel B). The exemption cutoff for wage earners was increased from PKR 150K to PKR 180K in 2008, to PKR 200K in 2009, to PKR 300K in 2010, and to PKR 350K in 2011. The net-of-tax rate changes implied by these reforms are plotted in Figure II.

Figure A.II: Self-Employment Income Trend



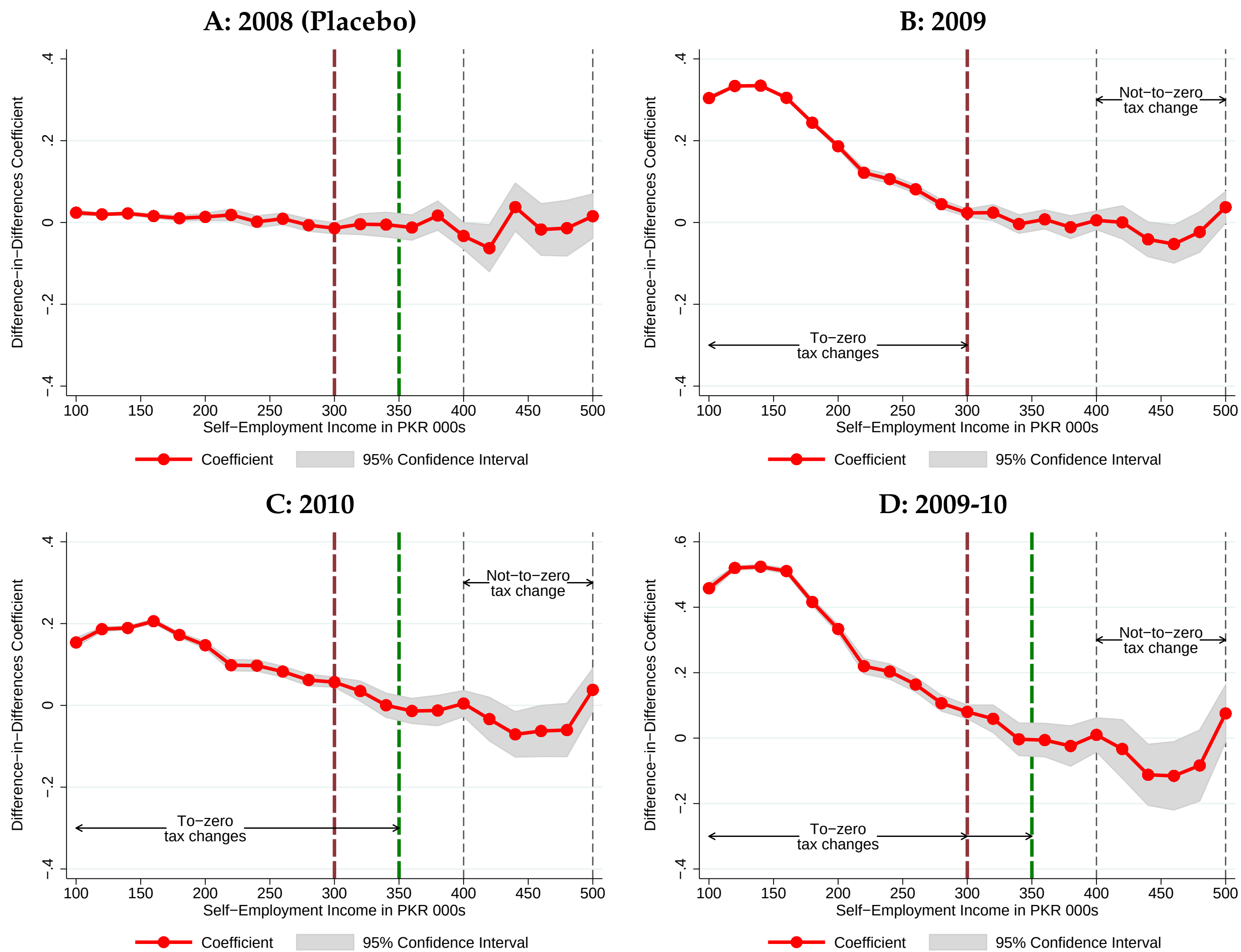
Notes: The figure plots the time path of self-employment income growth from 2006 to 2010. Each dot on the two curves illustrates log change in self-employment income from period t to $t+1$ averaged across taxpayers in the corresponding group. The treatment group comprises individuals whose self-employment income is subject to the schedule for self-employed, whereas the control group comprises individuals whose self-employment income is subject to the schedule for wage earners. Vertical dashed line demarcates the time from which the to-zero and not-to-zero tax changes applied to the schedule for self-employed in 2010–11, as shown in Figure IIA–B, affect income growth in the treatment group.

Figure A.III: Self-Employment Income Response (Balanced Panel)



Notes: The figure compares how self-employment income responds to the to-zero and not-to-zero rate changes shown in Figure IIA–B for a balanced panel of taxpayers who file in all years from 2006 to 2011, testing the main prediction of the model that responses produced by the to-zero rate changes would be uncharacteristically large. To construct the plots, taxpayers are grouped into bins of PKR 20,000 on the basis of their base period income. Then, average log changes in income from year t to $t+1$ are plotted as a function of the base period income in the bin. Self-employed with base period income up to PKR 300K (350K) experience the reduction of the rate to zero going from 2009 to 2010 (from 2010 to 2011), whereas self-employed with base period income more than PKR 400K but not more than PKR 500K experience a roughly similar sized, not-to-zero rate reduction going from 2009 to 2010 (Panel A). Panel B displays the corresponding curves for the self-employment income of taxpayers classified wage earners by the tax code. This group of taxpayers does not experience the rate changes.

Figure A.IV: Self-Employment Income Response (Balanced Panel)

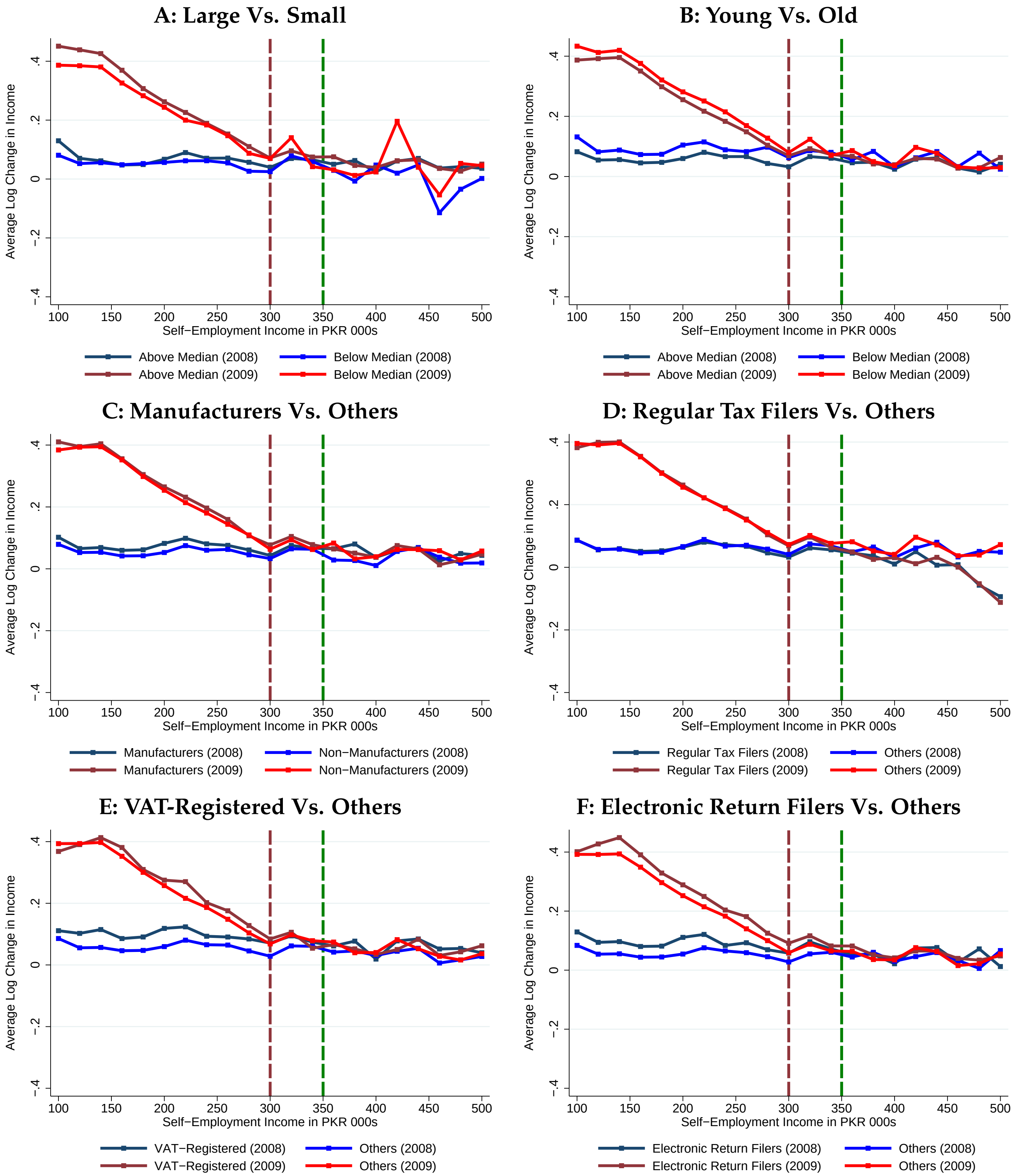


Notes: The figure displays a difference-in-differences analog of Figure A.III, illustrating self-employment income response to the tax reforms for a balanced panel of taxpayers who file in all years from 2006 to 2011. The panel for period t plots the corresponding coefficient δ_t from the following regressions

$$\Delta \log z_{it}^{S} = \alpha + \beta\, SE_i + \gamma\, \text{year}_t + \mathbf{SE_i} \times \mathbf{T_t}\, \boldsymbol{\delta} + u_{it},$$

where year_t is a linear time trend and $\mathbf{SE_i} \times \mathbf{T_t}$ is a vector of interaction dummies. For the placebo regressions (Panel A), these interactions comprise three treatment $\times$ year dummies, one each for 2008–2010. For the regressions in Panels B and C, the 2008 interaction is dropped. Panel D illustrate the sum of 2009 and 2010 coefficients, reflecting aggregate response to the tax changes. The regressions are run separately in each bin so that the displayed coefficients reflect average additional earnings growth from period t to period $t+1$ experienced by the treated taxpayers with the base period income in the bin relative to the control taxpayers in the same bin. The 95% confidence interval around the coefficient is represented by the gray area in the plots. The standard errors have been clustered at the individual level. Vertical lines demarcate the areas where the to-zero and not-to-zero tax changes shown in Figure IIA–B were applied to.

Figure A.V: Heterogeneity in Self-Employment Income Response



Notes: The figure explores heterogeneity in self-employment income response. Taxpayers are grouped into bins of PKR 20,000 on the basis of their base period self-employment income, and then average log change in income from period t to $t+1$ is plotted as a function of the base period income in the bin, stratifying the sample by the given trait j. The figure shows that there is hardly any difference in response across taxpayers with different characteristics but located in the same area of the income distribution. The details of taxpayer characteristics variables used here are given in Appendix A.

Table A.I: Summary Statistics

	Full Sample		Analysis Sample	
	Self-Employed	Wage Earners	Self-Employed	Wage Earners
	(1)	(2)	(3)	(4)
Observations				
1. Taxable Income > 0	2,187,943	985,293	1,279,516	274,478
2. Self-Employment Income > 0	100	4.30	100	6.40
3. Wage Income > 0	0.60	100	0.40	100
4. Partnership Income > 0	3.50	1.10	1.80	1.20
5. Switchers	0.90	1.30	0.90	2.70
Outcomes				
6. Taxable Income	3,825,812	3,509,781	160,154	403,456
	(3,334,397,184)	(1,408,862,848)	(69,773)	(153,240)
	[147,800]	[512,185]	[135,000]	[390,234]
7. Self-Employment Income	3,820,831	154,147	178,195	86,146
	(3,362,148,352)	(244,261)	(16,954,248)	(77,459)
	[146,000]	[88,889]	[135,000]	[65,500]
8. Wage Income	356,891	3,049,276	142,476	397,086
	(1,236,341)	(1,351,159,296)	(187,928)	(164,085)
	[160,000]	[512,887]	[113,460]	[382,493]
Characteristics				
9. Years Registered	7.04	7.80	7.39	7.79
	(4.95)	(5.19)	(4.83)	(4.99)
10. Location	0.39	0.52	0.38	0.47
	(0.49)	(0.50)	(0.49)	(0.50)
11. Male	0.98	0.97	0.99	0.97
	(0.12)	(0.18)	(0.11)	(0.18)
12. Age	39.55	43.41	40.39	43.27
	(15.48)	(12.23)	(15.22)	(13.08)

Notes*:* This table presents the summary statistics of the data. The analysis sample for period t contains taxpayers for whom *(i)* log $\frac{z_{it+1}}{z_{it}}$ is defined, *(ii)* log $\frac{z_{it+1}}{z_{it}}$ is greater than the first percentile and less than the 99th percentile of the corresponding pooled distribution, and *(iii)* $z_{it} \in (80\text{K } 500\text{K}]$ if i is a self-employed and $z_{it} \in (140\text{K } 700\text{K}]$ if i is a wage earner. The detailed description of variables used here are provided in Appendix A. The first row of the table reports the number of self-employed and wage earners in the two samples who report positive taxable income. Rows 2-5 of the table shows the share of taxpayers with the given characteristic in the corresponding sample. Rows 6–12 report the mean of the outcome / characteristic variable in the corresponding sample with the standard error in parenthesis and the median in square brackets. The variable location in row 10 of the table indicates that the taxpayer is located in one of the three big cities of Pakistan, Lahore, Karachi, and Islamabad.

Table A.II: Before-after Estimates of Self-Employment Income Response

	Dependent variable: Log change in self-employment income from period t to $t+1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
to-zero $\times$ 2009	0.292	0.295	0.288	0.288	0.287	0.283	0.259	0.255
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
to-zero $\times$ 2010	0.146	0.143	0.142	0.142	0.143	0.138	0.142	0.137
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
not-to-zero $\times$ 2009	-0.004	-0.001	-0.003	-0.004	-0.010	-0.009	-0.021	-0.020
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
post	0.014	-0.004	0.017	0.017	0.006	0.007	-0.061	0.103
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.061)	(0.074)
Fixed effects:								
Year	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Region	No	No	Yes	No	No	Yes	No	Yes
Tax Office	No	No	No	Yes	No	Yes	No	Yes
Industry	No	No	No	No	Yes	Yes	Yes	Yes
Industry $\times$ year	No	No	No	No	No	No	Yes	Yes
Time trend:								
Linear	No	Yes	No	No	No	No	No	No
Pre-reform mean of the dependent variable	0.049	0.049	0.049	0.049	0.068	0.068	0.068	0.068
Observations	1,214,538	1,214,538	1,214,524	1,152,314	255,153	255,140	255,153	255,140

Notes: This table reports the estimates from the before-after analog of equation (14). The first column in the table correspond to the baseline specification; column (2) replaces the year fixed effects with a linear time trend; and the rest of the columns add additional control variables. The definitions of the control variables used here are provided in Appendix A. I do not observe industry classification for all taxpayers, owing to which the numbers of observations in the last four columns are lower. Reassuringly, however, even for this restricted sample the results are similar to the baseline results. Standard errors are in parenthesis, which have been clustered at the individual level.

Table A.III: Difference-in-differences Estimates of Self-Employment Income Response (Balanced Panel)

	Dependent variable: Log change in self-employment income from period t to $t+1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
to-zero × *SE* × 2009	0.290	0.290	0.318	0.318	0.319	0.318	0.264	0.264
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)
to-zero × *SE* × 2010	0.191	0.191	0.168	0.168	0.168	0.168	0.187	0.187
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
not-to-zero × *SE* × 2009	-0.085	-0.086	-0.057	-0.057	-0.057	-0.057	-0.097	-0.097
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)
$SE \times post$	0.007	0.005	-0.029	-0.029	-0.029	-0.029	0.009	0.008
	(0.012)	(0.013)	(0.002)	(0.002)	(0.002)	(0.002)	(0.013)	(0.014)
Fixed effects:								
Year	Yes	Yes	No	No	No	No	Yes	Yes
Industry	No	No	No	No	No	No	Yes	Yes
Industry × year	No	No	No	No	No	No	Yes	Yes
Time trend:								
Linear	No	No	Yes	Yes	No	No	No	No
Separate linear	No	No	No	No	Yes	Yes	No	No
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Pre-reform mean of the dependent variable	0.050	0.050	0.050	0.050	0.050	0.050	0.067	0.066
Observations	527,835	527,595	527,835	527,595	527,835	527,595	94,705	94,575

Notes: This table reports the estimates from equation (14), restricting the sample to a balanced panel of taxpayers who file in all six year 2006-2011. The first two columns in the table correspond to the baseline specification; columns (3)-(6) replace year fixed effects with parametric time trends—linear in columns (3)-(4) and separate linear in columns (5)-(6); columns (7)-(8) include year, industry and industry × year fixed effects, allowing taxpayers in each industry their own earnings growth trend. The definition of the industry variable is provided in Appendix A. I do not observe industry classification for all taxpayers, owing to which the numbers of observations in the last two columns are lower. Reassuringly, however, even for this restricted sample the results are similar to the baseline results. The control group here (*SE*=0) comprises taxpayers whose self-employment income is subject to the schedule for wage earners. All even-numbered columns drop control group taxpayers for whom the mechanical change in tax rate from period t to $t+1$ is not zero. Standard errors are in parenthesis, which have been clustered at the individual level.

Table A.IV: Triple-difference Estimates of Self-Employment Income Response (Balanced Panel)

	Dependent variable: Log change in self-employment income from period t to $t+1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
to-zero × *SE* × 2009	0.321	0.318	0.315	0.312	0.308	0.296	0.269	0.266
	(0.041)	(0.041)	(0.012)	(0.013)	(0.021)	(0.022)	(0.043)	(0.043)
to-zero × *SE* × 2010	0.183	0.180	0.168	0.165	0.161	0.149	0.154	0.151
	(0.041)	(0.041)	(0.012)	(0.013)	(0.021)	(0.022)	(0.043)	(0.043)
to-zero × $post$	-0.042	-0.039	-0.030	-0.027	-0.023	-0.011	-0.047	-0.044
	(0.041)	(0.041)	(0.011)	(0.013)	(0.020)	(0.022)	(0.042)	(0.042)
to-zero × *SE*	-0.092	-0.093	-0.085	-0.085	-0.080	-0.074	-0.064	-0.064
	(0.033)	(0.033)	(0.026)	(0.026)	(0.028)	(0.029)	(0.034)	(0.034)
SE × $post$	-0.011	-0.011	-0.001	-0.001	-0.001	-0.001	0.014	0.014
	(0.040)	(0.040)	(0.004)	(0.004)	(0.004)	(0.004)	(0.042)	(0.042)
Fixed effects:								
Year	Yes	Yes	No	No	No	No	Yes	Yes
Industry	No	No	No	No	No	No	Yes	Yes
Industry × year	No	No	No	No	No	No	Yes	Yes
Time trend:								
Linear	No	No	Yes	Yes	No	No	No	No
Separate linear	No	No	No	No	Yes	Yes	No	No
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Pre-reform mean of the dependent variable	0.050	0.050	0.050	0.050	0.050	0.050	0.067	0.066
Observations	527,835	527,595	527,835	527,595	527,835	527,595	94,705	94,575

Notes*:* This table reports the estimates from equation (15), restricting the sample to a balanced panel of taxpayers who file in all six year 2006-2011. The first two columns in the table correspond to the baseline specification; columns (3)-(6) replace year fixed effects with parametric time trends—linear in columns (3)-(4) and separate linear in columns (5)-(6); columns (7)-(8) include year, industry and industry × year fixed effects, allowing taxpayers in each industry their own earnings growth trend. The definition of the industry variable is provided in Appendix A. I do not observe industry classification for all taxpayers, owing to which the numbers of observations in the last two columns are lower. Reassuringly, however, even for this restricted sample the results are similar to the baseline results. The control group here (*SE*=0) comprises taxpayers whose self-employment income is subject to the schedule for wage earners. All even-numbered columns drop control group taxpayers for whom the mechanical change in tax rate from period t to $t+1$ is not zero. Standard errors are in parenthesis, which have been clustered at the individual level.

Table A.V: Robustness of the Difference-in-differences Estimates of Self-Employment Income Response

	Dependent variable: Log change in self-employment income from period t to $t+1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
to-zero × *SE* × 2009	0.281	0.252	0.297	0.298	0.276	0.277	0.277	0.273
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
to-zero × *SE* × 2010	0.156	0.149	0.162	0.165	0.152	0.152	0.152	0.147
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
not-to-zero × *SE* × 2009	-0.017	-0.024	-0.017	-0.009	-0.021	-0.016	-0.017	-0.020
	(0.005)	(0.006)	(0.006)	(0.005)	(0.007)	(0.005)	(0.005)	(0.007)
SE × $post$	-0.012	-0.012	-0.010	-0.007	-0.023	-0.006	-0.006	-0.020
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)	(0.008)
Fixed effects:								
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry	No	No	No	No	Yes	No	No	Yes
Region	No	No	No	No	No	Yes	No	Yes
Tax office	No	No	No	No	No	No	Yes	Yes
Sample:								
$z_{it} \in (80\text{K } 500\text{K}]$	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Bunchers dropped	No	Yes	No	No	No	No	No	No
$z_{it} \in (0\ 500\text{K}]$	No	No	Yes	No	No	No	No	No
$z_{it} \in (0\ \infty)$	No	No	No	Yes	No	No	No	No
Pre-reform mean of the dependent variable	0.049	0.052	0.052	0.052	0.049	0.049	0.049	0.068
Observations	1,221,167	777,387	1,271,218	1,282,104	260,063	1,220,205	1,157,932	260,050

Notes: This table assesses the robustness of the estimates from equation (14). Column (2) drops taxpayers who *bunch* at the notches from concerns that their base period income might be low owing to strong, local incentives created by the notches or because these taxpayers might be special. Columns (3)–(4) increase the range of the data from $z_{it} \in (80\text{K } 500\text{K}]$ in the baseline results to $z_{it} \in (0\ 500\text{K}]$ in column (3) and $z_{it} > 0$ in column (4). Columns (5)–(8) add additional control variables into specification (14): industry fixed effects in column (5), region fixed effects in column (6), tax office fixed effects in column (7), and all these fixed effects together in column (8). The control group here (*SE*=0) comprises taxpayers whose self-employment income is subject to the schedule for wage earners. Standard errors are in parenthesis, which have been clustered at the individual level.

Table A.VI: Placebo Difference-in-differences Estimates of Self-Employment Income Response

	Dependent variable: Log change in self-employment income from period t to $t+1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
to-zero × *SE* × 2008	0.009	0.009	0.009	0.009	0.009	0.009	0.018	0.018
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
not-to-zero × *SE* × 2008	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021	-0.030	-0.030
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
SE × *post*	-0.031	-0.035	0.011	0.011	0.012	0.012	-0.036	-0.038
	(0.008)	(0.009)	(0.002)	(0.002)	(0.002)	(0.002)	(0.009)	(0.010)
Fixed effects:								
Year	Yes	Yes	No	No	No	No	Yes	Yes
Industry	No	No	No	No	No	No	Yes	Yes
Industry × year	No	No	No	No	No	No	Yes	Yes
Time trend:								
Linear	No	No	Yes	Yes	No	No	No	No
Separate linear	No	No	No	No	Yes	Yes	No	No
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Pre-reform mean of the dependent variable	0.041	0.041	0.041	0.041	0.041	0.041	0.052	0.052
Observations	757,029	756,637	757,029	756,637	757,029	756,637	138,593	138,341

Notes: This table reports the results from placebo regressions corresponding to equation (14). I restrict the sample for this analysis to pre-reform periods (2006-2009) only and pretend that the to-zero and not-to-zero changes took place one year earlier than they actually did. The first two columns in the table correspond to the baseline specification; columns (3)-(6) replace year fixed effects with parametric time trends—linear in columns (3)-(4) and separate linear in columns (5)-(6); columns (7)-(8) include year, industry and industry × year fixed effects, allowing taxpayers in each industry their own earnings growth trend. The definition of the industry variable is provided in Appendix A. I do not observe industry classification for all taxpayers, owing to which the numbers of observations in the last two columns are lower. Reassuringly, however, even for this restricted sample the results are similar to the baseline results. The control group here (*SE*=0) comprises taxpayers whose self-employment income is subject to the schedule for wage earners. All even-numbered columns drop control group taxpayers for whom the mechanical change in tax rate from period t to $t+1$ is not zero. Standard errors are in parenthesis, which have been clustered at the individual level.

Table A.VII: Heterogeneity in Self-Employment Income Response (Size)

	Dependent variable: Log change in self-employment income from period t to $t+1$					
	(1)	(2)	(3)	(4)	(5)	(6)
to-zero × 2009	0.295	0.267	0.293	0.252	0.294	0.241
	(0.009)	(0.022)	(0.010)	(0.021)	(0.013)	(0.022)
to-zero × 2010	0.160	0.165	0.168	0.142	0.162	0.115
	(0.009)	(0.024)	(0.010)	(0.022)	(0.012)	(0.025)
to-zero × *trait* × 2009	-0.038	0.058	-0.077	0.040	-0.120	0.066
	(0.010)	(0.036)	(0.011)	(0.042)	(0.013)	(0.048)
to-zero × *trait* × 2010	-0.043	0.037	-0.072	-0.015	-0.081	0.052
	(0.010)	(0.048)	(0.011)	(0.053)	(0.013)	(0.059)
to-zero × *trait*	0.012	-0.001	0.030	0.001	0.044	-0.059
	(0.007)	(0.019)	(0.007)	(0.021)	(0.009)	(0.027)
trait × *post*	-0.002	-0.039	0.008	-0.008	0.010	-0.073
	(0.010)	(0.036)	(0.010)	(0.042)	(0.013)	(0.050)
Trait = Size	Above Vs. Below Median		Top Vs. Bottom Quartile		Top Vs. Bottom Decile	
Controls for Base Period Income	No	Yes	No	Yes	No	Yes
Pre-reform mean of the dependent variable	0.048	0.048	0.052	0.052	0.053	0.053
Observations	823,552	823,552	414,183	414,183	165,099	165,099

Notes: This table explores heterogeneity in self-employment income response across taxpayers with different firm size. The size thresholds are defined on the basis of observed yearly firm-size distribution. The odd-numbered columns report the results from equation (17) with no controls for the base period income. The even-numbered columns control for the base-period income in a non-parametric way: *(i)* taxpayers are grouped into bins of PKR 20K on the basis of their base period income, *(ii)* the regression (17) is run separately in each bin, and *(iii)* the aggregate estimates are generated as the weighted average of the bin-level estimates, with the weights provided by the distribution of firm-size in the binned income distribution. I do not observe firm-size for all taxpayers, owing to which sample size here is lower than in Table I. The estimates show that there is hardly any difference in the self-employment income response across taxpayers with different firm size, once taxpayers in the same area of the income distribution are compared.

Table A.VIII: Heterogeneity in Self-Employment Income Response

	Dependent variable: Log change in self-employment income from period t to $t+1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
to-zero $\times$ 2009	0.280	0.274	0.270	0.262	0.282	0.271	0.288	0.260
	(0.004)	(0.010)	(0.004)	(0.011)	(0.003)	(0.010)	(0.004)	(0.015)
to-zero $\times$ 2010	0.141	0.165	0.131	0.152	0.147	0.164	0.153	0.194
	(0.003)	(0.012)	(0.003)	(0.013)	(0.003)	(0.012)	(0.003)	(0.027)
to-zero $\times$ *trait* $\times$ 2009	-0.019	-0.010	0.029	0.011	-0.100	-0.049	-0.065	0.014
	(0.005)	(0.028)	(0.005)	(0.017)	(0.007)	(0.039)	(0.005)	(0.017)
to-zero $\times$ *trait* $\times$ 2010	-0.000	-0.030	0.029	0.009	-0.075	-0.057	-0.062	-0.029
	(0.005)	(0.031)	(0.005)	(0.019)	(0.007)	(0.041)	(0.005)	(0.030)
to-zero $\times$ *trait*	0.021	0.067	0.029	-0.014	0.049	0.048	0.026	-0.002
	(0.004)	(0.022)	(0.004)	(0.010)	(0.004)	(0.031)	(0.004)	(0.013)
trait $\times$ *post*	-0.011	0.007	-0.037	-0.015	0.006	0.032	0.009	0.006
	(0.005)	(0.026)	(0.005)	(0.016)	(0.006)	(0.037)	(0.005)	(0.022)
Trait	Manufacturers		Regular Tax Filers		VAT-Registered		Electronic Filers	
Percent with Trait	24.1		41.8		4.9		11.7	
Controls for Base Period Income	No	Yes	No	Yes	No	Yes	No	Yes
Pre-reform mean of the dependent variable	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049
Observations	1,221,165	1,221,165	1,221,165	1,221,165	1,221,165	1,221,165	1,221,165	1,221,165

Notes: This table explores heterogeneity in self-employment income response across taxpayers with different characteristics. The odd-numbered columns report the results from equation (17) with no controls for the base period income. The even-numbered columns control for the base-period income in a non-parametric way: *(i)* taxpayers are grouped into bins of PKR 20K on the basis of their base period income, *(ii)* the regression (17) is run separately in each bin, and *(iii)* the aggregate estimates are generated as the weighted average of the bin-level estimates, with the weights provided by the distribution of characteristic j in the binned income distribution. The estimates show that there is hardly any difference in the self-employment income response across taxpayers with different characteristics, once taxpayers in the same area of income distribution are compared.

Table A.IX: Heterogeneity in Self-Employment Income Response (Age)

	Dependent variable: Log change in self-employment income from period t to $t+1$					
	(1)	(2)	(3)	(4)	(5)	(6)
to-zero × 2009	0.251	0.257	0.232	0.221	0.205	0.119
	(0.008)	(0.026)	(0.010)	(0.031)	(0.014)	(0.052)
to-zero × 2010	0.105	0.242	0.100	0.195	0.087	0.091
	(0.007)	(0.033)	(0.009)	(0.040)	(0.013)	(0.049)
to-zero × *trait* × 2009	0.003	-0.007	-0.009	-0.006	0.002	0.041
	(0.009)	(0.027)	(0.012)	(0.035)	(0.017)	(0.066)
to-zero × *trait* × 2010	0.011	-0.098	0.006	-0.032	0.013	0.013
	(0.009)	(0.037)	(0.012)	(0.046)	(0.017)	(0.071)
to-zero × *trait*	-0.046	-0.009	-0.043	-0.024	-0.036	0.064
	(0.007)	(0.021)	(0.009)	(0.029)	(0.012)	(0.057)
trait × *post*	-0.019	0.052	-0.016	0.013	-0.028	-0.044
	(0.009)	(0.029)	(0.011)	(0.038)	(0.016)	(0.065)
Trait = Age	Above Vs. Below Median		Top Vs. Bottom Quartile		Top Vs. Bottom Decile	
Controls for Base Period Income	No	Yes	No	Yes	No	Yes
Pre-reform mean of the dependent variable	0.065	0.065	0.064	0.064	0.062	0.062
Observations	252,750	252,750	133,063	133,063	55,004	55,004

Notes: This table explores heterogeneity in the self-employment income response across taxpayers of different ages. The age thresholds are defined on the basis of observed yearly age distribution of self-employed. The odd-numbered columns report the results from equation (17) with no controls for the base period income. The even-numbered columns control for the base-period income in a non-parametric way: *(i)* taxpayers are grouped into bins of PKR 20K on the basis of their base period income, *(ii)* the regression (17) is run separately in each bin, and *(iii)* the aggregate estimates are generated as the weighted average of the bin-level estimates, with the weights provided by the distribution of the taxpayer's age in the binned income distribution. I do not observe ages of all taxpayers, owing to which sample size here is lower than in Table I. The estimates show that there is hardly any difference in the self-employment income response across taxpayers of different ages, once taxpayers in the same area of the income distribution are compared.

Table A.X: Robustness of the Difference-in-differences Estimates of Wage Income Response

	Dependent variable: Log change in wage income from period t to $t+1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: Tax-Driven Response								
to-zero	0.019	0.019	0.019	0.016	0.016	0.015	0.018	0.016
	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
not-to-zero	0.004	0.003	0.003	0.010	0.004	0.003	0.002	0.004
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
B: Placebo								
to-zero	0.009	0.007	0.007	-0.005	0.008	0.006	0.008	-0.000
	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.002)
not-to-zero	0.008	0.007	0.006	0.003	0.006	0.005	0.007	0.003
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Fixed effects:								
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region	No	Yes	No	No	No	Yes	No	No
Tax office	No	No	Yes	No	No	Yes	No	No
Age deciles	No	No	No	Yes	No	No	No	No
Years-registered deciles	No	No	No	No	Yes	Yes	No	No
Sample:								
$z_{it} \in (140\text{K } 700\text{K}]$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Bunchers dropped	No	No	No	No	No	No	Yes	No
$z_{it} \in (0\ 1000\text{K}]$	No	No	No	No	No	No	No	Yes
Pre-reform mean of the dependent variable	0.162	0.162	0.162	0.162	0.162	0.162	0.162	0.166
Observations	240,804	240,798	236,636	85,619	236,644	236,622	227,039	310,716

Notes: This table reports the estimates from equation (16). The first column corresponds to the baseline specification; columns (2)–(6) add additional control variables; column (7) drops taxpayers who *bunch* at the notches from concerns that their base period income might be low owing to the strong, local incentives created by the notches, or because these taxpayers might be special; columns (8) increase the range of the data from $z_{it} \in (140\text{K } 700\text{K}]$ in other specifications to $z_{it} \in (0\ 1000\text{K}]$. The definitions of variables used here are in Appendix A. All regression include log base period income as a control to account for mean-reversion. Standard errors are in parenthesis, which have been clustered at the individual level.