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

Mazhar Waseem*

April 2021

Abstract

Value-added tax has seen phenomenal expansion in recent decades. Its appeal largely lies in its perceived robustness to tax evasion. But VAT's evasion has not been examined systematically so far. Exploiting a reform that cuts the tax rate from 15% to 0%, I estimate the size and nature of VAT evasion in Pakistan, finding that it ranges from 31–46% of the potential revenue. One important channel through which the evasion occurs is the overclaim of refunds, which constitute 11-23% of the potential revenue. Roughly two-fifths of the overclaimed refund is based on spurious invoices issued by invoice mills. Qualitatively, noncompliance is stronger in the latter stages of the supply chain, but it runs deep inside the chain, suggesting that on their own the self-enforcement mechanisms built into a VAT do not deter tax evasion fully.

Keywords: VAT, Tax Evasion, Firm Behavior

JEL Classification: H25, H26, H32

*I thank Miguel Almunia, Zareh Asatryan, Pierre Bachas, Anne Brockmeyer, Anders Jensen, Florian Neumeier, Juan Carlos Suarez Serrato, Joel Slemrod, Christian Traxler, David E. Wildasin for helpful insights. I also thank the seminar participants at the University of Manchester, the National Tax Association, the University of Freiberg, the Institute of Fiscal Studies, the University of Mannheim, Economic Research South Africa, CESifo Public Economics, RIDGE Annual Conference, and International Online Public Finance Seminar Series for helpful suggestions and comments. Financial support from the Economic and Social Research Council (ESRC), UK's Future Research Leaders Program is gratefully acknowledged. Email Address: Mazhar Waseem mazhar.waseem@manchester.ac.uk.

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 income tax in both rich and emerging economies. In fact, its share in government revenue has been rising steadily in recent decades (OECD, 2017; International Tax Dialogue, 2013). This apparent discord between theory and practice in large part reflects enforcement concerns: it is argued that among the class of production-efficient tax instruments the VAT is perhaps the most robust to tax evasion (Slemrod & Velayudhan, 2020). This belief has underpinned VAT's remarkable expansion 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 nearly did 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, 2020b). This last-mile problem 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 provides another channel through which government revenue can be misappropriated, a channel not present in alternative tax instruments.

While theoretical mechanisms underlying VAT noncompliance are 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, 2020a). This paper fills

[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.

this gap in literature. I leverage a novel tax reform from Pakistan to uncover the size and nature of VAT noncompliance in a representative emerging economy.

The reform I exploit reduced the tax rate on five major industries of the country from 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 reform reduced the rate applicable to both intermediates purchased and final goods sold by the treated industries to zero, thus eliminating the incentive to misreport. With the tax rate going to zero, profit maximizing firms must begin reporting the true level of activity to the government as evasion is costly but offers no benefits. This allows me to infer the extent of evasion as it existed at the baseline. I do so using a simple difference-in-differences framework, comparing outcomes across firms in the treated and untreated industries. Identification in this setup presumes that the trajectory 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. I run many robustness checks to rule out other identification concerns, including any spillovers between the two groups.

In a first-best world where firms report truthfully, reducing the VAT rate to zero should have no bearing on activity reported by them other than some positive, real effects, arising for example from the liquidity channel. I, however, find that reported activity falls sharply in the treatment group after the tax cut: purchases fall by 42 log-points, sales by 22 log-points, exports by 11 log-points, and non-export sales by 8 log-points. Not only are the elasticities implied by these responses of opposite sign, they are large in size, ranging from -0.5 to -2.6. Using a simple conceptual framework, I show that the size and direction of these changes is consistent with, and can only be reconciled by, the presence of large misreporting at the baseline. I back out the level of tax evasion implied by this misreporting, finding that it ranges between 31% and 46% of the potential revenue in the four baseline years. A large proportion of this evasion occurs through the channel of overclaimed refunds, which amount to 11–23% of the potential revenue. Qualitatively, the noncompliance is stronger at the later stages of the supply chain. But it runs deep inside the chain, where transactions are largely subject to third-party information, suggesting that the self-enforcement mechanisms built into a VAT on their own do not deter tax evasion fully.

In developing economies, VAT supply chains are rarely complete. When a 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 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. Invoice mills are a poorly understood phenomenon. There is little micro-based evidence in the existing literature on how they operate and how much revenue loss they cause.[2] I show that invoice mills are a key mechanism through which the overclaim of VAT refund takes place. On average, around 83% of their sales are booked to exporters, who in turn use these to claim money back from the government. I report above that in the baseline year the amount overclaimed by exporters was 23% of the potential revenue. Roughly 37% of this amount was expropriated using invoices issued by invoice mills.

Stiglitz (2010) argues that devising evasion-resistant tax structures must be the principal aim of the tax policy in developing economies. Understanding the nature and size of tax evasion is the first step toward devising such tax structures. The primary contribution of this paper is to do so in the context of an emerging-economy VAT. In this effort, the paper contributes to a developing literature that studies the enforcement properties of VAT in weak enforcement settings (see for example Pomeranz, 2015; Agrawal & Zimmermann, 2019; Almunia *et al.*, 2019; Shah, 2019; Waseem, 2020b). This paper contributes to two other literatures as well. First, the traditional public finance theory has been developed keeping in view the compliance and information environments of high-income economies. There is a growing realization now that the policy predictions of these models do not generalize to low-compliance, high-informality settings of emerging economies (Brockmeyer *et al.*, 2019b; Best *et al.*, 2015; Basri *et al.*, 2019). To adapt the optimal tax theory to these settings, one needs to take into account how the behavior of economic agents there differs from the developed world. This paper does so in the context of VAT compliance, introducing in the process new economic agents—invoice mills—whose behavior has not been studied so far. Finally, this paper adds to a 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.*, 2019a; Slemrod *et al.*, 2020).

[2] Mittal *et al.* (2018) develop a machine learning algorithm to identify bogus VAT firms. They implement their algorithm using VAT returns from the National Capital Territory of Delhi, India. The focus of this paper goes beyond identifying to documenting the behavior of invoice mills, in particular their role in facilitating noncompliance.

II Conceptual Framework

This section develops a simple framework that highlights key mechanisms through which VAT evasion occurs and derives a formula to compute its magnitude.

II.A Setup

Consider a supply chain consisting of J production stages indexed by $j \in 1, 2, ..., J$. Figure I shows three consecutive production stages of the supply chain, illustrating inputs and outputs of firms along with their tax liability. For simplicity, I ignore firm interactions within a production stage, assuming that it comprises a representative firm only. The firm in a given stage j uses intermediates c_j acquired from the preceding stage $j-1$ to produce output valuing s_j. Out of this output, an amount $s_{E,j}$ is exported out of the country, $s_{I,j}$ is supplied to the next production stage as intermediate, and $s_{F,j}$ is sold to the final consumer. The government implements the standard destination-based VAT, whereby exports are zero-rated, imports and domestic supplies are taxed at the standard rate τ, and a full credit of VAT paid on intermediates is given. Note that purchases or some types of sales of firms in the middle stages could be zero, which means that the structure I use is general, incorporating all types of firm interactions into it. Throughout this section, I focus solely on firms' reporting behavior. I show later that the reform does not affect firms' real production decisions significantly so that abstracting away from them is not too restrictive.

II.B First-Best Benchmark

Under the first-best benchmark, firms report their sales and purchases truthfully. Panel A of the figure illustrates the input-output linkages between firms under this scenario. In this case, the government receives a total revenue of

$$R \equiv \sum_j \tau.\left(s_j - s_{E,j} - c_j\right) \tag{1}$$

from the supply chain. The VAT remitted by each firm is the difference between its output tax $\tau.(s_j - s_{E,j})$ and input tax $\tau.c_j$, and the total government revenue R is the sum of VAT revenue across all firms.

II.C VAT with Weak Enforcement

I now consider a case closer to weak enforcement settings of developing economies, allowing firms to misreport their sales and purchases (see Figure I, Panel B). The sales and purchases reported by a firm are now denoted by $\hat{s}_{K,j}$; $K \in \{E, I, F\}$ and $\hat{c}_j$, distinguishing them from their true values $s_{K,j}$ and c_j. Total government revenue from the supply chain in this case is

$$\hat{R} \equiv \sum_j \tau. \left(\hat{s}_j - \hat{s}_{E,j} - \hat{c}_j \right). \tag{2}$$

This revenue could be lower than the first-best for two reasons. First, the firm may not report the true value of its sales and purchases, reducing its tax liability. Specifically, it faces a clear incentive to overreport purchases and exports $\hat{c}_j > c_j$; $\hat{s}_{E,j} > s_{E,j}$ and underreport domestic B2C sales $\hat{s}_{F,j} < s_{F,j}$. On the other hand, the incentive to misreport domestic B2B sales is ambiguous. A firm may overreport these if it is engaged in a diversion fraud: it is showing a portion of its B2C sales as B2B sales to enable its trading partner obtain excess input tax credit (see details below). Alternatively, it may underreport these sales if its trading partner intends to keep the transaction out of books. Second, the purchases of intermediates reported at a given production stage may not match the B2B sales reported at the preceding stage i.e. $\hat{c}_j \neq \hat{s}_{I,j-1}$. Such misreporting is referred to as one-sided evasion in the literature, signifying that the firm does not collude with its suppliers or buyers to achieve this evasion (see for example Pomeranz, 2015; Waseem, 2020b).[3] Comparing the government revenue in this setup with the first-best, we can write the total VAT evasion in the supply chain as

$$\Delta R \equiv R - \hat{R} \equiv \sum_j \tau. \left[(s_j - \hat{s}_j) - (s_{E,j} - \hat{s}_{E,j}) - (c_j - \hat{c}_j) \right]. \tag{3}$$

The expression within the sum here represents the VAT evaded by firm j. Summing the firm-level evasion across all firms, we obtain the aggregate VAT evasion in the supply chain.

In my empirical application, I exploit a large tax cut that reduces the standard VAT rate on five major supply chains of the country from 15% to 0%. The reduction

[3] It is a relatively crude form of evasion because it is not robust to cross-matching of sales-purchases records by the government.

of the rate to zero virtually eliminates the incentives to misreport sales and purchases, pushing them toward their true values

$$\begin{aligned} \hat{s}_{K,j}\,(\tau = 0) &\rightarrow s_{K,j} \\ \hat{c}_j\,(\tau = 0) &\rightarrow c_j. \end{aligned} \tag{4}$$

It means that we can use the reform driven changes in sales ($\hat{s}_j$), export ($\hat{s}_{E,j}$), and purchases ($\hat{c}_j$) to estimate the three terms in Formula (3). For example, indexing the post-reform outcomes by t' and pre-reform ones by t, the reform-driven change in sales can be used to estimate the first term in the formula

$$\Delta \hat{s}_j = (\hat{s}_{jt'} - \hat{s}_{jt}) \approx (s_j - \hat{s}_j)\,. \tag{5}$$

The VAT evasion at the baseline, thus, can be identified at both the micro and the aggregate level.

II.D Overclaimed Refunds

The destination-based design of VAT requires that any tax collected on intermediates used for export be refunded to firms. This creates incentives for firms to exaggerate purchases used for exports, thus increasing the value of refunds they obtain from the government. To quantify the value of overclaimed input tax on account of exports, I rewrite formula (3) in the following format,[4] separating the VAT noncompliance in the export and domestic sectors of the economy

$$\Delta R \equiv \sum_j \tau.\left[-\left(c_{E,j} - \hat{c}_{E,j}\right) + \left(s_{NE,j} - \hat{s}_{NE,j}\right) - \left(c_{NE,j} - \hat{c}_{NE,j}\right)\right]. \tag{6}$$

Subscripts E and *NE* here denote the export and non-export values of a variable. The first term in the above expression is the overclaimed input tax on exports: it is the amount by which firms overreport the value of intermediates used to produce exports. In my empirical setting, I do not observe $\hat{c}_{E,j}$ directly. Firms in Pakistan report the total value of their purchases without breaking it down by its use (domestic vs. exports). I will therefore estimate the amount of excess input tax claim on account

[4]To derive this formula, I simply substitute the two definitions $s - s_{E,j} \equiv s_{NE,j}$ and $c_j \equiv c_{E,j} + c_{NE,j}$ into (3) and rearrange terms.

of exports using the following formula

$$\Delta R_E = \Sigma_j -\tau . f^{-1}(s_{E,j} - \hat{s}_{E,j}), \quad (7)$$

where $f(.)$ is the production function of exports i.e. $s_{E,j} = f(c_{E,j})$.

Note that the overclaim of export-related input tax identified by (6) is already accounted for in formula (3). Intuitively, the overclaimed input tax consists of either diverted input tax from domestic consumption toward exports or one-sided evasion. In both cases, it is already covered in the third term of formula (3). The formula therefore continues to identify the aggregate amount of VAT noncompliance in the economy.

II.E Invoice Mills

VAT chains are rarely complete, especially in developing countries. When a VAT chain breaks, the tax remitted at the pre-break stages cannot be claimed at the post-break stage, becoming a part of the price. This creates arbitrage opportunities that are 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. Figure A.I shows how this process works. In this example, I focus only on the last three production stages of a supply chain. The final stage is composed of two firms: the top firm deals only in exports and the bottom only in domestic sales. The firm in the preceding stage is not registered and supplies its output to the two firms, a fraction α to the exporter and the rest to the retailer. Because the firm is not in the VAT chain, the tax remitted at the $J-2$ and preceding stages, amounting to $\tau.\hat{s}_{I,J-2}$, cannot be claimed at the final production stage. Exploiting this gap, the invoice mill places itself in the $J-1$ stage, transferring the credit of VAT remitted at the preceding stages to the final stage. Doing so creates a gross surplus of $\tau.\hat{s}_{I,J-2}$, which can be shared by the colluding firms depending upon their bargaining weights.

Note that the bottom firm in the final stage has a distinctly lower incentive to deal with the invoice mill. All of its sales are to final consumers and hence do not generate any information trails. Dealing with the invoice mills brings the transaction on to books, thus creating an information trail and forcing the firm to pay tax to the extent of value-added on it. In contrast, the top firm in the final stage has a greater incentive to deal with the invoice mill. Given that all of its output is exported, its tax liability is negative and by dealing with the invoice mill it can exaggerate its purchases, increasing the refund it obtains from the government. This coincidence of incentives means that the invoice mill would divert a proportion of input tax from domestic consumption toward the export sector ($\hat{\alpha} > \alpha$), helping exporters claim exaggerated refunds. This phenomenon is knows as the diversion fraud in the literature (Keen & Smith, 2006). It results from the interaction of the two design features of the VAT—the last mile problem and the destination-principle.

It is important to emphasize that the presence of invoice mills does not invalidate formula (3). Given the flexible structure used to derive the formula, it already accounts for the diversion fraud. Invoice mills are one of the mechanisms through which the VAT noncompliance identified by the formula takes place. Note further that the incentive to operate as an invoice mill is strengthened if the government implements a policy commonly observed in developing countries whereby the sales of intermediates to unregistered firms are taxed at a higher rate. In our example, the gross surplus generated by the invoice mill would be $(\tau + \tau_a).\hat{s}_{I,J-2}$ if the government operates such a policy, where τ_a denotes the *additional* rate applied to the sales of intermediates to unregistered firms.

II.F Comparative Statics

How will firms react to the reduction of the VAT rate from 15% to 0%? The incentive scheme I outline translates into following testable predictions.

Testable Prediction I: Reported purchases will fall $\Delta\hat{c}_j\,(\tau \to 0) < 0$;

Testable Prediction II: Reported exports will fall $\Delta\hat{s}_{E,j}\,(\tau \to 0) < 0$;

Testable Prediction III: Domestic B2C sales will rise $\Delta\hat{s}_{F,j}\,(\tau \to 0) > 0$;

Testable Prediction IV: Domestic B2B sales may rise or fall $\Delta\hat{s}_{I,j}\,(\tau \to 0) \lesseqgtr 0$.

Depending upon the strength of last three responses, total sales ($\hat{s}_j \equiv \hat{s}_{E,j} + \hat{s}_{F,j} + \hat{s}_{I,j}$) may rise or fall $\Delta\hat{s}_j\,(\tau \to 0) \lesseqgtr 0$. I do not observe the last two components of sales—domestic B2B and B2C sales—directly. Instead, I present responses of four variables reported by firms—purchases ($\hat{c}_j$); sales ($\hat{s}_j$); exports ($\hat{s}_{E,j}$); and non-export sales ($\hat{s}_j - \hat{s}_{E,j}$). Together, the responses of these four variables are sufficient to confirm that overall firm responses are consistent with predictions I–IV above.

III Contextual Background

III.A Pakistani VAT System

Pakistan introduced its 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, 2020b).[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.

The Pakistani VAT largely follows the standard design. Firms with annual turnover 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 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

[5] The energy sector was brought into the tax net in July 1999 and services in July 2000. Please see Waseem (2020b) 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.

and the payment of refund.

Panel A of Figure A.II plots the standard VAT rate in the country. The rate generally remained at 15% until 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—imposed by the policy. The rate remained between 1 and 3 percentage points before it was eliminated in 2004. The policy, as I explained above, strengthens the incentive of a firm to operate as an invoice mill.

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).[7] First, a large proportion of the output of these industries was exported out of the country in one form or another and hence was zero-rated anyway. 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 for the refund of VAT paid on their inputs.

Second, the VAT chains of these industries were rarely complete. The breaks had given rise to the phenomenon of spurious invoices. Over time, the volume of such

[7] 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.

invoices was growing, making it increasingly costly for the tax administration to distinguish between genuine and fraudulent refund claims. Ultimately the problems created by these two related issues became so severe that the government forsake the VAT revenue from the domestic consumption of these industries and zero-rated their entire supply chains.[8]

The reform was announced on June 06, 2005 and became applicable from July 01, 2005. Although there was no firm commitment from the government prior to its announcement, the reform to some extent was anticipated in the sense that the government had been in negotiation with the stakeholders few months leading up to the reform to finalize details, such as the list of inputs to be zero-rated. In its final form, the reform 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.[9] 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.[10] It thus seriously weakened, if not entirely eliminated, the incentives to misreport sales or purchases.

[8]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.

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

[10]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.

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 required 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 the VAT return and the annex was made mandatory for all firms. Together, these changes reduce the costs of cross-matching sales and purchase records, enabling the tax administration to detect one-sided misreporting at a relatively low cost. 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, a point I come back to in section V of the paper.

III.C Invoice Mills

One key focus of the paper is to document the role of invoice mills in VAT noncompliance. Like all fraudulent enterprises, invoice mills take great care in disguising themselves as legitimate businesses so that distinguishing them from the 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 against the firm could proceed. This inquiry was of a quasi-judicial nature, where the firm was confronted with the evidence against it and was afforded an 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. I observe both the suspension and blacklisting of a firm and use these to proxy if the firm is an invoice mill.

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.[11] 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 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 spurious invoices.

IV Empirical Methodology

Given that the reform I exploit affects a subset of firms only, the natural research design in this setting is the difference-in-differences framework. I exploit the fact 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

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

model

$$y_{it} = \alpha_i + \lambda_t + \boldsymbol{\beta X_{it}} + \gamma.\, zero\text{-}rated_i \times after_t + \varepsilon_{it}, \quad (8)$$

where α_i and λ_t are the 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}) introduced as logs. In parallel specifications, I replace the double-difference term with $\log(1 - \tau_{it})$, where τ_{it} is the standard VAT rate faced by firm i in period t. Given that the VAT is a linear tax system, the tax rate is not endogenous to a firm's reporting decision. These parallel specifications therefore deliver the elasticity of the corresponding outcome with respect to the net-of-tax rate. The baseline specification clusters standard errors at the firm level, but in one of robustness checks I report results where the standard errors are clustered at the industry level—the most aggregate level feasible in this setting (Abadie *et al.*, 2017).

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

$$y_{it} = \alpha_i + \sum_{j=2}^{N} \delta_j.\, zero\text{-}rated_i \times 1.(tax\ period{=}j)_t + \lambda_t + u_{it}, \quad (9)$$

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. 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.[12]. 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, 2020b). I present two sets of evidence to rule out this class of concerns (please see section VI).

I estimate the impact of the reform on the following six outcomes: (i) input tax; (ii) output tax; (iii) purchases; (iv) sales; (v) exports; and (vi) non-export sales. The definitions of these variables are provided in Appendix A.1. Table A.I 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 terms of 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 II shows the aggregate level evidence, plotting the amount of VAT refund paid in Pakistan as a proportion of gross VAT collected in the country. The figure

[12]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

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.[13]

To show that this large drop in refund was triggered by the zero-rating reform, I next turn to micro level evidence showing the reform's effects on output tax charged and input tax claimed by the treated firms. Figure III plots the coefficient $\hat{\eta}_j$'s from the following version of the event study equation (9)

$$y_{it} = \alpha_i + \sum_{j=2}^{N} \eta_j .\, 1.(tax\ period{=}j)_t + u_{it}, \tag{10}$$

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). Figure IV presents the DD version of these plots, where I display the coefficients $\hat{\delta}_j$'s from equation (9) 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. By definition these variables equal the tax rate times the non-export sales and purchases reported by the firm (see Appendix A.1). Given that the tax rate falls from 15% to 0%, I expect large fall in these variables in the treated group. The fall comprises both the mechanical and the behavioral effects. I isolate the latter effect below by estimating the sales and purchases response separately. The objective of showing the output tax and input tax responses here is to confirm that a strong first-stage exists in this setup. Clearly, a sharp fall occurs in both the treated outcomes exactly at 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

[13]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.

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 due likely to the time taken in zero-rating two important inputs—electricity and gas—of the treated firms.

The above two pieces of evidence show that the incentives to misreport collapse in the treatment group at the time of the reform as both their output tax liability and input tax credit fall to a near-zero level. Such a sharp change in incentives is likely to induce sharp behavioral responses to which I turn next.

V.A.2 Behavioral Responses

Panels C-F of Figures III and IV 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 fall sharply in the treatment group after the reform. The fall is the sharpest for purchases, which decline 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.

The direction of movement of these variables is consistent with the predictions of the model noted in section II.F. Both purchases and exports fall in accordance with predictions I and II. Non-export sales, by definition, equal the sum of domestic B2B and B2C sales $\hat{s}_{NE} \equiv \hat{s}_I + \hat{s}_F$. The model has a clear prediction on the latter of these two components (B2C sales will rise after the tax cut) but no clear prediction on the former. The fall of non-export sales therefore shows that B2B sales fall after the tax cut and that their response outweighs the rise in B2C sales. Note that B2B sales are mirror image of purchases: B2B sales reported by a seller are reported as purchases by the buyer. Being so, their larger fall relative to the rise in B2C sales is plausible given the sharp fall in purchases seen in Figures III–IV. The overall reduction in reported sales ($\hat{s}$) is a mere reflection of the fact that both its components—exports and non-export sales—fall.

One other noticeable feature of the event-study results is that the sales response is slower than the purchases response. Again, such a dynamics is expected. Purchases are expected to fall immediately after the incentives to overreport them end. 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 event study results also validate my empirical strategy. The preexisting trends are fairly parallel in the two groups for all the six outcomes I explore (see Figures III and IV). 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 I presents the results from the difference-in-differences model (8) for each of the six outcomes. Table A.II reports results from parallel specifications where the standard errors are clustered at the industry level. Unsurprisingly, 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. If firms report truthfully, the reduction of the rate to zero should induce no change in behavior other than some positive, real effects on activity operating through the demand or liquidity channels. But the observed responses reveal a different story. Not only are all elasticities of the opposite sign, they are also very large, ranging from -0.5 to -2.6. The large fall in reported activity in the treated industries is consistent with, and can only be reconciled by, a large misreporting at the baseline. In section V.D, I use formula (3) to estimate the extent of misreporting implied by these responses and study the mechanisms underlying it.

The event study results presented above show that the key identification assumption in my empirical setting—parallel trends—is plausible. Panel B of the table tests this formally. I estimate a parallel placebo specification, where I estimate equation (8) on the pre-reform panel only, pretending that the zero-rating reform occurred in July 2002. The results support the identification assumption. There is little to distinguish the treated and untreated groups in terms of the baseline evolution of the VAT outcomes.

Table II explores the dynamics of the responses. I estimate a flexible version of model (8) 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 III and IV. 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.[14] 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.

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.III shows this formally. I break down the aggregate response 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.[15] 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 III carries out this exercise. I restrict the treatment sample to the textile industry only and estimate a triple-difference version of model (8) by including inter-

[14] 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.

[15] 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.

actions 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—only, leaving the rest as the omitted category. The results reflect important heterogeneity in response across the supply chain. Responses 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 exhibit no significant change relative to the control group after the reform. This suggests that the real production activity did not change significantly in the treated industries.[16] In contrast, treated outcomes decline by a lot in the later production stages.[17] To the extent that this decline captures tax evasion, increasing response over the supply chain from the second stage onward suggests that the tax evasion is not limited to the later production stages but to some extent persists throughout the supply chain.

This finding goes against the conventional wisdom on VAT's enforcement properties in two important ways. First, it shows that VAT's noncompliance is not merely a last mile phenomenon but runs deep into the supply chain. The initial rungs of the production chain disproportionately consist of inter-firm transactions that are subject to third party reporting, generating information trails for the government. Theoretically, therefore, there is a strong case of lower evasion in these production stages. But this is not supported empirically as I find substantial evasion from the second production stage onward. The self-enforcement processes built into a VAT therefore do not seem to have an automatic inhibiting effect on misreporting. In the absence of strong enforcement capacity, firms seem to be able to get around these processes easily. To this extent, these results are consistent with the recent public finance literature that finds strong complementarity between the enforcement capacity of the revenue authority and efficacy of the self-enforcing processes (see for example Carrillo *et al.*, 2017; Waseem, 2020b).

[16] I assume here that the zero-rating response at the first production stage would largely capture the real production effect as the stage only produces intermediates used exclusively by large spinning mills, making misreporting less likely.

[17] 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.

V.C Robustness

One concern in this setting is that the zero-rating reform may create spillovers in the nontreated industries, violating the SUTVA assumption. Table A.IV shows that I get similar results from the balanced panel specification. This mitigates the spillover concern along the extensive margin, operating through the entry or exit. Figure III further illustrates that the outcomes of nontreated industries do not undergo a structural break at the time of the reform. Building on this evidence, Table A.V and A.VI explore spillovers more formally. If the reform creates significant spillovers, they would be stronger in industries producing substitutes or complements of the treated goods. The tables test this by looking at the evolution of outcomes in such industries. Pakistani tax administration, as I note above, uses the HS Code to classify firms into industries. The first two digits of this eight-digit code divide firms into broad categories with industries producing similar goods getting adjoining codes.[18] I exploit this coding scheme to explore the tax cut's spillovers on adjoining industries. Table A.V drops all treated industries from the sample and compares the evolution of outcomes in their adjoining industries with that of others. Here I estimate specifications parallel to ones in Table I, the only difference being that the variable *treat* now indicates firms belonging to the adjoining industries. I experiment with three definitions of adjoining. Table A.VI is structured similarly with the difference that it retains the treated industries in the sample, including the adjoining industries as an additional interaction term. Reassuringly, all specifications return trivial or insignificant spillover coefficient, putting to rest the concern that the reform might have affected outcomes of the nontreated industries as well.

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 in September 2008. Some of the negative responses documented above may reflect that these events differentially affect the treated industries. Figure A.III addresses this class of concerns. It is a truncated version of Figure IV, where I show the post-reform periods only. Zooming in on these periods illustrates that the reform started a slow, downward trend in the outcomes of the treated industries. This downward trend did not accelerate during

[18]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 similar industries producing close substitute or complement goods.

2008. In fact, the outcomes of the treated industries stabilized or started rising slowly 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.

I have noted above that the reform was anticipated in the sense that the government had been discussing it with the stakeholders before its formal announcement on June 6, 2005. To the extent that the reform was anticipated, it may trigger some intertemporal shifting of activity by forward looking firms to calibrate their response to the new incentives. I show later (section VI) that invoice mills indeed engaged in such intertemporal shifting. Here, I focus on firms other than invoice mills, which comprise 93% of the treatment sample, to see if they also engaged in such shifting. I begin with Figure A.IV, which plots the aggregate values of the VAT outcomes of these firms. Clearly, there are no signs of intertemporal shifting in either directions. This should not be surprising as the event study plots (Figures III–IV) do no show any concentration of activity on either sides of the reform: there is no peak or hole in the treated series close to the tax cut. Tables A.VII–A.VIII rule out anticipation formally. The former shows that dropping invoice mills—the subsample that did engage in the intertemporal shifting of activity—does not change the results. The latter defines the $after$ period from the date of announcement of the reform rather than the date of its coming into force. This specification also returns similar results.

Table A.IX 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 I are explained by lazy reporting 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 more negative among noncorporate firms whose quality of record maintenance is in general poorer. Contrary to this, the responses are on average less negative for noncorporate firms. Lazy reporting is therefore unlikely to be an explanation of the observed responses.

V.D Quantifying VAT Evasion

Section II.C shows that a lower bound on the amount of VAT evaded in the treated industries is provided by formula (3). I now use the results in Table I to estimate this lower bound. The calculations are in Table IV. Rows 1-3 compute the three terms in the formula for the four baseline years (2001-2004). To estimate each term, I multiply the percent change in the variable as implied by its response reported in Table I with its aggregate value in the baseline year. Row 4 adds the three terms to compute the total evaded amount in PKR billions using formula (3). Finally, the last row of the top panel expresses the evaded amount as a fraction of the potential revenue in the year.

Panel B of the table uses formula (7) to calculate the amount of overclaimed refund by treated firms. One important ingredient in this formula is the inverse mapping from exports to purchases of intermediates used in exports . Figure V estimates this mapping nonparametrically. To construct the plot, I group firms into small bins on the basis of log purchases reported by them and plot the average log exports in each bin. The sample for this binned scatter plot consists of all firm-month observations of the treated industries, excluding those with the log purchases less than the 5th or more than the 95th percentile of the corresponding distribution. The relationship between the two variables in the log-log space is approximately linear with a slope parameter of 0.5, which means that the production function of exports $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ $s_E = f(c_E)$ is Cobb-Douglas with $s_E = A \, . \, c_E^{0.5}$. Note that this relationship does not need to be causal as I only use it to predict $\Sigma_j \left(c_{E,j} - \hat{c}_{E,j}\right)$ implied by $\Sigma_j \left(s_{E,j} - \hat{s}_{E,j}\right)$, the term identified through the DD estimator. Rows 6–8 combine the estimated production function and exports response to compute the amount overclaimed as refund. The final row expresses it as a percentage of the potential revenue in the year.

The results reveal that a large amount of VAT is evaded in the country. The evaded amount varies from 31% of potential revenue in 2001 to 46% in 2004, with an increasing trend over time (Row 5 of Table IV). A large proportion of this evasion occurs through the channel of overclaimed refunds. The overclaimed amount, reported in the last row of the table, ranges between 11–23% of the potential revenue, again with an increasing trend over time. How do these evasion estimates compare with similar estimates from other countries? Gómez Sabaini & Jiménez (2012) compute the VAT evasion rate for many Latin American economies, finding similar levels of evasion. For example, the evasion rate is 37.5% in Guatemala; 38.1% in Nicaragua; 33.8% in Panama; and 37.7% in Peru. Such large levels of evasion suggest that VAT's robustness to tax evasion is somewhat overstated. Theoretically, the tax indeed contains mechanisms that must make evasion harder than under its alternatives. But it appears that firms in low tax capacity setting are able to get around these mechanisms through collusion and other strategies. One manifestation of such collusion is that the evasion I find runs deep into the initial stages of the supply chain, most transactions where are subject to third-party reporting (Table III). Evasion at these stages suggests that the third-party information on its own does not deter evasion meaningfully, a result that has important implications for both tax policy and optimal tax design in weak enforcement-capacity settings.

VI Invoice Mills and VAT Evasion

One important focus of this paper is to understand the role of invoice mills in facilitating VAT's noncompliance. As I note in section II.E, invoice mills arise naturally in the low-enforcement, high-informality setting of developing countries, bridging the gaps created by broken VAT chains. Figure A.V illustrates this empirically, plotting the stock of invoice mills in Pakistan's VAT over time. Invoice mills appeared soon after the destination-based VAT was implemented in the country. Their number grew sharply in the next few years, reaching a peak in 2003. It partly reversed then, as the new mechanism to blacklist firms came into effect. But a much 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 plummet as the tax rate on five major export-oriented industries falls to zero). To explore this point further, I next look at the *activity* reported by them.

Figure A.VI examines the trajectory of VAT outcomes of invoice mills, replicating the event-study analysis of Figure III. Since invoice mills comprise only 7% of the treatment sample, the results are noisier than the baseline results. Yet the pattern of responses is similar. All four outcomes fall sharply at the time of the reform, with purchases falling more than any other outcome. Table A.X formalizes these results. I estimate a triple-difference version of model (8), exploring any differential response of invoice mills relative to other treated firms. Unsurprisingly, mills' responses are stronger, but given their smaller number their stronger responses do not affect the overall average significantly: the double-difference coefficient is always within the 95% confidence interval around the baseline coefficient.

Importantly, the event study plots in Figure A.VI show that the outcomes spike just before the tax cut. This spike is particularly prominent in the top four panels of the figure. To probe this result further, Figure VI examines the aggregate values of these six outcomes in each month. These plots show the aggregate *level* of each outcome, highlighting the size of spurious input tax credit injected into the system by invoice mills. Strikingly, all outcomes feature a sharp spike just before the tax cut. For example, aggregate purchases jump from the prereform average of around PKR 5 billion to 30 billion one month before the tax cut (May 2005). This six-fold rise is followed by even sharper fall as purchases plummet to PKR 3.6 billion in July 2005 and to 1.8 billion in November 2005.[19] Such a large concentration of activity on the wrong side of the reform is puzzling. Under any standard model of tax behavior a large tax cut would induce some inter-temporal shifting of activity toward the low-tax period.[20] Shifting in the opposite direction strengthens the conclusion that invoice mills are primarily refund generating devices: they are injecting spurious input tax credit into the system while the goods are still standard-rated. Doing so will maximize the refunds their buyers can obtain in accordance with the mechanism outlined in section II.E.

How much excess refund did invoice mills inject into the system? I examine this question using the transaction-level data exporters file to claim refund, which are available from 2002 onward. Figure VII reports the results of this exercise. The blue curve plots the aggregate value of purchases exporters report each month from invoice mills. The red curve, on the other hand, plots the aggregate value of sales invoice mills report each month to all buyers including exporters. Although the two curves use data from two different sources—the blue curve from refund claims filed by exporters and the red from VAT returns filed by invoice mills, they line up tightly. This shows that most of the sales of invoice mills are booked to exporters, who use their invoices to claim refunds. A regression of blue series on the red estimated for the baseline years returns a coefficient of 0.83, suggesting that on average 83% of the output of invoice mills is claimed as refund by exporters.[21] This ratio shoots up to 97% (PKR 8.6 billion) in 2004, the last year such a refund could be claimed in. Table

[19]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.

[20]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.

[21]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.

IV shows that exporters overclaimed refund by nearly PKR 23 billion in 2004 (see Row 5 of the table). Roughly 37% of the overclaimed amount was based on invoices issued by mills.

Figure A.VII shows the evolution of outcomes of invoice mills around the event of their blacklisting. It illustrates that by the time an invoice mill is blacklisted its activity has already passed its peak. Being so, most of their sales occur while their invoices are legitimate. To see why invoice mills are detected so late in their life cycle, Figure A.VIII and Table A.XI compare their characteristics with other firms. Invoice mills are essentially hit-and-run enterprises, and consistent with this notion their most distinct marker is the level of activity they report soon after registration.[22] To catch them, the revenue authority therefore needs to move early when they are still distinguishable from other firms and when they have still not caused significant revenue damage. But doing so requires developed administrative capacity, which revenue authorities usually lack in developing and emerging economies.

VII Conclusion

The value-added tax has seen remarkable expansion in recent decades. Its popularity in large part stems from the belief that among the class of production-efficient tax instruments it perhaps 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, 2019). But the tax also contains built-in mechanisms that may worsen tax compliance: (1) the destination principle necessitates the payment of refund on exports; (2) the last-mile problem eliminates paper trail at the end of the supply chain; and (3) broken VAT chains give rise to invoice mills. These mechanisms are emphasized a great deal in the policy literature, 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 and strength of these mechanisms in a representative emerging economy.

I leverage a novel tax reform that eliminates firms' incentives to misreport outcomes. This allows me to infer the level of misreporting in the treated industries at the baseline and study mechanisms underlying it. I document four key facts. First,

[22] *Early* in Figure A.VIII denotes the level of activity reported in the first six months after registration

the evaded VAT constitutes a significant part of the tax base, amounting to nearly 31–46% of the potential revenue. Second, the evasion is not limited to the latter stages of the supply chain—where the last-mile problem is worse—but runs deep into the supply chain. This suggests that on their own the self-enforcement mechanisms built into a VAT do not deter tax evasion fully. Third, overclaiming of export-related refunds is a key channel through which VAT evasion occurs: the overclaimed amount constitutes 11–23% of the potential revenue. Finally, invoice mills are an important conduit through which the overclaim of refund takes place. More than 80% of their output is booked to exporters, who use it to obtain excessive refunds. Of the amount overclaimed as refund, nearly 37% is based on invoices issued by invoice mills. The results have important policy implications, highlighting the wide gulf between theory and practice of a modern tax instrument in weak enforcement setting. Optimal tax theory and policy advice to developing economies must take this difference into account.

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.

Basri, M Chatib, Felix, Mayara, Hanna, Rema, & Olken, Benjamin A. 2019. *Tax Administration vs. Tax Rates: Evidence from Corporate Taxation in Indonesia*. Tech. rept. National Bureau of Economic Research.

Best, Michael Carlos, Brockmeyer, Anne, Kleven, Henrik Jacobsen, Spinnewijn, Johannes, & Waseem, Mazhar. 2015. Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan. *Journal of Political Economy*, **123**(6), 1311–1355.

Brockmeyer, Anne, Smith, Spencer, Hernandez, Marco, & Kettle, Stewart. 2019a. Casting a Wider Tax Net: Experimental Evidence from Costa Rica. *American Economic Journal: Economic Policy*, **11**(3), 55–87.

Brockmeyer, Anne, Estefan, Alejandro Manuel, Arras, Karina Ramirez, & Serrato, Juan Carlos Suárez. 2019b. Taxing Property in Developing Countries: Theory and Evidence from Mexico. Working Paper.

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.

Gómez Sabaini, Juan Carlos, & Jiménez, Juan Pablo. 2012. Tax structure and tax evasion in Latin America. Macroeconomics of Development Series 118.

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.

MITTAL, SHEKHAR, REICH, OFIR, & MAHAJAN, APRAJIT. 2018. Who is bogus? using one-sided labels to identify fraudulent firms from tax returns. *Pages 1–11 of: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*.

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.

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

SLEMROD, 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, REHMAN, OBEID UR, & WASEEM, MAZHAR. 2020. How Do Taxpayers Respond to Public Disclosure and Social Recognition Programs? Evidence from Pakistan. *The Review of Economics and Statistics*, 1–44. https://doi.org/10.1162/rest_a_00929.

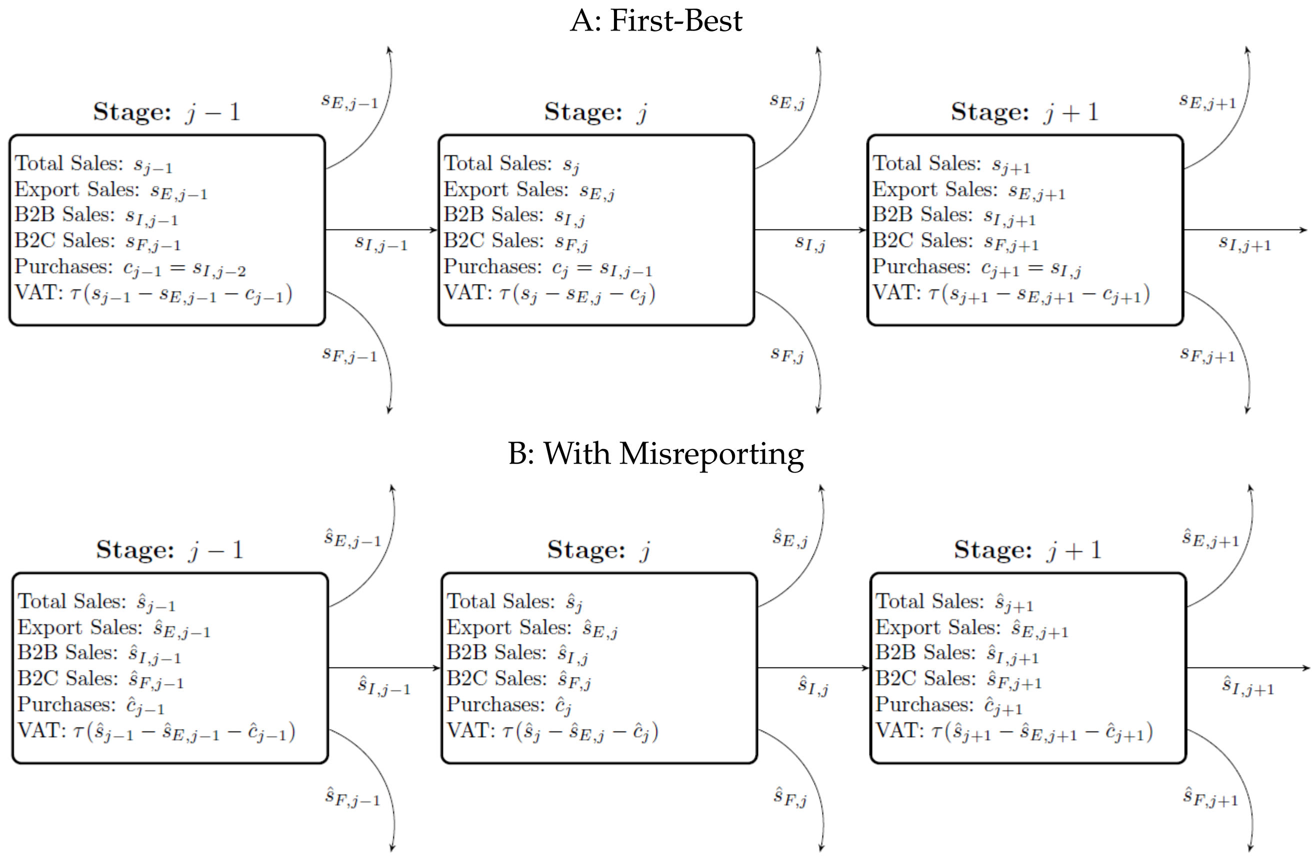
STIGLITZ, JOSEPH E. 2010. Development-oriented tax policy. *Taxation in developing countries: Six case studies and policy implications*, 11–36.

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

WASEEM, MAZHAR. 2020a. Does Cutting the Tax Rate to Zero Induce Behavior Different from Other Tax Cuts? Evidence from Pakistan. *The Review of Economics and Statistics*, **102**(3), 426–441.

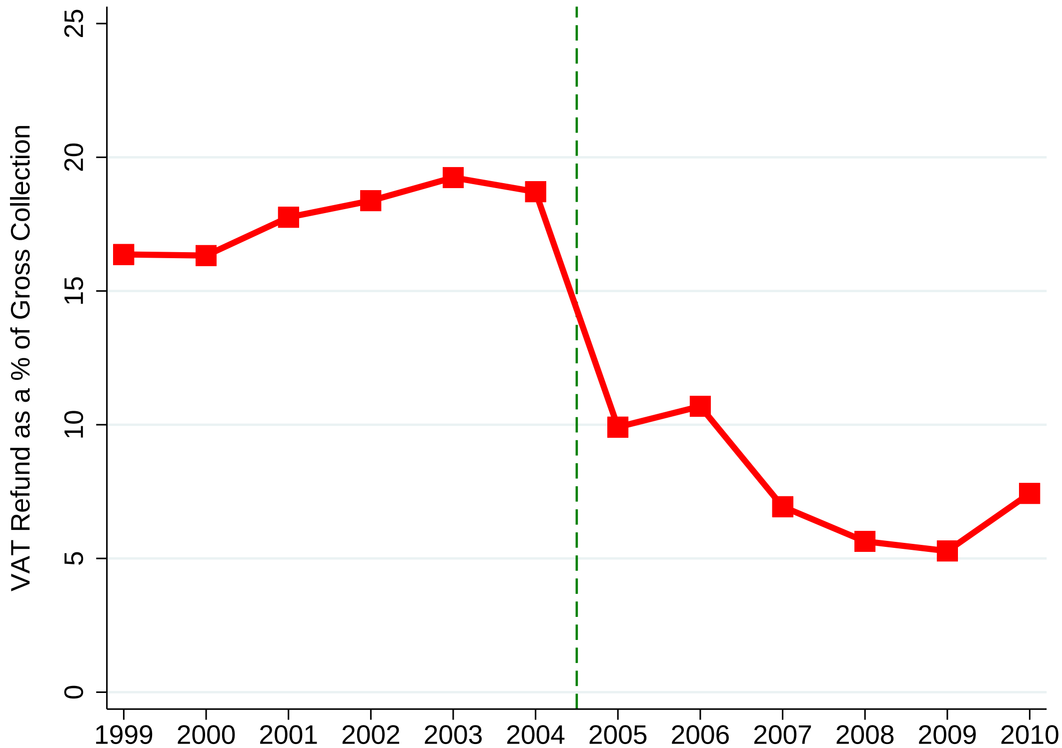
WASEEM, MAZHAR. 2020b. The Role of Withholding in the Self-Enforcement of a Value-Added Tax: Evidence from Pakistan. *The Review of Economics and Statistics*, 1–44. https://doi.org/10.1162/rest_a_00959.

FIGURE I: **CONCEPTUAL FRAMEWORK**



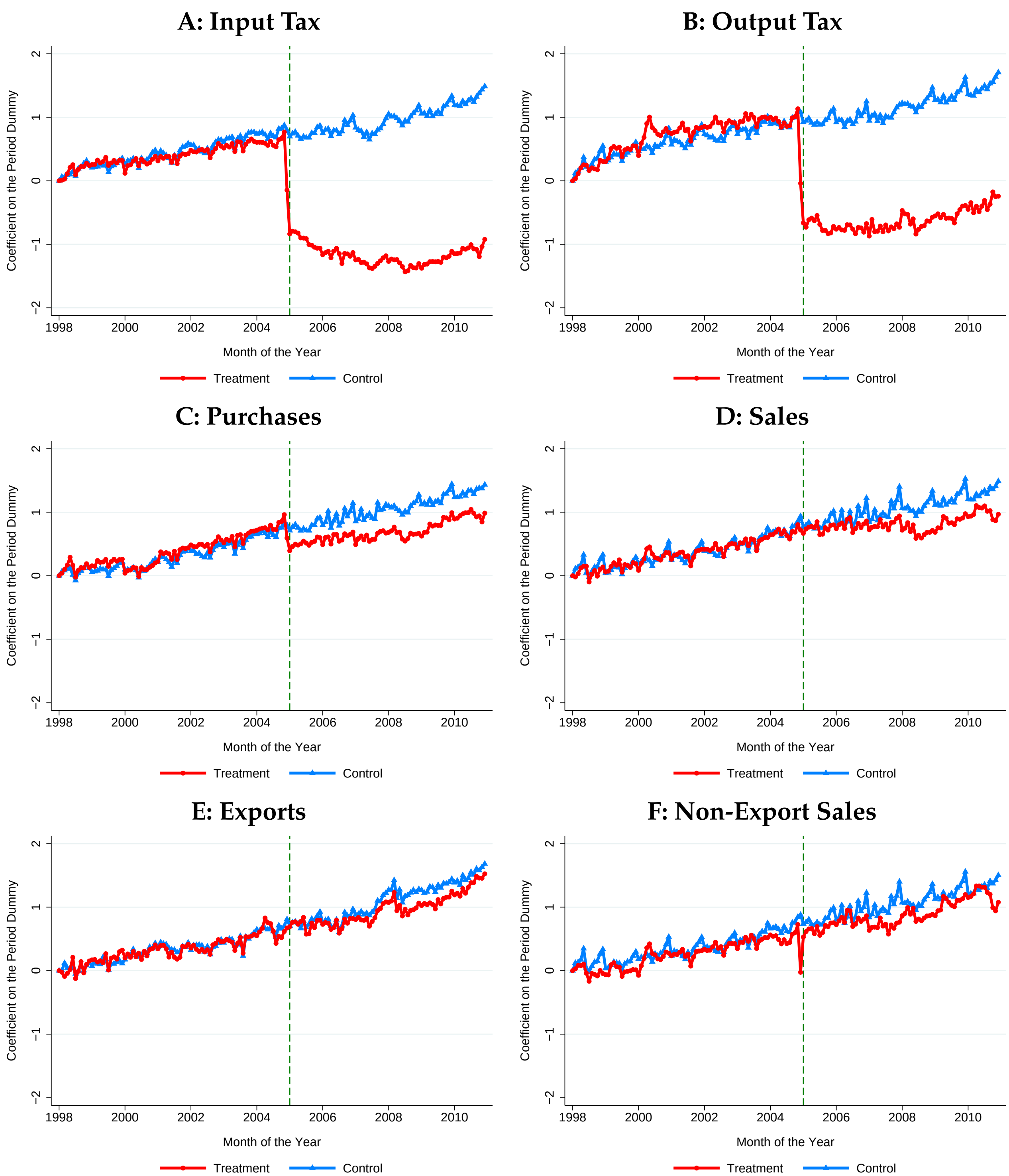
Notes: This figure displays three consecutive production stages of a supply chain. The top panel displays behavior under the first-best, where all firms report their sales and purchases truthfully. Total sales s_j of a firm consists of three components: (1) sales exported out of the country $s_{E,j}$; (2) sales made to other domestic firms $s_{I,j}$; and (3) sales made to final, domestic consumer $s_{F,j}$. Purchases here refer to the value of *taxable* intermediates, which here are all acquired from the previous production stage $c_j = s_{I,j-1}; \forall j \in 1, ..., J$. The VAT liability of a firm equals its output tax $\tau(s_j - s_{E,j})$ minus the input tax τc_j. Panel B displays the same supply chain, but firms are now allowed to misreport. Firms' sales and purchases in this second-best world are denoted by $\hat{s}_{K,j}; K \in \{E, I, F\}$ and $\hat{c}_j$ to distinguish them from their true values $s_{K,j}$ and c_j.

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



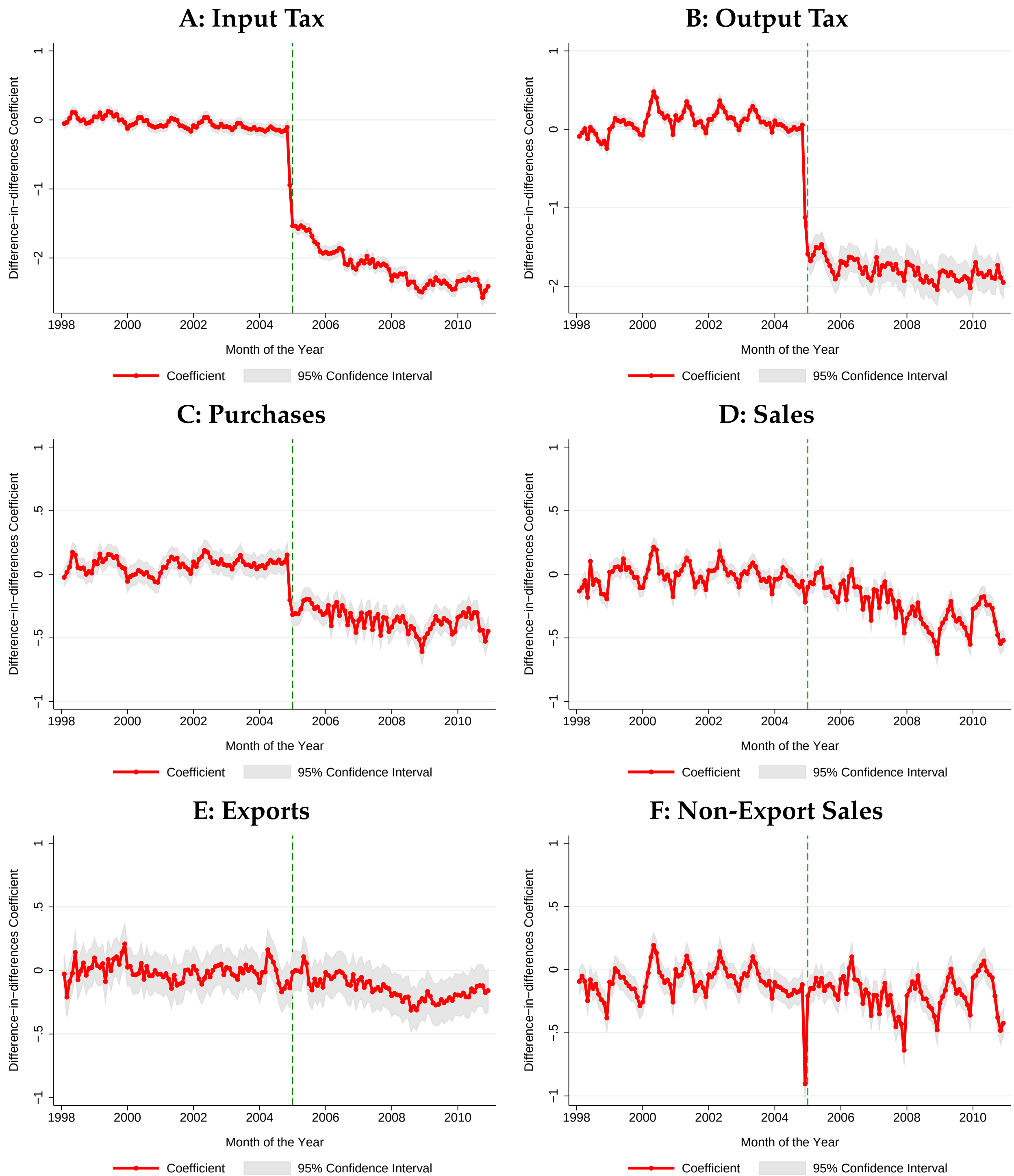
Notes: The figure shows the effects of the zero-rating reform on the VAT refund paid in Pakistan. Each marker in the plot denotes the aggregate VAT refund paid by the FBR to firms in *all* industries as a percentage of the gross VAT collection in that year. The data used for this plot are publicly available 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 III: **FIRM BEHAVIOR TO THE TAX CUT**



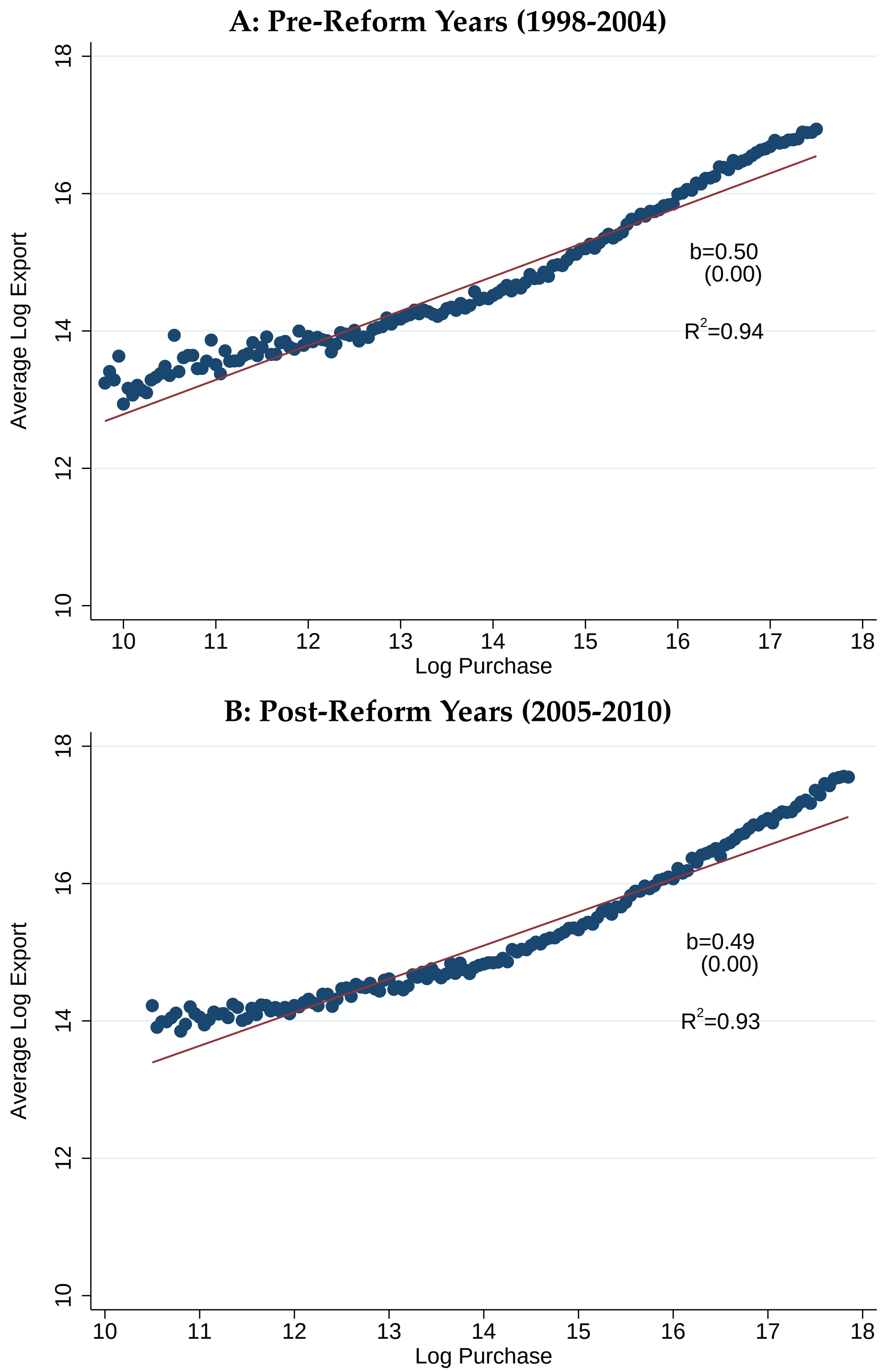
Notes: The figure compares the evolution of six VAT outcomes from the tax year 1998 to 2010 across the treatment and control groups . Treatment and control groups consist respectively of firms in the zero-rated and non-zero-rated industries. To construct these charts, I regress the log of the outcome 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 IV: **FIRM BEHAVIOR TO THE TAX CUT**



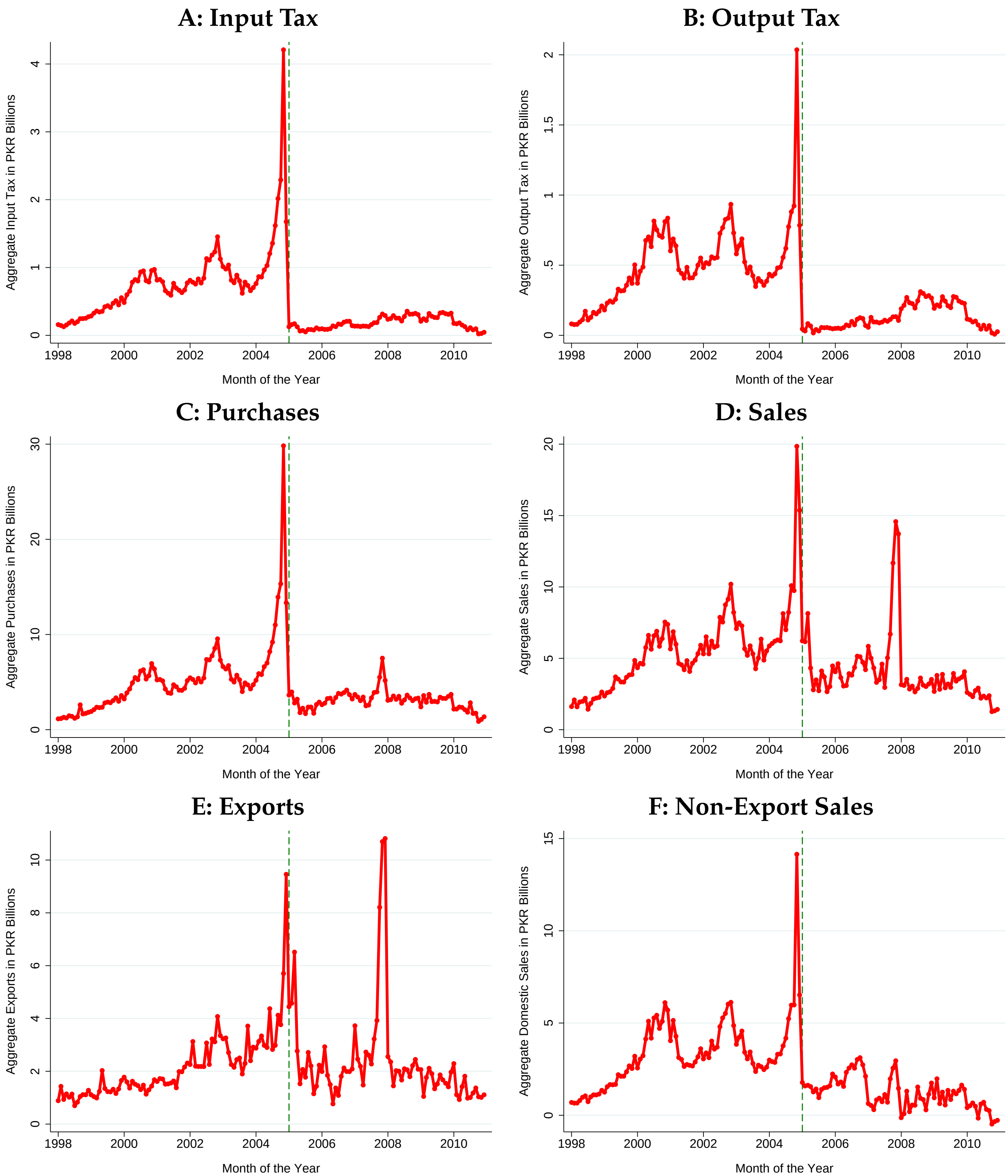
Notes: The figure shows the difference-in-differences version of the plots in Figure III. To construct these charts, I regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, and month×treat dummies, dropping the dummies for July 1998. I then plot the 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 V: **EXPORTS AS A FUNCTION OF PURCHASES**



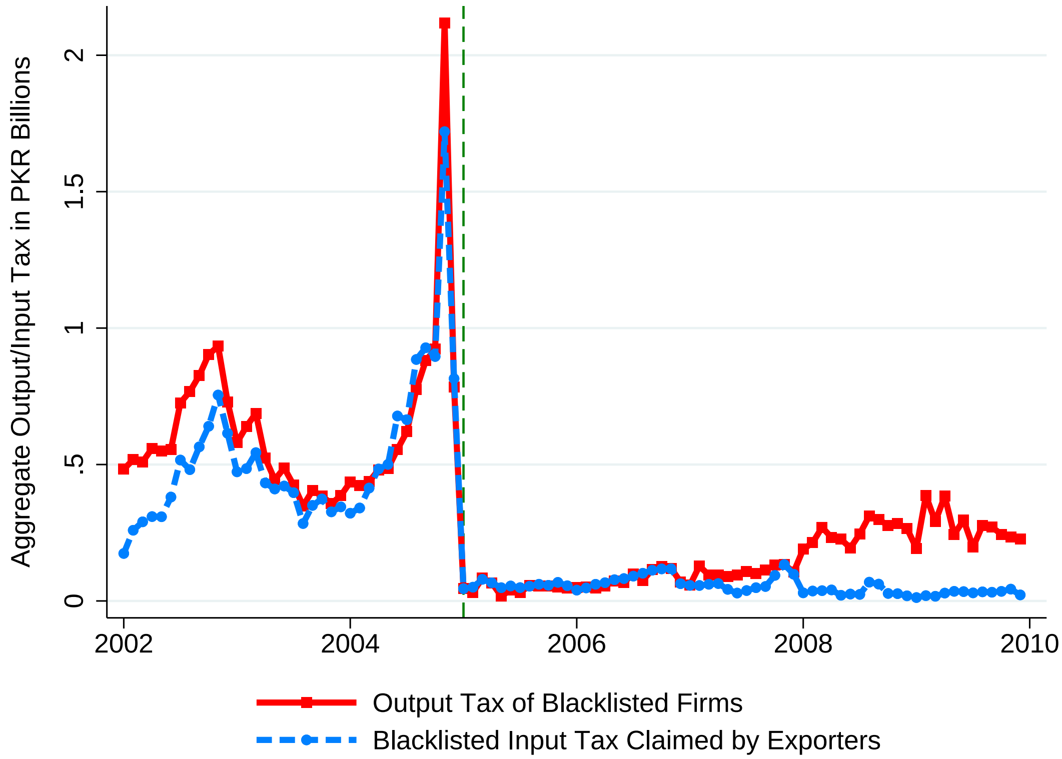
Notes: The figure explores the relationship between purchases of intermediates and exports reported by firms of the treated industries. The blue curve shows a binned scatter plot for the data. I group firms into bins on the basis of log of purchases reported by them. Each blue marker represents the average 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 onto the data using ordinary least squares. The slope of the fitted line and R^2 from the regression are reported in the panel. Panel A relates to the pre-reform years and Panel B to the post-reform years.

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



Notes: The figure illustrates how VAT outcomes reported by invoice mills 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 VII: **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 shows that the primary purpose of the existence of invoice mills is to help exporters claim exaggerated refunds. 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: **FIRM BEHAVIOR TO THE TAX CUT**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
A: Baseline Specification						
treat $\times$ 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)
Elasticity	-12.065	-11.334	-2.578	-1.369	-0.651	-0.504
	(0.158)	(0.439)	(0.103)	(0.076)	(0.154)	(0.083)
Observations	3,728,660	4,179,561	3,983,213	5,058,579	612,993	4,623,907
B: Placebo Specification						
treat $\times$ after	-0.004	0.001	0.052	0.019	-0.022	-0.014
	(0.012)	(0.011)	(0.012)	(0.010)	(0.026)	(0.011)
Elasticity	-0.026	0.005	0.318	0.117	-0.134	-0.083
	(0.073)	(0.069)	(0.076)	(0.062)	(0.162)	(0.067)
Observations	1,999,987	2,423,815	1,979,624	2,583,424	306,931	2,381,448
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table explores how firms respond to the reduction of the rate applicable to their supply chain to zero. The coefficient treat $\times$ after shows $\hat{\gamma}$ from model (8), where the dummy variable treat$_i$ denotes that firm i belongs to a zero-rated industry and the dummy variable after$_t$ that month t falls in the tax year 2005 and later. The coefficient $Elasticity$ shows $\hat{\gamma}$ from the same model (8) when I replace the double difference term with $\log(1 - \tau_{it})$. Panel B shows results from parallel placebo regressions, where I restrict the sample to pre-reform years only, defining the period beginning from July 2002 as the $after$ period. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE II: **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
Month 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 (8), 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 III: **FIRM BEHAVIOR TO THE TAX CUT – ACROSS THE SUPPLY CHAIN**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
A: Response						
treat $\times$ 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 $\times$ after $\times$ 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 $\times$ after $\times$ 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 $\times$ after $\times$ 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
B: Elasticities						
$\log{(1-\tau)}$	-12.550	-2.881	-2.641	-1.634	-0.666	0.568
	(0.325)	(0.454)	(0.207)	(0.166)	(0.213)	(0.201)
$\log{(1-\tau)} \times$ ginning	12.633	-13.441	2.394	1.876	1.187	-0.003
	(0.461)	(0.593)	(0.411)	(0.221)	(0.351)	(0.259)
$\log{(1-\tau)} \times$ spinning	-3.102	-22.230	0.675	0.186	1.294	-2.162
	(0.458)	(1.277)	(0.279)	(0.218)	(0.438)	(0.258)
$\log{(1-\tau)} \times$ weaving	0.111	-11.266	0.004	0.158	0.156	-1.255
	(0.378)	(0.907)	(0.243)	(0.192)	(0.271)	(0.225)
Baseline Coefficient	-12.169	-11.844	-2.425	-1.358	-0.398	-0.504
	(0.162)	(0.465)	(0.103)	(0.079)	(0.165)	(0.084)
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
Month 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 (8). 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 IV: **CALCULATING VAT EVASION**

	2004 (1)	2003 (2)	2002 (3)	2001 (4)
A: Underpaid VAT				
1. $\Sigma_j \, \tau.(s_j - \hat{s}_j)$	-47.6	-38.9	-31.6	-27.1
	(3.4)	(2.8)	(2.3)	(2.0)
2. $\Sigma_j \, \tau.(s_{E,j} - \hat{s}_{E,j})$	-11.5	-7.6	-6.2	-5.2
	(3.9)	(2.6)	(2.1)	(1.8)
3. $\Sigma_j \, \tau.(c_j - \hat{c}_j)$	-78.5	-61.9	-49.3	-40.8
	(3.9)	(3.1)	(2.4)	(2.0)
4. $\Delta R = R - \hat{R}$	42.4	30.5	23.9	18.9
	(6.5)	(4.9)	(3.9)	(3.3)
5. $\Delta R/R$	0.459	0.386	0.359	0.312
B: Overclaimed Refund				
6. $\Sigma_j \, \tau.(s_{E,j} - \hat{s}_{E,j})$	-11.5	-7.6	-6.2	-5.2
	(3.9)	(2.6)	(2.1)	(1.8)
7. $\Sigma_j \, \tau.(c_{E,j} - \hat{c}_{E,j})$	-23.0	-15.1	-12.3	-10.5
	(15.7)	(10.3)	(8.4)	(7.1)
8. ΔR_E	23.0	15.1	12.3	10.5
	(15.7)	(10.3)	(8.4)	(7.1)
9. $\Delta R_E/R$	0.249	0.191	0.186	0.172

Notes*:* The table uses formulas (3) and (6) to quantify VAT noncompliance implied by the estimates in Table I. Panel A computes the amount of VAT evaded by the treated industries in the four baseline years using formula (3). Rows 1-3 compute the three terms on the RHS of the formula. To compute each term, I proceed in two steps. First, I convert the response of the corresponding item in Table I into percentage terms (for example, sales response of -0.223 implies a percent change of $-[exp(0.223) - 1] = -25\%$). I multiply the percent change with the aggregate value of the variable reported in the corresponding year to estimate the amount by which it was overreported in the year. Row 4 displays the sum of Rows 1-3, and Row 5 expresses the evaded amount as a percent of the potential revenue in the year. I calculate the potential revenue as the tax due on the value added reported in the year plus the evaded amount given in row 4. Panel B estimates the first term on the RHS of formula (6), illustrating the level of refund overclaimed in the treated industries. Row 6 reproduces the exports response computed in Row 2. Using the production function of exports estimated in Figure V, Row 7 converts the exports response into the amount by which purchases meant for export were overreported. Row 8 shows the PKR value of the excess claimed refund (negative of Row7). Row 9 expresses the excess refund as a percent of the potential revenue in the year. All amounts 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).

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 non-export sales made by a registered firm in a given tax period (month). It equals $\tau.\left(\hat{s}_{it} - \hat{s}_{E,it}\right)$, where τ is the applicable VAT rate and $\left(\hat{s}_{it} - \hat{s}_{E,it}\right)$ is the value of non-export 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.

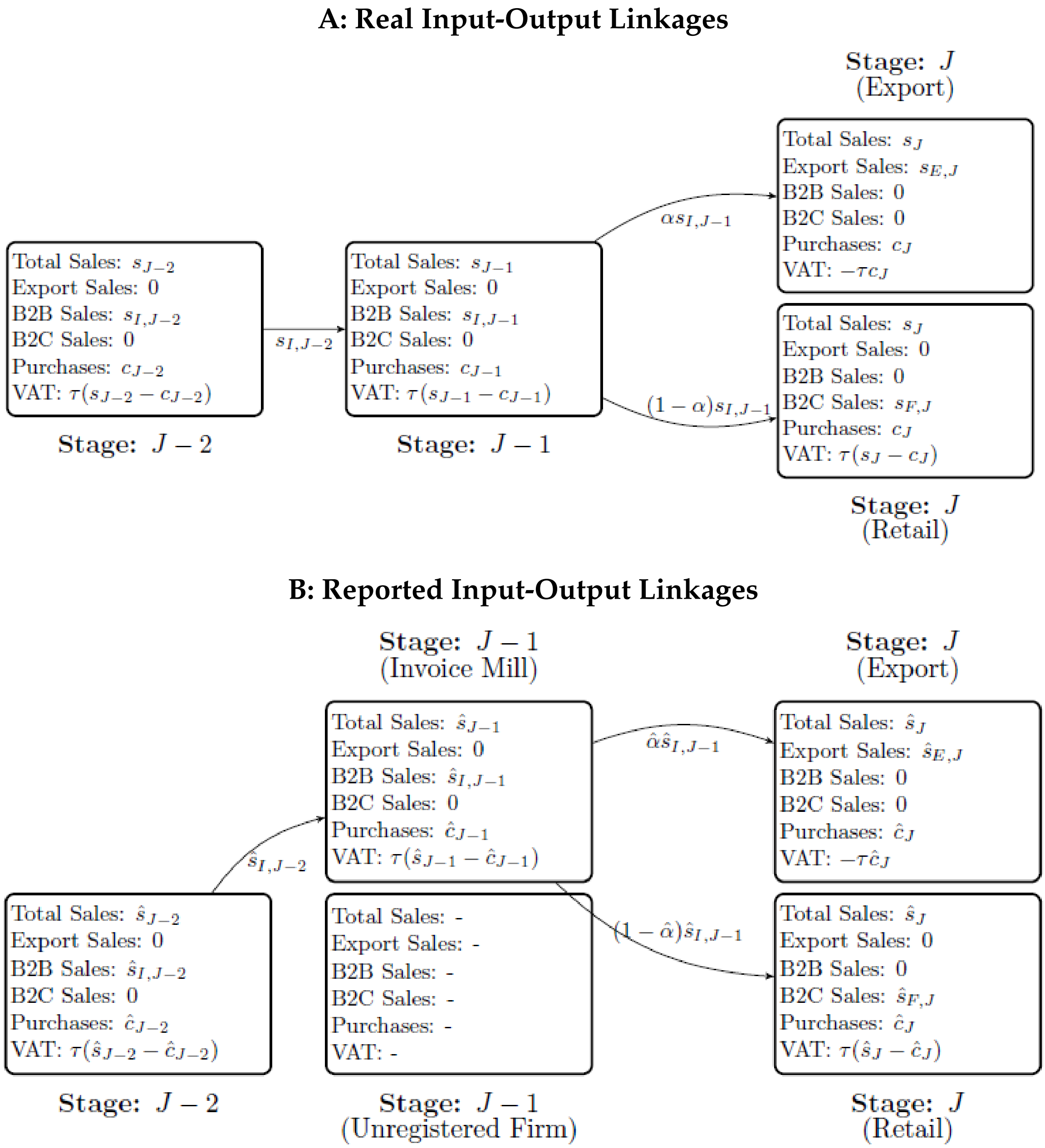
(*ix*) **Retailer.** A person, supplying goods to general public for the purpose of consumption.

(*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.

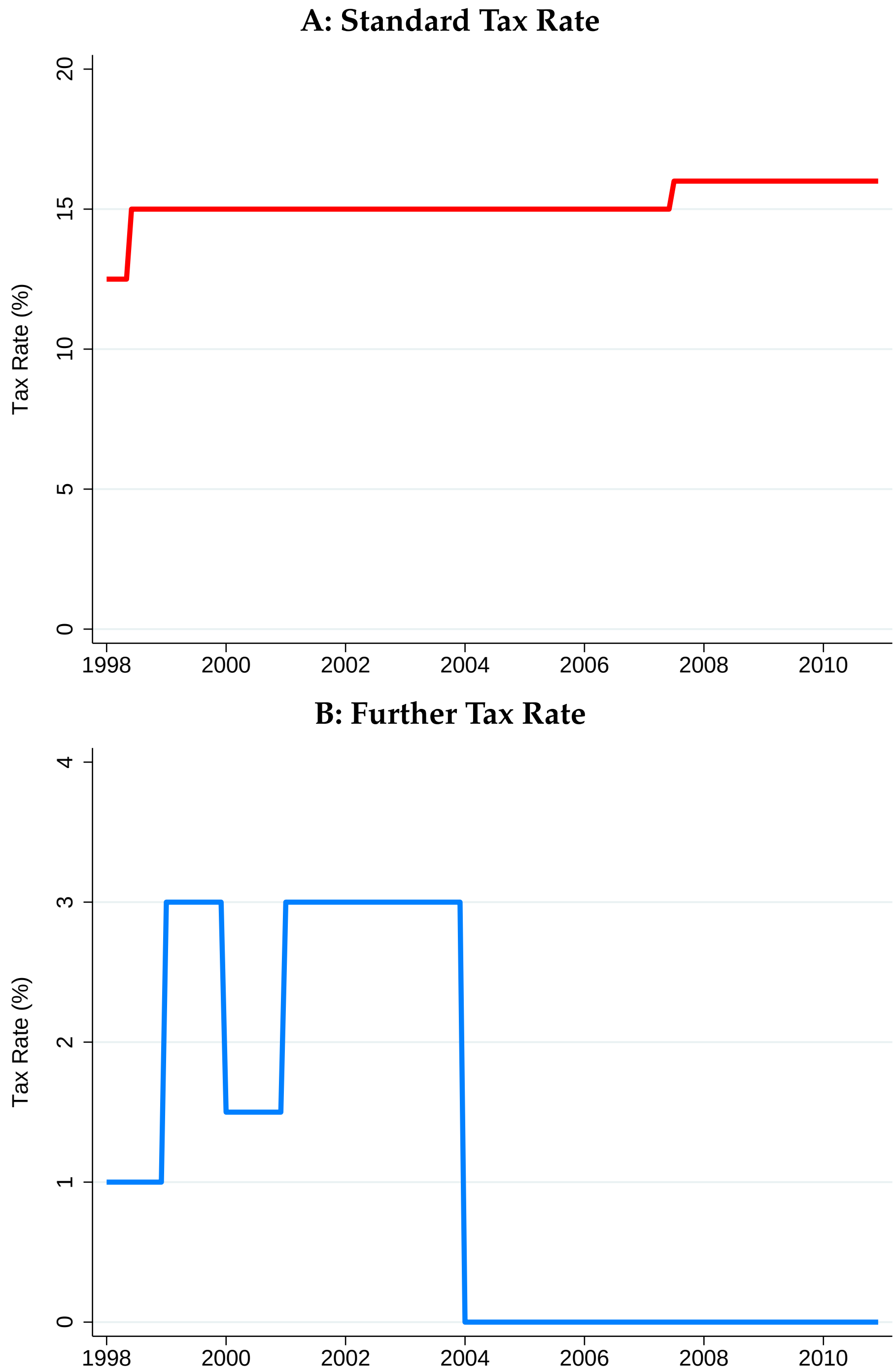
(*xii*) **Initial Capital.** The value of initial capital of the firm, as reported by it at the time of registration for VAT.

FIGURE A.I: **HOW INVOICE MILLS WORK**



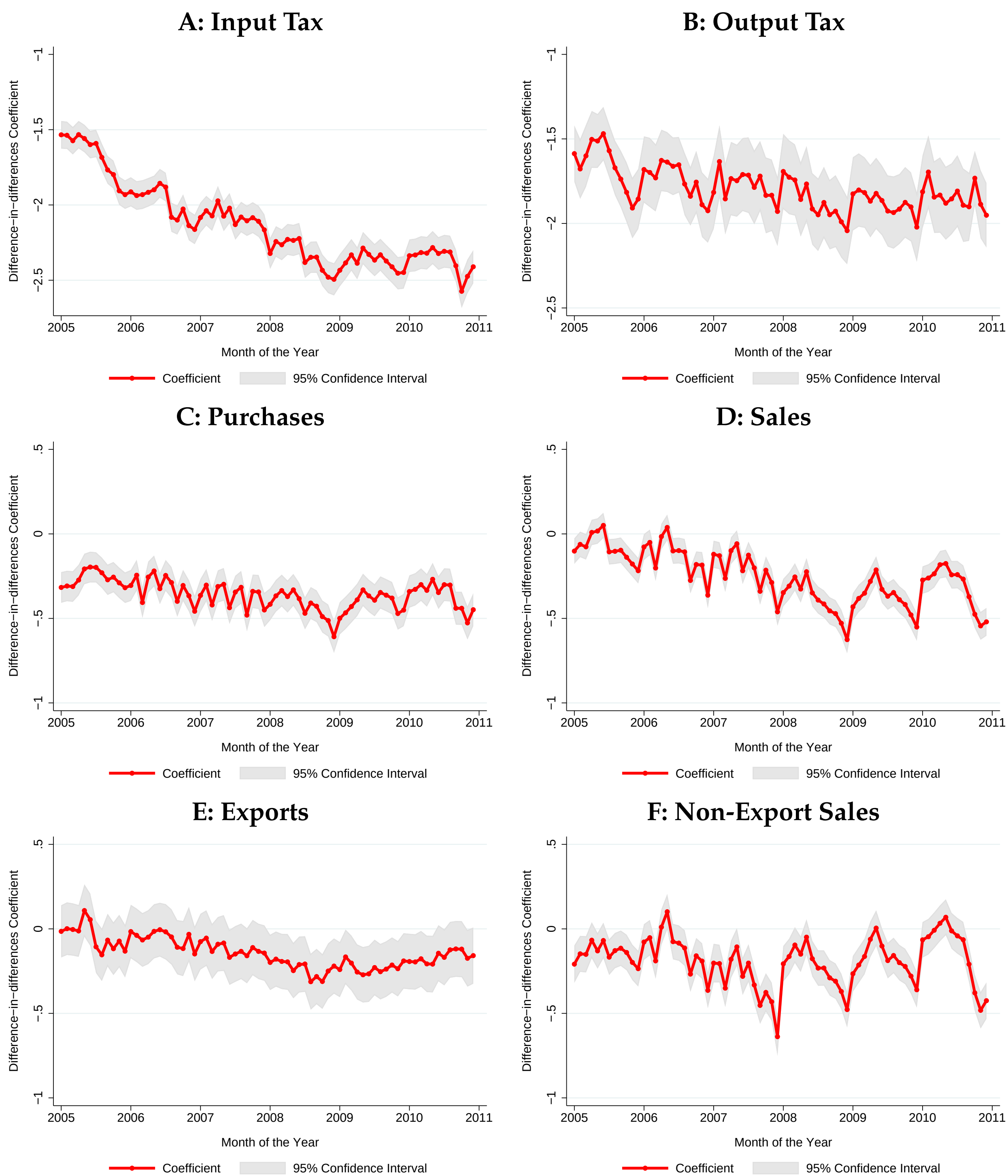
Notes: This figure shows how an invoice mills work. The top panel displays real input-output linkages in the last three stages of a supply chain. The final stage of the chain consists of two firms: the bottom firm deals exclusively in domestic sales and the top in exports. The middle firm sells a proportion α of its output to the exporter and the rest to the retailer. The middle firm is not in the VAT net. The two firms in the final stage thus cannot adjust input tax paid at the $J-2$ and previous stages. To exploit the gap, an invoice mills places itself in the stage $J-1$. It buys invoices from the bottom firm and sells them to the firms in the final stage. The bottom panel displays the reported input-output linkages in this setup. Given that the exporting firm has a greater incentive to exaggerate the value of its purchases, the invoice mills shifts a greater proportion of tax remitted at the bottom stages toward the exporting firms i.e. $\hat{\alpha} > \alpha$.

Figure A.II: **VAT Rates**



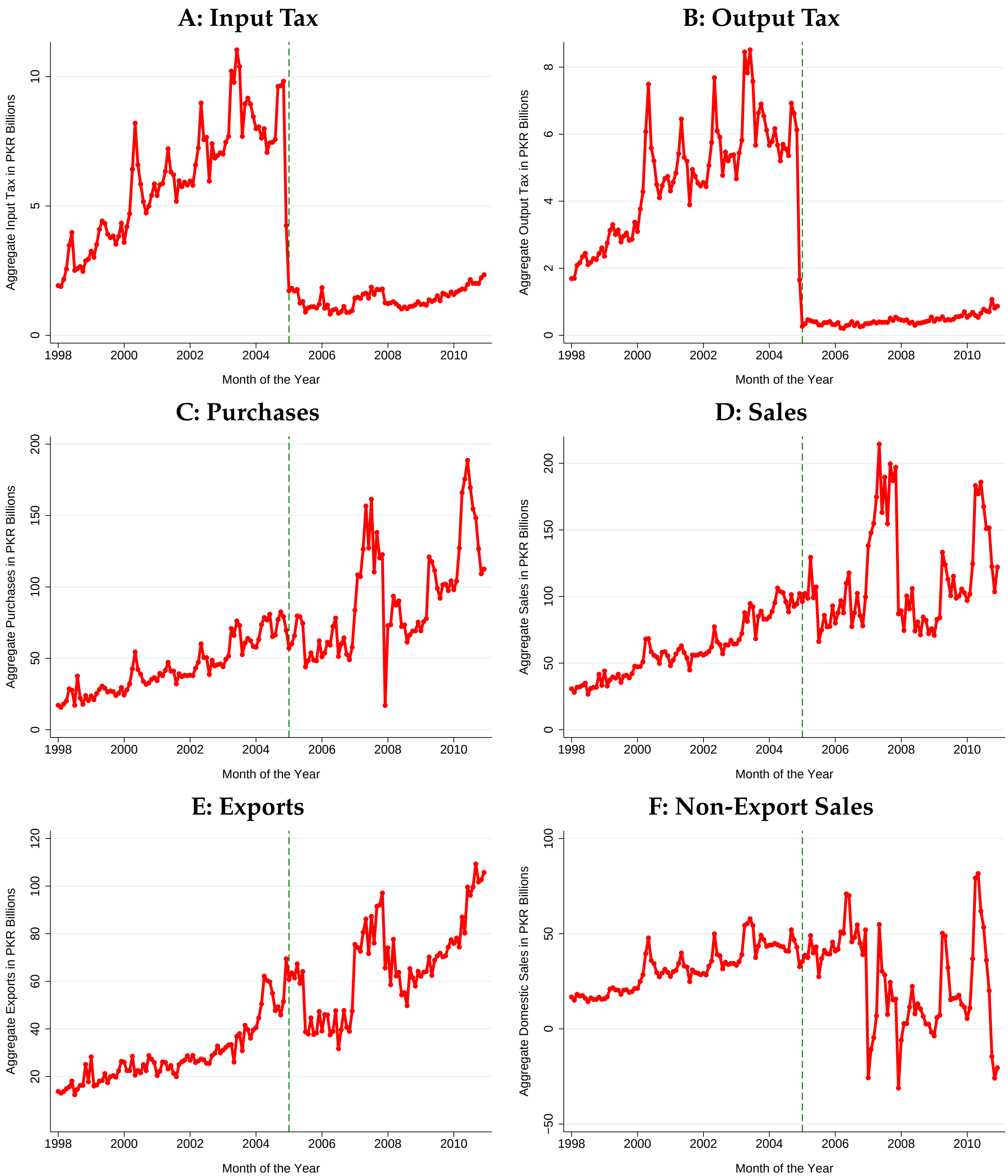
Notes: The figure shows the VAT rates applicable in Pakistan from July 1998 to June 2011. Panel A shows the standard VAT rate, which is applied to all sales made by a registered firm regardless of whether the recipient is registered or not. The rate largely stayed at 15%, but was increased to 16% from 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.III: **FIRM BEHAVIOR TO THE TAX CUT**



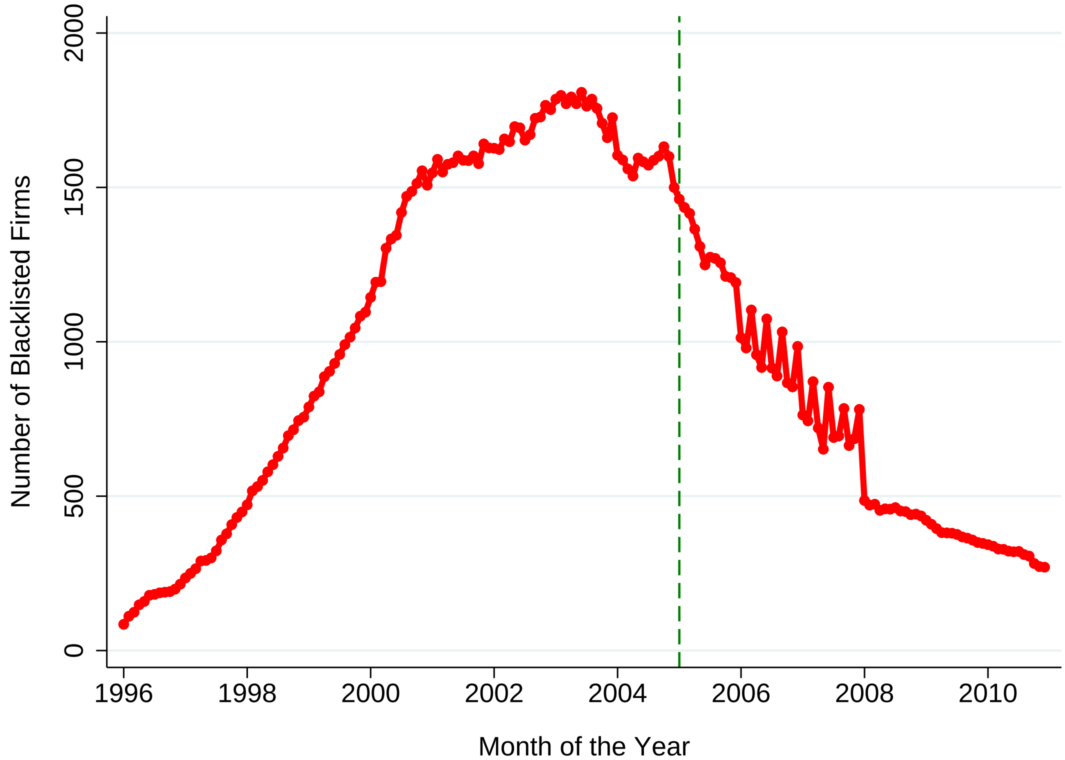
Notes: The figure reproduces the event study results reported in Figure III. The only difference between the two figure is that I show only the post-reform periods here. To construct these charts, I regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, 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.IV: **AGGREGATE VALUES OF VAT OUTCOMES – EXCLUDING INVOICE MILLS**



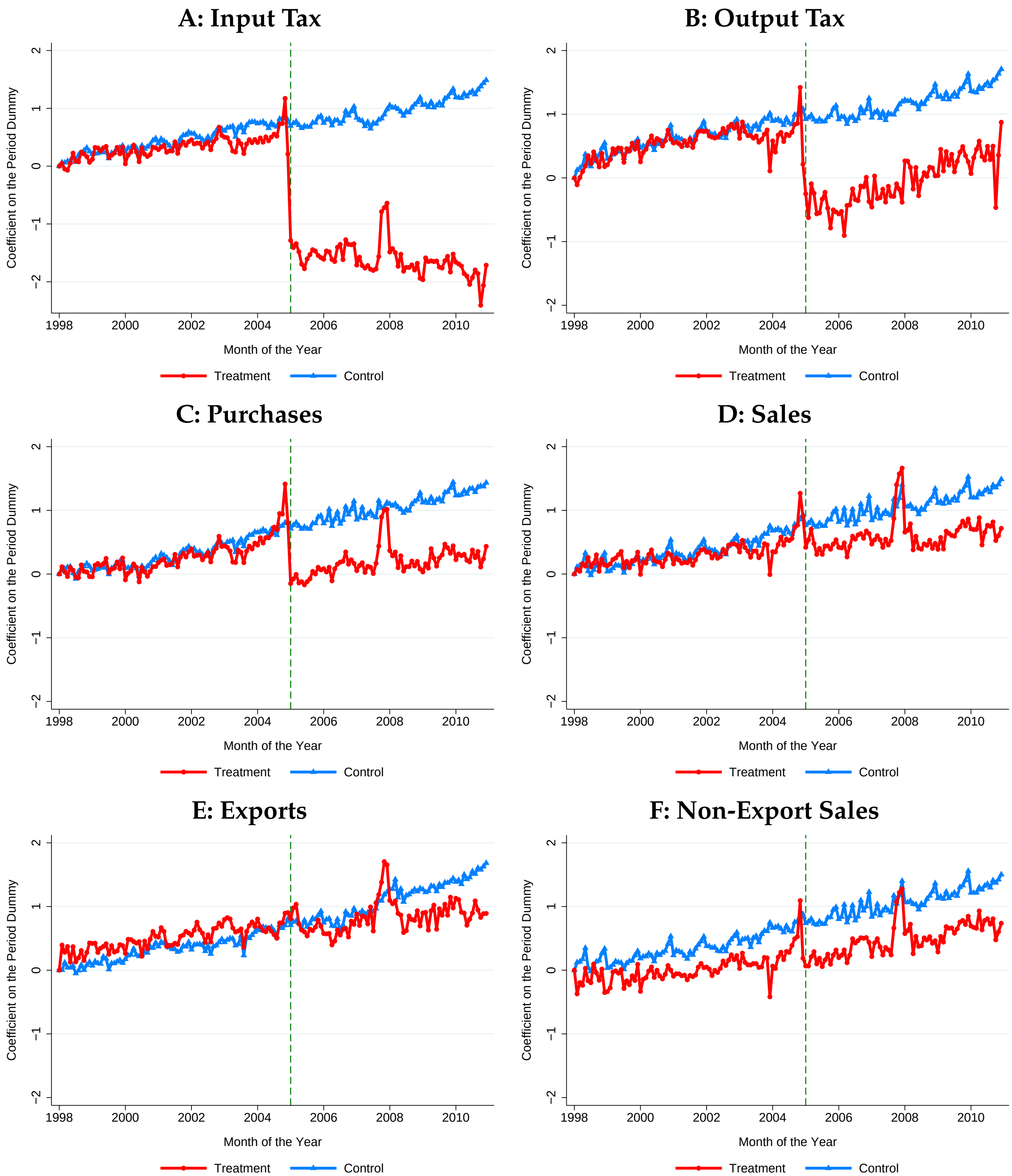
Notes: The figure illustrates how VAT outcomes evolve around the time of the zero-rating reform. The sample contains all firms operating in the zero-rated industries other than the blacklisted and suspended firms. Each panel of the figure shows the aggregate value of the outcome indicated in the title 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 A.V: **EMERGENCE, GROWTH, AND DECLINE OF INVOICE MILLS**



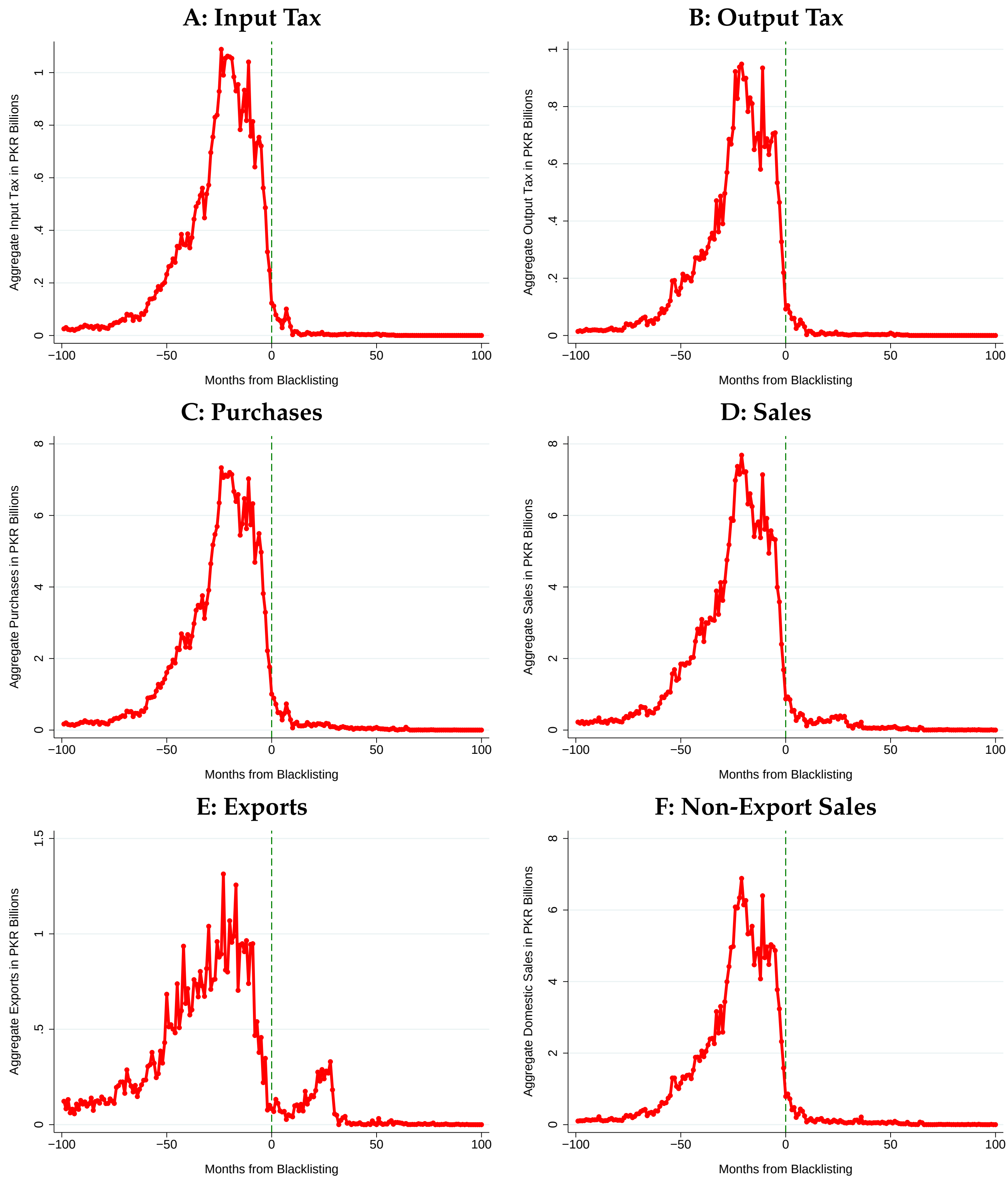
Notes: The figure shows the emergence, growth, and subsequent decline of invoice mills in the Pakistani setting. The sample begins from 1996, when a *broad-based* VAT with coverage extended to almost entire manufacturing and exports stages begun in the country. Each marker in the curve denotes the number 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 A.VI: **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.

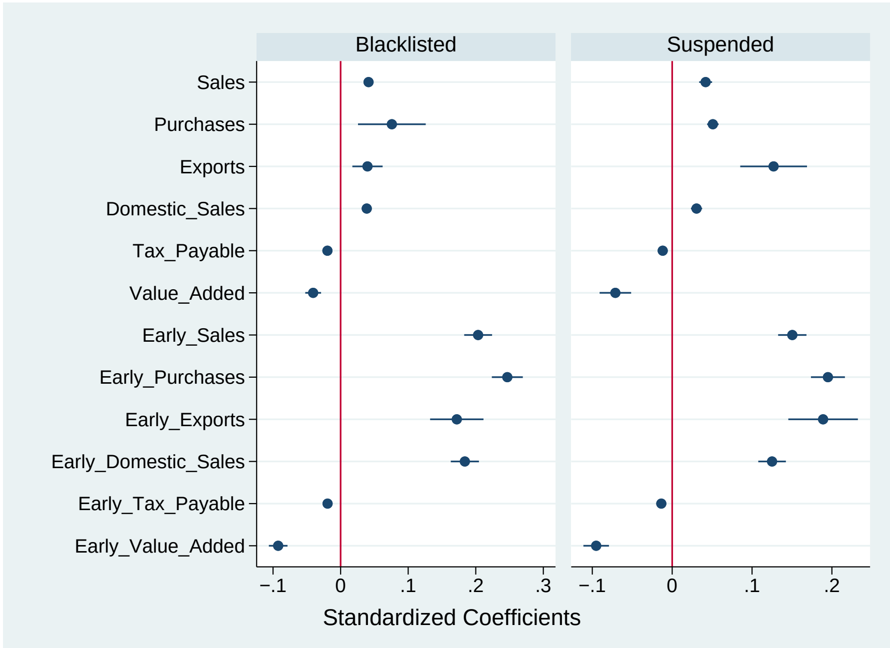
FIGURE A.VII: **AGGREGATE VALUES OF VAT OUTCOMES – INVOICE MILLS**



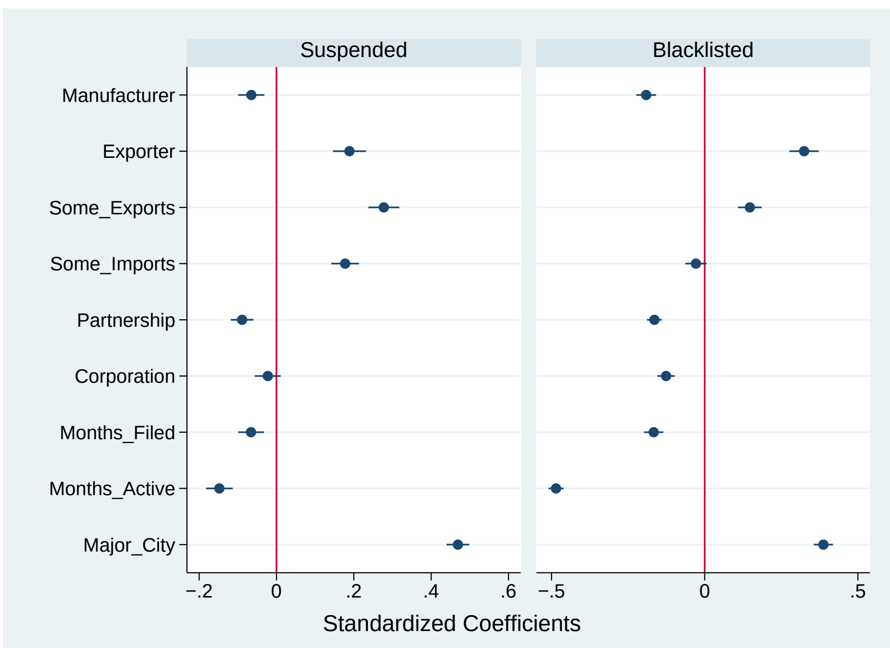
Notes: The figure explores the change in activity carried out by blacklisted firms around the event of blacklisting. The sample contains all blacklisted and suspended firms of the treated industries. Time 0 in the horizontal axis (marked by the vertical, dashed line) denotes the month in which the firm 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.

FIGURE A.VIII: **DISTINGUISHING MARKERS OF INVOICE MILLS**

A: VAT Outcomes



B: Firm Characteristics



Notes: The figures explores if invoice mills have any distinguishing markers that could help identify them. I compare the VAT outcomes reported by (Panel A) and and firm characteristics of (Panel B) blacklisted and suspended firms with other firms. The sample here consists of pre-reform years only (1998-2004). To construct these plots, I standardize each variable by subtracting the mean and dividing by the standard deviation of each variable. I then regress the standardized variable on two dummies, indicating if the firm is a blacklisted or a suspended firm. The coefficients on these dummy variables along with the standard errors are plotted with the heading of each row indicating the LHS variable in the regression. Early here denotes the first six months after the firm's registration For example Early_sales denotes the average sales reported by the firm in its first six returns after registration. The definitions of the compared variables are provided in Appendix A.1.

TABLE A.I: **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 A.II: **FIRM BEHAVIOR TO THE TAX CUT**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
A: Baseline Specification						
treat × after	-1.961	-1.842	-0.419	-0.223	-0.106	-0.082
	(0.176)	(0.380)	(0.036)	(0.034)	(0.041)	(0.046)
Elasticity	-12.065	-11.334	-2.578	-1.369	-0.651	-0.504
	(1.081)	(2.338)	(0.221)	(0.207)	(0.255)	(0.284)
Observations	3,728,658	4,179,560	3,983,211	5,058,578	612,993	4,623,906
B: Placebo Specification						
treat × after	-0.004	0.001	0.052	0.019	-0.022	-0.014
	(0.050)	(0.053)	(0.053)	(0.042)	(0.054)	(0.044)
Elasticity	-0.026	0.005	0.318	0.117	-0.134	-0.083
	(0.310)	(0.327)	(0.326)	(0.259)	(0.333)	(0.268)
Observations	1,999,985	2,423,814	1,979,622	2,583,423	306,931	2,381,447
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table explores how firms respond to the reduction of the rate applicable to their supply chain to zero. The coefficient treat × after shows $\hat{\gamma}$ from model (8), where the dummy variable treat$_i$ denotes that firm i belongs to a zero-rated industry and the dummy variable after$_t$ that month t falls in the tax year 2005 and later. The coefficient $Elasticity$ shows $\hat{\gamma}$ from the same model (8) when I replace the double difference term with $\log(1-\tau_{it})$. Panel B shows results from parallel placebo regressions, where I restrict the sample to pre-reform years only, defining the period beginning from July 2002 as the $after$ period. Standard errors are in parenthesis, which have been clustered at the industry level.

TABLE A.III: **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)
A: Response						
treat $\times$ 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 $\times$ after $\times$ 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
B: Elasticity						
log $(1-\tau)$	-12.169	-11.844	-2.425	-1.358	-0.401	-0.503
	(0.162)	(0.465)	(0.103)	(0.079)	(0.164)	(0.084)
log $(1-\tau)\times$ non-textile	1.602	5.893	-2.479	-0.141	-1.377	-0.009
	(0.628)	(1.208)	(0.481)	(0.247)	(0.258)	(0.499)
Baseline Coefficient	-12.065	-11.334	-2.578	-1.369	-0.651	-0.504
	(0.158)	(0.439)	(0.103)	(0.076)	(0.154)	(0.083)
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
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table decomposes the average response to the zero-rating reform estimated in Table I into its constituent textile and non-textile components. I estimate a triple-differences version of model (8), 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. Baseline coefficient reports the $treat \times after$ or log $(1-\tau)$ coefficient I obtain from estimating the baseline model (8). Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.IV: **FIRM BEHAVIOR TO THE TAX CUT – BALANCED PANEL**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
A: Baseline Specification						
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)
Elasticity	-14.382	-15.606	-2.977	-2.494	-0.973	-0.620
	(0.267)	(0.857)	(0.193)	(0.150)	(0.227)	(0.180)
Observations	948,385	877,354	981,954	1,126,539	264,719	960,697
B: Placebo Specification						
treat × after	-0.076	-0.009	-0.008	0.008	-0.031	-0.051
	(0.021)	(0.024)	(0.021)	(0.019)	(0.034)	(0.024)
Elasticity	-0.468	-0.055	-0.048	0.047	-0.189	-0.314
	(0.128)	(0.146)	(0.131)	(0.118)	(0.212)	(0.147)
Observations	560,180	562,860	564,044	646,637	142,345	564,505
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table explores how firms respond to the reduction of the rate applicable to their supply chain to zero. The sample here is restricted to a balanced panel of firms that file at least once in every quarter included in the sample. The coefficient treat × after shows $\hat{\gamma}$ from model (8), where the dummy variable treat_i denotes that firm i belongs to a zero-rated industry and the dummy variable after_t denotes that month t falls in the tax year 2005 and later. The coefficient $Elasticity$ shows $\hat{\gamma}$ from the same model (8) when I replace the double difference term with $\log(1-\tau)$. Panel B shows results from parallel placebo regressions, where I restrict the sample to pre-reform years only, defining the period beginning from July 2002 as the $after$ period. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.V: **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)
<u>A: Complete Panel</u>									
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
<u>B: Balanced Panel</u>									
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
Month 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 (8). 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.VI: **SPILLOVER EFFECTS ON THE NON-TREATED INDUSTRIES – ALTERNATIVE SPECIFICATION**

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)
<u>A: Complete Panel</u>									
treat × after	-0.420	-0.222	-0.105	-0.420	-0.222	-0.105	-0.426	-0.222	-0.103
	(0.017)	(0.012)	(0.025)	(0.017)	(0.012)	(0.025)	(0.017)	(0.012)	(0.025)
adjoining × after	-0.242	0.015	0.068	-0.065	0.044	0.015	-0.067	0.012	0.046
	(0.119)	(0.082)	(0.165)	(0.082)	(0.062)	(0.125)	(0.022)	(0.019)	(0.075)
Observations	3,983,213	5,058,579	612,993	3,983,213	5,058,579	612,993	3,983,213	5,058,579	612,993
<u>B: Balanced Panel</u>									
treat × after	-0.485	-0.405	-0.154	-0.483	-0.405	-0.155	-0.496	-0.404	-0.155
	(0.031)	(0.024)	(0.037)	(0.031)	(0.024)	(0.037)	(0.032)	(0.025)	(0.037)
adjoining × after	-0.154	-0.020	0.428	0.047	0.039	0.186	-0.100	0.012	0.070
	(0.168)	(0.131)	(0.318)	(0.121)	(0.108)	(0.226)	(0.037)	(0.031)	(0.119)
Observations	981,954	1,126,539	264,719	981,954	1,126,539	264,719	981,954	1,126,539	264,719
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month 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 estimate an augmented version of the difference-in-differences model (8), including the interaction term adjoining × after into the model. The dummy variable adjoining_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 *adjoining*. 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.VII: **FIRM BEHAVIOR TO THE TAX CUT – WITHOUT BLACKLISTED FIRMS**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
A: Complete Panel						
treat $\times$ after	-1.953	-1.871	-0.406	-0.221	-0.098	-0.079
	(0.026)	(0.074)	(0.017)	(0.013)	(0.026)	(0.014)
Elasticity	-12.016	-11.515	-2.499	-1.357	-0.603	-0.487
	(0.161)	(0.453)	(0.105)	(0.077)	(0.158)	(0.084)
Observations	3,562,585	4,020,573	3,808,066	4,865,461	584,747	4,451,965
B: Balanced Panel						
treat $\times$ after	-2.316	-2.530	-0.469	-0.402	-0.152	-0.101
	(0.044)	(0.141)	(0.032)	(0.025)	(0.038)	(0.029)
Elasticity	-14.250	-15.569	-2.886	-2.473	-0.938	-0.622
	(0.270)	(0.868)	(0.196)	(0.152)	(0.231)	(0.181)
Observations	923,512	857,195	955,603	1,098,880	256,765	938,355
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table explores how firms respond to the reduction of the rate applicable to their supply chain to zero. The sample here excludes blacklisted and suspended firms. The coefficient treat $\times$ after shows $\hat{\gamma}$ from model (8), where the dummy variable treat$_i$ denotes that firm i belongs to a zero-rated industry and the dummy variable after$_t$ that month t falls in the tax year 2005 and later. The coefficient $Elasticity$ shows $\hat{\gamma}$ from the same model (8) when I replace the double difference term with $\log(1-\tau_{it})$. Panel B shows results from parallel placebo regressions, where I restrict the sample to pre-reform years only, defining the period beginning from July 2002 as the $after$ period. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.VIII: **FIRM BEHAVIOR TO THE TAX CUT – FROM DATE OF ANNOUNCEMENT**

	Input Tax	Output Tax	Purchases	Sales	Exports	Non-Export Sales
	(1)	(2)	(3)	(4)	(5)	(6)
A: Complete Panel						
treat × after	-1.944	-1.793	-0.422	-0.227	-0.114	-0.113
	(0.025)	(0.064)	(0.017)	(0.012)	(0.026)	(0.013)
Elasticity	-11.964	-11.034	-2.594	-1.395	-0.699	-0.697
	(0.156)	(0.395)	(0.102)	(0.075)	(0.157)	(0.083)
Observations	3,728,660	4,179,561	3,983,213	5,058,579	612,993	4,623,907
B: Balanced Panel						
treat × after	-2.325	-2.476	-0.485	-0.404	-0.156	-0.121
	(0.043)	(0.131)	(0.031)	(0.024)	(0.037)	(0.029)
Elasticity	-14.304	-15.234	-2.984	-2.485	-0.963	-0.744
	(0.264)	(0.806)	(0.192)	(0.149)	(0.228)	(0.179)
Observations	948,385	877,354	981,954	1,126,539	264,719	960,697
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table explores how firms respond to the reduction of the rate applicable to their supply chain to zero. The only difference from the baseline specification (results reported in Table I) is that the $after$ period here begins from the date of announcement of the reform i.e. from June 2005. The coefficient treat × after shows $\hat{\gamma}$ from model (8), where the dummy variable treat$_i$ denotes that firm i belongs to a zero-rated industry and the dummy variable after$_t$ that month t is June 2005 or a later period. The coefficient $Elasticity$ shows $\hat{\gamma}$ from the same model (8) when I replace the double difference term with $\log(1 - \tau_{it})$. Panel B shows results from parallel regressions, where I restrict the sample to a balanced panel of firms that report in every quarter included in the sample. Standard errors are in parenthesis, which have been clustered at the firm level.

Table A.IX: **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)
A: Response						
treat $\times$ 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 $\times$ after $\times$ 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
A: Response						
$\log(1-\tau)$	-10.740	-6.981	-2.505	-0.910	-0.606	-0.347
	(0.187)	(0.365)	(0.111)	(0.081)	(0.173)	(0.086)
$\log(1-\tau) \times$ corporation	-3.683	-12.141	-0.248	-1.670	-0.106	-0.611
	(0.311)	(0.949)	(0.228)	(0.165)	(0.228)	(0.202)
Baseline Coefficient	-12.065	-11.334	-2.578	-1.369	-0.651	-0.504
	(0.158)	(0.439)	(0.103)	(0.076)	(0.154)	(0.083)
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
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table rules out lazy reporting as an alternative explanation of the responses documented in Table I. I divide the treatment sample into corporate and non-corporate firms. I then estimate a triple-differences version of model (8), including all double interactions of the binary variables treat$_i$, corporation$_i$, and after$_t$ and their triple-interaction. Baseline coefficient reports the $treat \times after$ or $\log(1-\tau_{it})$ coefficient I obtain from estimating the baseline model (8). Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.X: **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)
A: Response						
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
B: Elasticity						
log $(1-\tau)$	-11.989	-11.506	-2.467	-1.339	-0.586	-0.471
	(0.161)	(0.453)	(0.104)	(0.077)	(0.157)	(0.084)
log $(1-\tau)$ × blacklisted	1.277	4.289	-2.923	-1.884	-1.920	-2.738
	(1.290)	(1.798)	(0.771)	(0.598)	(1.411)	(0.633)
log $(1-\tau)$ × suspended	-2.678	4.271	-2.299	-0.423	-1.343	-0.262
	(0.873)	(2.055)	(0.553)	(0.424)	(0.407)	(0.564)
Baseline Coefficient	-12.065	-11.334	-2.578	-1.369	-0.651	-0.504
	(0.158)	(0.439)	(0.103)	(0.076)	(0.154)	(0.083)
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
Month 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 (8), 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.XI: **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.