DOES CUTTING THE TAX RATE TO ZERO INDUCE BEHAVIOR DIFFERENT FROM OTHER TAX CUTS? EVIDENCE FROM PAKISTAN

Mazhar Waseem*

Abstract—Using a series of Pakistani tax reforms and administrative records, I document that taxable income responses induced by to-zero tax cuts are orders of magnitude larger than ones induced by similar-sized other cuts. This finding is remarkably robust to alternative specifications and holds for both the self-employed and wage earners. I explore salience, selective enforcement, and discontinuous evasion costs as explanations of the observed behavior. I find that the data favor the last explanation. The difference between the two sets of responses is primarily driven by a large, discrete tax evasion response, which is included in the former but not in the latter behavior. I estimate the difference as a lower bound on tax evasion, showing that at least 70% of the income of low- and middle-income self-employed and 1% of low-income wage earners goes unreported.

I. Introduction

IMPORTANT policy questions such as how high the tax rate can be and how wide the tax base needs to be depend critically on how agents react to tax changes (Feldstein, 1999; Saez, 2004). A rich body of literature leverages changes in the income tax schedule to estimate these reactions (Saez, Slemrod, & Giertz, 2012). The changes exploited in this literature, however, are exclusively of the type where the tax rate moves within the positive region. A common feature of income tax systems around the world is that incomes below a given cutoff are not taxed. Upward revisions of the exemption cutoff create tax reforms where the rate moves from a positive value to zero. A priori, agents may not react to these to-zero reforms the same way they do to others. Tax evasion offers no tangible benefit when the rate is zero. To-zero reforms may be more salient than others, and the authorities may audit zero-rated incomes lightly. If behavior differs substantially across to-zero and not-to-zero reforms, it would have important policy implications. Yet there is little work in the literature that examines the question either theoretically or empirically.

In this paper, I exploit a series of sharp changes in the Pakistani income tax system to study this question. Pakistan has two income tax schedules; one for the self-employed and one for wage earners. The schedules are not indexed to inflation, and bracket boundaries, in particular the exemption cutoff, need to be moved every few years to avoid bracket creep. During the period considered in this paper (2006–2011), the schedule for self-employed was revised once, in 2010, but the exemption cutoff was moved twice, in 2010 and 2011.[1] Similarly, the wage earners' schedule was revised once, in 2008, but the exemption cutoff was moved four times, in 2008, 2009, 2010, and 2011. These movements create plausibly exogenous to-zero and not-to-zero rate changes that are particularly suited to the requirements of this paper because they are similar in size and are applied to a similar area of the income distribution.

I use data from the Federal Board of Revenue (FBR) that comprise the universe of income tax returns filed between 2006 and 2011. Using the data, I present nonparametric evidence establishing that the behavioral responses produced by to-zero rate changes are orders of magnitude larger than ones produced by similar-sized not-to-zero changes. The elasticities underlying the former responses are larger than fifteen, while those underlying the latter are close to zero. I formalize this result using the difference-in-differences framework, comparing the outcomes across taxpayers affected and not affected by the tax changes. Identification requires that reported earnings of the compared groups would have followed a common trend in the absence of tax changes. I confirm this using both visual and regression-based analysis. I also demonstrate that the result is robust to a series of specification checks.

Why do to-zero reforms produce much larger earnings responses than others? I consider three potential explanations. It could be that the costs of evading certain categories of income are small and of others large. Income entailing little evasion cost would be reported at a zero rate but not otherwise. Income entailing large evasion cost would always be reported. Response to a to-zero tax cut would include both categories of income and, hence, would be larger. This evasion-costs-based explanation generates three testable predictions. First, to-zero responses would be substantially larger than not-to-zero responses. Second, the difference between the two would represent tax evasion. And finally, tax evasion would be nontrivial even at very low rates. The first of these predictions is validated by the data. To test the second prediction, I compare the evolution of easy-to-evade line items on the tax return form with hard-to-evade items. The easy-to-evade items respond much more aggressively than others, demonstrating that the to-zero responses in large part comprise changes in tax evasion and not effort. The evidence validates the third prediction too. I show that tax evasion is large (more than 70% of reported earnings) even when the tax rate is very low (just half a percent). The evasion-costs-based explanation thus fits the observed evidence quite well.

I next explore a salience-based explanation. There is growing evidence that agents do not optimize fully to taxes. Tax

Received for publication September 15, 2017. Revision accepted for publication February 20, 2019. Editor: Asim I. Khwaja.

*Waseem: University of Manchester.

I thank the editor and three anonymous referees for helpful comments. This research has greatly benefited from suggestions and comments by 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, the Centre for European Economic Research, and the National Tax Association Annual Conference.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00832.

[1] The Pakistani tax year runs from July to June. A year t in this paper refers to the tax year from July t to June $t + 1$.

The Review of Economics and Statistics, July 2020, 102(3): 426–441

https://doi.org/10.1162/rest_a_00832

schedules are complex, many decision-relevant attributes of taxes are shrouded, and attention is a depletable resource. Together, this implies that agents may not pay full attention to less-salient taxes, underreacting to them (see, e.g., Chetty, Looney, & Kroft, 2009; Finkelstein, 2009; Taubinsky & Rees-Jones, 2017). This salience-based explanation can reconcile the observed behavior, meaning that the to-zero responses reflect true and not-to-zero responses attenuated behavior so that the difference between the two represents optimization errors. The evidence, however, does not support this explanation. The to-zero responses are too large to be taken as true responses to a typical tax change, and the not-to-zero responses are comparable to salience-adjusted, structural responses estimated for the same set of taxpayers using other sources of variation (Kleven & Waseem, 2013).

The final explanation I consider is that the enforcement function may not be neutral across incomes in various brackets. For example, it may treat zero-rated incomes favorably considering that no tax is payable. But this mechanism is completely absent in the Pakistani setting. The Pakistani tax administration audits around 2% to 5% of tax returns annually. These returns are selected at random through a publicly held ballot. The audit and enforcement functions in my empirical setting are therefore independent of income brackets or any other taxpayer trait and cannot explain the observed responses.

Having concluded that the empirical evidence favors the evasion-costs-based explanation, I proceed to show that the large difference between the observed to-zero and not-to-zero responses identifies a lower bound on tax evasion. The intuition for this result is simple. At a zero tax rate, it is optimal for a taxpayer to report her true income, as evasion offers no pecuniary benefit but still entails costs. As the rate increases marginally above zero, evading the component of income that entails trivial evasion cost becomes optimal. Reported income thus jumps as the rate moves to or away from zero. Because this jump represents the component of income that will not be reported at any positive tax rate, it identifies a lower bound on tax evasion. Comparing the to-zero and not-to-zero responses, I estimate this lower bound to be 70% for self-employment income and 1% for wage income, meaning that at least 70% of reported self-employment and 1% of reported wage income are evaded by zero-rated taxpayers in Pakistan. In the most parsimonious formulation of the model, these lower bounds reflect actual evasion rates tightly. But in richer settings, the baseline result—true incomes are reported at the zero tax rate—may not hold and the lower bounds may not be tight. I explore three such settings: (1) the possibility of downward revision of the exemption cutoff, (2) cross-checks in other tax bases, and (3) the threat of future audits. In each case, the evidence suggests that incorporating the richer element of behavior is unlikely to take us too far from the baseline results.

It is important to emphasize that this tight lower-bound interpretation is primarily relevant to the Pakistani setting only. In order to recover the level of evasion by comparing to-zero and not-to-zero responses, it is crucial that both actual and perceived enforcement functions do not change discontinuously at the point the tax rate rises from zero to a positive value. This requirement is satisfied in the Pakistani setting but may not be satisfied in other settings for one or more of the reasons already mentioned.[2] One other factor limiting the generalizability of the result is that while the exemption cutoff in most of the developing countries—similar to Pakistan—is located high up in the income distribution (near the 80th percentile), it is located quite low in rich countries (below the 20th percentile). In both cases (discontinuous enforcement function and low exemption cutoff), the difference between to-zero and not-to-zero responses would continue to recover a lower bound on tax evasion, but this lower bound would not be informative on the actual evasion level in the economy.

This paper contributes to a growing literature that uses quasi-experimental variation created by tax reforms to estimate behavioral responses to taxation (see Saez et al., 2012, for a survey). In particular, it adds to a recent strand of this literature that uses administrative microdata to study tax compliance in low-enforcement-capacity environments, emphasizing the role of information in compliance (see, e.g., Pomeranz, 2015; Best et al., 2015; Naritomi, 2018; Carrillo, Pomeranz, & Singhah, 2017; Waseem, 2018a). Nonincremental, sizable to-zero reforms are frequent in both rich and developing countries, and their policy implications are potentially serious.[3] Yet there is very little work that examines the distinction between to-zero and not-to-zero reforms either theoretically or empirically.[4] This paper fills the gap, documenting how behavior differs substantially depending on whether the taxpayer faces a zero or positive tax rate.

II. Context, Data, and Research Design

This section describes important features of the Pakistani income tax system and the research design I use for the empirical analysis.

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 the self-employed and wage earners. A taxpayer is classified as self-employed (wage earner) if her wage income does not exceed (exceeds) 50% of the taxable

[2] Indeed, there is some evidence from another developing country context that taxpayers are worried about reporting true income even when facing a zero tax rate due to both changes in the audit function and dynamic enforcement considerations (see Tourek, 2019).

[3] I have mentioned the Pakistani case. See Piketty and Qian (2009) for China and India. Since 2010, the U.K. exemption cutoff has moved every year, almost doubling from £6,475 in 2010 to £11,850 in 2018.

[4] Besides this paper, one other work that I am aware of that examines taxpayer behavior at a zero tax rate is Tourek (2019).

income and is then taxed according to the assigned schedule on the entire taxable income. The two schedules, shown in figure A1 in the online appendix, specify average tax rate as a function of taxable income. The Pakistani tax system is quite simple. 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.[5]

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 need revision every few years to avoid bracket creep.[6] During the period considered in this study (2006–2011), the schedule for the self-employed was comprehensively revised once, in 2010, but the exemption cutoff was moved twice, in 2010 and 2011. Similarly, the wage earners' schedule was comprehensively revised once, in 2008, but the exemption cutoff was moved four times, in 2008, 2009, 2010, and 2011.[7] These reforms create plausibly exogenous tax variation, which for at least two reasons is particularly suited to the requirements of this paper. First, the to-zero and not-to-zero rate changes resulting from the reforms, illustrated in figure 1, are almost of the same size and are applied roughly to a similar area of the income distribution.[8] Second, as the main motivation behind these 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.

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. Table A1 assesses compliance with these filing requirements (also see figure A2 for a nonparametric counterpart of this exercise). I regress an indicator that a tax filer in period t also files in period $t+1$ on a dummy indicating if the tax filer experiences a to-zero rate change. The regression is run separately for the self-employed and wage earners, and I also report results from placebo regressions where the rate changes are predated by one year. Overall, around 80% of the self-employed continue to file in the next period, but, more important, this probability does not drop for tax filers whose rate gets reduced to zero. In fact, the placebo exercise shows that such taxpayers are slightly more likely to file a return. This should not be surprising as filing is an easily verifiable, discrete variable, and most tax administrations, including the FBR, use automated processes to identify and penalize nonfiling.

The Pakistani income tax system is based on the principle of self-assessment. Returns filed in a tax year are considered final unless they are selected for audit. Audit therefore is the only mechanism through which compliance can be secured. The FBR, like its counterparts in other countries, have limited resources for audit, which means they can audit only a small fraction of returns filed every year.[9] The Pakistani tax code provides that the selection of returns for audit can only be based on objective criteria. Over the years, the superior courts of the country have narrowed the definition of *objective* considerably. In fact, the FBR has been finding it difficult to defend any parametric selection criterion as objective. To avoid further litigation on the issue, it has adopted the practice of selecting audit cases randomly through a computer ballot. These ballots are carried out publicly, and the results are displayed on the FBR website. The audit function faced by taxpayers in my sample is therefore quite simple: every tax filer faces a small (around 2% to 5%), exogenous probability of audit. The probability does not change discontinuously at the exemption cutoff, nor does it increase or decrease on declaring income in the zero-rated region.

B. Data

I use administrative data from the FBR that include income tax returns filed by the self-employed and wage earners from 2006 to 2011 and a set of taxpayer characteristics. The tax-return data set 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 data set contains information captured at the time of registration, such as the date of registration, gender, and location of a taxpayer. Appendix A.1 provides a detailed description of the variables used in the empirical analysis.

Table A2 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 is restricted to taxpayers for whom $\log \frac{z_{it+1}}{z_{it}}$ is defined. Second, because the main focus of the paper is earnings responses to the movement of the exemption cutoff, I do not include

[5] 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.

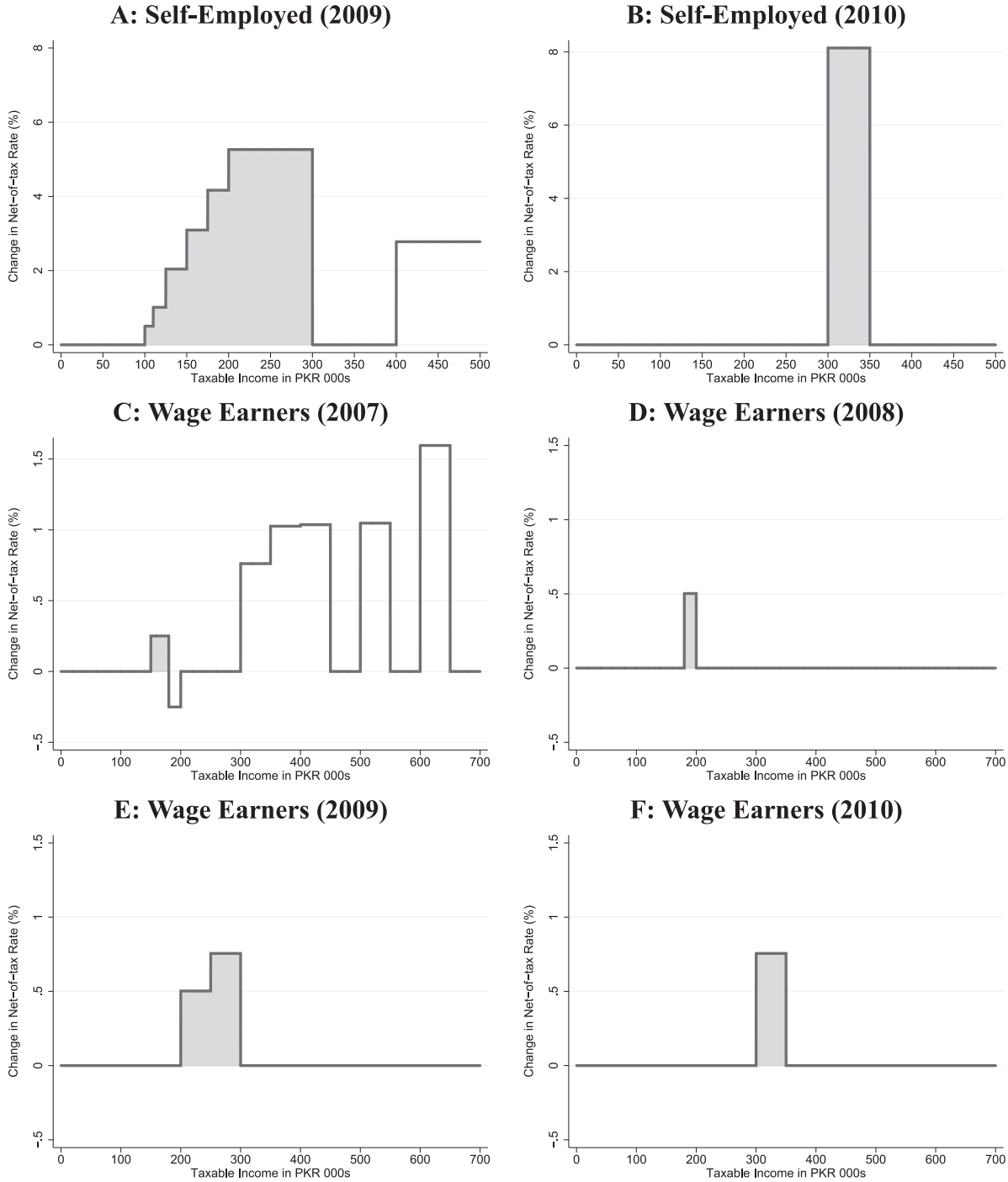
[6] Inflation is generally high in Pakistan and hovered around 10% during the periods considered in this study.

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

[8] For example, the 2008 not-to-zero change and 2011 to-zero change for wage earners are exactly similar other than that the latter reduces the rate to zero whereas the former does not.

[9] For example, only 10,271 (2.1%) of the 485,420 nonsalaried returns were audited in the tax year 2010.

FIGURE 1.—TAX VARIATION



The figure displays tax variation created by the Pakistani reforms from 2006 to 2011. All curves plot percentage change in the 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 2007–2008 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 positive. The to-zero changes, which reduce the rate to zero, are shaded in gray.

taxpayers who have base period earnings (z_{it}) too far away from the exemption cutoff.[10] For the self-employed, the analysis sample, accordingly, includes taxpayers with $z_{it} \in$ (PKR 80k, PKR 500k], which constitutes around 94% of the population. The wage income distribution is more dispersed, and the analysis sample therefore includes all taxpayers with $z_{it} \in$ (PKR 140k, PKR 700k], 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 the log change in earnings ($\log \frac{z_{it+1}}{z_{it}}$) is less than the first percentile or is in excess of the 99th percentile of the

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

corresponding pooled distribution. Such winsorizing is common in the literature to deal with the extreme outliers (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 for male taxpayers in 2006 to 2009,[11] (2) partners in partnership firms, as their earnings are subject to a different tax regime (Waseem, 2018b), and (3) taxpayers who switch from self-employed to wage earner, and vice versa, from concerns that such switching may be endogenous to tax changes.[12] 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.

C. Research Design

I use a simple difference-in-differences research design to estimate earnings responses generated by the to-zero and not-to-zero rate changes. The research design, based on the workhorse empirical model in the tax responsiveness literature (Saez et al., 2012), leverages the fact that taxpayers in different brackets of the two tax schedules experience differential rate changes over time. It is particularly suited to the Pakistani setting as taxpayers in a few brackets undergo no tax change at all and can therefore serve as a clean control group. I estimate the following model,

$$\Delta \log z_{it}^{k} = \alpha + \mathbf{treat}_i \boldsymbol{\beta} + \mathbf{year}_t \boldsymbol{\gamma} + \mathbf{treat}_i \times \mathbf{post}_t \boldsymbol{\delta} + \boldsymbol{X}_{it} \boldsymbol{\mu} + u_{it}, \quad (1)$$

where $\Delta \log z_{it}^{k}$ is the log change in income of type k from period t to $t+1$ for taxpayer i, $\mathbf{treat}_i$ is a vector of two dummies [$to\text{-}zero_i$ $not\text{-}to\text{-}zero_i$] which turn on whenever the corresponding tax change is experienced, $\mathbf{year}_t$ is a vector of year fixed effects, $\mathbf{post}_t$ are dummies indicating the year in which the particular change takes place, and $\boldsymbol{X}_{it}$ are a set of controls. Given that the identification here comes from the comparison of taxpayers in different areas of the income distribution, the major threat to identification is mean reversion. I take three steps to rule out this and related concerns. First, I provide nonparametric evidence showing that the earnings growth rate ($\Delta \log z_{it}^{k}$) remains remarkably uniform throughout the income distribution during the periods of no tax change. Second, I also estimate augmented specifications corresponding to equation (1), where controls for mean reversion suggested in the literature (see Saez et al., 2012)—the log base period income and a ten-piece spline of log base period income—are included in the model. The results with and without these controls are very similar. Third, I conduct placebo analysis pretending that each reform took place one year prior to its actual implementation. If areas of the income distribution affected by the reforms experience significantly different earnings growth for nontax reasons, it would show up in the placebo regressions.

III. Does Behavior Differ across To-Zero and Other Tax Cuts?

In this section, I estimate behavioral responses produced by the Pakistani to-zero and not-to-zero reforms to see if they differ substantially from each other. I begin by presenting nonparametric evidence. The results are then formalized through the regression-based framework. Finally, I show that the results are robust to a series of specification checks.

A. Self-Employment Income

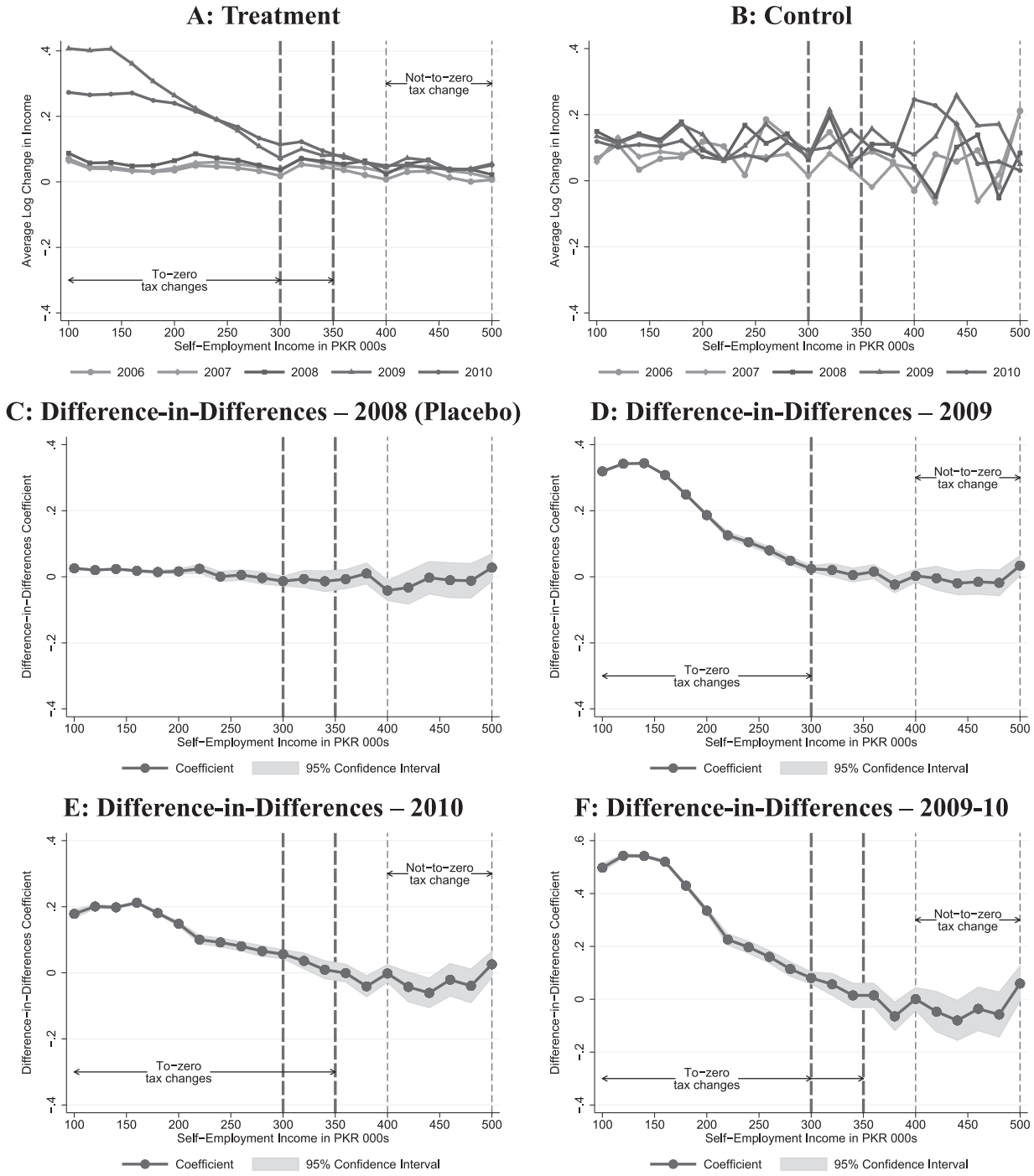
Nonparametric evidence. Figure 2A plots the evolution of self-employment income from 2006 to 2011 for taxpayers classified as self-employed by the tax code. To construct the diagram, I group taxpayers into PKR 20,000 bins on the basis of their base period income (z_{it}^{S}) and plot the mean log change in income from year t to $t+1$, $\mathbb{E}\left[\log \frac{z_{it+1}^{S}}{z_{it}^{S}} | z_{it}^{S} \in b\right]$, in each bin b. These plots show how self-employment income growth in various areas of the income distribution responds to the 2010–2011 rate changes. Two features of the evidence are noteworthy. First, the growth rate is remarkably stable over time and homogeneous across the income distribution in periods of no tax change. Second, the responses produced by the two types of rate changes are strikingly different from each other: while reported 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 2010–2011 caused by the reduction of the rate to zero? That it is concentrated precisely in the region between the old and new cutoffs, very strong at the bottom and then tapers off monotonically as we move toward the new cutoffs, and indistinguishable from the prereform level just above the new cutoffs strongly suggests that it was. To further reinforce the causal link, figure 2B looks at the evolution of the self-employment income of taxpayers classified as wage earners by the tax code. These taxpayers do not experience the 2010–2011 rate changes. Therefore, to the extent that

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

[12] Table A3 assesses if the switching probability differs across years. Switching between the two bases is rare, and there are no meaningful differences in switching across years. I also investigate if switchers are concentrated disproportionately around the exemption cutoff. The regression of an indicator variable that self-employed worker in period t becomes a wage earner in period $t+1$ on a dummy indicating that the taxpayer is located within PKR 50,000 of the exemption threshold (PKR 300,000) returns a coefficient of 0.0064 with a standard error of 0.0012. This coefficient is in fact smaller than the one I obtain from the placebo regression, which is run on the prereform periods only (when PKR 300,000 is not the exemption cutoff).

FIGURE 2.—SELF-EMPLOYMENT INCOME RESPONSE



The figure compares the self-employment income response to the two types of rate changes. To construct the top two 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. The bottom four panels display a difference-in-differences version of the top two plots. Panel F illustrates the sum of 2009 and 2010 coefficients. The standard errors have been clustered at the individual level.

self-employment income is subject to common macroshocks, any nontax factors affecting it in 2010–2011 could be detected here. However, in sharp contrast to panel A, all curves in panel B are tight to each other, establishing that the 2010–2011 responses are indeed driven by the tax changes. Panels C to F, which are the difference-in-differences versions of panels A and B, formalize this conclusion by demonstrating that there are no significant preexisting differences across the two groups (panel C); the reduction of the rate to zero causes a sharp jump in income reported by the treated group (panels D and E); and the additional reported income is as large as 70% of the base period income at the bottom (panel F).

TABLE 1.—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)
To-zero × 2009	0.273	0.287	0.258	0.236	0.271	0.263	0.229
	(0.003)	(0.001)	(0.006)	(0.006)	(0.003)	(0.003)	(0.006)
To-zero × 2010	0.126	0.136	0.109	0.091	0.127	0.124	0.089
	(0.004)	(0.001)	(0.006)	(0.006)	(0.004)	(0.004)	(0.006)
Not-to-zero × 2009	−0.014	−0.007	−0.006	−0.003	−0.017	−0.015	−0.003
	(0.008)	(0.007)	(0.012)	(0.012)	(0.008)	(0.008)	(0.012)
Fixed effects							
Year	Yes	No	Yes	Yes	Yes	Yes	Yes
Industry	No	No	Yes	Yes	No	No	Yes
Industry × year	No	No	No	Yes	No	No	Yes
Region	No	No	No	No	Yes	Yes	Yes
Region × year	No	No	No	No	No	Yes	Yes
Time trend Linear	No	Yes	No	No	No	No	No
Prereform mean of the dependent variable	0.050	0.050	0.066	0.066	0.050	0.050	0.066
Observations	526,150	526,150	93,525	93,525	525,260	525,260	93,319

The table reports the estimates from equation (1). I restrict the sample to a balanced panel of taxpayers who file in all six years 2006 to 2011. The first columns correspond to the baseline specification; the second column replaces the year fixed effects with a linear time trend; and the rest of the columns add year, industry, industry × year, region, and region × year fixed effects. The details of the industry and region classifications are provided in appendix A1. I do not observe industry classification for all taxpayers, owing to which the numbers of observations in the last two columns are lower. The control group here comprises self-employed who do not experience any rate change. Standard errors are in parentheses, which have been clustered at the individual level.

Figure A3 shows that the results are indistinguishable if I restrict the sample to a balanced panel of taxpayers who file in all six years, 2006 to 2011.

Regression-based estimates. Table 1 reports the results from equation (1). The first column contains the estimates from the baseline specification, while the rest of the columns add control variables or experiment the replacement of year fixed effects with a parametric time trend. To address any concerns from a change in the composition of the sample, I restrict it to a balanced panel of taxpayers who file every year included in the estimation. Table A4 drops this restriction, replicating the results for the complete panel. The two sets of results are indistinguishable. Two 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 remarkably robust to alternative specifications. Column 1 of the table, for example, shows that the reduction of the rate to zero causes a 27 log-point additional income growth in the treatment group in the first year after the reform. This is around five times larger than the prereform average of 5 log points per year. Considering that the average net-of-tax-rate change behind the response is only 1.7 log points, the estimate translates into an enormous elasticity of greater than 15. Second, the similar-sized, not-to-zero tax cut generates no response at all. The corresponding estimate is always of opposite sign, small, and statistically insignificant in all but one specification.

Robustness. Given the difference-in-differences research design, the key identification requirement in this setup is of parallel trends: reported earnings of the compared groups (tax brackets that experience and do not experience the tax rate changes) must have followed the same path in the absence of the rate changes. The nonparametric evidence in figure 2 illustrates that this is indeed true. The earnings growth rate remains remarkably stable and uniform over time and across the income distribution during the periods of no tax change. Figure A4 provides an aggregate counterpart to this result. It shows that the average earnings growth was very similar in the prereform years for both groups but surged sharply in the treated group exactly at the time of the reform. In fact, the preexisting earnings trends were so flat and stable that the time-series estimates, reported in table A5, are indistinguishable from the corresponding difference-in-differences estimates. Tables 1, A4, and A5 further confirm that the results are insensitive to (a) including additional control variables; (b) replacing year fixed effects with the parametric time trend; (c) adding a full set of year, industry, region, industry × year, and region × year fixed effects; and (d) keeping the composition of the estimation sample fixed.

Table A6 conducts an additional set of robustness checks. 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. Column 3 drops taxpayers around the income-composition notch, where the classification of a taxpayer switches from self-employed to wage earner, and vice versa. Columns 4 and 5 increase the range of the data from $z_{it} \in$ (PKR 80k, PKR 500k] in the baseline results to $z_{it} \in$ (0, PKR 500k] in column 4 and $z_{it} > 0$ in column 5. Columns 6 to 9 add more control variables into specification (1). Reassuringly, the results from all these alternative specifications are very similar to the baseline results.

Another common concern in the tax responsiveness studies is mean reversion. I take two steps to alleviate this concern. First, I test if mean reversion is a significant problem in this setting by estimating equation (1) on the prereform periods only. Table A7 shows the results. The coefficients on the two interaction dummies are extremely small and statistically indistinguishable from zero in all but one specification. This demonstrates that the relative difference in reported income

from one year to the next does not change significantly across the treated and untreated groups for any nontax reason, including mean reversion. I complement this test with another strategy (see table A8). I reestimate equation (1) after including the standard controls for mean reversion: log of base period income and a ten-piece spline of log base period income (Saez et al., 2012). I obtain very similar results with and without these controls, reinforcing the conclusion that mean reversion is not a significant concern in this setting.[13]

Finally, I rule out one alternative explanation of the observed behavior. Suppose that agents do not like to pay taxes (or there are some fixed costs of actually making the payments), but they also do not like to cheat more than they have to. In this setting, agents would evade only up to the point of zero tax liability, reporting income just below the exemption cutoff. They would simply move to the new cutoff after a reform that increases the cutoff. This alternative model generates two testable predictions: bunching would be stronger at the exemption cutoff than a similar other notch, and it would shift to the new cutoff after the reform. Figure A5 tests the latter prediction. It compares the earnings growth rate needed to hit the new exemption cutoff with the actual earnings growth rate observed in the data. For example, earnings of a taxpayer bunching at the baseline exemption cutoff of PKR 100,000 have to increase by 200% if it, in accordance with this explanation, simply moves to the new exemption cutoff of PKR 300,000 after the reform. The data clearly reject this alternative explanation. The earnings growth rate around the old cutoff is roughly one-third of the rate needed to hit the new cutoff. In fact, the two curves do not coincide in any area of the income distribution. The first prediction of this alternative model is also rejected by the data: the bunching at the exemption cutoff is not significantly larger than the one at similar other notches.[14]

B. Wage Income

Nonparametric evidence. Figures 3A and 3B show the evolution of wage income from 2006 to 2011 for taxpayers classified as wage earners by the tax code. The diagram is constructed analogous to figures 2A and 2B 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 20,000. The comparison of figures 3 and 2 shows that the growth of wage income, in distinction to that of self-employment income, is not homogeneous across years. Because of this, it is hard to differentiate between the tax-induced behavior and yearly shocks in the simple plot of raw data.

I follow a simple strategy to obtain first-pass evidence on the tax-induced behavior, regressing the log change 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 20,000 bins, and the estimated coefficients, and 95% confidence intervals are plotted in panels C to F. Clearly, once a common year effect is partialed out, the residual income growth is homogeneous across years and over the income distribution. It, however, spikes in the areas of the distribution where the rate was brought to zero. Though this spike is not as prominent as that of self-employment income, the overall pattern is consistent with the earlier result that to-zero and not-to-zero reforms elicit substantially different behavior. The next section formalizes this analysis by presenting the regression-based estimates.

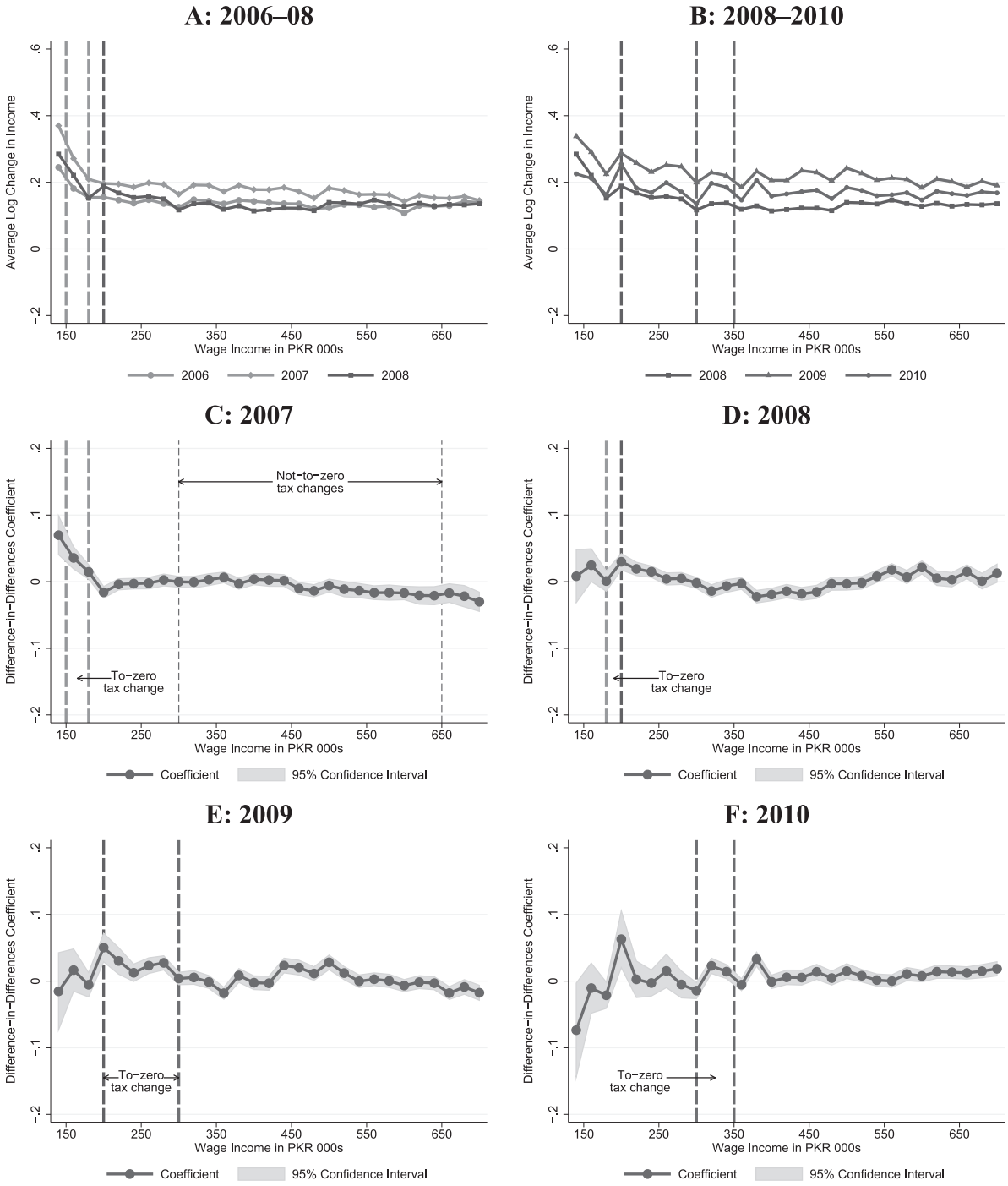
Regression-based estimates. Table 2 reports the results from equation (1). I begin with the baseline specification in column 1 and then successively add more control variables, permuting among the combinations of controls for mean reversion—log base period income and a ten-piece spline of log base period 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. Table A9 runs additional robustness checks.

The main findings are the following. First, the to-zero coefficient is always economically meaningful, statistically significant, and considerably stable across specifications. Given that the rate changes underlying the response are extremely small (always less than 1%; see figure 1C–F), the estimate translates into a huge elasticity of more than 3. This elasticity is orders of magnitude larger than the one, 0.04, estimated by Kleven and Waseem (2013) for wage earners in Pakistan. Second, despite the fact that the not-to-zero changes are on average twice the size of the to-zero changes, the earnings responses they generated are statistically and economically insignificant. Third, the placebo coefficient corresponding to the to-zero estimate is always trivial and statistically insignificant in most of the specifications. This shows that mean reversion is not much of a concern in this setting and that the base period income controls are able to account for it adequately.

[13] In addition to these, a working paper version of this paper (Waseem, 2019) carries out further robustness tests. These include, inter alia, reporting estimates from two variants of equation (1), where I use wage earners with positive self-employment income as the control group. These alternative research designs compare the self-employment income of taxpayers classified as self-employed by the tax code with the self-employment income of taxpayers classified as wage earners by the tax code, who do not undergo the tax rate changes in 2010–2011. These research designs allow additional set of robustness checks, including built-in tests for the parallel trends assumption. The results from these double- and triple-difference specifications are strictly consistent with those in this version of the paper.

[14] In a working paper version, Kleven and Waseem (2013) find that the elasticity implied by the bunching at the baseline exemption cutoff is 0.077. In comparison, the elasticity implied by the bunching at the next three notches is 0.097, 0.083, and 0.091 (see table 1 of the 2012 version of the paper). The exemption cutoff during these years was at PKR 100,000 and the next three notches were at PKR 10,000, 125,000, and 150,000. The notches were also of a similar size, involving a jump in average tax rate of 0.5 (for the first two) and 1 (for the others) percentage points.

FIGURE 3.—WAGE INCOME RESPONSE



The figure shows wage income growth from 2006 to 2011. For the top two panels, taxpayers are grouped into bins of PKR 20,000 on the basis of their base period income. Then the 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. For the bottom four panels, I regress the 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. The figure plots the estimated coefficients and 95% confidence intervals on the four yearly dummies from these regression. The standard errors have been clustered at the individual level.

IV. Why Does Behavior Differ across to-Zero and Other Tax Cuts?

The evidence convincingly shows that reported income responses produced by to-zero tax cuts differ substantially from those produced by not-to-zero tax cuts. In this section, I explore three potential explanations of the finding.

A. *Reverse-L-Shaped Evasion Costs*

The standard way to think about evasion costs is that they are expected fine and penalty payments that would be recovered in case the evasion is detected by the government (Allingham & Sandmo, 1972; Slemrod, 2001). Recent empirical evidence shows that the probability that evasion gets

TABLE 2.—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	0.016	0.016	0.013	0.014	0.014	0.011	0.014	0.014	0.011
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Not-to-zero	0.000	0.000	0.002	0.001	0.002	0.003	0.001	0.002	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
B. Placebo									
To-zero	0.006	0.004	0.003	0.001	−0.000	−0.002	0.001	−0.000	−0.002
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Not-to-zero	0.005	0.006	0.006	0.010	0.010	0.011	0.010	0.010	0.011
	(0.002)	(0.002)	(0.002)	(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
Prereform mean of the dependent variable	0.162	0.162	0.162	0.162	0.162	0.162	0.162	0.162	0.162
Observations	240,649	240,643	240,643	240,649	240,643	240,643	240,649	240,643	240,643

This table reports the estimates from equation (1). 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 A1. 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 parentheses, which have been clustered at the individual level.

detected is quite high if reported income is covered by the third-party information the government obtains from sources such as employers and financial institutions and quite low otherwise (see Slemrod, 2019, for a recent survey of the evidence).[15] In appendix A2, I propose a simple model that incorporates this government's information problem into the standard model of tax reporting under imperfect enforcement (Chetty, 2009). The model assumes that $\underline{e}$ units of income of the agent are not covered by any third-party information, while the rest are covered. Note that this assumption is without any loss of generality as $\underline{e}$ potentially varies from 0 to the maximum. The assumption means that the agent faces a reverse-L-shaped evasion costs function, as shown in figure 4A. The costs of evasion are low at the bottom because earning or consumption of income up to $\underline{e}$ leaves no verifiable information trails, and therefore its evasion entails little detection probability. The costs turn sharply afterward once the third-party-reported units of income begin. An optimizing agent facing such evasion costs would report true income when the tax rate is zero but would evade $\underline{e}$ as the rate increases marginally above zero, producing a discontinuous earnings supply function of the form shown in figure 4B.

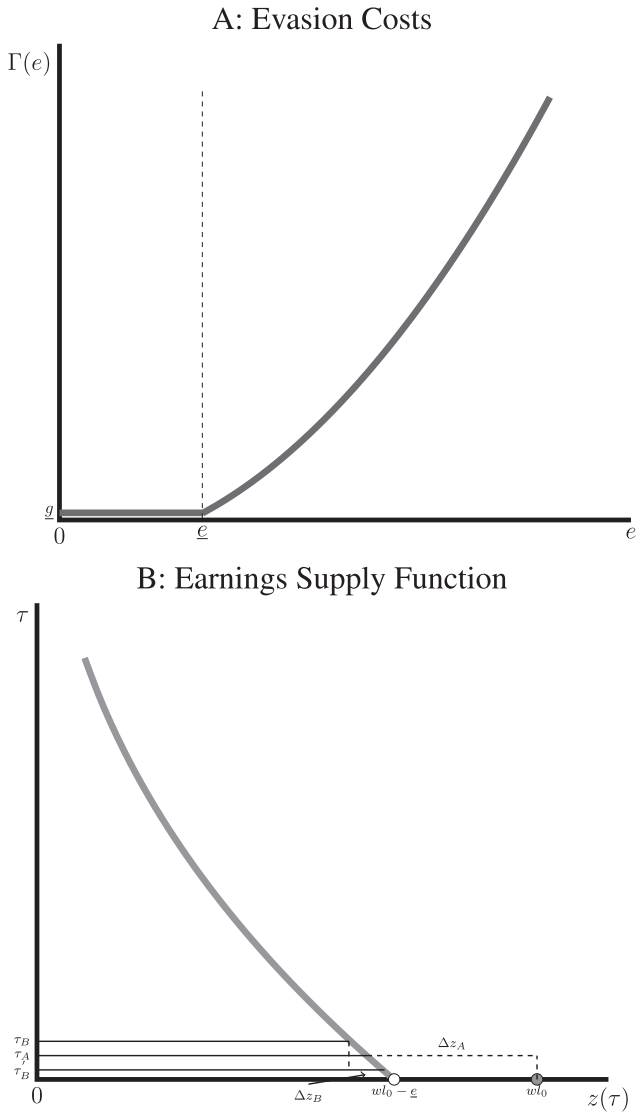
The discontinuity in the earnings supply function can explain the large, substantive difference observed between the two types of responses. The intuition for this result is provided in figure 4B, which considers the effects of two equal-sized rate cuts on income reported by the agent. The response triggered by the to-zero cut $\Delta z_A(\tau_A \to 0)$ is considerably larger, as it consists of both the discrete change in tax evasion (movement along the horizontal axis) and the continuous change in reported income (movement along the supply curve). By contrast, the response induced by the not-to-zero tax cut $\Delta z_B(\tau_B \to \tau_B')$ is smaller, as it consists of the latter component only. The model thus fits the observed pattern of response quite well. To probe this point further, I take the other two predictions of the model to the data.

If the difference between the two sets of response reflects evasion costs, then it must be that the to-zero response largely arises from a changes in tax evasion and not effort. Figure A7 tests this prediction of the model. I look at how individual line items on the tax return form react to the two types of rate cuts. The idea behind the exercise is to see which factor—adjustments in tax evasion or effort—drives the larger response to the to-zero tax cuts. 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 more easily than others. My focus here is to identify any differential response between the easy- and hard-to-misreport items. Each panel of the figure plots the 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 2A, the analysis should be seen as the decomposition of the response depicted there. Figures A8 and A9 formalize this analysis, showing the difference-in-differences version of these plots. Clearly, the line items do not respond uniformly: annual sales and costs respond aggressively, profit-and-loss expenses (in part third-party-reported and therefore harder to misreport) respond moderately,[16] and imports do not respond at all. Of all the items, imports is perhaps the hardest to misreport because

[15] Throughout the paper, I maintain the assumption that the third-party reports the government receives are complete and not themselves distorted by evasion. Otherwise, they would not deter tax evasion as effectively (Brockmeyer & Hernandez, 2017).

[16] 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 overreported, it is difficult to do so considering that these can potentially be verified at the time of audit.

FIGURE 4.—CONCEPTUAL FRAMEWORK



Panel A illustrates the evasion cost function (7) in the online supplement. The agent can evade $\underline{e}$ units of income on paying a small, fixed cost of $\underline{g}$. The costs turn sharply at the cutoff $\underline{e}$ from where the third-party reported units of income begin. Panel B displays the earnings supply function (12) in the online supplement, illustrating how an optimizing taxpayer facing such evasion costs 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 $\Delta z_A(\tau_A \to 0)$ and a not-to-zero reform $\Delta z_B(\tau_B \to \tau'_B)$ identifies $\underline{e}$. 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).

such misreporting can easily be detected through the customs and excise records. Its nonresponsiveness therefore provides the cleanest evidence that the large jump in reported earnings is driven by a drop in tax evasion. Panels E to F of figure A7 strengthen this conclusion. A surge in real activity triggered by an 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 rise sharply.[17] Thus, overall behavior of line items is consistent with a tax evasion explanation of the observed behavior. 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 others.

The third prediction of the model is that tax evasion would be high even at a very low tax rate. This prediction simply reflects that the cost of evading income up to $\underline{e}$ is quite low, and therefore not reporting this component of income becomes optimal even at a very low tax rate (this can be seen from figure 4). The evidence presented above is consistent with this prediction of the model as well. Specifically, figure 2A shows that the reported earnings of taxpayers with baseline income in the range (PKR 100k, PKR 110k] on average rise by around 70% as their tax rate reduces from 0.5% to 0%. This demonstrates that consistent with the structure in the model, a large component of tax evasion is fixed in nature: around 70% of reported income is evaded even when the tax rate is as low as 0.5%.

B. Salience

There is growing evidence in the literature that agents do not optimize fully to taxes. Tax schedules are complex, many decision-relevant attributes of taxes are shrouded, and attention is a depletable resource. Together, this implies that agents may not pay full attention to less salient taxes, underreacting to them (Chetty et al., 2009; Finkelstein, 2009; Taubinsky & Rees-Jones, 2018). Salience can also explain the behavior I have documented. It would mean that the to-zero tax cuts are fully salient and hence elicit true behavior, while the not-to-zero cuts are less salient and hence elicit attenuated behavior. Representing true earnings response by $\hat{\Delta}z$ and observed response by Δz, this statement can be translated into two testable conditions:

$$\hat{\Delta}z_A(\tau_A \to 0) \approx \hat{\Delta}z_B(\tau_B \to \tau'_B),$$
$$\Delta z_B(\tau_B \to \tau'_B) = \theta.\ \hat{\Delta}z_B(\tau_B \to \tau'_B). \quad (2)$$

True responses of both types are nearly equal,[18] but the observed not-to-zero response is attenuated by a factor $\theta \in [0, 1]$. In the extreme case, agents completely ignore a not-to-zero rate change $\theta = 0$ so that the difference between the two captures optimization error only.

Figure A6 assesses the first of these conditions. It replicates figure 2A but plots the elasticity of taxable income on the

[17] The Pakistani tax cuts of 2010 were announced on June 6, 2010, but took effect from the beginning of the new financial year on July 1, 2010. This gave taxpayers a window of around three weeks to plan for the next year. Forward-looking taxpayers would have beefed up inventories had they anticipated a large increase in output in the coming year. But this is not what we observe. The start-of-year inventories in 2010 are exactly similar to those in 2009 (panel E shows no growth in 2009). If start-of-year inventories do not increase and production increases enormously during the year, the end-of-year inventories must go down as firms have limited resources to invest. But again, this is not what we observe. The end-of-year inventories in fact go up significantly (panel F shows significant growth in 2009).

[18] Note that I assume throughout this section that both types of rate changes are roughly equal in magnitude and are applied to very similar taxpayers.

vertical axis in place of the earnings response. The elasticity implied by the to-zero response is implausibly large (above 60 at the bottom). It is in fact orders of magnitude larger than the corresponding estimate reported in the existing literature for both developed and developing countries (see Chetty, 2012, for the meta-analysis and Kleven & Waseem, 2013, for Pakistan). It therefore cannot represent the true responsiveness to a typical rate change. To assess the second condition, I compare the observed not-to-zero response with the structural elasticity estimated in Kleven and Waseem (2013). The structural elasticity reflects true responsiveness—free of all frictions including salience—of the same set of Pakistani taxpayers. For taxpayers with income in the range PKR 400,000 to 500,000, it is estimated to be 0.06, a response not different from the one observed here. Both conditions in equation (2) therefore lack empirical support, suggesting that salience is not the first-order mechanism driving the large difference between the two behaviors. I establish this further in section V of the paper, where I show that the difference remains unchanged even if the not-to-zero response is corrected of the salience bias by assuming an extreme value of the attenuation factor θ, such as 0.1.

C. *Enforcement Environment*

If the audit function has a discontinuity at the exemption cutoff, it may make an agent's decision of how much to report contingent on the tax rate (0 or positive). We have seen in section IIA that this is not the case in Pakistan. The audit probability faced by a tax filer in the country is independent of the tax rate, income, or any other characteristic. But even more generally, a discontinuity in the audit function at the exemption cutoff is unlikely in the Pakistani setting. Figure 2 illustrates that taxpayers just below the baseline exemption cutoff of PKR 100,000 respond to the to-zero tax cut exactly similar to ones just above the exemption cutoff. The first two bins of the figure contain taxpayers in the income range (PKR 80k, PKR 100k) and (PKR 100k, PKR 120k), respectively, and the average earnings response is virtually the same in both bins. To the extent that these responses capture tax evasion, the evidence thus shows that the evaded amount does not change discontinuously at the exemption cutoff. Being so, auditing taxpayer below the exemption cutoff less (or more) aggressively would not have been optimal.[19]

V. Lower Bound on Tax Evasion

Figure 4B shows that the difference in responses induced by to-zero and not-to-zero tax cuts identifies $\underline{e}$. (This result is formally shown in appendix A2; see equation [15] in the appendix.) Because $\underline{e}$ represents the component of income earning or consumption of which leaves no verifiable information trail and evasion of which is therefore optimal even at a very low tax rate, it represents a lower bound on tax evasion. In this section, I use formula (15) in the online appendix to estimate this lower bound for the zero-rated Pakistani taxpayers. Table 3 shows the results for the self-employed. 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 terms in the formula: 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.

TABLE 3.—EVASION RATE OF SELF-EMPLOYMENT INCOME (IN PKR)

Income (1)	Earnings Response ($\tau \to 0$) (2)	Earnings Response ($\tau \not\to 0$) (3)	Difference (4)	Evasion Rate (%) (5)
80–100k	71,754	333	71,421	74.2
	(1,251)	(188)	(1,265)	(1.3)
100–150k	92,280	603	91,676	71.9
	(784)	(341)	(855)	(0.7)
150–200k	92,440	573	91,867	52.2
	(938)	(209)	(961)	(0.5)
200–250k	70,537	763	69,774	30.7
	(1,116)	(278)	(1,150)	(0.5)
250–300k	28,868	692	28,176	10.0
	(1,172)	(252)	(1,198)	(0.4)
300–350k	20,124	419	19,704	6.0
	(918)	(153)	(931)	(0.3)

The table presents the estimates of the rates of evasion of self-employment income from equation (15) in the online supplement. 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 2010–2011 shown in table 1. The estimates in column 3 are computed using elasticities estimated for the same group of taxpayers in Kleven and Waseem (2013). The details on how the estimates in columns 2 and 3 have been computed are in section V. The difference between the two set of estimates represents average $\underline{e}$ for taxpayers in the segment. I obtain the percentage evasion rate implied by the average $\underline{e}$ by dividing it with the average income in the segment (see column 5). The standard errors are in parentheses.

To compute the estimates in column 2, I follow three steps. I first replace the two *to-zero* × *post* dummies in equation (1) with twelve *segment* × *post* dummies to estimate the self-employment income response separately in each segment in 2010 and 2011. I next 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 the average income in the segment to convert it into rupees. The estimates in column 3 are computed using the taxable income elasticities reported in Kleven and Waseem (2013).[20] Using these elasticities instead of the not-to-zero estimates in table 1 is preferable for two reasons. First, these elasticities are estimated from the

[19] In an optimal plan, a revenue-maximizing tax authority would allocate audit resources toward cases where the enforcement action is likely to be most productive. This means that among similar taxpayers, ones evading the most would be targeted. This rule is unlikely to generate discontinuous enforcement given that there is no difference in the amount evaded on both sides of the cutoff.

[20] In computing the estimates for column 3, I keep the magnitude of the tax rate change the same as for column 2. More specifically, 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 taxpayers experience in the segment because of the reform.

same area of the income distribution where the 2010–2011 to-zero rate changes were applied to, making the comparison between the two more like-for-like.[21] Second, Kleven and Waseem (2013) are able to account for all frictions including inattention nonparametrically. Using the structural elasticities estimated by them mitigates any salience-related concerns in my lower bounds. More specifically, Kleven and Waseem (2013) adjust their estimates of observed elasticities to compute structural elasticities that are free from salience and other related frictions. This adjustment is based on the attenuation factor they estimate from the variation created by the tax notches. The attenuation factor estimated by them for the zero-rated self-employed ranges between 0.3 and 0.5. The estimates in column 3 are therefore already inflated by a factor of 2 or 3 to alleviate any frictions, including inattention. To establish that salience plays little role in my setting, I replicate table 3 assuming an extreme value of 0.1 for the attenuation factor θ. The results remain virtually unaffected (compare tables 3 and A10). This should not be surprising given that the large difference between the to-zero and not-to-zero responses documented here reflects not that the latter responses are too small but rather that the former responses are too large.

Two features of the results need emphasizing. First, tax evasion is large even at very low rates, such as 0.5% at the bottom. This behavior, as I note above, is consistent with the predictions 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 150,000 and then declines monotonically. This pattern of response is consistent with an extended version of the model where the postreform tax schedule—similar to the Pakistani setting—takes a nonlinear form, containing a notch at the earnings level z_N (see appendix A2.3 for details). With the nonlinear tax schedule, earnings responses of taxpayers close to the new cutoff (notch) are constrained, as reporting true income would take them into the positive-rate region. Formally, it is not optimal for a taxpayer experiencing the reduction of the rate from $\tau > \underline{\tau}$ to $\tau = 0$ to report true income as long as $z(\tau) + \underline{e} \geq z_N$, where z_N is the new cutoff. There thus exists an interval $[z_N - \underline{e},\ z_N]$ below the new cutoff where evasion does not approach zero even when the rate drops to zero. This interval is larger if taxpayers have dynamic considerations so that they try 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 to be influenced by it. On the basis of this consideration, I conclude that the lower bound on the evasion rate of zero-rated self-employed, as implied by the first four rows of the table, is around 70%.

Table 2 shows that wage income also behaves according to the predictions of the discontinuous evasion-costs-based model. Of course, its response is much smaller than that of the self-employment income because it is third-party-reported. But to the extent that formula (15) in the online appendix holds for wage income as well, the difference between the to-zero and not-to-zero responses identifies the lower bound on the evasion of wage income. The extent of such evasion, however, is small at around 1% of the reported income.

A. *Applicability*

The Pakistani to-zero reforms bite deep into the self-employment income distribution. Figure A10 shows this formally by superimposing the baseline CDF of self-employment income on figure 2F. The exercise shows that the minimum evasion rate I report in section V applies to more than 50% of the population of self-employed tax filers (panel A). Note, however, that although broadly applicable, this estimate is essentially local in nature. To claim global applicability, I need to assume that the evasion technology available to top-income taxpayers is the same as the one available to low- and middle-income taxpayers. In other words, top-income taxpayers face the same evasion costs as do low- and middle-income taxpayers. It is, however, a strong assumption and the one I cannot test,[22] as the to-zero reforms do not extend further to the right of the distribution. The estimates in table 3 therefore apply to both low- and middle-income taxpayers but not to top-income taxpayers. The evasion rate of wage income, however, applies to the bottom 20% of taxpayers only (see panel B).

B. *Heterogeneity*

To explore heterogeneity in the self-employed evasion rate, I estimate the following triple-difference version of equation (1),

$$\begin{aligned}\Delta \log z_{it}^{S} = {} & \alpha_0 + \alpha_1\, \textit{to-zero}_{it} + \alpha_2\, \textit{trait}_i + \mathbf{year}_t\, \boldsymbol{\gamma} \\ & + \mathbf{to\text{-}zero}_{it} \times \mathbf{post}_t\, \boldsymbol{\delta} + \beta_1\, \textit{to-zero}_{it} \\ & \times \textit{trait}_i + \beta_2\, \textit{trait}_i \times \textit{post}_t + \mathbf{to\text{-}zero}_{it} \\ & \times \boldsymbol{trait}_i \times \mathbf{post}_t\, \boldsymbol{\eta} + \boldsymbol{X}_{it}\, \boldsymbol{\mu} + u_{it}, \end{aligned} \tag{3}$$

where $\textit{trait}_i$ is a taxpayer characteristic indicator. To avoid making strong functional-form assumptions, all traits are introduced into the equation nonparametrically. The

[21] Kleven and Waseem (2013) use notches in the baseline Pakistani tax system (2006–2009) to identify these elasticities. These notches were in the same area of the income distribution where the 2010 and 2011 to-zero reforms were applied. Columns 2 and 3 thus compare how very similar taxpayers react to equal-sized to-zero and a not-to-zero rate changes.

[22] High-income taxpayers may have access to more sophisticated methods of evasion (e.g., they can hire accountants, move income offshore). But they may also be subject to stricter enforcement (a higher proportion of digital/third-party-reported transactions). Thus, a priori it is not clear if the evasion rate among them would be higher or lower than the low- and middle-income taxpayers.

coefficients on the triple-interaction dummies capture differential responses of taxpayers with the given trait j. Since traits are not randomly assigned, one problem with this specification is that the estimates might simply reflect that taxpayers with different traits are located in different areas of the income distribution.[23] To make the comparison more meaningful, I also show results from an alternative approach, where I group taxpayers into bins of PKR 20,000 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 trait j in the binned income distribution. This approach is similar to matching and compares taxpayers of different traits by matching them on the basis of their base period income. Figure A11 and tables A11 to A13 show the results. Clearly, once the position of a taxpayer in the income distribution is controlled for nonparametrically, the response does not vary much across small versus large taxpayers (table A11), manufacturers versus nonmanufacturers, regular tax filers versus irregular tax filers, VAT-registered taxpayers versus other taxpayers, electronic return filers versus manual return filers (table A12), and young taxpayers versus old taxpayers (table A13). The evidence thus suggests that within the population of the self-employed, there is little variation in the the extent of evasion.

C. *How Tight Is the Lower Bound?*

In the baseline model, taxpayers report their true income as the tax rate approaches zero: evasion offers no benefit at the zero rate but entails a strictly positive cost. Although quite intuitive, the assumption might not be satisfied in richer settings. In this section, I take up three such settings, examining in each case if the extension can cause a significant departure from the baseline result.

The first setting is the downward revision of the exemption cutoff. Taxpayers may fear that future reforms could lead to a strictly positive rate in the currently zero-rated region, thereby not revealing their true income as the rate approaches zero. While such downward revision of the exemption cutoff is possible and does occur in some countries, it is not common. In the seventy-year history of Pakistan, the cutoff has never been revised downward, although its upward revisions are quite frequent. Piketty and Qian (2009) study the income tax exemption cutoff and its consequences in India and China. Between 1986 and 2008, the cutoff increased from Rs 18,000 to Rs 150,000 in India and from yuan 10,000 to yuan 20,000 in China; it was never revised downward in either country. Similarly, the exemption cutoff in the United Kingdom, known as the personal allowance, has increased from £1,165 in 1979–1980 to £11,500 in 2017 and has never been revised downward. There is thus some evidence that the downward revision of the exemption cutoff is a relatively rare event. Note, however, that even if the exemption threshold is nominally rigid, the real exemption threshold may move downward if the nominal threshold is underindexed to inflation and economic growth (see the comparison of China and India in Piketty and Qian, 2009). It is one of the reasons that I estimate my lower bound on the evasion rate from the behavior of taxpayers who are reasonably away from the nominal threshold (see the discussion at the end of this section).

The second setting is a cross-check in other bases. Reporting true income in one tax base may cause taxpayers difficulties in other tax bases. Pakistan raises almost all of its revenue through five tax instruments: income tax, VAT, customs duty, excise duty, and petroleum levy. Of these instruments, VAT has the greatest overlap with income tax. If the difficulty of cross-matching declared outcomes in other tax bases is really important, the to-zero responses of VAT-registered taxpayers must be substantially smaller than taxpayers who are not registered for VAT. But the responses of the two sets of taxpayers—VAT registered and others—are not significantly different from each other (see figure A11, panel E, and table A12 for the evidence). That the possibility of cross-verification across multiple bases does not worry taxpayers too much should not be surprising, as the evidence from other contexts also shows that when taxpayers report a higher base in one tax, they can leave the base reported in an overlapping tax unchanged by adjusting along some other margin (Carrillo et al., 2017; Waseem, 2018b).

The third setting is the threat of future audits. Taxpayers may not report their true income even at a zero tax rate if they fear that it would make them more likely to face an audit or that it would be used against them in future audits. Note, however, that the first of these two considerations is entirely absent in the Pakistani setting. As explained in section IIA, the probability of audit faced by Pakistani tax filers is small and exogenous. The probability does not increase or decrease with a change in reported income. On the second of these considerations, note that the ability of an audit to assess the legitimacy of a change in the reported income categorized $\underline{e}$ in this paper is limited. The earning or consumption of this component of income leaves no verifiable information trail, making it extremely difficult for an auditor to claim, much less prove, that the change in income is not legitimate.

This above discussion shows that none of the above three extension is likely to push us too far away from the baseline setting. The lower bounds reported in table 3 therefore must be close to the actual evasion rate of these taxpayers. One argument reinforcing the conclusion is that any disadvantage of revealing true income, even if it exists, must diminish as we move left within the zero-rated region. It is because all the forces mentioned above, and similar other forces, must decline as we move toward the lower end of the distribution.[24] Empirical evidence is consistent with this argument. As we move left of the new exemption cutoff, the

[23] This is especially problematic in the current setup because the responses, as shown in figure 2A, decline monotonically along the income distribution.

[24] Taxpayers at the lower end of the distribution become increasingly less likely to face a positive tax rate in future (the first point here), be registered

to-zero responses initially grow stronger. They reach their peak at PKR 150,000 and remain static thereafter (see figure 2). This pattern provides perhaps the strongest evidence that the responses of taxpayers with baseline income up to PKR 150,000 are free from the above considerations and therefore capture the amount evaded by these taxpayers reasonably.

VI. Conclusion

I leverage a series of sharp changes in the Pakistani income tax schedule to document that behavioral responses induced by to-zero tax reforms—reforms that move the rate to or away from zero—are orders of magnitude larger than those induced by others. This result is remarkably robust across specifications and is true for both the self-employed and wage earners, although as expected, the responses of wage earners are much smaller. I explore three explanations of the observed behavior. It can be argued that taxpayers pay full attention when the rate reduces to zero but not otherwise. While a salience-based model can explain the weak and insignificant not-to-zero responses, it cannot explain the large to-zero responses documented in this paper (taxable income elasticity as large as 60). I also reject discontinuous enforcement—zero-rated incomes receive lighter enforcement—as an explanation of the result. The likelihood of audit faced by Pakistani income tax filers is small and independent of the tax rate they face or any other of their traits. I conclude that the divergence in behavior is driven primarily by tax evasion. The costs of not reporting some categories of income are small and of others large. Income entailing little evasion cost is reported at the zero rate but not otherwise. On the other hand, income entailing large evasion cost is always reported. Responses to to-zero tax reforms are larger because they include both categories of reported income. In the evasion-cost-based model, the difference between the to-zero and not-to-zero responses provides a lower bound on tax evasion. Exploiting the Pakistani tax reforms, I estimate that at least 70% of self-employment and 1% of wage income is not reported by the low- and middle-income taxpayers of the country.

The reporting behavior I uncover has important policy implications. First, the elasticity of taxable income is an important parameter for key tax policy choices such as setting the optimal tax rates. The literature estimates this elasticity from not-to-zero reforms only. I show that the most important reaction to taxes takes place as the rate moves from zero to a small, positive value. Ignoring this can result in suboptimal policy choices. For example, the shape of the optimal income tax schedule, especially at the bottom of the income distribution, that takes these reactions into account would be substantially different from the one that does not. Second, I find that a substantial proportion of taxable income goes unreported. This suggests that the return from investment in the enforcement capacity could be large (Besley & Persson, 2013). Research shows that altering the tax mix (Best et al., 2015), redesigning the incentives of tax collectors (Khan, Khwaja, & Olken, 2016), promoting documentation (Naritomi, 2018), strengthening traditional enforcement methods (Almunia & Lopez-Rodriguez, 2018; Waseem, 2018a), and priming social and psychological factors (Waseem, Slemrod, & Rehman, 2020) can promote tax compliance. Exploiting such measures is even more desirable in developing economies where the third-party information, which drives tax compliance in rich countries, is limited in both its scope and effectiveness (Jensen, 2019; Carrillo et al., 2017; Waseem, 2018a). Third, I show that tax evasion lowers the effective tax rate faced by the low- and middle-income self-employed to one-half the statutory rate. This creates significant horizontal equity concerns across the self-employed and wage earners that need to be addressed to build trust in the tax system and to avoid misallocation of physical and human capital.

for other taxes (the second point here), or face a heightened risk of audit (the third and fourth point here).

REFERENCES

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

Almunia, Miguel, and David Lopez-Rodriguez, "Under the Radar: The Effects of Monitoring Firms on Tax Compliance," *American Economic Journal: Economic Policy* 10:1 (2018), 1–38.

Besley, Timothy, and Torsten Persson, "Taxation and Development," in Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, eds., *Handbook of Public Economics*, vol. 5 (Amsterdam: Elsevier, 2013).

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem, "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan," *Journal of Political Economy* 123 (2015), 1311–1355.

Brockmeyer, Anne, and Marco Hernandez, "Taxation, Information and Withholding: Evidence from Costa Rica," World Bank working paper (2017).

Carrillo, Paul, Dina Pomeranz, and Monica Singhal, "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," *American Economic Journal: Applied Economics* 9 (2017), 144–164.

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

——— "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica* 80 (2012), 969–1018.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," *American Economic Review* 99 (2009), 1145–1177.

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

Feldstein, Martin, "Tax Avoidance and the Deadweight Loss of the Income Tax," this REVIEW 81 (1999), 674–680.

Finkelstein, Amy, "E-ztax: Tax Salience and Tax Rates," *Quarterly Journal of Economics* 124 (2009), 969–1010.

Gruber, Jon, and Emmanuel Saez, "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics* 84 (2002), 1–32.

Jensen, Anders, "Employment Structure and the Rise of the Modern Tax System," NBER working paper 25502 (2019).

Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken, "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *Quarterly Journal of Economics* 131 (2016), 219–271.

Kleven, Henrik, and Mazhar Waseem, "Behaviroal Responses to Notches: Evidence from Pakistan Tax Records," London School of Economics working paper (2012).

——— "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *Quarterly Journal of Economics* 128 (2013), 669–723.

Naritomi, Joana, "Consumers as Tax Auditors," London School of Economics working paper (2018).

Piketty, Thomas, and Nancy Qian, "Income Inequality and Progressive Income Taxation in China and India, 1986–2015," *American Economic Journal: Applied Economics* 1:2 (2009), 53–63.
Pomeranz, Dina, "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review* 105 (2015), 2539–2569.
Saez, Emmanuel, "Reported Incomes and Marginal Tax Rates, 1960–2000: Evidence and Policy Implications," *Tax Policy and the Economy* 18 (2004), 117–173.
Saez, Emmanuel, Joel Slemrod, and Seth Giertz, "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature* 50 (2012), 3–50.
Slemrod, Joel, "A General Model of the Behavioral Response to Taxation," *International Tax and Public Finance* 8 (2001), 119–128.
——— "Tax Compliance and Enforcement," *Journal of Economic Literature* 54 (2019), 904–954.
Taubinsky, Dmitry, and Alex Rees-Jones, "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *Review of Economic Studies* 85 (2018), 2462–2496.
Tourek, Gabriel, "Targeting in Tax Behavior: Evidence from Rwandan Firms," Harvard University Kennedy School working paper (2019).
Waseem, Mazhar, "Information, Asymmetric Incentives, or Withholding: Understanding the Self-Enforcment of Value-Added Tax," University of Manchester working paper (2018a).
——— "Taxes, Informality and Income Shifting: Evidence from a Recent Pakistani Tax Reform," *Journal of Public Economics* 157 (2018b), 41–77.
——— "Using the Movement of Exemption Cutoff to Estimate Tax Evasion: Evidence from Pakistan," University of Manchester working paper (2019).
Waseem, Mazhar, Joel Slemrod, and Obeid Ur Rehman, "How Do Taxpayers Respond to Public Disclosure and Social Recognition Programs? Evidence from Pakistan," CEPR working paper DP 14463 (2020).