CESifo WORKING PAPERS

8231
2020

April 2020

Overclaimed Refunds, Undeclared Sales, and Invoice Mills: Nature and Extent of Noncompliance in a Value-Added Tax

Mazhar Waseem

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8231*

Overclaimed Refunds, Undeclared Sales, and Invoice Mills: Nature and Extent of Noncompliance in a Value-Added Tax

Abstract

I leverage a Pakistani tax reform that cuts the tax rate on the supply chains of five major industries of the country from 15% to 0% to cast light on the extent of, and mechanisms driving, VAT noncompliance in a representative emerging economy. I find that firms overclaim refunds by 22% and underreport domestic B2C sales by 43.5%. Together, this implies an evasion rate of 77% in the treated industries and 38% in the population. I explore the role of three mechanisms (1) the destination principle, (2) the last-mile problem, and (3) invoice mills in driving this noncompliance.

JEL-Codes: H250, H260, H320.

Keywords: VAT, tax evasion, firm behavior.

Mazhar Waseem
Department of Economics
University of Manchester
United Kingdom - M13 9PL, Manchester
mazhar.waseem@manchester.ac.uk

April 2020
I thank Miguel Almunia, Zareh Asatryan, Pierre Bachas, Juan Carlos Suarez Serrato, Joel Slemrod, and Christian Traxler for helpful insights. I also thank the seminar participants at the University of Manchester, the National Tax Association, and the University of Freiberg for helpful suggestions and comments. Financial support from the Economic and Social Research Council (ESRC), UK's Future Research Leaders Program is gratefully acknowledged.

I Introduction

In one of the most influential results in public finance, Atkinson & Stiglitz (1976) show that under a fairly broad set of conditions employing a differential commodity tax when the government has access to nonlinear income taxation is not optimal. Notwithstanding this canonical result, a broad-based VAT based often on nonuniform rates continues to be applied together with the income tax in both rich and emerging economies. In fact, its share in government revenue is rising steadily (OECD, 2017; International Tax Dialogue, 2013). This apparent discord between theory and practice in large part reflects the almost universally held belief that among the class of production-efficient tax instruments VAT has the best enforcement properties. This belief has underpinned a remarkable expansion of VAT over the last half century.

Recently, however, a few cracks have begun to appear in this consensus. For example, Malaysia has replaced its VAT (called Goods and Services Tax) with a turnover tax (called Sales and Services Tax) from September 2018 and Zambia came very close to doing so in 2019-2020.[1] The discontent with the VAT in these and similar other settings stems from its two well-known vulnerabilities (Keen, 2007). First, the *self-enforcement* forces built into a VAT work only on firm-to-firm transactions and break down at the final production stage, where sales to consumers take place (Pomeranz, 2015; Naritomi, 2019; Waseem, 2020). This last mile problem of the VAT is particularly severe in developing economies where the final production stage is often fragmented, being composed of small, informal firms. Second, the destination-based design of VAT necessitates that any tax collected on intermediates be refunded to exporters. Refund payment makes the VAT vulnerable to distinct forms of frauds not seen in other tax instruments, such as the missing trader fraud (Keen & Smith, 2006).

While the theoretical mechanisms underlying VAT noncompliance are well-known, we still do not understand fully how important they are empirically. In fact, VAT's evasion in general is much less understood than income tax's. For example, there is little micro-based evidence from tax return data on how much VAT gets evaded and what mechanisms underpin it. In contrast, such evidence on income tax has been available for some time from both policy studies and academic research (see for example Slemrod, 2007; Kleven *et al.*, 2011; Artavanis *et al.*, 2016; Waseem, 2019a). This paper fills this gap in literature. It leverages a novel tax reform from Pakistan to es-

[1] Please see Richard Asquith's blog at avalara.com for the policy changes in Malaysia and Zambia. Specifically, the Malaysian change is documented here and the Zambian here.

timate the extent of VAT noncompliance in a representative emerging economy. It also estimates the individual contribution of the two key mechanisms—the last mile problem and the destination-based design—in driving this noncompliance.

The reform I exploit reduced the VAT rate applicable on five major industries of the country from the standard rate of 15% to 0%. Before the reform, in accordance with the destination principle, only exports of these industries were zero-rated and their imports and domestic supplies were subject to the standard rate. The purpose of the reform was two-fold. First, a major proportion of the output of these industries—textile, leather, carpets, sports, and surgical goods—was exported out of the country and thus was zero-rated anyway. Their long domestic supply chains, however, meant that VAT was to be remitted at each production stage and claimed back at the next till a major proportion of it (nearly three-fourths) was refunded at the export stage. Not only was this arrangement cumbersome, creating compliance costs throughout the supply chain; it also resulted in liquidity problems for exporters who had to wait—sometimes for long time—for the payment of refunds. Second, over time fraud element had crept into the refund system, whereby exaggerated refunds were being claimed on the basis of spurious invoices. Ultimately, the problems created by these two related issues became so severe that the government forsake the VAT due on the domestic consumption of goods produced by these industries and zero-rated their entire supply chains.

VAT can be evaded by overreporting purchases or underreporting sales. The reform I exploit is novel in the sense that it cuts the rate applicable on both intermediates acquired and final goods sold by firms in the treated industries to zero. The reduction of the rate to zero weakens considerably, if not removes entirely, the incentive to misreport. As a result, sales and purchases reported by the treated firms would approach their true values in the post-reform periods. This allows me to infer the extent of evasion as it existed in the treated industries at the baseline. I put this empirical strategy to use, comparing outcomes across firms in the five treated industries with the rest. All specifications include a full set of firm and time fixed effects. Identification in this setup requires that conditional on the firm fixed effects the time path of outcomes would have been similar across the compared groups absent the tax reform. Using the standard event study plots spanning 84 pre-reform and 72 post-reform periods, I show that this indeed is highly plausible in my setting. I run a battery of additional tests to rule out other identification concerns, including any spillovers between the two groups.

To see how firms react to the reduction of the rate to zero, I look at six outcomes reported in the VAT return: (1) input tax, (2) output tax, (3) purchases, (4) sales, (5) exports, and (6) non-export sales. The first two of these are a direct function of the tax rate and therefore capture the mechanical effect of the policy change—the first stage in my empirical framework. I find that both input tax and output tax plummet in the treatment group at the time of the reform. No discernible contemporaneous change occurs in the untreated group. The other four outcomes are not mechanically affected by the tax rate. Indeed, if firms report truthfully, the reduction of the rate to zero should have no bearing on their sales and purchases other than through positive, *real* effects, arising for example from the liquidity channel. I, however, find that all these outcomes reduce sharply in the treatment group after the reform: reported purchases by 42 log-points, sales by 22 log-points, exports by 11 log-points, and non-export sales by 8 log-points. The observed behavior therefore can only be rationalized by the presence of significant misreporting at the baseline.

Using a simple conceptual framework, I show that ignoring one-sided evasion[2] aggregate VAT noncompliance in an economy is pinned down by the extent to which (1) purchases of intermediates used for exports are overreported, and (2) domestic B2C sales are underreported. These two terms capture the effects of the two revenue-depleting mechanisms noted above: the destination principle and the last mile problem. Using the responses documented above, I estimate that at the baseline export-related purchases were overreported by nearly 22%. Roughly twice this amount was evaded on B2C sales, which were underreported by around 43.5%. Together, this caused a total revenue loss of PKR 38 billion in 2004, which amounts to 11.5% (77% of the statutory rate of 15%) of the *true* B2C sales. These noncompliance rates are relevant to the treated industries only. To estimate the average noncompliance among the universe of VAT filers, I divide firms into 32 cells on the basis of their three baseline characteristics: firm size, location, and principle business activity. I re-estimate my model in these cells, and by combining the cell-level estimates with the baseline distribution of firm characteristics compute the average evasion rate among the universe of VAT filers, finding it to be 38% of the true base.

VAT supply chains are rarely complete, especially in developing economies. The profusion of small, informal firms in such economies means that for almost every supply chain some part of the production process occurs in the informal sector. When

[2]One-sided evasion means that a firm underreports its sales or overreport its purchases unilaterally, that is not in collusion with the buyer or the supplier.

the VAT chain breaks, the tax charged at the pre-break stages cannot be claimed back in the post-break stage. This creates arbitrage opportunities, which are sometimes exploited by firms called invoice mills (Keen & Smith, 2006). These firms engage in no real business activity and exist solely to trade in VAT invoices. They buy invoices from firms whose real buyers do not need them because they are either consumers or informal firms. They sell invoices to firms that either buy their intermediates from the unregistered sector or intend to overreport purchases.

Invoice mills are a poorly understood phenomenon. To my knowledge, there is no micro-based evidence in the existing literature on how they operate and how much revenue loss they cause by injecting spurious invoices into a VAT system. One key reason for the lack of evidence is that being fraudulent enterprises invoice mills are hard to identify in the data. In the Pakistani context, a legal process used by the tax administration helps me overcome this challenge. The legal process consists of two steps. In the first step, the tax administration suspends the registration of a firm it suspects of being an invoice mill. While suspended, the firm is given an opportunity to defend itself in quasi-judicial proceedings, at the culmination of which either its registration is restored or it is blacklisted permanently. I observe both suspension and blacklisting in my data and use them to proxy if a firm is an invoice mill. The data show that invoice mills appeared soon after the destination-based VAT was implemented in the country. Their numbers grew steadily over time before crashing once the zero-rating reform took effect. I further show that invoice mills exist primarily to serve exporters, helping them claim exaggerated refund. For example, in the baseline year of 2004, almost 97% of their sales were to exporters, ending up as refund claims. Roughly two-fifths of the refund overclaimed by exporters in that year, amounting to PKR 8.6 billion, was claimed on their invoices. Invoice mills are thus a critical channel through which fraudulent overclaim of VAT refund takes place.

Stiglitz (2010) argues that the central goal of a development-oriented tax policy should be to ensure that the tax structure is resistant to evasion. Understanding the nature and magnitude of tax evasion is the first step in evolving such a tax design. As I note above, exaggerated refunds and the last mile problem are known vulnerabilities of the VAT. Refund, for example, has been termed the Achilles heel of the VAT (Bird & Gendron, 2007), and the last mile problem has been argued to be the key reasons the enforcement of a VAT can unravel (Keen, 2007). The primary contribution of this paper is to empirically demonstrate the nature of these revenue-worsening mechanisms built into the VAT and to quantify noncompliance resulting from them. To my

knowledge, this has not been done earlier, although the enforcement properties of the VAT in emerging and transition economies have been studies in a nascent strand of literature (see for example Pomeranz, 2015; Agrawal & Zimmermann, 2019; Almunia *et al.*, 2019; Shah, 2019; Waseem, 2020 and Slemroad & Velayudhan, 2020 for a survey). This paper also adds to a series of works that uses observational or experimental data to estimate the extent of tax evasion in a variety of contexts including the UK, the US, Russia, Denmark, Greece, and Pakistan (see for example Feldman & Slemrod, 2007; Gorodnichenko *et al.*, 2009; Kleven *et al.*, 2011; Artavanis *et al.*, 2016; Waseem, 2018b). None of these papers, however, relates specifically to VAT. Documenting the nature and extent of VAT noncompliance is important not least because its claim of superiority among the class of production-efficient taxes rests primarily on its compliance properties. Finally, this paper adds to a growing body of empirical literature that uses administrative tax return data to study tax compliance as a key constraint on the development of fiscal capacity in developing and emerging economies (see for example Bachas & Soto, 2019; Brockmeyer *et al.*, 2019; Carrillo *et al.*, 2017; Slemrod *et al.*, 2019; Waseem, 2018a).

The rest of this paper is organized as follows. Section II develops a simple framework to guide the empirical analysis; section III describes institutional features of the Pakistani context; section IV develops empirical methodology used to tease out the effects of the zero-rating on firm behavior; section V presents the results for all firms; section VI computes the extent of noncompliance using behavioral responses to the zero-rating reform; section VII documents the role of invoice mills in facilitating noncompliance, and section VIII concludes.

II Conceptual Framework

This section develops a simple framework that characterizes key mechanisms underlying VAT noncompliance and derives a formula to compute the revenue loss caused by them. The framework is developed through three examples of firm behavior under different settings. Although very stylized, the examples cover a broad range of VAT noncompliance witnessed in emerging economies.

II.A Setup

Consider a simple supply chain consisting of three production stages shown in Figure I. For simplicity, I ignore firm interactions within a production stage, assuming that it is populated by a representative firm only. The firm in the middle stage uses the intermediate produced by the bottom-stage firm to produce a consumption good, a proportion α of which is sold to the export stage for onward supply to foreign buyers. The rest of the good serves the domestic market through the final production stage (retail; B2C sale). I denote true sales and purchases of the firm in stage $J \in \{B, M, E, R\}$ by s_J and c_J and its reported sales and purchases by $\hat{s}_J$ and $\hat{c}_J$. Purchases here represent the value of *taxable* intermediates acquired by the firm.

II.B First-Best Benchmark

I first consider behavior under the first-best scenario, where all firms report their sales and purchases truthfully ($\hat{s}_J = s_J;\ \hat{c}_J = c_J;\ \forall J$). Panel A of Table I illustrates the input-output linkages between firms under this and other scenarios considered later. In the first-best, the government receives a total revenue of $\tau.s_R$ from the entire supply chain, which intuitively equals the tax rate times the retail sales: the value of consumption in the domestic market (B2C sales). The tax paid on intermediates (B2B sales) balances out against the input tax credit and refund. It is the standard result that absent misreporting a VAT is equivalent to a retail sales tax in terms of revenue given that only B2C sales matter in such setting.

II.C Undeclared Sales and Overclaimed Refunds

The second example considers a case closer to low-tax-capacity settings of developing economies, allowing for two forms of noncompliance. First, I assume that firms can underreport sales $\hat{s}_J < s_J$ and overreport purchases $\hat{c}_J < c_J$ on paying a resource costs of $g(s_J - \hat{s}_J, \hat{c}_J - c_J)$. Second, I assume that firms can shift some of their input tax entitlement arising from making domestic sales toward exports. Since this latter form of noncompliance known as the diversion fraud (Keen & Smith, 2006) is unique to VAT, it requires further elaboration.

The VAT, almost as a principle, is implemented as a destination-based tax (Ebrill *et al.*, 2001). Under the destination principle, commodities are taxed in the country

they are consumed rather than where they are produced. This necessitates that imports into the country be taxed at the standard rate and exports at a zero rate. Zero-rating means that the seller does not pay any VAT on output but can still claim the tax paid on inputs as refund. While zero-rating serves the important purpose of ensuring that exports remain free of any domestic tax, it makes the VAT vulnerable to a distinct form of noncompliance not seen in other tax instruments (Keen & Smith, 2006). In this form of noncompliance, exporting firms take advantage of the destination based design to overclaim refunds.

To see how this process works, note that the exporting firm in our example faces an incentive to overreport its purchases. Given that the firm's tax liability is negative, overreporting purchase would increase the refund the firm obtains from the government. The retail firm, on the other hand, would like to underreport its purchases. While doing so would lower the firm's tax credit, keeping a few transactions out of books would in turn enable the firm to underreport its sales, thus avoiding the tax due on its value-added. Because these transactions are to end-consumers, keeping them out of books would not be as costly as inter-firm transactions, a process known as the last mile problem (Pomeranz, 2015; Naritomi, 2019; Waseem, 2020). This scheme of incentives can give rise to a collusive equilibrium in which the middle-stage firm books some of its sales to the retailer as sales to the exporter in its VAT records. Relabeling one rupee of sales in this manner generates a monetary benefit of $\tau + \tau.v_R$, where τ is the amount of overclaimed refund at the export stage and $\tau.v_R$ is the underpaid tax at the retail stage (v_R denotes the value-added at the retail stage). I assume that this monetary benefit exceeds the collusion costs so that a collusive equilibrium ensues in which the middle-stage firm misreports a proportion $(\hat{\alpha} - \alpha)$ of its sales to retailer as sales to exporter. The government's revenue in this case is lower than the first-best by an amount

$$\Delta R = \underbrace{\tau\left(\hat{c}_E - c_E\right)}_{\text{overclaimed refund}} + \underbrace{\tau\left(s_R - \hat{s}_R\right)}_{\text{underpaid tax}}. \tag{1}$$

Note that in this example the bottom-stage firm reports truthfully and the middle-stage firm remits the correct amount of VAT. The revenue loss therefore stems from two sources: (1) the middle firm falsifying the destination of its sales from domestic consumption to exports, and (2) the retail firm underreporting its sales. This form of misreporting is sometimes referred to as evasion through *flying* invoices by VAT

administrators. It is an example of two-sided evasion meaning that the seller and buyer are in collusion, their VAT records match each other, and the input tax claimed on the VAT invoice has been remitted into the treasury. It is only that the firm claiming the credit is not the one which has consumed the inputs the *flying* invoice relates to. This form of evasion needs to be distinguished from the more serious case of one-sided evasion through *fake* invoices, where the tax credit claimed by the buyer has not even been remitted into the treasury.

It is important to emphasize that not allowing tax evasion at the bottom or middle stage though unrealistic does not make any material difference to formula (1). Ignoring one-sided evasion,[3] what matters for VAT revenue is how much B2C sales and purchases used for export are reported in the economy. To see it formally, note that for given values of $\hat{s}_R$ and $\hat{c}_E$ allowing misreporting at the lower stages does not change the above formula. For instance, if the bottom firm misreports its sales $\hat{s}_B < s_B$, its effect would be canceled out as long as the middle firm reports purchases matching with the bottom firm's sales $\hat{c}_M = \hat{s}_B$ (no one-sided evasion). Real world supply chains are longer and more complicated than the one in our example. In such supply chains, firms at each production stage may simultaneously engage in B2B, B2C, and export transactions. But the intuition developed here still applies. Any misreporting on B2B transactions would cancel out so that the formula (1) would remain valid. It means that while our example is stylized, the result it produces is quite general.

As I note above, the first term in formula (1) is the overclaimed refund at the export stage. In its simplest form, the overclaim may mean that exporters overreport purchases only ($\hat{s}_E = s_E$ and $\hat{c}_E > c_E$). But there is a limit to which purchases can be overreported as persistent small or negative value-addition is difficult to justify, especially over a longer period of time. In a more serious version of such noncompliance, therefore, both exports and purchases are overreported ($\hat{s}_E > s_E;\ \hat{c}_E > c_E$). Overreporting exports allows exporters to overclaim refunds while remaining within a justifiable range of value addition.[4]

[3]One-sided evasion results when a firm underreports its sales or overreport its purchases unilaterally, that is not in collusion with its buyer or supplier. It is a crude form of evasion, which can be detected through cross-matching of VAT invoices. Two-sided evasion, on the other hand, is robust to cross matching.

[4]There is some anecdotal evidence that overreporting of exports occurs routinely. Ebrill *et al.* (2001), for example, write "Overstatement of exports, sometimes with the support of falsified customs documents, can be a particularly attractive way to do this [overclaim refunds] ... Temporary shell corporations, false export documentation with respect to actual production sold domestically, falsification of

II.D Invoice Mills

In our third example, the middle tier of the production chain is outside the VAT net. Such breaks in the VAT chain are common in developing countries, where high exemption threshold and profusion of small, informal firms mean that some part of the production process often occurs outside the VAT chain. Because of the break, VAT charged at the pre-break stages cannot be claimed at the post-break stage, becoming a part of the price. This creates an arbitrage opportunity that is sometimes exploited by firms called invoice mills (Keen & Smith, 2006). These firms do not carry out any real business activity and exist solely to trade in VAT invoices. In our example, the invoice mill exploits the gap created by the unregistered middle-tier firm, transferring the credit of VAT charged at the bottom stage to the export stage. Routing one rupee of sale through the invoice mill in this manner generates a monetary benefit of τ (the refund claimed by the exporter). This benefit is higher if the government implements a policy commonly observed in developing countries whereby sales of intermediates to unregistered firms are taxed at a higher rate (the benefit in that case would be $\tau+\tau_a$, where τ_a is the *additional* rate applied on sales to unregistered firms). Note that the mere presence of an invoice mill does not lower government revenue relative to the first-best case as long as the exporter claims VAT credit to the extent of inputs used by it. But invoice mills by their very nature are fraudulent enterprises and often engage in activities more pernicious than simply passing on the input tax credit from one stage to the other. Such activities include diverting the input tax credit from domestic consumption toward exports (as in our previous example). Note that formula (1) remains valid even in this case. The invoice mill simply acts as a device to replicate actions taken in the second example by the middle firm on its own.

Although very stylized, the above three examples broadly capture key mechanisms underlying VAT noncompliance. To reiterate: *(i)* negative tax liability at the export stage; *(ii)* the last-mile problem; and *(iii)* broken VAT chains create incentives for firms to (1) overreport exports; (2) overreport purchases; (3) underreport B2C sales; and (4) operate as invoice mills. My empirical application exploits a large tax reform that cuts the rate applicable on five major industries of the country from 15% to 0%. After the change, no tax is payable on the domestic consumption of output of these industries and very little refund is to be claimed on their exports. The reform thus substantially weakens, if not removes entirely, the incentives to misreport doc-

invoices ... are all used extensively to support claims of excess credits to be refunded".

umented above, moving the affected firms roughly from the environment in our last two examples to that of the first. By looking at how firms react to the change, we can infer the extent of noncompliance as it existed at the baseline and the mechanisms driving it.

It is important to emphasize that the reform would also induce some real responses. By eliminating the remit-credit-refund cycle, it improves cash-flow of the treated firms. It also eliminates other distortions, arising for example from broken VAT chains. The rise in real activity caused by such factors would act in opposite direction to the reporting responses mentioned above, attenuating them. To this extent, my estimates of VAT noncompliance should be seen as lower bounds. I, however, present empirical evidence suggesting that the size of these real responses in my setting is not significant.

III Contextual Background

III.A Pakistani VAT System

Pakistan introduced the VAT in the 1990s. The legislation for this purpose was passed in July 1990, and although it envisaged the VAT to be a broad-based tax with standard features, the scope of the new tax was kept limited in the initial few years through large-scale exemptions. These exemptions were withdrawn rapidly from 1996 so that by the year 1998 the tax had been extended to all notable industries of the country other than the energy and services sectors Waseem (2020).[5] In my empirical analysis, I focus primarily on the post-1998 period during which the VAT remained applicable to almost the entire goods sector of the country.

As I note above, the Pakistani VAT follows the standard design. Firms whose annual turnover falls above the exemption threshold are required to register with the tax administration.[6] Firms not required to register can do so voluntarily. While registered, whether voluntarily or otherwise, firms are required to charge VAT on their sales and are allowed to adjust the tax paid on inputs. In case the adjustment

[5]The energy sector was brought into the tax net in July 1999 and services in July 2000. Please see Waseem (2020) for more details on the introduction and growth of VAT in Pakistan

[6]Exemption threshold is applicable to manufacturers and retailers only. For manufacturers, it was PKR 1 million in 1998, and was increased to 2.5 million in 1999 and to 5 million in 2004. For retailers, it remained at PKR 5 million throughout the sample period.

exceeds the output tax, they can carry forward or obtain the refund of the balance amount. A seller is required to issue a tax invoice for each sale transaction, and the buyer can claim the tax credit only if it possesses the invoice issued in its name. Firms are required to file a return and remit the tax due every month. The filing is based on the principle of self-assessment and there is no preaudit contact between taxpayers and tax collectors. Filed returns are considered final unless selected for audit.

The tax is destination-based: imports into the country are taxed at the standard rate and exports are zero-rated. Any tax charged on inputs used for exports is therefore refunded. To obtain refund, the exporter needs to file supporting documents in addition to the VAT return, which is treated as the refund claim. The supporting documents can be filed within reasonable time after the return has been filed, and include the customs and shipping documents showing the export of goods and VAT invoices showing the purchase of intermediates. No refund is sanctioned before an audit of the claim has been completed, and hence there is a natural delay between the claim and the payment of refund. The delay largely consists of the time taken by the tax administration to scrutinize and sanction the claim, although some of it could arise from the exporter taking time to file the supporting documents.

Till 2008, the tax administration did not have the capacity to cross-match input invoices presented by an exporter with the output invoices of the exporter's sellers. They could, however, see that total purchases claimed by an exporter from a seller do not exceed total sales declared by the seller in its VAT return. They could also cross-match import of inputs and exports of outputs claimed by the exporter with the customs records. The preaudit of a refund claim was therefore partly based on electronic verification and partly on the professional judgment of the auditor. There is some anecdotal evidence of corruption in the process, in particular that exporters had to pay bribes to obtain timely payment of their refund claims.

Panel A of Figure A.I plots the standard VAT rate in the country. The rate generally remained at 15% till 2008 when it was increased to 16%. Pakistan introduced a policy in 1998 through which supplies made to unregistered *firms* were taxed at a higher rate. Of course, the higher rate was not applicable on supplies made to end consumers. Panel B of the figure plots the *additional* rate—called Further Tax by the tax code—imposed by the policy. The rate remained between 1 and 3 percentage points before it was eliminated in 2004.[7] The policy, as I explained in the third stylized example

[7]For example, with a standard VAT rate of 15% and a further tax rate of 1%, supplies to registered firms would be charged to tax at a rate of 15% and to unregistered firms at a rate of 16%.

above, strengthens the incentive of a seller to falsify the destination of goods from the real to a fictitious buyer by issuing flying invoices. More specific, in a sale transaction to an unregistered firm the seller can avoid the additional rate applicable by issuing the invoice in the name of a registered firm.

III.B Zero-rating Reform

In July 2005, Pakistan introduced a novel tax reform through which the VAT rate applicable to supply chains of five major industries of the country—textile, leather, carpets, sports goods, and surgical goods—was reduced from 15% to 0%. Before the reform, in accordance with the standard destination-based design, only exports of these industries were zero-rated and their domestic supplies were taxed at the standard rate of 15%. The reform zero-rated not only the supplies of *final* goods produced by these industries but also of their major inputs. For example, in addition to the finished goods produced by the textile industry (fabric, garments, etc.) all its major inputs including ginned cotton, polyester, yarn, undyed fabric, and important dyes and chemicals were zero-rated. The purpose of the change was two-fold (FBR, 2005).[8] First, nearly three-quarters of the output of these industries was exported out of the country in one form or another and hence was already zero-rated. Long within-country supply chains of these industries, however, meant that VAT was to be remitted and claimed back whenever the goods moved from one production stage to the next. This created cash-flow problems for exporters who had to wait, sometimes for long times, for the refund of VAT paid on their inputs.

Second, the VAT chains of these industries were rarely complete. The breaks together with the need to pay a higher rate on transactions to unregistered firms (see Figure A.I.B) had given rise to the phenomenon of fake and flying VAT invoices. Over time, the volume of such invoices in the system was growing, making it increasingly costly for the tax administration to distinguish between genuine and fraudulent refund claims and process them within reasonable time. Ultimately the problems created by these two related issues became so severe that the government forsake the

[8]The two purposes of the reform were described by the Federal Board of Revenue in the following words: "Delays in refunds payments has been a source of anxiety for the taxpayers. ... The measure was also necessary due to the rampant use of fake and flying invoices by unscrupulous agents to claim illegitimate refunds". The FBR hoped that the reform will lead to two benefits: "Firstly, the refund payments would be reduced considerably, and secondly there would be an improvement in the liquidity position of textile sector leading to investment and boosting export and growth further". Please see page 25-26 of FBR (2005) for details.

VAT revenue from the domestic consumption of these industries and zero-rated their entire supply chains.[9]

The reform was announced on June 06, 2005 (applicable from July 01, 2005).[10] It zero-rated 152 items, which included both finished goods and major inputs of the five industries. Where an input was included in the list, its supply became zero-rated regardless of whether it was used in the production of the five industries or otherwise. For this reason, only inputs *predominantly* used by the five industries were zero-rated.[11] The list of 152 items did not include electricity and gas, two important inputs of these industries. These two inputs were also zero rated, but their zero-rating—unlike that of others—was made conditional on their use in the production of the five industries. Legislative instruments zero-rating electricity and gas were therefore issued at the firm level after verification that the firm was indeed operating in one of the five treated industries. This exercise caused some delay in the zero-rating of these two inputs. The first set of orders granting such zero-rating were issued in August 2005 but the exercise was completed only in April 2007.

The reform moved the treated firms from a standard VAT regime to a novel, new regime, where both their output tax liability and input tax entitlement reduced to *nearly* zero.[12] It thus seriously weakened, if not entirely eliminated, any incentive to misreport sales or purchases.

As noted above, the Pakistani tax administration began obtaining transaction-level data from firms from July 2008. This requirement was introduced by adding an annex to the VAT return wherein firms were to provide the details of their sales and purchases during the month, aggregating them up to the level of each supplier and buyer. Simultaneously, electronic filing of both VAT return and the annex was

[9]Limited forms of such schemes have been implemented in Ireland and South Korea as well. In Ireland, for example, firms that export more than 75% of their output can obtain an authorization that allows their suppliers not to charge VAT. In South Korea, those who supply exporters are zero-rated in respect of selected transactions (see Ebrill *et al.*, 2001 for details of these two schemes). The Pakistani tax reform is novel in the sense that it zero-rates the entire supply chain rather than just the pre-export production stage.

[10]Although no firm decision was taken before the June 06, to some extent the reform was anticipated in the sense that the government had been in negotiation with the business associations representing the zero-rated industries in few months leading up to the reform to chart out the details of the reform such as what inputs to be zero-rated.

[11]Otherwise, the loss of revenue from zero-rating would have been unsustainable for the government.

[12]Note that the reform would not reduce the output tax and input tax of the treated firms to zero, although it would reduce both these variables sharply. The output tax charged would not go down to zero if the firm sells a byproduct not included in the list of zero-rated items. And the input tax would not go down to zero because not all inputs used by the treated firms were zero-rated.

made mandatory for all firms. Together, these changes reduced the costs of cross-matching sales and purchase records considerably, enabling the tax administration to detect one-sided misreporting across the supply chain costlessly. This change in enforcement technology in 2008 affects both treated and untreated industries similarly and therefore should not matter in my empirical setting unless there is some interaction between it and the zero-rating reform.

III.C Invoice Mills

The third example in section II shows that the forces created by broken chains can give rise to the emergence of invoice mills in a VAT system. In this section, I document the emergence, growth, and subsequent decline of invoice mills in Pakistan. The key challenge in this exercise is how to identify invoice mills. Like all fraudulent enterprises, these firms take great care in disguising themselves as legitimate businesses so that distinguishing them from others is not easy. In the Pakistani setting, however, I am able to exploit a legal mechanism used by the tax authority to identify these firms.

In the initial few years after the adoption of VAT in Pakistan, the tax authority did not have any specific mechanism to deal with invoice mills. Fraudulent input tax claims based on invoices issued by such firms were dealt with generically, like any other form of noncompliance. Over time, however, the volume of such transactions grew, making it necessary to have a tailored mechanism to deal with the issue. The new mechanism, which came into force in July 2003, empowered the tax authority to suspend the registration of a firm it suspected of being involved in the issuance of fake or flying invoices. The suspension was meant to be a temporary measure aimed at protecting revenue while an inquiry of the firm's tax dealing could proceed. This inquiry was of a quasi-judicial nature, where the firm was confronted with the evidence against it and was afforded the opportunity to present its case. On completion of the inquiry, either the registration of the firm was restored or it was blacklisted permanently. Once a firm was blacklisted, its invoices no longer remained valid and could not be used to claim input tax credit.

Figure II tracks the stock of blacklisted firms in the country. As I mention above, blacklisting connotes that the firm is likely an invoice mill so that the plot roughly captures the evolution of invoice mills in the Pakistani VAT system. Blacklisted firms appeared soon after the destination-based VAT was implemented in the country.

Their number rose sharply in the next few years, reaching a peak of around 1750 in 2003 (roughly 2-3% of all filed returns). The rising trend was partly reversed in 2003, when the new mechanism to blacklist firms came into effect.[13] But a sharper decline occurred after 2005, when the zero-rating reform became applicable. The initial evidence thus suggests that invoice mills primarily exist to serve the export refund market (their numbers declined as the tax rate on five major export-oriented industries declined to zero). I examine this point formally in section VII of the paper.

III.D Data

I use administrative data from Pakistan, which include the universe of VAT returns filed in the country. The VAT return consists of three main sections. In the first section, firms report the aggregate value of their sales, breaking it down into three—domestic taxable, domestic exempt, and exports—components. In the second section, the aggregate value of inputs purchased are reported, divided likewise into the three components. In the final section, firms calculate their tax liability, indicating the tax charged on sales, the tax credited on inputs, and the final tax payable. They select one of the two options—carry forward or refund—in case the tax payable is negative. Each firm in the VAT net is assigned a unique registration number and is expected to file every tax period (month). The data, therefore, have a panel structure. In addition to the return data, I use information on firm characteristics from the tax register. This information includes the 4-digit industry, date of registration, and current registration status (suspended, blacklisted, or otherwise) of the firm. The 4-digit industry coding corresponds to the Harmonized Commodity Description and Coding System (HS Code) and classifies firms on the basis of goods or services they supply.[14] The industry coding allows me to determine if a firm belongs to one of the zero-rated industries.

As I note above, exporters have to file additional documents in support of their refund claims. These documents include supplier-wise details of purchases of intermediates acquired by them. These transaction-level data are available from the tax

[13] Note that blacklisting began from the tax year 2003, and therefore most of the returns included in this plot were filed at the time the firms had still not been declared blacklist and invoices issued by them were still legal tender. In section VII below, I break down these firms' behavior around the event of blacklisting, looking at the pre- and post-blacklisting periods separately.

[14] This system is commonly used by customs administrations around the world to classify traded goods and services.

year 2002 onward, and I use them to construct linkages between exporters and invoice mills to see what proportion of a refund claim is based on invoices issued by invoice mills.

IV Empirical Methodology

Given that the reform I exploit affects a subset of firms in the sample only, the natural research design in this setting is the difference-in-differences framework. I exploit that the incentives to misreport reduce sharply at the time of the reform if the firm belongs to a zero-rated industry and remain unchanged otherwise, estimating the following model

$$y_{it} = \alpha_i + \lambda_t + \boldsymbol{\beta X_{it}} + \gamma.\, zero\text{-}rated_i \times after_t + \varepsilon_{it}, \tag{2}$$

where α_i and λ_t are firm and time fixed effect, $\boldsymbol{X_{it}}$ is a vector of control variables, $zero\text{-}rated_i$ denotes that firm i belongs to an industry whose rate was cut to zero by the reform, and $after_t$ indicates a post-reform tax period (July 2005 or after). I use the model to estimate the impact of the reform on six outcomes (y_{it}) discussed below.

Identification in this setup requires that a given outcome would have evolved similarly in the treatment and control groups in the absence of the tax reform. I exploit the long panel of VAT records to show that this assumption is indeed reasonable in this setting. Specifically, I plot results from the following event-study model

$$y_{it} = \alpha_i + \sum_{j=2}^{N} \delta_j.\, zero\text{-}rated_i \times 1.(tax\ period{=}j)_t + \lambda_t + u_{it}, \tag{3}$$

where j indexes the set of tax periods (months) included in the sample. I estimate the equation on a sample from July 1998 to June 2011, dropping the dummy for the first tax period. I then plot the coefficients on the interaction terms from these regressions for all six outcomes. A given coefficient $\hat{\delta_j}$ captures the average difference in the outcome y_{it} between the two groups in the tax period j relative to the reference period (July 1998). Using these regressions, I show that all six outcomes evolve fairly similarly across the two groups in the 84 pre-reform periods, validating the key identification assumption.

Notwithstanding parallel trends, identification in this setting may fail if the zero-rating reform creates significant spillovers in the non-zero-rated industries, violating the SUTVA assumption (see for example Imbens & Rubin, 2015). These spillovers can take two broad forms. At the extensive margin, the reform could distort the entry decision of a firm: the differential tax treatment may force a firm which would otherwise have entered into a non-zero-rated industry to switch to a zero-rated industry or vice versa. I address this concern by reporting parallel results from balanced panel samples, where the composition of the sample is held fixed. Shutting down the entry and exit, however, does not rule out spillovers along the intensive margin. These spillovers may arise from general equilibrium considerations (the zero-rating reform affecting prices or the cost of compliance generally) or from demand and supply linkages of control firms with the zero-rated industries (Waseem, 2020). I present two sets of evidence to rule out this class of concerns (please see section VII).

I estimate the impact of the reform on the following six outcomes.

(i) Output Tax ($\tau.\hat{s}_{it}$): By definition output tax equals the tax rate times the reported sales. Given that the reform reduces the rate applicable to the treatment group from 15% to 0%, I expect a large, negative $\hat{\beta}$ from these regressions. This decline in output tax is a combination of the mechanical effect (reduction of the rate to zero) and the behavioral effect (changes in reported sales). I isolate the behavioral effect by estimating the sales response separately (see below). The objective of showing the output tax response is to demonstrate that a strong first-stage exists in this setup, whereby the output tax liability of treated firms reduces sharply relative to the control firms as a result of the reform.

(ii) Input Tax ($\tau.\hat{c}_{it}$): The case of input tax is exactly similar to that of output tax outlined above. Recall, however, that there was some delay in the zero-rating of two important inputs (electricity and gas) used by the treated firms.[15] I therefore expect the immediate impact of the reform on input tax to be smaller than that on output tax. This dynamics of the response can be seen clearly in the event study plots.

(iii) Sales ($\hat{s}_{it}$): The variable denotes *total* sales—sum of exports and domestic (both B2B and B2C) sales—of firm i in tax period t. Because the variable is not directly affected by the tax rate, its response represents pure behavioral effect of the reform, as do the responses of the three next outcomes outlined below.

[15]Orders zero-rating electricity and gas were issued at the firm level. The first set of these orders were issued in August 2005 and the last in April 2007. See section III.B for details.

(iv) Purchases ($\hat{c}_{it}$): The variable denotes the reported value of taxable inputs purchased by firm i in tax period t. These inputs include raw materials and intermediates acquired by the firm from other firms and do not include non-taxable inputs such as labor.

(v) Exports: The variable represents the total value of exports made by firm i in tax period t.

(vi) Non-Export Sales: The variable denotes the value of non-export sales of firm i in period t. I construct this variable by taking away exports from aggregate sales reported by the firm. It therefore includes both B2B and B2C domestic sales.

Table II presents summary statistics of the data. The treatment group here contains firms belonging to the five zero-rated industries; all other firms are included in the control group. The first row of the table reports firm-month observations of the two groups in the sample for the two baseline years, 2003 and 2004. The next rows present mean of the six VAT outcomes and other firm characteristics for the two baseline years. On average, the treated firms are larger and are more likely to be engaged in exports. But they are not much different from the control sample in other characteristics such as location and age.

V Firm Responses to the Reform

V.A Nonparametric Evidence

I first present visual evidence on how firms respond to the reduction of the VAT rate to zero. The analysis is then formalized using the regression based framework.

V.A.1 First Stage

Before documenting firm behavior to the zero-rating, it is important to show that the reform creates large tax variation between the treated and untreated firms. I do so by presenting both aggregate and micro level evidence.

Figure III plots the amount of VAT refund paid in Pakistan as a proportion of gross VAT collected in the country during the tax years 1999 to 2010. The figure is based on annual aggregate statistics reported by the FBR on its website, which

include both treated and untreated industries. The refund-to-gross-collection ratio in the country was roughly 20% at the baseline. It fell by nearly 10 percentage points in the first year after the reform. It fell even further in the later years as the backlog of pending refunds was cleared and more inputs of the treated industries were zero-rated (electricity and gas), settling at around the 5% level. The refund paid in the country thus dropped to one-fourth of the baseline level within three years of the reform; in terms of absolute numbers, the amount refunded reduced from PKR 55 billion in 2004 to PKR 27 billion in 2008.[16]

To show that this large drop in refund was triggered by the zero-rating reform, I next turn to the micro level evidence showing the reform's effects on output tax charged and input tax claimed by the treated firms. Figure IV plots the coefficient $\hat{\eta}_t$s from the following version of the event study equation (3)

$$y_{it} = \alpha_i + \sum_{j=2}^{N} \eta_j . 1.(tax\ period{=}j)_t + u_{it}, \tag{4}$$

where j indexes the tax periods (months) included in my estimation frame. I estimate the equation separately for the treatment and control groups, omitting the dummy for the first tax period (July 1998). A given $\hat{\eta}_t$, therefore, denotes the average within-firm change in the outcome from July 1998 to the period t for the corresponding group of firms. Figure V presents DD version of the plots, where I display the coefficients $\hat{\delta}_t$s from equation (3) along with the 95% confidence intervals around them. Panels A-B of the two figures together comprise the first-stage of the empirical setting, depicting the responses of output tax and input tax to the reform. Clearly, a very sharp drop occurs in the treated outcomes exactly from the time of the reform, while the two control outcomes continue to evolve on the preexisting trend. The dynamics of the two responses is also consistent with our expectations. Both output tax and input tax decline sharply as the reform comes into effect, but unlike the output tax the input tax continues to drop, stabilizing only around the beginning of the tax year 2008. The continuing drop of input tax, as noted in section III.B above, is very likely due to the time taken in zero-rating of the two important inputs—electricity and gas—of the treated firms.

Taken together, the above two pieces of evidence demonstrate that a very strong

[16]For these statistics, see FBR Year Books from 1999 to 2009 available here. Year books prior to 1999 are unfortunately not available on the FBR's website.

first stage exists in this setting. The incentives to misreport collapse in the treatment group at the time of the reform as both their output tax liability and input tax entitlement crash down to a near-zero level. Such a salient drop in incentives is likely to induce sharp behavioral responses to which I turn next.

V.A.2 Behavioral Responses

Panels C-F of Figures IV and V illustrate these responses. None of the four outcomes shown in these panels is directly influenced by the tax rate, and their responses therefore isolate *pure* behavioral effects induced by the rate cut. These behavioral effects are negative for all four outcomes. Sales, purchases, exports, and non-export sales decline clearly in the treatment group after the reform. This decline is the sharpest for purchases, which fall by 30-35 log points immediately after the reform. Compared to purchases, the decline in other outcomes is slow and gradual, materializing fully in the next two years or so. This pattern of responses is expected. Purchases are expected to go down immediately once the incentives to overreport them cease to exist. Past overreporting of purchases, however, would have left firms with large inventories even if only in books, which would not let the volume of reported sales and exports drop to the new equilibrium in the few periods following the reform.

The conceptual framework presented in section II predicts that the zero-rating reform would lead to (1) a drop in purchases, (2) a drop in exports, and (3) a rise in domestic B2C sales reported by them. The results in Figures IV and V are consistent with the first two of these predictions. On the third prediction, note that I do not observe domestic B2C sales in the data directly, but using two simple accounting identities I show in section VI that at the aggregate level the variable nearly equals the difference between the non-export sales and purchases reported by firms. The larger, negative response of purchases relative to non-export sales in the two figures shows that domestic B2C sales reported by treated industries indeed rose after the reform.

The above event study results also validate my empirical strategy. The preexisting trends were fairly parallel in the two groups for all the six outcomes I explore (see Figures IV and V). The reform causes sharp changes in the treatment outcomes, while the control outcomes continue to evolve on the preexisting trend with no appreciable break at the time of the reform.

V.B Regression Results

Table III presents the results from the difference-in-differences model. I estimate equation (2) for each of the six outcomes using both complete and balanced panel samples. The balanced panel sample includes only those firms which file their VAT return at least once in every quarter included in the estimation frame (July 1998 to June 2011). I always include the full set of firm and tax period (month) fixed effects and cluster standard errors at the firm level (Abadie *et al.*, 2017; Bertrand *et al.*, 2004). The results are consistent with the visual evidence presented above. Both output tax and input tax drop sharply as a result of the reform, showing that a strong first-stage exists in this setting. Columns 3-6 report pure behavioral responses induced by the rate cut. All four DD coefficients are large and negative for both estimation samples, confirming that the zero-rating causes a large drop in sales, purchases, exports, and non-export sales reported by the treated firms. I use these results to back out the baseline VAT noncompliance in section VI of the paper.

Table IV explores the dynamics of the responses. I estimate a flexible version of model (2) by replacing the double-difference term with six double-interactions, one each for every post-reform tax year. The results confirm the time pattern of response seen visually in Figures IV and V. The first year response as a proportion of the average post-reform response is 79% for input tax, 94% for output tax, 80% for purchases, 32% for sales, 18% for exports, and 56% for non-export sales. I have already noted the likely reasons for this pattern. The first two outcomes capture the mechanical impact of the reform, which unsurprisingly is immediate.[17] Of the other four items, purchases were likely to be impacted first given that once input tax credit available on purchases drove down to zero there was no incentive to overreport them. In contrast, sales, exports, and non-export sales would return to the new equilibrium only after inventories built up in the books through past overreporting of purchases have been cleared.

Two important events occur in 2008 that may influence the interpretation of my results. First, as I note in section III.B, Pakistan introduced new filing requirements from July 2008, which mandated firms to file transaction-level data along with their returns. Second, the financial crisis hit the world markets, initiating the Great Recession. Some of the negative responses documented above may reflect that these events

[17]The slightly lower first-year response of input tax, as I note above, was in large part due to the delay in the zero-rating of electricity and gas used by the treated firms.

affect the treated industries worse than the nontreated industries. For example, the negative export response in 2008 and later years might reflect that the Great Recession reduced the demand of Pakistani exports of treated industries more than those of others. Figure A.II addresses this class of concerns. It is a truncated version of Figure V, where I show the post-reform periods only. Zooming in on these periods shows that the reform started a slow, downward trend in the outcomes of the treated industries. This downward trend did not accelerate during 2008. In fact, the exports of the treated industries started rising again from the mid of the financial year 2008-2009. The evolution of the responses thus rules out any significant *differential* impact of the two events on the treated industries.

Recall that the reform applied to five major industries. Of these, textile is the largest and the most important in terms of its VAT impact. Table A.I shows this formally. I break down the aggregate response reported in Table III into its constituent textile and non-textile components. The response of the textile industry roughly equals the average response for all the outcomes: all textile coefficients are within the 95% confidence interval of the corresponding baseline coefficient. The finding is significant in one important respect. The textile industry has a very well-defined supply chain comprising five distinct production stages: ginning, spinning, weaving, processing, and the made-up stage.[18] Given that I observe the production stage a textile firm operates in, I can explore any heterogeneity in response across the supply chain. One key difference between various production stages in the supply chain is that the upstream stages produce intermediates rather than consumer goods. For example, the outputs of the ginning and spinning industries—pressed cotton and cotton yarn—have no significant non-industrial use. To the extent that upstream firms engage primarily in B2B transactions, they have distinctly lower incentives to misreport their outcomes. The heterogeneity analysis can thus help us uncover the nature of the observed responses. Specifically, any real responses produced by the reform would be roughly symmetric throughout the supply chain. Reporting responses, on the other hand, would be stronger in the later stages.

Table V carries out this exercise. I restrict the treatment sample to the textile industry only and estimate a triple-difference version of model (2) by including interactions

[18]Cotton ginning is the first production stage of the textile industry. In it, cotton fiber is separated from the seed and is compressed into bales. Spinning converts these cotton bales into cotton yarn, which then is converted into gray fabric by the weaving industry. Processing converts gray fabric into colored and printed fabric, which finally is converted into garments and other textile made-ups by the final production stage.

of the double-difference term with dummies indicating the production stage. I include dummies for the three upper-most production stages—ginning, spinning, and weaving, leaving the downstream stages as the omitted category. Clearly, responses are heterogeneous across the supply chain. Specifically, they are considerably weaker in the upstream stages, becoming progressively stronger as one moves down the supply chain. For example, both sales and purchases of the first production stage are not significantly different from those of the control group. In contrast, these outcomes decline by a lot in the later production stages of the treated group.[19] Overall, the analysis thus suggests that any real effects induced by the reform are not significant so that the responses documented above in large part capture changes in misreporting.

One important concern in this setting is that the zero-rating reform may create spillovers in the nontreated industries, violating the SUTVA assumption. Table III shows that I get fairly similar results when the sample is reduced to a balanced panel of firms that file throughout the sample period. This mitigates the spillover concern along the extensive margin. Figure IV further shows that the outcomes of nontreated industries do not exhibit any signs of a structural break at the time of the reform. Building on this evidence, Table A.II explores spillovers more formally. If the reform creates any spillovers, they would be stronger in industries whose products are close substitutes or complements of the treated goods. The tables tests this by looking at the evolution of outcomes of industries similar to the treated industries. I have mentioned in section III.D that the Pakistani tax administration follows the HS Code to classify firms into industries. The first two digits of this eight-digit code divide firms into broad industry categories with similar industries getting adjoining codes. For example the code 08 is assigned to edible fruit and nuts; 09 to coffee, tea, mate and spices; and 10 to cereals. The adjoining codes thus contain fairly similar industries with their products being close substitutes or complements. The table exploits this scheme of classifying industries. I drop all treated industries from the sample and compare the evolution of outcomes in their adjoining industries with that of others. The table reports results from specifications parallel to ones in Table III, the only difference being that the variable *treat* now indicates firms belonging to the adjoining

[19] Another feature of the results is that the input tax drops more in the downstream stages, while the output tax does so in the upstream stages. This is expected because overreporting inputs becomes more feasible as one moves down the value-added chain with both the number and share of taxable inputs increasing in the downstream stages. For example, the first production stage—ginning—primarily uses two inputs only. Both these inputs—labor and raw cotton—are not taxable. Compared to this almost all non-labor inputs used by later production stages are taxable, increasing the margins along which overreporting of purchases can take place.

industries. I experiment with three definitions of adjoining. The first three columns of the table, for example, regard the two 2-digit industries next to the zero-rated industries as adjoining. For space considerations, I report results for the three main outcomes—sales, purchases, and exports—only. Reassuringly, all specification return trivial or insignificant DD coefficient, putting to rest the concern that the reform might have affected outcomes of the nontreated industries as well.

Table A.III rules out one alternative explanation of the results. It can be argued that the incentives to maintain records (receipts, invoices, etc.) and to report them correctly go down once the tax rate goes down to zero. In this world, the negative responses documented in Table III are explained by lazy reporting in the post-reform periods rather than a reduction in misreporting. To rule out this concern, the table looks at the responses of treated corporate and non-corporate firms separately. The idea behind the exercise is that lazy reporting is expected to be worse and the observed responses hence strongly negative among noncorporate firms whose quality of record maintenance is in general poorer. Contrary to this, the responses are in general less negative for the noncorporate firms. More generally, note that all VAT firms are expected to file and pay income tax on their profits. Lazy reporting in the VAT regime can have costly consequences in the income tax regime. For example, if firms underreport purchases because of lazy reporting, it would raise their profits and hence tax liability in the income tax regime. These real consequences mean that lazy reporting is unlikely to be a significant explanation of the observed responses.[20]

VI Quantifying VAT Evasion

Section II shows that misreporting by firms causes the following loss in VAT revenue relative to the first-best case

$$\Delta R = \underbrace{\tau\left(\hat{C}_E - C_E\right)}_{\text{overclaimed refund}} + \underbrace{\tau\left(S_R - \hat{S}_R\right)}_{\text{underpaid tax}}. \tag{5}$$

This formula is an aggregate version of formula (1), where the capital letter notation stands for the sum of the corresponding variable over all firms; for example $\hat{S}_R$

[20] Also note that reporting different figures in the VAT and income tax records is also costly as it exposes firms to the charge of wrong reporting at the time of audit when VAT and income tax records are reconciled.

stands for $\int_i \hat{s}_{i,R}\, d\nu(i)$, where $d\nu(i)$ is the distribution of firms. In this section, I use the formula to compute a lower bound on VAT evasion as it existed at the baseline.

This exercise is based on the idea that the reduction of the rate to zero would eliminate any incentive to misreport, driving firm behavior in the treated industries toward the first-best case. It means that we can proxy the two terms in the above formula by the corresponding changes caused by the zero-rating reform. For example, the first term in the above formula can be represented as $\tau(\hat{C}_E - C_E) \approx \tau(\hat{C}_{E,t} - \hat{C}_{E,t'}) \equiv \tau\Delta\hat{C}_E$, where t and t' here index pre- and post-reform outcomes. Using this notation, the above formula can be rewritten as the following

$$\Delta R \approx \underbrace{-\tau\,\Delta\hat{C}_E}_{\text{overclaimed refund}} + \underbrace{\tau\,\Delta\hat{S}_R}_{\text{underpaid tax}}, \tag{6}$$

where $\Delta\hat{C}_E$ and $\Delta\hat{S}_R$ are the reform-driven changes in these variables.

As I note earlier, the two terms in the above formula are not directly observed. The first of these two terms, $\Delta\hat{C}_E$, is the excess purchases claimed by firms on goods exported by them. The difficulty in computing this term is that firms in their VAT returns do not apportion purchases by their use (domestic sales vs. exports). I therefore need to infer $\Delta\hat{C}_E$ from the observed response of exports to the zero-rating reform ($\Delta\hat{S}_E$). Figure VI shows the baseline relationship between the two variables, plotting a nonparametric representation of the mapping $\hat{s}_E = f(\hat{c})$. To construct this plot, I group firms into small bins on the basis of log of purchases reported by them. I then plot the average log exports of firms in each bin. The sample for this binned scatter plot consists of all firm-month observations of the treated industries for the pre-reform years (1998-2004), excluding those with the log purchases less than the 5th or more than the 95th percentile of the aggregate distribution. I also superimpose a linear regression line on the scatter plot. The relationship between the two variables is fairly linear with a slope parameter of 0.5. Note that this relationship does not need to be causal as I use it solely to predict $\Delta\hat{C}_E$ implied by the $\Delta\hat{S}_E$ estimated in the last section using the DD estimator. Column (5) of Table III shows that the zero-rating caused on average a 10.6 log-points or an 11.2% decrease in exports reported by the treated industries. Firms in these industries reported total exports of PKR 687 billions in 2004. The DD estimator therefore implies that at the baseline exports were over-reported by nearly PKR 76 billions. Using the relationship shown in the binned scatter plot, this translates into excess purchases of nearly PKR 154 billion claimed on

exports at the baseline and an overclaim of refund of nearly PKR 23 billions. These calculations are shown in greater details in Table VI.

The second term in formula (6), $\Delta\hat{S}_R$, denotes the change in B2C sales reported by firms, which as I note earlier is not directly observed. More specific, because the data are not transaction-level, I cannot divide the total non-export sale of a firm into its constituent B2B and B2C components. I can, however, estimate the average under-reporting of B2C sales at the baseline using the following accounting identities

$$\int_i (\hat{s}_i - \hat{s}_{i,E})\ d\nu(i) \equiv \int_i (\hat{s}_{i,B2B} + \hat{s}_{i,B2C})\ d\nu(i) \tag{7}$$

$$\int_i \hat{c}_i\, d\nu(i) \equiv \int_i (\hat{s}_{i,B2B} + \hat{c}_{i,OS})\ d\nu(i). \tag{8}$$

The first of these two identities simply reflects that the sum of non-export sale reported by all firms in the economy must equal the sum of their B2B and B2C sales. Because B2B sales of a firm are reported as purchases of intermediates by other firms, the sum of all B2B sales plus any fake purchases claimed by firms ($\hat{c}_{i,OS}$, where OS denotes one-sided evasion) must equal the sum of all purchases of intermediates reported in the economy. This is captured by the second identity. Substituting this second identity into the first and rearranging terms, I obtain the following expression

$$\int_i \hat{s}_{i,B2C}\, d\nu(i) \equiv \int_i [(\hat{s}_i - \hat{s}_{i,E}) - \hat{c}_i + \hat{c}_{i,OS}]\ d\nu(i), \tag{9}$$

which in our usual notation in terms of changes can be rewritten as

$$\Delta\hat{S}_R \equiv \Delta\left(\hat{S} - \hat{S}_E\right) - \Delta\hat{C} + \Delta\hat{C}_{OS}. \tag{10}$$

The LHS of this expression is our variable of interest—the change in B2C sales reported in the economy. It is a sum of three components: (1) the change in non-export sales, (2) the negative of the change in purchases, and (3) the change in fake purchases reported in the economy. I do not observe the last term in this expression and ignore it from my subsequent analysis, rewriting the above equation as

$$\Delta\hat{S}_R \approx \Delta\left(\hat{S} - \hat{S}_E\right) - \Delta\hat{C}. \tag{11}$$

Ignoring the term means that I estimate a lower bound on the underreporting of B2C

sales $\Delta\hat{S}_R$.[21]

Columns (6) and (4) of Table III show that the zero-rating caused an 8.3 log-points (8.7%) decrease in non-export sales and a 42 log-points (52%) decrease in purchases of intermediates reported by the treated firms. Together, these responses imply that at the baseline domestic B2C sales of treated firms were on average underreported by at least 43.5%. Assuming that domestic B2C sales of treated firms constitute just 25% of their total sales,[22] this underreporting translates into a revenue loss of PKR 15 billion. Panel B of Table VI gives the details of these calculations. The last two rows of the table computes overall evasion rate at the baseline. Misreporting in the treated industries caused an aggregate VAT revenue loss of PKR 38 billion in 2004. Roughly three-fifths of this loss resulted from the overclaiming of VAT refunds and the rest from the underreporting of B2C sales. The PKR 38 billion revenue loss amounts to nearly 13% of the total VAT collected in the country in that year and translates into 11.5% (77% of the statutory rate of 15%) of the *true* B2C sales reported in the treated industries in the year.

The last column of the table repeats these calculations for the balanced panel sample. Firms in this sample remain active throughout the fairly long period of thirteen years included in the estimation frame. Their responses to the zero-rating reform therefore capture tax evasion in a more permanent component of the tax base. Clearly, both the magnitude of, and the mechanisms underlying, noncompliance are fairly similar for this set of firms. It shows that the noncompliance I document here captures the tax evasion of *regular* firms rather than that of "missing traders", a phenomenon commonly discussed with relevance to the noncompliance of VAT.

VAT compliance is particularly worse in export-oriented industries. These industries are particularly prone to the diversion fraud, whereby input tax credit is shifted from domestic consumption to exports. As a result, the government loses revenue from both sides as it refunds more than the due amount on exports and receives less than the due amount on domestic consumption. In my setting, this implies that the

[21] While I do not observe the extent of one-sided misreporting in the economy, it is important to emphasize that it entails greater costs relative to the two-sided misreporting. Importantly, one-sided misreporting is not robust to cross-matching and can be uncovered easily by reconciling the sales and purchase records. I therefore expect its extent to be considerably smaller than the other form of misreporting.

[22] Note that I cannot compute the exact value of the aggregate B2C sales of the treated industries because the two accounting identities (7) hold at the aggregate and not industry level. That the B2C sales of the treated industries constitute 25% of their aggregate output is based on my own calculation using statistics from the Economic Survey of Pakistan for the year 2004-05 (GOP, 2004). Specifically, I deduct export of these industries from total production to arrive at the figure.

noncompliance rate I report above being relevant to the five export-oriented industries only may overstate the average noncompliance among the universe of VAT filers. I follow a simple strategy to estimate this population-wide effect. I divide firms into 32 ($4 \times 4 \times 2$) cells on the basis of their baseline turnover, location, and principle activity.[23] I then use model (2) to estimate the average response of the treated firms within each cell. Extrapolating these results to *all* firms in the cell, I estimate the extent by which VAT is overclaimed as refund on exports and is underpaid on domestic consumption. The approach relies on the assumption that a firm's ability to misreport is a function only of its size, location, and principle business activity so that it does not differ across firms within a cell. In other words, treated and untreated firms within a given cell may differ in terms of how much of their output is exported but not in terms of their ability to overreport purchases or underreport sales. With this assumption, the population-wide effect can be estimated by summing the treatment effect on treated over the baseline distribution of firm characteristics. For example, suppose that the cell-level regressions show that the treated firms in a cell j on average overreport their exports by $\hat{\delta}_{j,E}$. The estimate can be used to compute the total amount by which exports are overreported by all firms in the cell

$$\Delta S_{j,E} = \int_{i \in J} \hat{\delta}_{j,E}\, \hat{s}_{i,E}\, d\nu(i), \tag{12}$$

where J is the set of firms in the cell. Using the approach, I find that a total VAT of PKR 182 billion was evaded in the country in 2004. Of this amount, PKR 43 billion was overclaimed as refund and the rest underpaid on B2C sales. Pakistan reported a net VAT collection of PKR 293 billion in 2004. The evaded amount thus translates into an evasion rate of 38% of the true VAT base. As I note above, the two terms in formula (6) roughly capture the two key mechanisms that drive the evasion of VAT: (1) destination-based design that necessitates the payment of refunds and (2) the last-mile problem that reduces the costs of underreporting B2C sales. The results show that these two mechanisms contribute roughly in the ratio of 1:3 to the total evasion. Of course, this ratio reflect the structure of the aggregate economy; for example the first mechanism would be far more important in economies that export a greater share

[23] Specifically, I divide firms along three dimensions. First, I divide firms into four size quartiles on the basis of their average turnover in the six baseline years (1998-2004). Second, I divide firms into four regions: Karachi, Lahore, Faisalabad and others; the first three of these are the three major cities of the country. Finally, I divide firms into two categories based on if their principle business activity is manufacturing or not. This gives me $4 \times 4 \times 2 = 32$ cells.

of their output.

VII Invoice Mills and VAT Evasion

One important focus of this paper is to understand the role played by invoice mills in the noncompliance of a VAT. The third example in the conceptual framework shows that invoice mills arise naturally in the low-enforcement, high-informality setting of developing countries, bridging the gap created by broken VAT chains. Figure II illustrated this empirically: consistent with the conceptual framework invoice mills did appear and grow in the Pakistani VAT system before declining sharply as the zero-rating reform took effect. I now turn to documenting the behavior of these firms in more details to see if mills are an important channel through which the tax evasion documented in Table VI takes place.

I begin by showing the responses of invoice mills to the zero-rating reform. Figure A.III replicates the analysis in Figure IV, restricting the treatment sample to invoice mills only. Invoice mills here refer to all blacklisted and suspended firms. Together the two categories of firms form only 7% of the treatment sample. The results are therefore noisier than the complete sample results. Yet the pattern of responses is very similar. All four outcomes—sales, purchases, exports, non-export sales—drop sharply at the time of the reform, with purchases dropping more than any other outcome. Table A.IV formalizes these results. I estimate a triple-difference version of model (2), exploring any differential response of invoice mills from other treated firms. Two results in the table are noteworthy. First, clearly mills respond more aggressively than other treated firms. Second, because they constitute a small proportion of the overall sample, mills' larger responses do not affect the average responses too much: the double-difference $treat \times after$ coefficient is always within the 95% confidence interval around the baseline coefficient.

One notable feature of the event study plots in Figure A.III is that the outcomes of mills spike just before the reform. This spike is particularly prominent in the top four panels of the figure. To explore this finding further, Figure VII plots the aggregate values of the six outcomes of mills in each month. Relative to the within-firm average changes shown in Figure A.III, this figure displays the aggregate *level* of each outcome, illustrating more clearly the magnitude of spurious input tax credit injected into the system by invoice mills. Strikingly, all outcomes exhibit a sharp spike just

before the zero-rating reform took effect. For example, aggregate purchases jump from the prereform average level of around PKR 5 billion to 30 billion one tax period before the reform, *i.e.* in May 2005. This sharp jump is followed by an even sharper drop, whereby purchases reduce to PKR 3.6 billion in July 2005 and to 1.8 billion in November 2005.[24]

This large concentration of activity on the wrong side of the reform is puzzling. Under any standard model of tax behavior, a large tax cut reducing the rate from 15% to 0% would induce some inter-temporal shifting of activity from the high-tax periods to the low-tax periods.[25] The behavior I find is polar opposite to this. The only plausible explanation of the behavior therefore is that it captures invoice mills injecting input tax credit into the system while the goods are still standard-rated. Doing so maximizes spurious refunds in accordance with the mechanism laid out in the third stylized example.

How much of excess refund did invoice mills inject into the system? As I note in section III.A, exporters file supplier-wise details of their purchases at the time of claiming VAT refund. These transaction-level data are available from the tax year 2002 onward, and they allow me to quantify the role of invoice mills in the overclaim of refunds. Figure VIII reports the results of this exercise. The blue curve plots the aggregate value of refund claimed by exporters each month on the invoices of black-listed firms. The red curve, on the other hand, plots the aggregate value of output tax in all invoices issued by blacklisted firms each month. Although the two curves are based on data from two different sources—the blue from refund claims filed by exporters and the red from VAT returns filed by blacklisted firms, they line up quite well. This shows that almost all invoices issued by mills end up in the refund claims of exporters. For example, in the baseline year of 2004, out of a total output tax of PKR 8.9 billion reported by blacklisted firms nearly 97% (PKR 8.6 billion) was claimed as refund by exporters. I estimate in Table VI that exporters overclaimed VAT refund of nearly PKR 23 billion in 2004 (see Row 5 of the table). Roughly 37% of this amount was based on invoices of blacklisted firms.

Filing a refund claim does not imply that the amount claimed would necessarily

[24]One other important feature of the plots is a very large spike in exports just before the beginning of the tax year 2008. This is very likely driven by an effort by invoice mills to clear their inventories before the requirement of filing transaction-level data comes into effect from the tax year 2008.

[25]For instance, booking a transaction that occurs just before the reform to a date just after the reform could save the seller the costs of remitting the tax, the buyer the costs of claiming the input tax credit, and any associated cash-flow costs.

be paid by the government. As I note in section III.A, the tax administration preaudits all refund claims, assessing among other things the validity of purchase invoices. To rule out that the numbers I report above represent real and not potential revenue loss, Figure A.III tracks the evolution of VAT outcomes of blacklisted firms around the event of blacklisting. It illustrates that by the time a firm is blacklisted, it is already past its peak activity and both its turnover and purchases are declining. In addition, almost all the activity of such firms occurs while their invoices are still legitimate. To see why invoice mills are detected so late in their life cycle, Table A.V compares their baseline characteristics with other firms. Invoice mills are essentially hit-and-run enterprises, and consistent with this notion their first-year turnover and purchases are much higher than a typical firm. They also display other markers that can help the tax authority distinguish them from other firms such as high turnover, low value-addition, and tax payment. It therefore appears that weak enforcement capacity is perhaps the primary reason invoice mills are detected too late, when their invoices have already been used to obtain refunds.

In sum, the above analysis suggests that invoice mills exist primarily to facilitate overclaim of refunds by exporters. Nearly two-fifths of the excess refund claimed by exporters results from their invoices. In contrast, their role in the VAT evaded on domestic B2C transactions is limited. This finding, however, is subject to an important caveat. I identify mills using the blacklisting procedure employed by the Pakistani tax administration. It is possible that this procedure is more effective in identifying mills connected to exporters either because the government is more concerned about the overclaim of refund than the underpayment of VAT on domestic transactions or because of any data limitation. Notwithstanding the reason, this measurement error means that I underestimate the VAT loss caused by mills. Specifically, mills issuing invoices solely to non-export firms may escape blacklisting and thus may not feature in my empirical framework.

VIII Conclusion

The value-added tax has seen remarkable expansion in recent decades. Its popularity in large part is driven by the belief that among the class of production-efficient tax instruments it has the best enforcement properties. By creating (1) third-party information on firm-to-firm transactions; (2) tax withholding at the upstream production

stages; and (3) asymmetric cheating incentives between sellers and buyers, a VAT facilitates tax compliance (Waseem, 2019b). The revenue gains from these mechanisms, however, need to be weighed up against the losses arising from two revenue-worsening mechanisms built into a VAT: (1) the destination principle that necessitates the payment of refund on exports, and (2) the last-mile problems that lowers the costs of misreporting B2C transactions. These mechanisms are emphasized a lot in the policy literature (see, for example, Bird & Gendron, 2007), but due mainly to a lack of empirical evidence are largely absent from the economic literature. This paper fills the gap by casting light on the nature of these mechanisms and identifying the amount of evasion caused by them in a representative emerging economy.

For this purpose, I leverage a novel tax reform that seriously weakens the incentives of firms to misreport their sales or purchases. The variation allows me to estimate the level of misreporting in the country and study the mechanisms driving it. I find that firms overreport export-related purchases by nearly 22% and underreport domestic B2C sales by nearly 43.5%. Together, this results in a total revenue loss of PKR 38 billion at the baseline, which is nearly 11.5% (77% of the statutory rate of 15%) of the *true* B2C sales—the real tax base. I combine these estimates of average treatment effect on treated with the baseline distribution of firm characteristics to estimate the noncompliance rate among the population of VAT filers, finding it to be 38% of the true base. The two mechanisms—the destination principle and the last mile problem—roughly contribute in the ratio of 1:3 to the total revenue loss.

I next explore the role of invoice mills in the observed noncompliance. I find that they are an important channel through which the overclaim of refunds takes place. Almost all of their sales are booked to exporters. In 2004, for example, 97% of their sales invoices were used by exporters to claim refunds. Refund claimed on these invoices amounted to nearly 37% of all *excess* refund claimed by exporters in that year. Overall, my results confirm the concerns frequently expressed in the policy literature on the weaknesses of the VAT in emerging and transition economies. Significant amount of VAT evasion occurs in both export refunds and domestic B2C sales. Future research may take up its consequences for both optimal taxation and enforcement policies.

References

ABADIE, ALBERTO, ATHEY, SUSAN, IMBENS, GUIDO W, & WOOLDRIDGE, JEFFREY. 2017. *When should you adjust standard errors for clustering?* Tech. rept. National Bureau of Economic Research.

AGRAWAL, DAVID R, & ZIMMERMANN, LAURA. 2019. Production and evasion responses with limited state capacity. International Growth Centre Working Paper.

ALMUNIA, MIGUEL, HJORT, JONAS, KNEBELMANN, JUSTINE, & TIAN, LIN. 2019. *Strategic or Confused Firms? Evidence from "Missing" Transactions in Uganda*. Mimeo.

ARTAVANIS, NIKOLAOS, MORSE, ADAIR, & TSOUTSOURA, MARGARITA. 2016. Measuring Income Tax Evasion Using Bank Credit: Evidence from Greece. *The Quarterly Journal of Economics*, **131**(2), 739–798.

ATKINSON, ANTHONY BARNES, & STIGLITZ, JOSEPH E. 1976. The design of tax structure: direct versus indirect taxation. *Journal of public Economics*, **6**(1-2), 55–75.

BACHAS, PIERRE, & SOTO, MAURICIO. 2019. Not (ch) your average tax system: corporate taxation under weak enforcement. World Bank Policy Research Working Paper.

BERTRAND, MARIANNE, DUFLO, ESTHER, & MULLAINATHAN, SENDHIL. 2004. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, **119**(1), 249–275.

BIRD, RICHARD, & GENDRON, PIERRE-PASCAL. 2007. *The VAT in Developing and Transitional Countries*. Cambridge University Press.

BROCKMEYER, ANNE, SMITH, SPENCER, HERNANDEZ, MARCO, & KETTLE, STEWART. 2019. Casting a Wider Tax Net: Experimental Evidence from Costa Rica. *American Economic Journal: Economic Policy*, **11**(3), 55–87.

CARRILLO, PAUL, POMERANZ, DINA, & SINGHAL, MONICA. 2017. Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. *American Economic Journal: Applied Economics*, **9**(2), 144–64.

EBRILL, LIAM P, KEEN, MICHAEL, BODIN, JEAN-PAUL, SUMMERS, VICTORIA P, *et al.* 2001. *The modern VAT*. International Monetary Fund.

FBR. 2005. CBR Quarterly Review. **5**(1). Federal Board of Revenue, Pakistan, https://tinyurl.com/y4tuxp3q.

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

GOP. 2004. *Economic Survey of Pakistan, 2004-05*. Government of Pakistan, Finance Division.

GORODNICHENKO, YURIY, MARTINEZ-VAZQUEZ, JORGE, & PETER, KLARA SABIRIANOVA. 2009. Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effects in Russia. *Journal of political economy*, **117**(3), 504–554.

IMBENS, GUIDO W, & RUBIN, DONALD B. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

INTERNATIONAL TAX DIALOGUE. 2013. International Tax Dialogue, Key Issues and Debates in VAT, SME Taxation and the Tax Treatment of the Financial Sector. *International Tax Dialogue*.

KEEN, MICHAEL. 2007. VAT attacks! *International Tax and Public Finance*, **14**(4), 365–381.

KEEN, MICHAEL, & SMITH, STEPHEN. 2006. VAT fraud and evasion: What do we know and what can be done? *National Tax Journal*, 861–887.

KLEVEN, HENRIK J., 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.

NARITOMI, JOANA. 2019. Consumers as tax auditors. *American Economic Review*, **109**(9), 3031–72.

OECD. 2017. Revenue Statistics, Organization for Economic Cooperation and Development. URL, https://goo.gl/WC1M1L.

POMERANZ, DINA. 2015. No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review*, **105**(8), 2539–2569.

SHAH, JAWAD A. 2019. *Using Computerized Information to Enforce VAT: Evidence from Pakistan*. Mimeo.

SLEMROAD, JOEL, & VELAYUDHAN, TEJASWI. 2020 (February). *The VAT at 100: A Comprehensive Health Assessment, With a Plan of Care and Suggestions for Diagnostic Tests*. University of Michigan Working Paper.

SLEMROD, JOEL. 2007. Cheating Ourselves: The Economics of Tax Evasion. *Journal of Economic Perspectives*, **21(1)**, 25–48.

SLEMROD, JOEL, REHMAN, OBEID UR, & WASEEM, MAZHAR. 2019. Pecuniary and non-pecuniary motivations for tax compliance: Evidence from Pakistan.

STIGLITZ, JOSEPH E. 2010. Development-oriented tax policy. *Taxation in developing countries: Six case studies and policy implications*, 11–36.

WASEEM, MAZHAR. 2018a. Taxes, Informality and Income Shifting: Evidence from a Recent Pakistani Tax Reform. *Journal of Public Economics*, **157**, 41–77.

WASEEM, MAZHAR. 2018b. Using Movement of Exemption Cutoff to Estimate Tax Evasion: Evidence from Pakistan. Working Paper, University of Manchester.

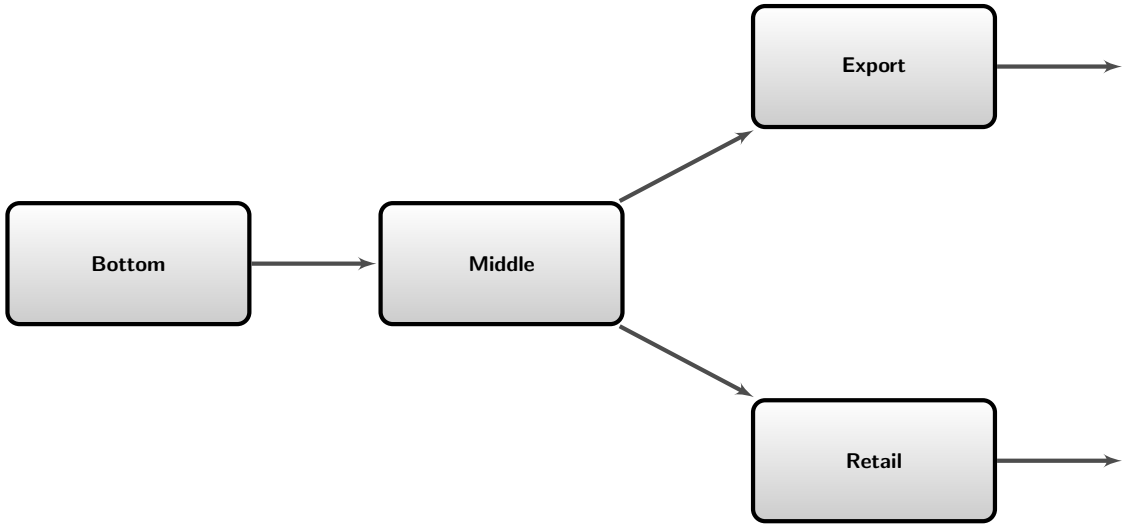
WASEEM, MAZHAR. 2019a. Does Cutting the Tax Rate to Zero Induce Behavior Different from Other Tax Cuts? Evidence from Pakistan. Working Paper, University of Manchester.

WASEEM, MAZHAR. 2019b. Information, Asymmetric Incentives, or Withholding: Understanding the Self-Enforcment of Value-added Tax. Working Paper, University of Manchester.

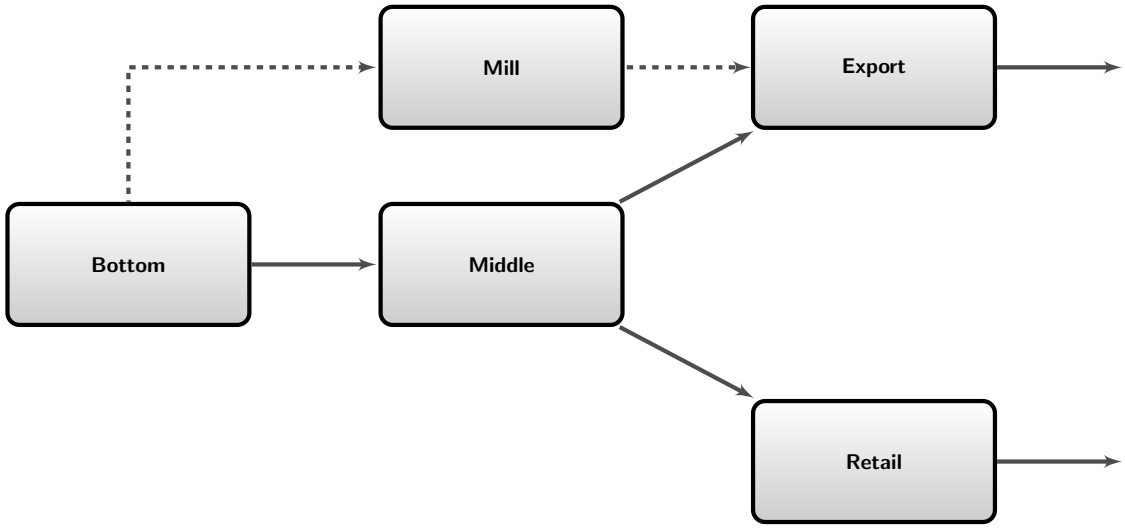
WASEEM, MAZHAR. 2020. The Role of Withholding in the Self-Enforcement of a Value-Added Tax: Evidence from Pakistan. Working Paper, University of Manchester.

FIGURE I: **INPUT OUTPUT LINKAGES**

A: Real Supply Chain

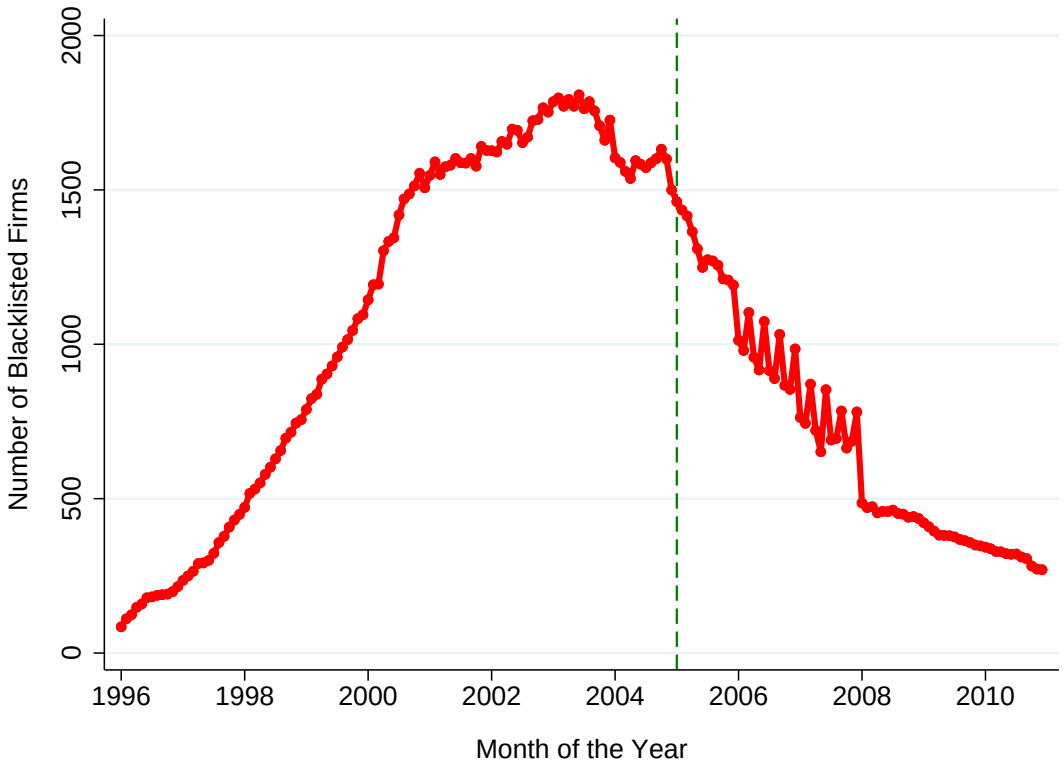


B: Supply Chain With an Invoice Mill



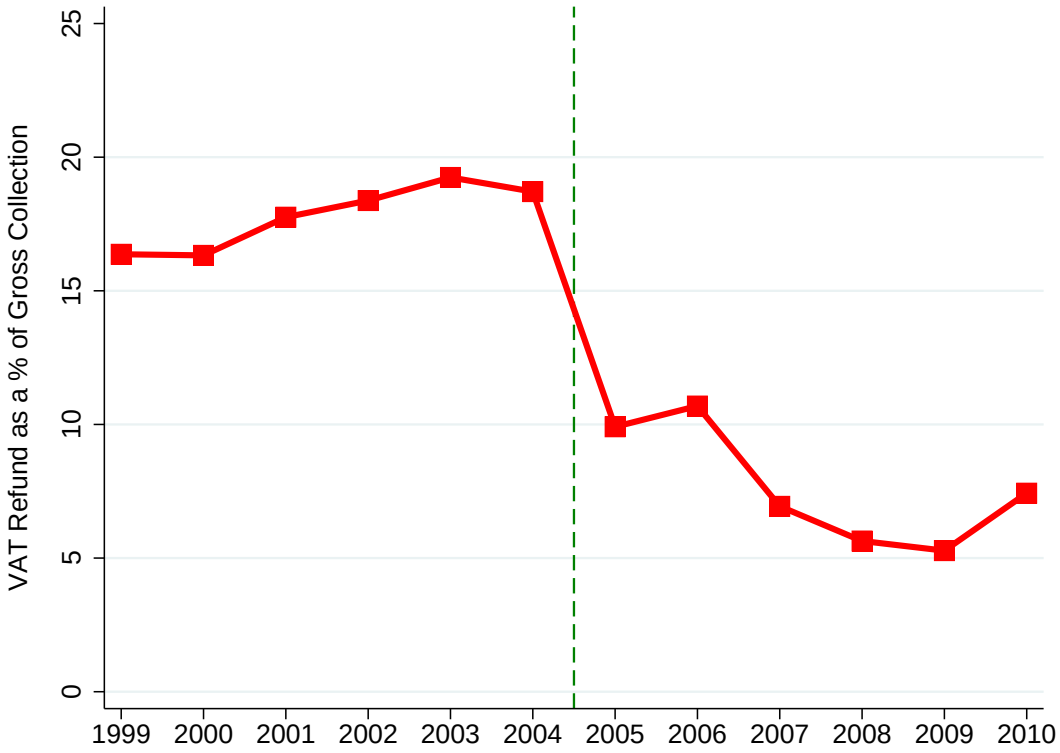
Notes: This figure displays a typical supply chain comprising three production stages. The top panel displays the *real* input output linkages between firms, showing the path the consumption good travels before it is exported or consumed in the domestic market. The bottom panel displays the role of an Invoice Mill in this setting. The Invoice Mill acts as a conduit between the exporter and the bottom-tier firm, transmitting the VAT credit to the exporter. The dashed line, thus, shows the path on which the VAT invoice travels.

FIGURE II: **EMERGENCE, GROWTH, AND DECLINE OF INVOICE MILLS**



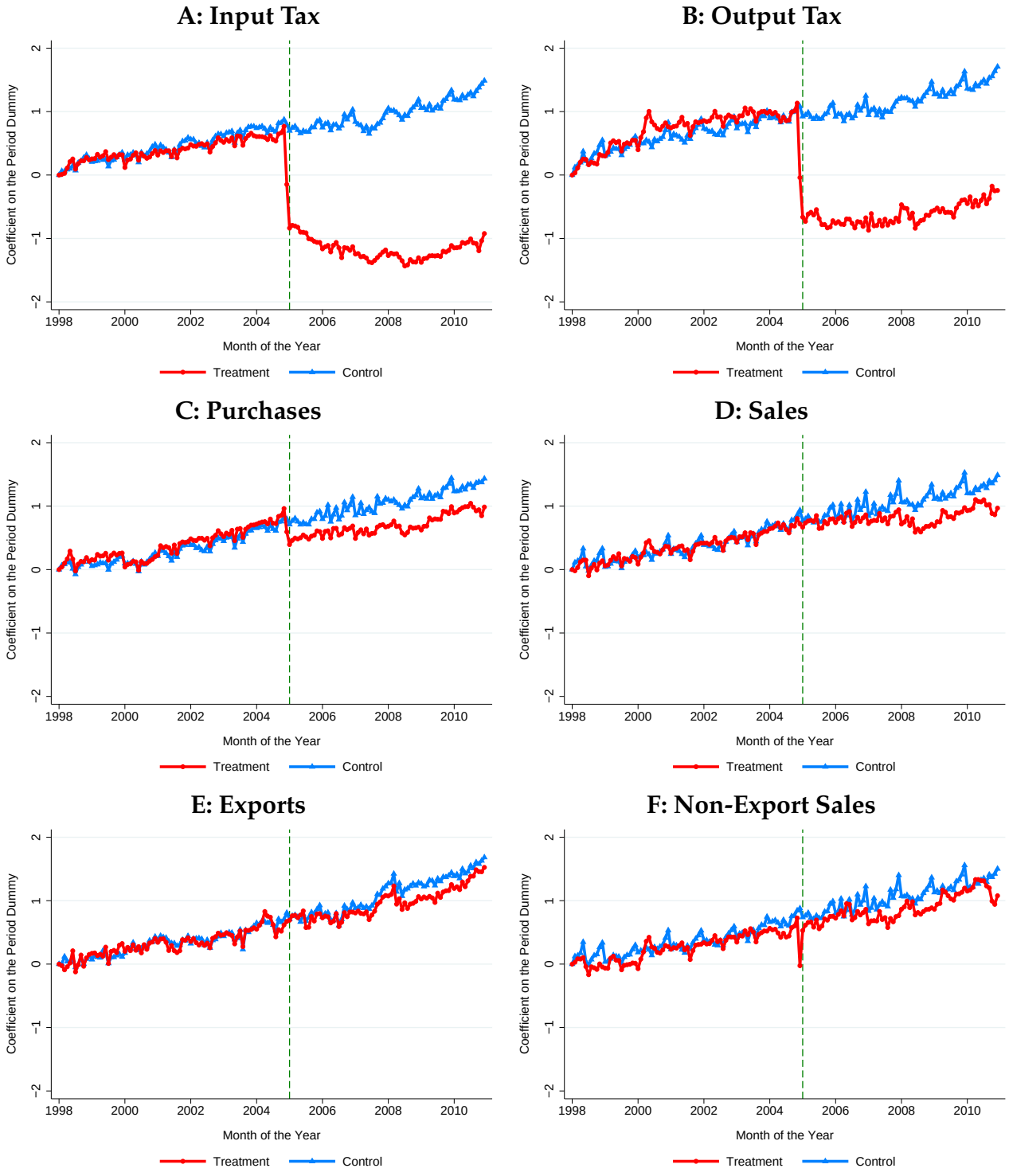
Notes: The figure shows the emergence, growth, and subsequent decline of invoice mills in the Pakistani setting. The sample begins from 1996, when a *broad-based* VAT with coverage extended to almost entire manufacturing and exports stages begun in the country. Each marker in the curve denotes the number of blacklisted firms that file a return in the given month. The year t in the horizontal axis denotes the month July of year t. The dashed, vertical line represents the time from which the zero-rating reform became applicable.

FIGURE III: **REFUND AS A PROPORTION OF GROSS COLLECTION**



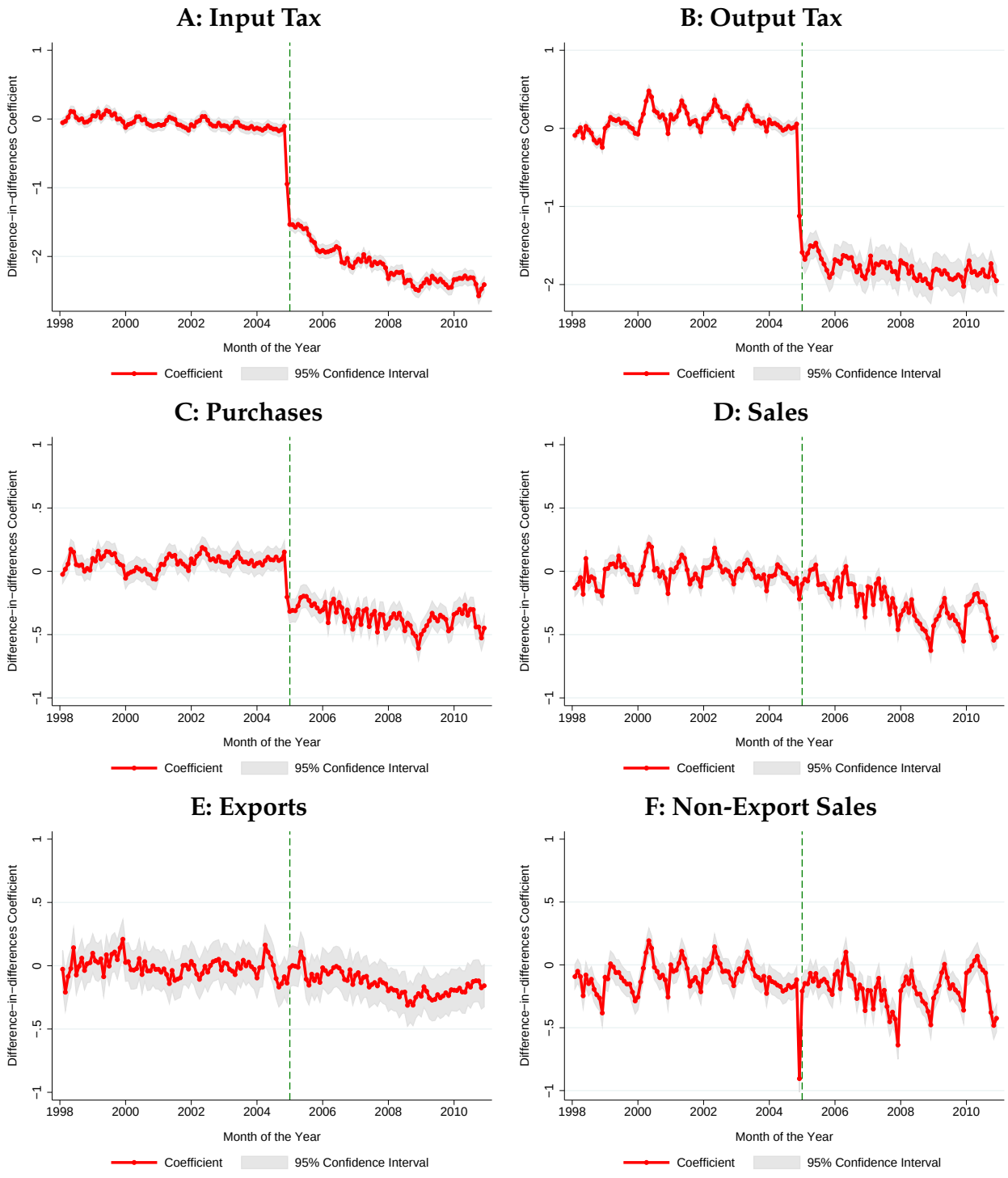
Notes: The figure shows the effects of the zero-rating reform on the VAT refund paid in Pakistan. Each marker in the plot denotes the aggregate VAT refund paid by the FBR to firms in *all* industries as a percentage of the gross VAT collection in that year. The data used for this plot are publicly available and have been compiled from the FBR yearbooks, containing annual tax collection statistics. These yearbooks are available here. The data are available from the tax year 1999 only. The year t in the horizontal axis denotes the month July of year t. The dashed, vertical line represents the time from which the zero-rating reform became applicable.

FIGURE IV: **FIRM BEHAVIOR TO THE TAX CUT**



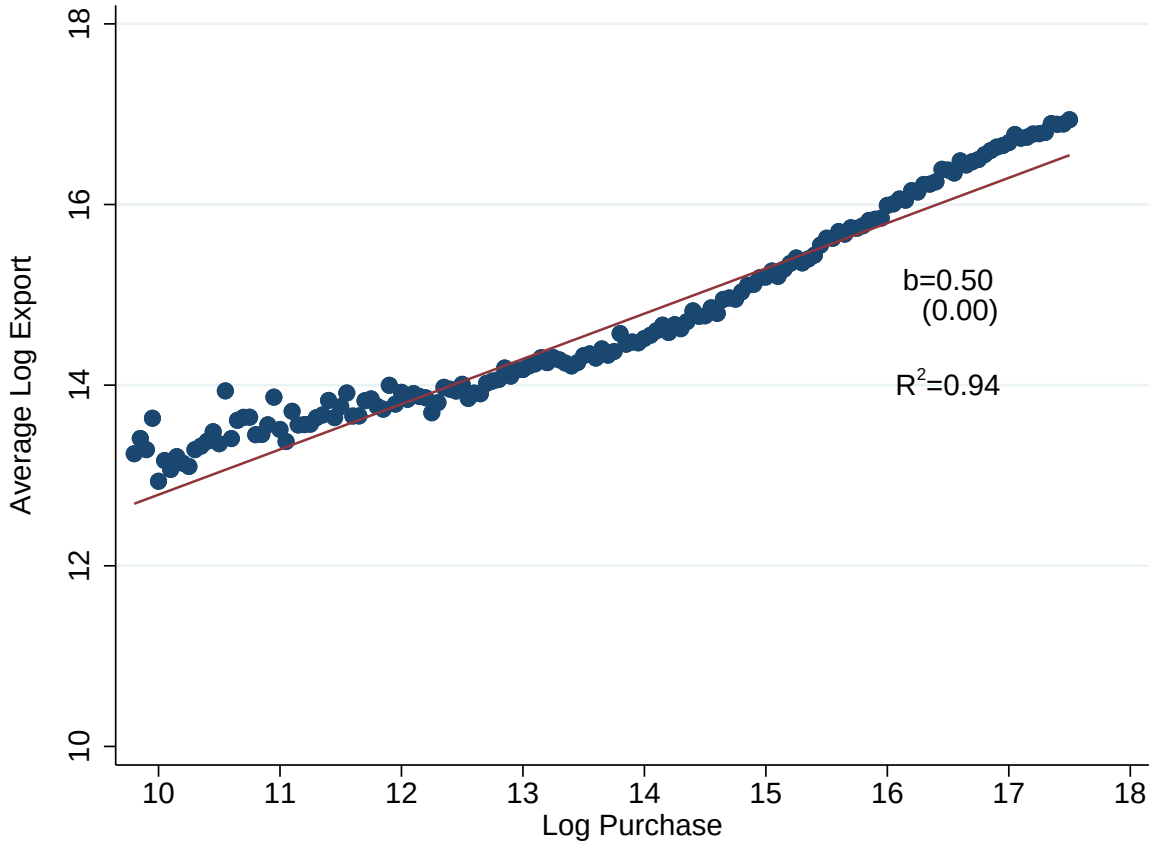
Notes: The figure compares the evolution of six VAT outcomes from the tax year 1998 to 2010 across the treatment and control groups . Treatment and control groups consist respectively of firms in the zero-rated and non-zero-rated industries. To construct these charts, I regress the log of the outcome variable shown in the title of each panel on the full set of firm and month fixed effects, dropping the dummy for July 1998. I then plot the coefficients on the time dummies of these regressions. The regressions are run separately for the two groups of firms. Year t on the horizontal axis indicates July of the corresponding year. Vertical dashed lines demarcate the time from which the zero-rating reform became applicable.

FIGURE V: **FIRM BEHAVIOR TO THE TAX CUT – DIFFERENCE-IN-DIFFERENCES**



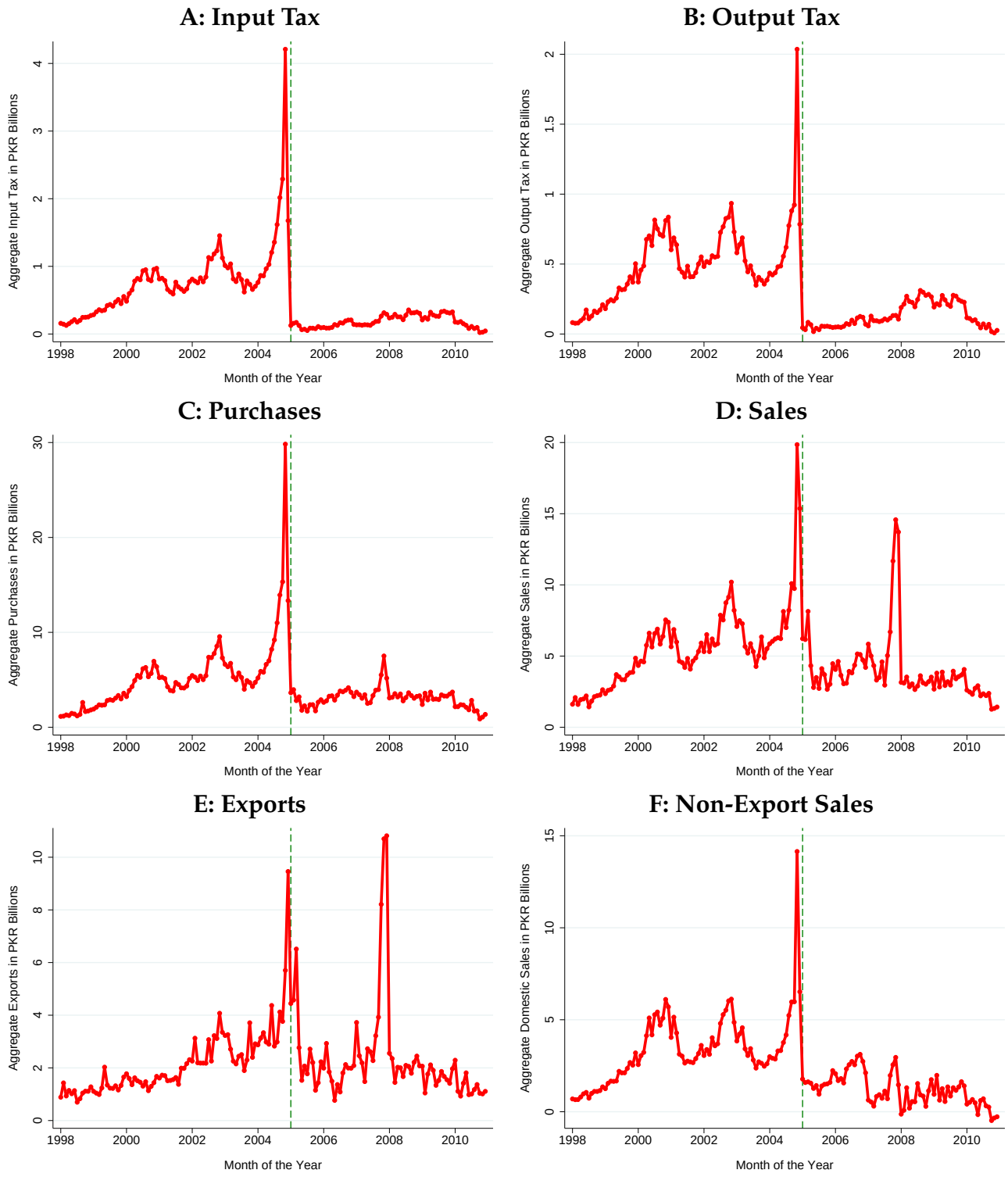
Notes: The figure shows the difference-in-differences version of the plots in Figure IV. To construct these charts, I regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, and month×treat dummies, dropping the dummies for July 1998. I then plot the coefficients on the month×treat dummies from these regressions, where treat_i denotes that firm i belongs to a zero-rated industry. The gray surface plot shows the 95% confidence interval around the coefficient. I cluster standard errors at the firm level. Year t on the horizontal axis indicates July of the corresponding year. The vertical, dashed lines demarcate the time from which the zero-rating reform became applicable.

FIGURE VI: **EXPORTS AS A FUNCTION OF PURCHASES**



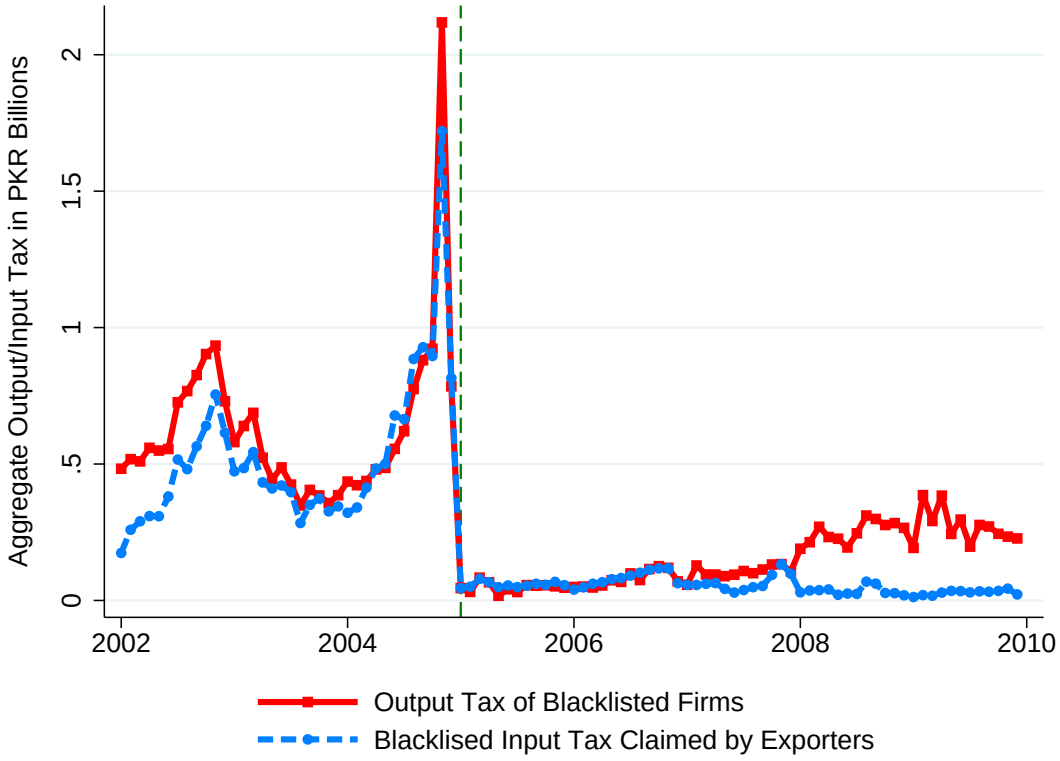
Notes: The figure explores the relationship between purchases of intermediates and exports reported by firms of the treated industries in the baseline years (1998-2004). The blue curve shows a binned scatter plot for the data. I group firms into bins on the basis of log of purchases reported by them. Each blue marker represents the average log export of firms within the bin. I restrict the sample to firms with log purchases within the fifth and 95th percentile of the aggregate log purchase distribution. The bin width is 0.05. The red curve is the straight line fitted into the data using ordinary least squares. The slope of the fitted line and R^2 from the regression are reported in the panel.

FIGURE VII: **AGGREGATE VALUES OF VAT OUTCOMES – INVOICE MILLS**



Notes: The figure illustrates how VAT outcomes evolve around the time of the zero-rating reform. The sample contains blacklisted and suspended firms operating in the zero-rated industries. Each panel of the figure shows the aggregate value of the outcome indicated in the title of the panel for the given month. Year t indicated in the horizontal axis denotes the month July of the corresponding year. To deal with outliers, I drop ten observations with the highest values of the given outcome in the entire sample. For example, for constructing Panel A, I sort all firm-month observations on the basis of Output Tax in a descending order and drop the top-ten observations. The dashed, vertical lines in the plots demarcate the time from which the zero-rated reform takes effect.

FIGURE VIII: **INVOICE MILLS AND VAT REFUND**



Notes: The figure explores the linkages between invoice mills and exporters. The solid, red curve in the figure plots the aggregate value of output tax involved in VAT returns filed by all blacklisted and suspended firms in the given tax period (month). The dashed, blue curve, on the other hand, shows the aggregate value of input tax claimed by exporters on the invoices of blacklisted firm for the given tax period (month). This curve has been plotted using transaction-level data filed by exporters in support of their refund claims, which provides the supplier-wise details of all purchases of intermediates made by them in the corresponding tax period. That these two curves almost lie above each other up to the year 2008 shows that to some extent the sole purpose of the existence of invoice mills is to abet the claim of exaggerated refunds by exporters. Year t indicated in the horizontal axis denotes the month July of the year. Dashed vertical lines in the plots demarcate the time from which the zero-rated reform takes effect.

TABLE I: **INPUT OUTPUT LINKAGES UNDER THREE STYLIZED EXAMPLES**

	Bottom	Middle	Invoice Mill	Export	Retail
	(1)	(2)	(3)	(4)	(5)
Example 1: Evasionless Benchmark					
Purchases:	-	$c_M = s_B$	-	$c_E = \alpha s_M$	$c_R = (1-\alpha)s_M$
Sales:	s_B	s_M	-	s_E	s_R
VAT Liability:	τs_B	$\tau(s_M - c_M)$ $= \tau(s_M - s_B)$	-	$-\tau c_E$ $= -\tau\alpha s_M$	$\tau(s_R - c_R)$ $= \tau[s_R - (1-\alpha)s_M]$
Example 2: Evasion Without Invoice Mill					
Purchases:	-	$c_M = s_B$	-	$\uparrow\ \hat{c}_E = \hat{\alpha} s_M$	$\downarrow\ \hat{c}_R = (1-\hat{\alpha})s_M$
Sales:	s_B	s_M	-	$\uparrow\ \hat{s}_E$	$\downarrow\ \hat{s}_R$
VAT Liability:	τs_B	$\tau(s_M - c_M)$ $= \tau(s_M - s_B)$	-	$\uparrow\ -\tau\hat{c}_E$ $\uparrow = -\tau\hat{\alpha}\hat{s}_M$	$\downarrow\ \tau(\hat{s}_R - \hat{c}_R)$ $\downarrow = \tau[\hat{s}_R - (1-\hat{\alpha})\hat{s}_M]$
Example 3: Evasion With Invoice Mill					
Purchases:	-	-	$\hat{c}_M = \hat{s}_B$	$\uparrow\ \hat{c}_E = \hat{\alpha} s_M$	$\downarrow\ \hat{c}_R = (1-\hat{\alpha})s_M$
Sales:	s_B	-	$\hat{s}_M$	$\uparrow\ \hat{s}_E$	$\downarrow\ \hat{s}_R$
VAT Liability:	τs_B	-	$\tau(\hat{s}_M - \hat{c}_M)$ $= \tau(\hat{s}_M - \hat{s}_B)$	$\uparrow\ -\tau\hat{c}_E$ $\uparrow = -\tau\hat{\alpha}\hat{s}_M$	$\downarrow\ \tau(\hat{s}_R - \hat{c}_R)$ $\downarrow = \tau[\hat{s}_R - (1-\hat{\alpha})\hat{s}_M]$

Notes*:* The table shows the input-output linkages across firms in the supply chain shown in Figure I under three alternative scenarios. Panel A shows the linkages under the first-best scenario, where all firms are registered and report truthfully, and there are no invoice mills. Panel B allows tax evasion by both underreporting of sales and overreporting of input costs, but all firms are registered and there are no invoice mills. The bottom panel considers a scenario where the firm in the middle production stage is not registered and therefore cannot pass on the VAT paid at the bottom stage to the export stage. This vacuum is filled by the invoice mill, which enables the exporter to claim VAT refund by routing the VAT invoice from the bottom stage to the export stage.

TABLE II: **SUMMARY STATISTICS**

	2003		2004	
	Treatment	Control	Treatment	Control
	(1)	(2)	(3)	(4)
1. # Observations	172,321	743,281	163,327	709,560
2. Input Tax	0.686	0.431	0.707	0.639
	(0.014)	(0.017)	(0.013)	(0.031)
3. Output Tax	0.514	0.602	0.479	0.850
	(0.013)	(0.028)	(0.012)	(0.045)
4. Purchases	4.600	2.678	6.158	4.223
	(0.084)	(0.096)	(0.116)	(0.175)
5. Sales	6.919	3.782	7.776	5.623
	(0.895)	(0.122)	(0.116)	(0.202)
6. Exports	2.614	0.311	4.207	0.503
	(0.054)	(0.010)	(0.065)	(0.016)
7. Domestic Sales	4.306	3.471	3.569	5.120
	(0.893)	(0.118)	(0.082)	(0.194)
8. Major City	0.486	0.563	0.507	0.576
	(0.001)	(0.001)	(0.001)	(0.001)
9. # Years Registered	3.695	3.592	3.238	3.108
	(0.009)	(0.004)	(0.010)	(0.004)
10. # Years Active	8.083	8.608	8.401	8.883
	(0.009)	(0.004)	(0.008)	(0.004)
11. Some Export	0.472	0.135	0.545	0.152
	(0.001)	(0.000)	(0.001)	(0.000)
12. Some Import	0.416	0.369	0.476	0.415
	(0.001)	(0.001)	(0.001)	(0.001)

Notes*:* The table presents summary statistics for the treatment and control groups. Treatment group comprises firms whose supplies were zero-rated by the zero-rating reform from 2005. The control group comprises all other firms. The first row of the table compares the number of firm-month observations for the two groups in the two prereform years. Subsequent rows compare the mean of eleven VAT outcomes and firm characteristics across the two groups. Major City denotes that the firm is registered in Karachi or Lahore, the two major cities of Pakistan. The variable # Years Registered reports the number of years up to 2003 since the firm's registration; # Years Active reports the number of years the firm remained active, filing its VAT return. Standard errors of the mean are in parenthesis.

TABLE III: **FIRM BEHAVIOR TO THE TAX CUT**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
A: Complete Panel						
treat × after	-1.961	-1.842	-0.419	-0.223	-0.106	-0.082
	(0.026)	(0.071)	(0.017)	(0.012)	(0.025)	(0.014)
Observations	3,728,660	4,179,561	3,983,213	5,058,579	612,993	4,623,907
B: Balanced Panel						
treat × after	-2.337	-2.536	-0.484	-0.405	-0.158	-0.101
	(0.043)	(0.139)	(0.031)	(0.024)	(0.037)	(0.029)
Observations	948,385	877,354	981,954	1,126,539	264,719	960,697
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table explores how firms respond to the reduction of the rate applicable on their supplies and major inputs to zero. I estimate the difference-in-differences model (2). The dummy variable treat_i denotes that firm i belongs to a zero-rated industry; the dummy variable after_t denotes that month t falls in the tax year 2005 and later. Panel B restricts the sample to a balance panel, including only the firms who file at least once in every quarter included in the sample. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE IV: **FIRM BEHAVIOR TO THE TAX CUT – DYNAMICS**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
treat × 2005	-1.568	-1.732	-0.335	-0.072	-0.019	-0.046
	(0.022)	(0.062)	(0.015)	(0.010)	(0.022)	(0.013)
treat × 2006	-1.891	-1.816	-0.391	-0.126	-0.035	-0.028
	(0.027)	(0.070)	(0.017)	(0.012)	(0.027)	(0.014)
treat × 2007	-1.984	-1.855	-0.440	-0.203	-0.103	-0.220
	(0.029)	(0.088)	(0.021)	(0.016)	(0.030)	(0.022)
treat × 2008	-2.240	-1.949	-0.498	-0.383	-0.216	-0.135
	(0.033)	(0.090)	(0.021)	(0.018)	(0.033)	(0.017)
treat × 2009	-2.283	-1.956	-0.475	-0.365	-0.212	-0.083
	(0.035)	(0.090)	(0.023)	(0.019)	(0.036)	(0.018)
treat × 2010	-2.270	-1.911	-0.432	-0.295	-0.148	-0.021
	(0.036)	(0.084)	(0.024)	(0.020)	(0.038)	(0.019)
Observations	3,728,660	4,179,561	3,983,213	5,058,579	612,993	4,623,907
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table explores how firm response to the zero-rating reform evolves over time. I estimate an augmented version of the difference-in-differences model (2), including interactions of the treatment variable with all post-reform years. The dummy variable treat_i denotes that firm i belongs to a zero-rated industry. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE V: **FIRM BEHAVIOR TO THE TAX CUT – ACROSS THE TEXTILE SUPPLY CHAIN**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
treat × after	-2.040	-0.468	-0.429	-0.266	-0.108	0.092
	(0.053)	(0.074)	(0.034)	(0.027)	(0.035)	(0.033)
treat × after × ginning	2.053	-2.184	0.389	0.305	0.193	-0.001
	(0.075)	(0.096)	(0.067)	(0.036)	(0.057)	(0.042)
treat × after × spinning	-0.504	-3.613	0.110	0.030	0.210	-0.351
	(0.074)	(0.207)	(0.045)	(0.035)	(0.071)	(0.042)
treat × after × weaving	0.018	-1.831	0.001	0.026	0.025	-0.204
	(0.061)	(0.147)	(0.039)	(0.031)	(0.044)	(0.037)
Baseline Coefficient	-1.978	-1.925	-0.394	-0.221	-0.065	-0.082
	(0.026)	(0.076)	(0.017)	(0.013)	(0.027)	(0.014)
Observations	3,685,909	4,158,258	3,934,914	4,987,477	563,822	4,597,087
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table explores heterogeneity in firm-response to the zero-rating reform within the textile industry. I restrict the treatment sample to firms in the textile industry only and estimate a triple-difference version of model (2). The dummy variables $ginning_i$, $spinning_i$ and $weaving_i$ denote that the firm i belongs to the corresponding production stage within the textile industry. Baseline coefficient reports the $treat \times after$ coefficient I obtain from estimating the model without the triple-interaction terms. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE VI: **CALCULATING VAT EVASION**

	Complete Panel (1)	Balanced Panel (2)
A: Overclaimed Refunds		
1. Percent Decrease in $\hat{S}_E$	0.112	0.171
	(0.025)	(0.038)
2. $\hat{S}_E$ in 2004	687.038	389.506
3. $\Delta\hat{S}_E$	76.826	66.670
	(17.392)	(14.682)
4. $\Delta\hat{C}_E$	153.652	133.340
	(34.785)	(29.363)
5. Overclaimed Refunds in 2004	23.048	20.001
	(5.218)	(4.404)
B: Underpaid Tax on B2C Sales		
6. Percent Decrease in $(\hat{S} - \hat{S}_E)$	0.087	0.111
	(0.013)	(0.029)
7. Percent Decrease in $\hat{C}$	0.522	0.626
	(0.017)	(0.031)
8. Percent Decrease in $\hat{S}_{B2B}$	0.435	0.515
	(0.022)	(0.043)
9. $\hat{S}_{B2C}$ in 2004	229.013	129.835
10. Under-reported $\hat{S}_{B2B}$ in 2004	99.717	66.877
	(4.939)	(5.595)
11. Under-paid VAT on $\hat{S}_{B2B}$ in 2004	14.957	10.032
	(0.741)	(0.839)
C: Total Evasion		
12. Total Tax Evasion	38.005	30.033
	(5.270)	(4.484)

Notes: The table computes the extent of VAT evasion as it existed in the treated industries at the baseline. The first row reports the percentage decrease in reported exports as estimated by the difference-in-differences model (see the fifth column of Table III). Total exports of the treated firms in 2004 are reported in Row 2. The third row computes the amount by which exports were over-reported at the baseline by multiplying the first two rows. Row 4 converts the overreported exports into overreported purchases used in exports by employing the relationship estimated in Figure VI. Row 5 converts overreported purchases into overclaimed refund by multiplying with the tax rate. Rows 6-11 do similar calculations to compute the underreporting of B2C sales. Row 12 computes total VAT evasion by adding rows 5 and 11. All amounts in this table are in PKR billions. To deal with outliers, I exclude ten observations with the highest values of exports and non-export sales in the entire sample (1998-2010). Column (2) contains the corresponding estimates for the balanced panel sample. Standard errors are in parenthesis.

A Online Appendix

A.1 Definition of Variables

(*i*) **Input Tax.** The value of VAT credit claimed on purchases of intermediates made by a registered firm in a given tax period (month). It equals $\tau.\hat{c}_{it}$, where τ is the applicable VAT rate and $\hat{c}_{it}$ is the value of purchases of intermediates claimed by firm i in period t.

(*ii*) **Output Tax.** The value of VAT charged on sales made by a registered firm in a given tax period (month). It equals $\tau.\hat{s}_{it}$, where τ is the applicable VAT rate and $\hat{s}_{it}$ is the value of sales reported by firm i in period t.

(*iii*) **Purchases.** The value of all taxable intermediates acquired by a firm in a given tax period (month).

(*iv*) **Sales.** The value of all goods and services supplied by a firm in a given tax period (month).

(*v*) **Exports.** The value of all goods and services exported by a firm in a given tax period (month).

(*vi*) **Non-Export Sales.** The value of all goods and services supplied by a firm minus the value of all goods and services exported by a firm in a given tax period (month).

(*vii*) **Manufacturer.** A firm whose principal business activity is the manufacture of goods. Manufacturing is the process whereby a firm converts inputs into a distinct article capable of being put to use differently than inputs and includes any process incidental or ancillary to it.

(*viii*) **Wholesaler.** Wholesaler' includes a dealer and means any person who carries on, whether regularly or otherwise, the business of buying and selling goods by wholesale or of supplying or distributing goods, directly or indirectly, by wholesale for cash or deferred payment or for commission or other valuable consideration or stores such goods belonging to others as an agent for the purpose of sale; and includes a person supplying taxable goods to a person who deducts income tax at source under the Income Tax Ordinance, 2001.

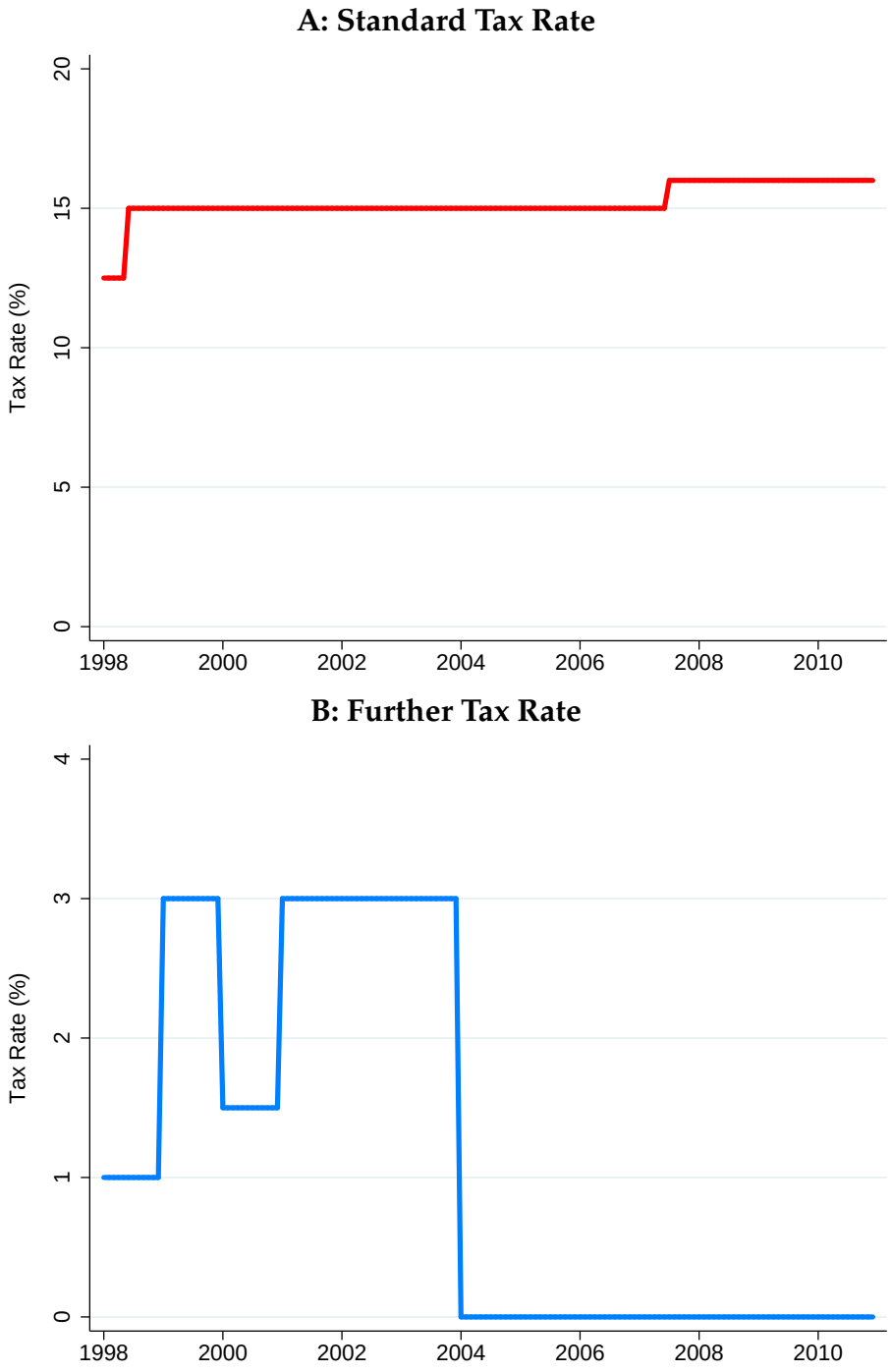
(*ix*) **Retailer.** A person, supplying goods to general public for the purpose of consumption.

(*x*) **Industry.** The Pakistani tax administration uses 4-digit Harmonized Commodity Description and Coding System (HS code) to classify firms into industry. The code, used by customs administrations throughout the world, divides all goods and services into 99 chapters (the first two digits in the code) and 21 sections. The sections broadly correspond to major industries in the country. I take the section a firm falls in as its industry. Table shows the sections, HS code, and description of these industries.

(*xi*) **Major City** A firm registered either in Karachi or Lahore, the two largest cities in Pakistan on the basis of both population and GDP.

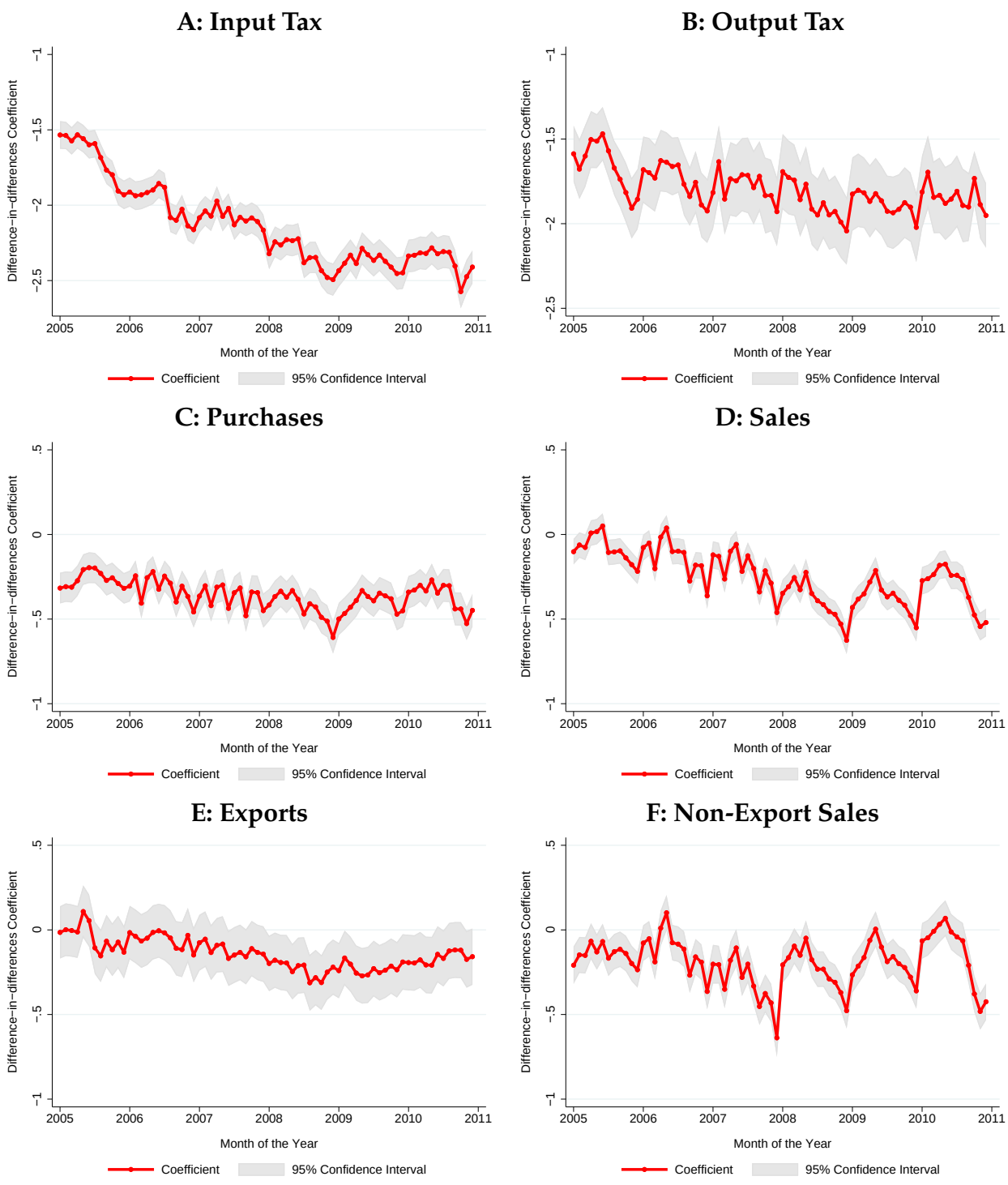
(*xii*) **Initial Capital.** The value of initial capital of the firm, as reported by it at the time of registration for VAT.

FIGURE A.I: **VAT RATES**



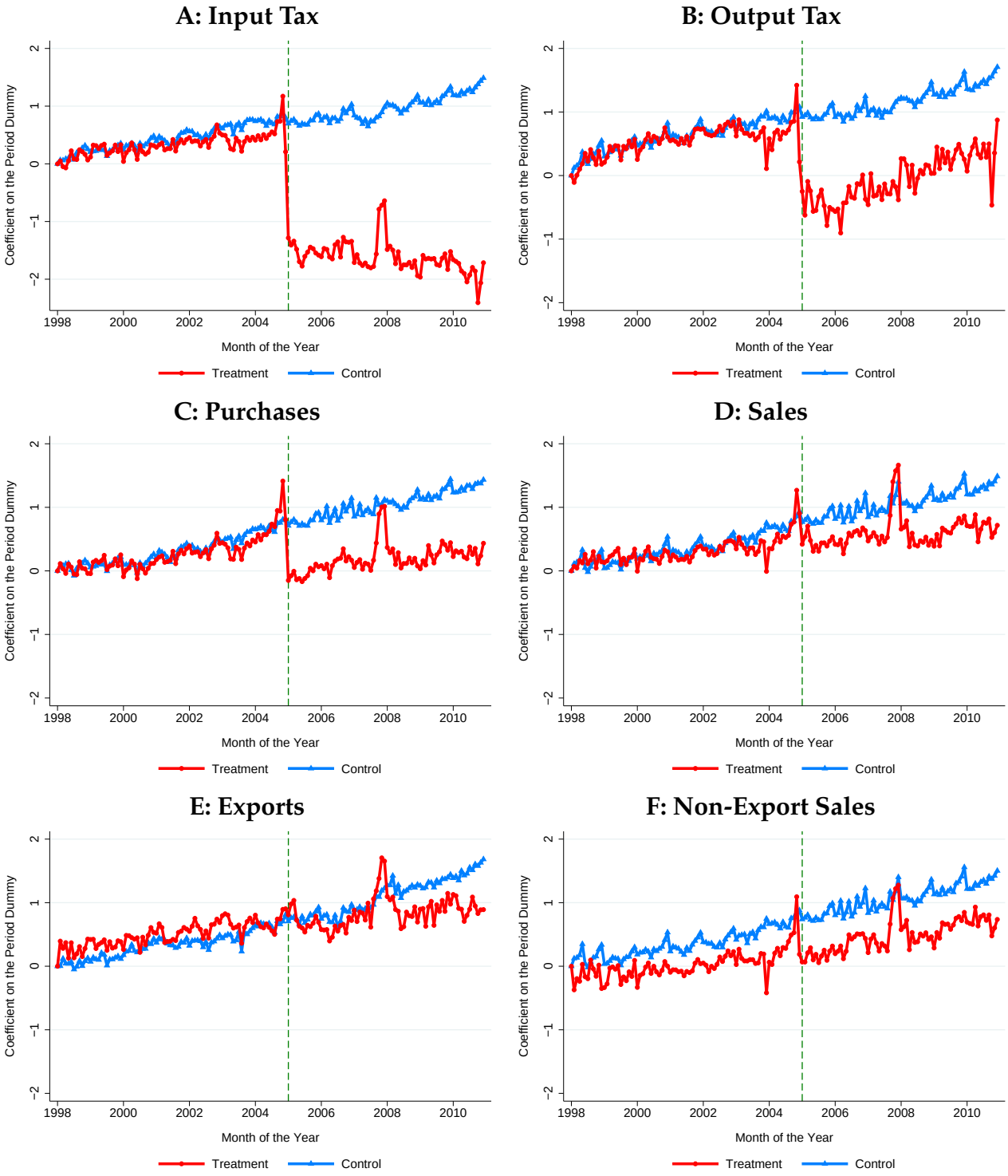
Notes: The figure shows the VAT rates applicable in Pakistan from July 1998 to June 2011. Panel A shows the standard VAT rate, which is applied to all sales made by a registered firm regardless of whether the recipient is registered or not. The rate largely stayed at 15%, but was increased to 16% from July 2008. Panel B reports the Further Tax Rate. This rate is added to the standard rate whenever the sale is made to an unregistered firm. For example, supplies made by a registered firm in July 1998 were subject to a rate of 15% if the recipient was a registered firm or an end-consumer and 16% if the recipient was an unregistered firm. To claim that a sale has been made to end-consumer, the selling firm must be registered as a retailer.

FIGURE A.II: **FIRM BEHAVIOR TO THE TAX CUT – DIFFERENCE-IN-DIFFERENCES**



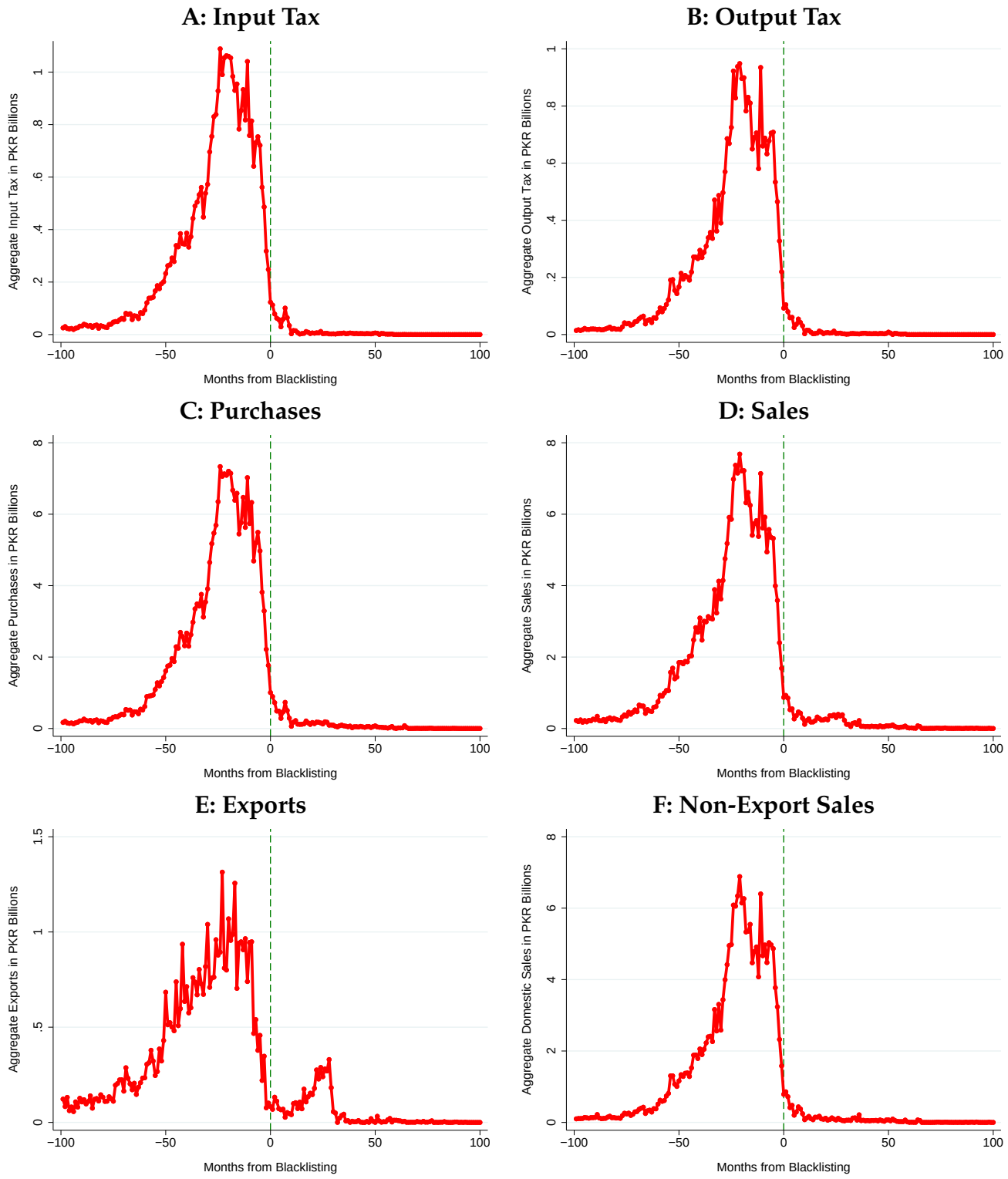
Notes: The figure reproduces the event study results reported in Figure IV. The only difference between the two figure is that I show only the post-reform periods here. To construct these charts, I regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, and month×treat dummies, dropping the dummies for July 1998. I then plot the coefficients on the month×treat dummies from these regressions, where treat_i denotes that firm i belongs to a zero-rated industry. The gray surface plot shows the 95% confidence interval around the coefficient. I cluster standard errors at the firm level. Year t on the horizontal axis indicates July of the corresponding year.

FIGURE A.III: **FIRM BEHAVIOR TO THE TAX CUT – INVOICE MILLS**



Notes: The figure compares the evolution of six VAT outcomes from the tax year 1998 to 2010 across the treatment and control groups . The treatment group here consist of blacklisted and suspended firms in the zero-rated industries only. The control group, as earlier, comprises all firms of the non-zero-rated industries. To construct these charts, I regress the log of the outcome variable shown in the title of each panel on the full set of firm and month fixed effects, dropping the dummy for July 1998. I then plot the coefficients on the time dummies of these regressions. The regressions are run separately for the two groups of firms. Year t on the horizontal axis indicates July of the year. The vertical, dashed lines demarcate the time from which the zero-rating reform became applicable.

FIGURE A.IV: **AGGREGATE VALUES OF VAT OUTCOMES – INVOICE MILLS**



Notes: The figure explores the change in activity carried out by blacklisted firms around the event of blacklisting. The sample contains all blacklisted and suspended firms of the treated industries. Time 0 in the horizontal axis (marked by the vertical, dashed line) denotes the month in which the firm was declared blacklist or its registration was suspended. Each panel of the figure shows the aggregate value of the outcome in 100 months prior to and 100 months after the event of blacklisting. To deal with outliers, I exclude ten observations with the highest values of the given outcome in the entire sample. For example, for constructing Panel A, I sort all firm-month observations on the basis of Output Tax in a descending order and exclude the top-ten observations.

TABLE A.I: **FIRM BEHAVIOR TO THE TAX CUT – TEXTILE VS. OTHERS**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
treat × after	-1.978	-1.925	-0.394	-0.221	-0.065	-0.082
	(0.026)	(0.076)	(0.017)	(0.013)	(0.027)	(0.014)
treat × after × Non-textile	0.260	0.958	-0.403	-0.023	-0.224	-0.001
	(0.102)	(0.196)	(0.078)	(0.040)	(0.042)	(0.081)
Baseline Coefficient	-1.961	-1.842	-0.419	-0.223	-0.106	-0.082
	(0.026)	(0.071)	(0.017)	(0.012)	(0.025)	(0.014)
Observations	3,728,660	4,179,561	3,983,213	5,058,579	612,993	4,623,907
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table decomposes the average response to the zero-rating reform estimated in Table III into its constituent textile and non-textile components. I estimate a triple-differences version of model (2), including all double interactions of the binary variables treat_i, Non-textile_i, and after_t and their triple-interaction. The dummy variable treat_i denotes that firm i belongs to a zero-rated industry; the dummy variable Non-textile_i denotes that the treated firm i does not belong to the textile industry; and the dummy variable after_t denotes that month t falls in the tax year 2005 and later. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.II: **SPILLOVER EFFECTS ON THE NON-TREATED INDUSTRIES**

Industries Within:	Next Two Digits			Next Five Digits			Next Ten Digits		
	Purchases	Sales	Exports	Purchases	Sales	Exports	Purchases	Sales	Exports
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A: Complete Panel									
treat × after	-0.234	0.002	0.067	-0.059	0.038	0.012	-0.066	0.012	0.044
	(0.121)	(0.083)	(0.172)	(0.082)	(0.062)	(0.128)	(0.022)	(0.019)	(0.077)
Observations	3,123,769	4,075,853	312,320	3,123,769	4,075,853	312,320	3,123,769	4,075,853	312,320
B: Balanced Panel									
treat × after	-0.152	-0.018	0.430	0.048	0.041	0.187	-0.100	0.014	0.072
	(0.168)	(0.130)	(0.319)	(0.121)	(0.108)	(0.228)	(0.037)	(0.031)	(0.119)
Observations	764,271	892,064	133,421	764,271	892,064	133,421	764,271	892,064	133,421
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table explores if the reform produces any spillovers on the nontreated industries. I drop all industries zero-rated by the reform from the sample. I then estimate the difference-in-differences model (2). The dummy variable treat_i here denotes that firm i belongs to an industry indicated in the title of each column. For example, the first three columns regard the two 2-digit industries immediately succeeding the zero-rated ones as treated. Panel B restricts the sample to a balance panel, including only the firms who file at least once in every quarter included in the sample. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.III: **FIRM BEHAVIOR TO THE TAX CUT – CORPORATIONS VS. OTHERS**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
treat × after	-1.745	-1.135	-0.407	-0.148	-0.099	-0.056
	(0.030)	(0.059)	(0.018)	(0.013)	(0.028)	(0.014)
treat × after × corporation	-0.599	-1.973	-0.040	-0.271	-0.017	-0.099
	(0.050)	(0.154)	(0.037)	(0.027)	(0.037)	(0.033)
Baseline Coefficient	-1.961	-1.842	-0.419	-0.223	-0.106	-0.082
	(0.026)	(0.071)	(0.017)	(0.012)	(0.025)	(0.014)
Observations	3,728,660	4,179,561	3,983,213	5,058,579	612,993	4,623,907
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table rules out lazy reporting as an alternative explanation of the responses documented in III. I divide the treatment sample into corporate and non-corporate firms. I then estimate a triple-differences version of model (2), including all double interactions of the binary variables treat_i, corporation_i, and after_t and their triple-interaction. The dummy variable treat_i denotes that firm i belongs to a zero-rated industry; the dummy variable corporation_i denotes that the treated firm i is a corporation; and the dummy variable after_t denotes that month t falls in the tax year 2005 and later. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.IV: **FIRM BEHAVIOR TO THE TAX CUT – BLACKLISTED VS. OTHERS**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
treat × after	-1.948	-1.870	-0.401	-0.218	-0.095	-0.077
	(0.026)	(0.074)	(0.017)	(0.013)	(0.026)	(0.014)
treat × after × blacklisted	0.208	0.697	-0.475	-0.306	-0.312	-0.445
	(0.210)	(0.292)	(0.125)	(0.097)	(0.229)	(0.103)
treat × after × suspended	-0.435	0.694	-0.374	-0.069	-0.218	-0.043
	(0.142)	(0.334)	(0.090)	(0.069)	(0.066)	(0.092)
Baseline Coefficient	-1.961	-1.842	-0.419	-0.223	-0.106	-0.082
	(0.026)	(0.071)	(0.017)	(0.012)	(0.025)	(0.014)
Observations	3,728,660	4,179,561	3,983,213	5,058,579	612,993	4,623,907
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table compares the responses of blacklisted and suspended firms with other treated firms. I estimate a triple-difference version of model (2), partitioning the treatment dummy into three dummies treat_i, blacklisted_i, and suspended_i. The dummy variables blacklisted_i and suspended_i denote that a treated firm i is blacklisted or its registration has been suspended. Baseline coefficient reports the $treat \times after$ coefficient I obtain from estimating the model without the triple-interaction terms. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.V: **SUMMARY STATISTICS – BLACKLISTED FIRMS**

	Blacklisted (1)	Suspended (2)	Others (3)
A: VAT Outcomes (PKR Millions)			
1. Purchases	7.043	5.781	1.875
2. Sales	4.609	5.465	2.183
3. Exports	0.649	1.187	0.388
4. Domestic Sales	3.960	4.277	1.795
5. Tax Payable	0.010	0.038	0.082
6. First Year Sales	6.105	5.636	1.484
7. First Year Purchases	6.246	6.019	1.071
8. First Year Exports	0.899	1.069	0.309
9. First Year Domestic Sales	5.206	4.566	1.175
10. First Year Tax Payable	0.013	0.024	0.053
11. Sales Minus Purchases	-3.422	0.633	0.861
12. Output Tax Minus Input Tax	-0.121	-0.152	0.105
B: Firm Characteristics			
13. Manufacturer	0.267	0.286	0.373
14. Wholesaler	0.338	0.339	0.253
15. Exporter	0.175	0.146	0.080
16. Some Export	0.236	0.262	0.160
17. Some Import	0.303	0.381	0.307
18. Company	0.062	0.092	0.092
19. Partnership	0.036	0.054	0.069
20. # Months Filed	38.329	40.580	41.450
21. # Months Active	15.572	23.662	27.763
22. Major City	0.718	0.752	0.503
23. Initial Capital	0.713	1.111	7.749

Notes: The table compares VAT outcomes and firm characteristics of blacklisted and suspended firms with other firms. Each row reports the mean value of the variable for the corresponding group of firms. The definitions of the compared variables are provided in Appendix A.1.