

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

Mazhar Waseem\*

October 2019

## Abstract

I leverage a Pakistani tax reform that cuts the tax rate on the supply chains of five major industries of the country from 15% to 0% to estimate the extent of VAT noncompliance in the country. I find that at the baseline the treated industries overclaimed refunds by nearly 22% and underreported domestic B2C sales by nearly 43.5%. Together, this caused a total VAT revenue loss of PKR 38 billion, which translates into 11.5% (77% of the statutory rate of 15%) of the *true* taxable sales of these industries. I also explore the role of invoice mills in facilitating this noncompliance. Invoice mills are firms that engage in no real business activity and exist solely to trade in VAT invoices. I find that the amount of excess refund claimed by exporters nearly equals the amount of input tax involved on invoices issued to them by invoice mills, suggesting that almost all of revenue loss on export refunds takes place through the medium of invoice mills.

**Keywords:** VAT, Tax Evasion, Firm Behavior

**JEL Classification:** H25, H26, H32

---

\*Email Address: Mazhar Waseem [mazhar.waseem@manchester.ac.uk](mailto: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 the income tax in both rich and emerging economies. In fact, its share in government revenue is rising steadily ([OECD, 2017](#); [International Tax Dialogue, 2013](#)). This apparent discord between theory and practice in large part reflects the almost universally held belief that VAT is a harder tax to evade. The perceived robustness of VAT to tax evasion has been a driving force behind its remarkable expansion over the last half century.

Recently, however, a few cracks have begun to appear in this consensus on the desirability of VAT and its superiority among consumption taxes in terms of enforcement. 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 similar change is scheduled to take place in Zambia in 2019-2020.<sup>1</sup> The discontent with 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, 2019b](#)). This final mile problem of the VAT is more severe in countries where the final production stage is fragmented, being dominated by small informal firms. Second, the destination-based design of VAT requires that any tax collected on inputs be refunded to exporters. Refund payment makes the VAT vulnerable to distinct forms of frauds not seen in other tax instruments, such as the missing trader fraud ([Keen & Smith, 2006](#)).

While the theoretical mechanisms underlying VAT noncompliance are well-known, we still do not understand fully how important they are empirically. In fact, VAT's evasion in general is much less understood than income tax's. For example, there is little micro-based evidence from tax return data on how much VAT gets evaded and what mechanisms underlie it. In comparison, such evidence on income tax has been available for some time from both policy studies and academic research (see for example [Slemrod, 2007](#); [Kleven et al., 2011](#); [Artavanis et al., 2016](#); [Waseem, 2019a](#)).

---

<sup>1</sup>Please see Richard Asquith's blog at [avalara.com](http://avalara.com) for the policy changes in Malaysia and Zambia. Specifically, the Malaysian change is documented [here](#) and the Zambian [here](#).

This paper fills this gap in literature. It leverages a novel tax reform from Pakistan to estimate the extent of VAT revenue loss caused by tax evasion. It also estimates the individual contribution of the two key mechanisms—overclaimed refunds and undeclared sales—in the aggregate revenue loss.

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

VAT can be evaded by overreporting purchases or underreporting sales. The reform I exploit is novel in the sense that it reduces the rate applicable on both intermediates acquired and final goods sold by these industries to zero. The reduction of the rate to zero weakens considerably if not removes entirely the incentive to misreport. As a result sales and purchases reported in the treated industries would approach their true values in the post-reform periods. This allows me to infer the extent of evasion as it existed in the treated industries at the baseline by comparing firm behavior to the reform across treated and untreated industries. I put this empirical strategy to use, exploiting administrative data comprising the universe of VAT returns and the tax register. The data have a panel structure: the VAT return is filed each month and contains a unique firm identifier. The tax register indicates, among other things, the industry a firm belongs to, which lets me divide firms into treated and untreated groups. Both sets of data are available for a large number of pre- and post-reform years.

I use the standard difference-in-differences methodology for my empirical analy-

sis, comparing outcomes across firms in the five treated industries with the rest. All specifications include a full set of firm and time fixed effects. Identification in this setup requires that conditional on the firm fixed effects the time path of outcomes would have been similar across the compared groups absent the tax reform. Using the standard event study plots spanning 84 pre-reform and 72 post-reform periods, I show that this indeed is highly plausible in my setting. I run a battery of additional tests to rule out other identification concerns, including any spillovers between the two groups.

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

Given that exports have negative VAT liability, total revenue loss caused by misreporting in a VAT system is the sum of overclaimed refund on exports and underpaid tax on domestic consumption. To compute the first component one needs to estimate the extent by which purchases of intermediates used in the production of exports are overreported and for the second the extent by which domestic B2C sales are underreported. Using the responses documented above, I estimate that purchases used for exports were overreported by nearly 22%. The extent of evasion on domestic B2C sales was roughly twice this amount, which were underreported by around 43.5%. Together, this caused a total VAT revenue loss of PKR 38 billion in 2004, which is nearly 13% of the total VAT collected in the country in that year. The evaded amount translates into 11.5% (77% of the statutory rate of 15%) of the *true* aggregate B2C sales reported at the baseline in the treated industries.

VAT supply chains are rarely complete. The presence of small informal firms, especially in developing economies, means that in almost every supply chain some part of the production process occurs in the unregistered sector. When the VAT chain breaks, the tax charged at the pre-break stages cannot be claimed back in the post-break stage. This creates arbitrage opportunities, which are sometimes exploited by firms called invoice mills (Keen & Smith, 2006). These firms exist solely to trade in VAT invoices. They buy invoices from firms whose real buyers do not need them because they are either consumers or unregistered firms. They sell invoices to firms who either buy their intermediates from the unregistered sector or who want to reduce their tax liability by overreporting purchases.

Invoice mills are a poorly understood phenomenon. To my knowledge, there is no micro-based evidence in the existing literature on how they operate and how much revenue loss they cause by pumping spurious invoices into a VAT system. The key reason behind this lack of evidence is that invoice mills are difficult to identify. In the Pakistani context, a legal process used by the tax administration helps me identify such firms cleanly. The legal process arose out of practical challenges posed by such firms and consists of a two-pronged strategy. The tax administration suspends the registration of a firm whom it suspects of being involved in the purchase or sale of invoices. Post-suspension, the firm is afforded an opportunity to defend itself in quasi-judicial proceedings, at the culmination of which either the firm's registration is restored or it is blacklisted permanently. I observe both suspension and blacklisting in my data and use them to determine if a firm is an invoice mill. The data show that invoice mills appeared soon after the destination-based VAT was implemented in the country. Their numbers grew steadily over time before crashing once the zero-rating reform came into effect. The analysis further show that 97% of their sales are to exporters, ending up as refund claims. In the baseline year of 2004, nearly two-fifths of the overclaimed refund amounting to PKR 8.6 billion was claimed on their invoices. The evidence thus shows that invoice mills are a critical channel through which fraudulent overclaim of VAT refund takes place.

This paper contributes to three strands of literature. First, starting with Pissarides & Weber (1989), there have been a series of papers that use observational or experimental data to estimate the extent of tax evasion in a variety of contexts including the UK, the US, Russia, Denmark, Greece, and Pakistan (see for example Feldman & Slemrod, 2007; Gorodnichenko *et al.*, 2009; Kleven *et al.*, 2011; Artavanis *et al.*, 2016; Waseem, 2018b). None of these papers, however, relates specifically to VAT. Under-

standing the nature and extent of VAT evasion is important both because mechanisms underlying it—given the destination-based design—are different and because the claim of its superiority among consumption taxes rests largely on its compliance properties. This paper is an effort in this direction, contributing well-identified estimates on the extent and nature of VAT noncompliance in a representative emerging economy. Second, this paper adds to a nascent literature begun by [Pomeranz \(2015\)](#) that studies the enforcement properties of VAT in weak tax capacity setting (see for example [Agrawal & Zimmermann, 2019](#); [Almunia \*et al.\*, 2019](#); [Eissa & Zeitlin, 2014](#); [Gadenne \*et al.\*, 2019](#); [Waseem, 2019b](#) ). In this line of literature, this paper is closest to [Shah \(2019\)](#), who uses the introduction of a software-based risk analysis program in Pakistan in 2013 to show that the enhanced monitoring of firms afforded by the program causes a nearly 50% reduction in input tax claimed by nonexporting firms of the country. Relative to [Shah \(2019\)](#), the focus of this paper is to identify aggregate noncompliance of VAT arising from all channels including the overclaim of input tax credit and to understand the mechanisms underlying it. Third, this paper adds to a growing body of empirical literature that uses administrative tax return data to identify the drivers of low tax compliance in developing and emerging economies (see for example [Bachas & Soto, 2019](#); [Brockmeyer \*et al.\*, 2019](#); [Carrillo \*et al.\*, 2017](#); [Slemrod \*et al.\*, 2019](#); [Waseem, 2018a](#)).

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

## II Conceptual Framework

This paper uses a large rate cut from Pakistan to identify the nature and extent of tax evasion under a VAT. This section develops a simple framework to guide this exercise. The framework is developed through examples of firm behavior under three different settings. Although very stylized, the examples cover a broad spectrum of noncompliance witnessed in an emerging economy VAT. The examples are used to

characterize the key forces driving VAT noncompliance and develop a formula to compute the revenue loss arising from it.

## II.A Setup

Consider a simple supply chain consisting of three production stages shown in Figure I. For simplicity, I assume that each of these production stages contains one firm only. The firm in the middle stage uses the intermediate produced by the bottom-stage firm to produce a consumption good, a proportion  $\alpha$  of which is sold to the export stage for onward supply to foreign buyers. The rest of the good serves the domestic market through the final production stage (retail). I denote true sales and purchases of the firm in stage  $J \in \{B, M, E, R\}$  by  $s_J$  and  $c_J$  and its reported sales and purchases by  $\hat{s}_J$  and  $\hat{c}_J$ . Purchases here represent the value of taxable intermediates acquired by the firm, which do not include nontaxable inputs such as labor.

## II.B First-Best Benchmark

I first consider behavior under the first-best scenario, where all firm report their sales and purchases truthfully ( $\hat{s}_J = s_J$ ;  $\hat{c}_J = c_J$ ;  $\forall J$ ). Panel A of Table I illustrates the input-output linkages between firms under this and other scenarios considered below. Since there is no misreporting in the benchmark case, the theoretical result of equivalence between the VAT and the retail sale tax applies. The government receives a total revenue of  $\tau \cdot s_R$  from the entire supply chain, which intuitively equals the tax rate times the retail sales (B2C sales): the value of consumption sales in the domestic market.

## II.C Undeclared Sales and Overclaimed Refunds

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

VAT, almost as a principle, is implemented as a destination-based tax (Ebrill *et al.*, 2001). Under the destination principle, commodities are taxed in the country they are consumed instead of where they are produced. This necessitates that imports into the country be taxed at the standard rate and exports at a zero rate. Zero-rating means that the seller does not pay any VAT on output but can still claim the tax paid on inputs as refund. While zero-rating serves the important purpose of ensuring that exports remain free of any domestic tax, it makes a VAT vulnerable to a distinct form of noncompliance not seen in other tax instruments (Keen & Smith, 2006). In this form of noncompliance, exporting firms take advantage of the destination based design to overclaim refunds.

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

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

Note that in this example the bottom-stage firm reports truthfully and the middle-stage firm remits the correct amount of VAT. The revenue loss therefore stems from two sources: (1) the middle firm falsifying the destination of its sales from domestic

consumption to exports, and (2) the retail firm underreporting its sales. The former type of misreporting is sometimes referred to as evasion through *flying* invoices by VAT administrators. It is an example of two-sided evasion meaning that the seller and buyer are in collusion, their VAT records match each other, and the input tax claimed on the VAT invoice has been remitted into the treasury. It is only that the firm claiming the credit is not the one which has consumed the inputs the *flying* invoice relates to. This form of evasion needs to be distinguished from the more serious case of one-sided evasion through *fake* invoices, where the tax credit claimed by the buyer has not even been remitted into the treasury.

It is important to emphasize that not allowing tax evasion at the bottom or middle stage though unrealistic does not make any material difference to formula (1). Ignoring one-sided evasion,<sup>2</sup> what matters for VAT revenue is how much final domestic sales (B2C sales) and purchases for export are reported in the economy. In our example, as long as reported retail sales and purchases claimed for exports remain  $\hat{s}_R$  and  $\hat{c}_E$ , allowing misreporting at the bottom or middle stage would not cause any change in the formula. For instance, if the bottom firm misreports its sales as  $\hat{s}_B < s_B$ , its effect would be canceled out as long as the middle firm reports purchases matching with the bottom firm's sales  $\hat{c}_M = \hat{s}_B$  (no one-sided evasion). Our stylized example therefore produces a more general result that VAT revenue loss caused by any form of noncompliance other than one-sided evasion equals the sum of underpaid tax on final domestic sales and overclaimed refund on export sales. Real world supply chains are longer and more complicated than the one in our example. In such supply chains, firms at each production stage may simultaneously engage in B2B, B2C, and export transactions. But the intuition developed in this example would still apply. Any misreporting on B2B transactions would cancel out so that the net tax evasion would still equal the sum of overclaimed refund on exports and underpaid tax on B2C transactions, as captured by the two terms in formula (1).

As I note above, the first term in formula (1) is the overclaimed VAT credit at the export stage. In its simplest form, the overclaim may mean that the exporter overreports its purchases only ( $\hat{s}_E = s_E$  and  $\hat{c}_E > c_E$ ). But there is a limit to which purchases can be overreported as persistent small or negative value-addition is difficult to justify, especially over a longer period of time. In a more serious version of such noncom-

---

<sup>2</sup>One-sided evasion results when a firm underreports its sales or overreports its purchases unilaterally, that is not in collusion with its buyer or supplier. It is a crude form of evasion, which can be detected through cross-matching of VAT invoices. Two-sided evasion, on the other hand, is robust to cross matching.

pliance, therefore, both exports and purchases are overreported ( $\hat{s}_E > s_E$ ;  $\hat{c}_E > c_E$ ). Overreporting exports allows the firm to enhance its VAT refund while remaining within a justifiable range of value addition.

## II.D Invoice Mills

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

Although very stylized, the above three examples broadly capture key mechanisms underlying VAT noncompliance. To reiterate, (i) negative tax liability at the export stage; (ii) the final-mile problem of VAT, whereby B2C transactions can be kept out of books at a relatively low cost; and (iii) broken VAT chains create incentives for firms to (1) overreport exports; (2) underreport final domestic sales (B2C trans-

actions); (3) overreport purchases; and (4) operate as invoice mills. My empirical application exploits a large tax reform that reduces the tax rate applicable to supply chains of five major industries of the country from 15% to 0%. As a result of this reform, no tax is payable on the domestic consumption of goods supplied by these industries and very little refund is to be claimed on their exports. The reform thus substantially weakens, if not removes entirely, any incentive for noncompliance documented above, moving the treated firms roughly from the environment in our last two examples to that of the first example. By looking at how the treated firms react to the reform, we can infer the extent of VAT noncompliance as it existed in the treated industries in the pre-reform years. This exercise is explained in greater details in section VI of the paper.

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

## III Contextual Background

### III.A Pakistani VAT System

Pakistan introduced the VAT in the 1990s. The legislation for this purpose was passed in July 1990, and although it envisaged the VAT to be a broad-based tax with standard features, the scope of the new tax in the initial few years was kept limited 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, 2019b).<sup>3</sup> 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.

---

<sup>3</sup>The energy sector was brought into the tax net in July 1999 and services in July 2000. Please see Waseem (2019b) for more details on the introduction and growth of VAT in Pakistan

As I note above, the Pakistani VAT follows the standard design. Firms whose annual turnover falls above the exemption threshold are required to register with the tax administration.<sup>4</sup> 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 and the payment of refund. The delay largely consists of the time taken by the tax administration to scrutinize and sanction the claim, although some of it could arise from the exporter taking time to file the supporting documents.

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

Panel A of Figure A.I plots the standard VAT rate in the country. The rate generally

---

<sup>4</sup>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.

remained at 15% till 2008 when it was increased to 16%. Pakistan introduced a policy in 1998 through which supplies made to unregistered *firms* were taxed at a higher rate. Of course, the higher rate was not applicable on supplies made to end consumers. Panel B of the figure plots the *additional* rate—called Further Tax by the tax code—imposed by the policy. The rate remained between 1 and 3 percentage points before it was eliminated in 2004.<sup>5</sup> The policy, as I explained in the third stylized example above, strengthens the incentive of a seller to falsify the destination of goods from the real to a fictitious buyer by issuing flying invoices. More specifically, in a sale transaction to an unregistered firm the seller can avoid the additional rate applicable by issuing the invoice in the name of a registered firm.

### III.B Zero-rating Reform

In July 2005, Pakistan introduced a novel tax reform through which the VAT rate applicable to supply chains of five major industries of the country—textile, leather, carpets, sports goods, and surgical goods—was reduced from 15% to 0%. Before the reform, in accordance with the standard destination-based design, only exports of these industries were zero-rated and their domestic supplies were taxed at the standard rate of 15%. The reform zero-rated not only the supplies of *final* goods produced by these industries but also of their major inputs. For example, in addition to the finished goods produced by the textile industry (fabric, garments, etc.) all its major inputs including ginned cotton, polyester, yarn, undyed fabric, and important dyes and chemicals were zero-rated. The purpose of the change was two-fold (FBR, 2005).<sup>6</sup> First, nearly three-quarters of the output of these industries was exported out of the country in one form or another and hence was already zero-rated. Long within-country supply chains of these industries, however, meant that VAT was to be remitted and claimed back whenever the goods moved from one production stage to the next. This created cash-flow problems for exporters who had to wait, sometimes

---

<sup>5</sup>For example, with a standard VAT rate of 15% and a further tax rate of 1%, supplies to registered firms would be charged to tax at a rate of 15% and to unregistered firms at a rate of 16%.

<sup>6</sup>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.

for long times, for the refund of VAT paid on their inputs.

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

The reform was announced on June 06, 2005 (applicable from July 01, 2005). It zero-rated 152 items, which included both finished goods and major inputs of the five industries. Where an input was included in the list, its supply became zero-rated regardless of whether it was used in the production of the five industries or otherwise. For this reason, only inputs *predominantly* used by the five industries were zero-rated.<sup>7</sup> 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*.<sup>8</sup> It thus seriously weakened, if not entirely eliminated, any incentive to misreport sales or purchases as documented in the three stylized examples above. As a consequence, firm behavior would approach the first-best case so that by comparing the post- and pre-reform behavior we can estimate the extent of VAT noncompliance as it existed at the baseline in the treated industries using formula (1).

---

<sup>7</sup>Otherwise, the loss of revenue from zero-rating would have been unsustainable for the government.

<sup>8</sup>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.

As noted above, the Pakistani tax administration began obtaining transaction-level data from firms from July 2008. This requirement was introduced by adding an annex to the VAT return wherein the firm was to provide the details of all its sales and purchases during the month, aggregating them up to the level of each supplier and buyer. Simultaneously, electronic filing of both VAT return and the annex was made mandatory for all firms. Together, these changes reduced the costs of cross-matching sales and purchase records considerably, enabling the tax administration to uncover any suspicious activity such as fake or flying invoice deep into the supply chain. This change in enforcement technology in 2008 affects both treated and untreated industries similarly and therefore should not matter in my empirical setting unless there is some interaction between it and the zero-rating reform.

### **III.C Invoice Mills**

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

In the initial few years after the adoption of VAT, Pakistan 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 non-compliance. Over time, however, the volume of such transactions in the system grew markedly, making it necessary to have a specific tailored mechanism to deal with the issue. The new mechanism, which came into force in July 2003, empowered the tax administration to suspend the registration of a firm it suspected of being involved in the issuance of fake or flying invoices. The suspension was meant to be a temporary measure aimed at protecting revenue while an inquiry of the firm's tax dealing could proceed. This inquiry was of a quasi-judicial nature, where the firm was confronted with the evidence against it and was afforded the opportunity to present its case. On completion of the inquiry, either the registration of the firm was restored or it was blacklisted permanently. Once a firm was blacklisted, its invoices no longer remained

valid and could not be used to claim input tax credit.

Figure II tracks the stock of blacklisted firms in the country, plotting the number of such firms that file their VAT return in the given tax period (indicated on the horizontal axis). As I mention above, blacklisting connotes that the firm is likely an invoice mill so that the plot roughly captures the evolution of invoice mills in the Pakistani VAT system. Blacklisted firms appeared soon after the destination-based VAT was implemented in the country. Their number rose sharply in the next few years, reaching a peak of around 1750 in 2003 (roughly 2-3% of all filed returns). The rising trend was partly reversed in 2003, when the new, dedicated mechanism to blacklist firms came into effect. But a sharper decline occurred after 2005, when the zero-rating reform, which eliminated the demand for fake and flying invoices in five major industries of the country, came into effect. Note that blacklisting began from the tax year 2003, and therefore most of the returns included in this plot were filed at the time the firms had still not been declared blacklist and invoices issued by them were still legal tender. In section VII below, I break down these firms' behavior around the event of blacklisting, looking at the pre- and post-blacklisting periods separately. The purpose of the exercise here is limited to showing that consistent with the incentive scheme documented in section II invoice mills did appear in the Pakistani VAT system and that their number declined dramatically once the zero-rating reform came into effect.

### 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 registra-

tion 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.<sup>9</sup> 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 with a view to see how much of a refund claim is based on invoices issued by invoice mills.

## IV Empirical Methodology

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

$$(2) \quad y_{it} = \alpha_i + \lambda_t + \beta \mathbf{X}_{it} + \gamma \cdot \text{zero-rated}_i \times \text{after}_t + \varepsilon_{it},$$

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

Identification in this setup requires that a given outcome  $y_{it}$  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 regressions

---

<sup>9</sup> This system is commonly used by customs administrations around the world to classify traded goods and services.

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

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

Notwithstanding parallel trends, identification in this setting may fail if the zero-rating reform creates significant spillovers in the non-zero-rated industries, violating the SUTVA assumption (see for example [Imbens & Rubin, 2015](#)). These spillovers can take two broad forms. At the extensive margin, the reform could distort the entry decision of a firm: the differential tax treatment may force a firm which would otherwise have entered into a non-zero-rated industry to switch to a zero-rated industry or vice versa. I address this concern by reporting parallel results from balanced panel samples, where the composition of the sample is held fixed. Shutting down the entry and exit, however, does not rule out spillovers along the intensive margin. These spillovers may arise from general equilibrium considerations (the zero-rating reform affecting prices or the cost of compliance generally) or from demand and supply linkages of control firms with the zero-rated industries ([Waseem, 2019b](#)). To rule out this class of concerns, I estimate the event-study equation (3) separately on the treatment and control groups. The results show that the control outcomes continue to evolve on the preexisting trend with no visible break at the time of the reform. In contrast, the treatment outcomes undergo sharp changes at the time of the reform.

I estimate the impact of the reform on the following six outcomes ( $y_{it}$ ) reported by firms.

**(i) Output Tax ( $\tau \cdot \hat{s}_{it}$ ):** By definition output tax equals the tax rate times the reported sales. Given that the reform reduces the rate applicable to the treatment group from 15% to 0%, I expect a large, negative  $\hat{\beta}$  from these regressions. This decline in output tax is a combination of the mechanical effect (reduction of the rate to zero) and the

behavioral effect (changes in reported sales). I isolate the behavioral effect by estimating the sales response separately (see below). The objective of showing the output tax response is to demonstrate that a strong first-stage exists in this setup, whereby the output tax liability of treated firms reduces sharply relative to the control firms as a result of the reform.

**(ii) Input Tax ( $\tau \cdot \hat{c}_{it}$ ):** The case of input tax is exactly similar to that of output tax outlined above. Recall, however, that there was some delay in the zero-rating of two important inputs (electricity and gas) used by the treated firms.<sup>10</sup> I therefore expect the immediate impact of the reform on input tax to be smaller than that on output tax. This dynamics of the response can be seen clearly in the event study plots.

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

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

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

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

Table II presents summary statistics of the data. 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

---

<sup>10</sup>Orders zero-rating electricity and gas were issued at the firm level. The first set of these orders were issued in August 2005 and the last in April 2007. See section III.B for details.

exports. But they are not much different from the control sample in other characteristics such as location and age.

## V Firm Responses to the Reform

### V.A Nonparametric Evidence

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

#### V.A.1 First Stage

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

Figure III plots the amount of VAT refund paid in Pakistan as a proportion of gross VAT collection in the country during the tax year 1999 to 2009. The figure is based on annual aggregate statistics reported by the FBR on its website, which include both treated and untreated industries. The refund-to-gross-collection ratio in the country was roughly 20% at the baseline. It fell by nearly 10 percentage points in the first year after the reform. It fell even further in 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. Thus, within three years of the reform, the aggregate VAT refund paid in the country had dropped to one-fourth of its baseline level; in terms of absolute numbers, the amount refunded reduced from PKR 55 billion in 2004 to PKR 27 billion in 2008.<sup>11</sup>

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

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

---

<sup>11</sup>For these statistics, see FBR Year Books from 1999 to 2009 available [here](#).

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

Taken together, the above two pieces of evidence demonstrate that a very strong first stage exists in this setting. The incentives to misreport collapse in the treatment group at the time of the reform as both their tax liability and input tax entitlement crash down to a near-zero level. Such a salient drop in incentives is likely to induce sharp behavioral responses to which I turn next.

## V.A.2 Behavioral Responses

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

exports drop to the new equilibrium in the few periods following the reform.

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

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

## V.B Regression Results

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

Table IV explores the dynamics of the responses documented above. I estimate an augmented version of model (2) by replacing the double-difference interaction term

with six double-interactions, one each for every post-reform tax year. The results confirm the time pattern of response seen visually in Figures IV and V. The first year response as a proportion of the average post-reform response is 94% for output tax, 79% for input tax, 32% for sales, 80% for purchases, 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 as the reform—being a tax cut—was likely to induce sharp compliance among the treated population.<sup>12</sup> 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 levels 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 perhaps the most important in terms of its VAT impact. Table V shows this formally. I break down the aggregate response reported in Table III into its constituent textile and non-textile components. The response of the textile industry roughly equals the average response for all the outcomes: all textile coefficients are within the 95% confidence interval of the corresponding baseline coefficient. The finding is significant in one important respect. The textile industry has a very well-defined supply chain comprising five distinct production stages: ginning, spinning, weaving, processing, and made-up stages.<sup>13</sup> Given that I observe the production stage a textile firm operates in, I can explore any heterogeneity in response across the supply chain, comparing in particular the upstream production stages with the downstream ones. This exercise will help me determine the extent to which observed responses are attenuated by real responses, which act in opposite direction to the reporting responses.

Table VI carries out this exercise. I restrict the treatment sample to the textile industry only and estimate a triple-difference version of model (2) by including interactions of the double-difference term with dummies indicating the production stage. I include dummies for the three upper-most production stages—ginning, spinning,

---

<sup>12</sup>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.

<sup>13</sup>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.

and weaving, leaving the downstream stages as the omitted category. The response varies across the supply chain along important margins. Specifically, output tax drops more in the upstream stages and input tax drops more in the downstream 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.<sup>14</sup> But perhaps more importantly, the other four outcomes, which reflect pure behavioral effects of the reform, decline less in the upstream stages. In fact, these four outcomes—sales, purchases, exports, non-export sales—either do not change or rise for the first production stage i.e. ginning. This mitigates the concern that the responses documented above are significantly attenuated by real responses to the tax cut. Both sales and purchases reported by the first production stage of the textile industry evolve similarly to the control group in the post-reform years, ruling out any significant real increase in production in the textile industry caused by the zero-rating reform.

## VI Quantifying VAT Evasion

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

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

This formula is an aggregate version of formula (1), where the capital letter notation stands for the sum of the corresponding variable over all firms; for example  $\hat{S}_R = \sum_{i \in I} \hat{s}_{Ri}$ , where  $I = 1 \dots N$  is the set of firms in the VAT net. As noted earlier, the first term in this formula is the overclaimed refund on exports and the second the underpaid tax on domestic B2C sales. In this section, I use the formula and the estimates from section V to compute a lower bound on VAT evasion as it existed in the treated industries at the baseline.

The estimation is based on the consideration that firm behavior in the treated in-

---

<sup>14</sup>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.

dustries would approach the first-best case as the tax rate reduces from 15% to 0% in the post-reform periods, reducing sharply if not eliminating entirely any incentive to misreport sales or purchases. As a result, any reform-driven changes in a given variable would capture the misreporting of the variable as it existed at the baseline. For example, indexing post-reform outcomes by  $t'$  and pre-reform ones by  $t$ , the first term in the above formula can be represented as  $\tau(\hat{C}_E - C_E) \approx \tau(\hat{C}_{E,t} - \hat{C}_{E,t'}) \equiv \tau\Delta\hat{C}_E$ . Using this notation, the above formula can be rewritten as the following

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

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

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

billion claimed on exports at the baseline and an overclaim of refund of nearly PKR 23 billions.

The second term in formula (6),  $\Delta \hat{S}_R$ , denotes the change in B2C sales of the treated industries. Again I do not directly observe this term. To be specific, because the data are not transaction-level, I cannot divide the total non-export sale of a firm into its constituent B2B and B2C components. I can, however, estimate the average under-reporting of B2C sales at the baseline using the following accounting identities

$$(7) \quad \begin{aligned} \sum_{i \in I} (\hat{s}_i - \hat{s}_{i,E}) &\equiv \sum_{i \in I} (\hat{s}_{i,B2B} + \hat{s}_{i,B2C}) \\ \sum_{i \in I} \hat{c}_i &\equiv \sum_{i \in I} (\hat{s}_{i,B2B} + \hat{c}_{i,OS}). \end{aligned}$$

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

$$(8) \quad \sum_{i \in I} \hat{s}_{i,B2B} \equiv \sum_{i \in I} (\hat{s}_i - \hat{s}_{i,E}) - \sum_{i \in I} \hat{c}_i + \sum_{i \in I} \hat{c}_{i,OS},$$

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

$$(9) \quad \Delta \hat{S}_R \equiv \Delta (\hat{S} - \hat{S}_E) - \Delta \hat{C} + \Delta \hat{C}_{OS}.$$

The LHS of this expression is our variable of interest—the change in B2C sales reported by firms. It is a sum of three components: (1) the change in non-export sales of firms, (2) the negative of the change in purchases of intermediates claimed by firms, and (3) the change in fake purchases of intermediates claimed by firms. I do not observe the last term in this expression, but note that the extent of one-sided evasion even if it existed would become vanishingly small from 2008 when the tax administration acquires the ability to costlessly cross-match sales and purchase transactions reported by sellers and buyers. I therefore ignore it from my subsequent analysis,

rewriting the above equation as

$$(10) \quad \Delta \hat{S}_R \approx \Delta (\hat{S} - \hat{S}_E) - \Delta \hat{C}.$$

Columns (6) and (4) of Table III show that the zero-rating caused an 8.3 log-points (8.7%) decrease in non-export sales and a 42 log-points (52%) decrease in purchases of intermediates reported by the treated firms. Together, these responses imply that at the baseline domestic B2C sales of treated firms were on average underreported by at least 43.5%. Assuming that domestic B2C sales of treated firms constitute just 25% of their total sales, this underreporting translates into a revenue loss of PKR 15 billion. These calculations are shown in greater details in Table VII. Noncompliance caused an aggregate VAT revenue loss of PKR 38 billion in the treated industries in 2004. Roughly three-fifths of this loss was caused by overclaiming of VAT refunds and the rest by underreporting B2C sales. The PKR 38 billion revenue loss in 2004 amounts to nearly 13% of the total VAT collected in the country in that year and translates into 11.5% (77% of the statutory rate of 15%) of the *true* B2C sales reported in the treated industries in the year.

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

## VII Invoice Mills and VAT Evasion

One important focus of this paper is to understand the role played by invoice mills in facilitating VAT noncompliance. The third example in the conceptual framework shows that invoice mills arise naturally in the low-enforcement, high-informality setting of developing countries, bridging the gap created by broken VAT chains. Figure II illustrated this empirically: consistent with the conceptual framework invoice mills did appear and grow in the Pakistani VAT system before declining sharply as the

zero-rating reform took effect. I now turn to documenting the behavior of these firms in more details to see if invoice mills are an important channel through which tax evasion documented in Table VII takes place.

I begin by showing the evolution of VAT outcomes for the invoice mills, focusing especially on their reaction to the zero-rating reform. Figure VII replicates the analysis in Figure IV after restricting the treatment sample to blacklisted and suspended firms only. Blacklisted firms constitute 2% and suspended firms 5% of the treatment sample only. To maximize power of this analysis, I combine both categories of firms here. Given that even together these firms form only 7% of the treatment sample, the results are noisier than the baseline. Yet the pattern of response is very similar. All four outcomes—sales, purchases, exports, non-export sales—drop sharply at the time of the reform, with purchases dropping more than any other outcome. Table VIII formalizes these results. I estimate a triple-difference version of model (2), exploring any differential response of blacklisted and suspended firms from other treated firms. Two results in the table are noteworthy. First, clearly blacklisted and suspended firms respond more aggressively than other treated firms. The triple-difference coefficients in the last four columns are always negative, with the coefficient of blacklisted firms always larger in magnitude than that of suspended firms. Second, because blacklisted and suspended firms constitute a small proportion of the overall sample, their larger responses do not affect the average responses too much: the double-difference  $treat \times after$  coefficient is always within the 95% confidence interval around the baseline coefficient.

One notable feature of the event study plots shown in Figure VII is that the outcomes spike just before the reform. This spike is particularly prominent in the top four panels of the figure. To explore this feature of the results in greater details, Figure VIII plots the aggregate values of the six outcomes reported by blacklisted and suspended firms in each month. Relative to the within-firm average changes shown in Figure VII, this figure displays the aggregate *level* of each outcome, illustrating more clearly the magnitude of the spurious input tax credit pumped into the system by invoice mills. The striking feature of the plots is that all outcomes exhibit a very large spike just before the zero-rating reform comes into effect. For example, aggregate purchases jump from a prereform average level of around PKR 5 billion to 30 billion one tax period before the reform, *i.e.* in May 2005. This large jump is followed by an equally sharp drop, whereby purchases reduce to PKR 3.6 billion in July 2005 and to 1.8 billion in November 2005.

This large concentration of activity on the wrong side of the zero-rating reform is surprising. Under any standard model of tax behavior, a large tax cut reducing the rate from 15% to 0% would cause some inter-temporal shifting of activity from high-tax periods to low-tax periods.<sup>15</sup> The behavior I find here is polar opposite to this intuitive and straight-forward prediction of the model. Only way to reconcile the concentration of activity on the wrong side of the tax cut therefore is that it represents an effort by invoice mills to pump input tax credit into the system while it is still taxable at the standard rate to maximize spurious refund claims in a process similar to one detailed in our third stylized example.

As I note in section III.A, exporter file supplier-wise details of their purchases of intermediates at the time of claiming VAT refund. These transaction-level data are available from the tax year 2002 onward, and they allow me to construct linkages between exporters and invoice mills to study the latter's role in the overclaim of refunds. Figure IX reports the results of this exercise. The blue, dashed curve plots the aggregate value of refund claimed by exporters on the invoices of blacklisted firms. To compare these purchases from blacklisted firms reported by exporters with the total volume of sales of blacklisted firms, the red, solid curve in the figure plots the aggregate value of output tax involved in all invoices issued by blacklisted firms each month. Although the two series are from two entirely different data sources—the blue from refund claims filed by exporters, and the red from VAT returns filed by blacklisted firms, they line up quite well. This shows that almost all invoices issued by invoice mills end up in the refund claims of exporters. For example, in the baseline year of 2004, out of total output tax of PKR 8.9 billion reported by blacklisted firms nearly 97% (PKR 8.6 billion) was claimed as refund by exporters. I estimate in Table VII that exporters overclaimed VAT refund of nearly PKR 23 billion in 2004 (see Row 5 of the table). Refund claimed on the invoices of blacklisted firms comprises 37% of this amount.

Filing a refund claim on the basis of spurious invoices does not mean that this refund claim would be sanctioned. As I note in section III.A, the tax administration preaudits all refund claims, assessing among other things the validity of purchase invoices. To rule out that the numbers I report above represent actual and not potential revenue loss, Figure VII plots the aggregate values of VAT outcomes reported by

---

<sup>15</sup>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.

blacklisted firms around the event of blacklisting. It illustrates that by the time a firm is blacklisted, it is already past its peak activity and both its turnover and purchases are declining sharply. In addition, almost all of the activity carried out by such firms occurs in periods while their invoices are still legitimate. To see why invoice mills are detected so late in their life cycle, Table IX compares the characteristics of blacklisted and suspended firms with other firms. The sample here is restricted to pre-reform years only (1998-2004). Invoice mills are essentially hit-and-run enterprises, and consistent with this notion the first-year turnover and purchases of blacklisted firms are much higher and their value-added much lower than a typical normal firm. But other than these there are no clear markers that can distinguish invoice mills from other firms. It is therefore not surprising that invoice mills are detected too late, when invoices issued by them have already been used to claim input tax credit or refund.

In sum, the above analysis suggests that invoice mills exist primarily to facilitate overclaim of refunds by exporters. Nearly two-fifths of the excess refund claimed by exporters is based on their invoices. In contrast, their role in VAT evaded on domestic B2C transactions is negligible. This finding is, however, subject to an important caveat. I identify invoice mills in the data on the basis of the blacklisting procedure used by the tax administration. It is quite possible that this procedure is more effective in identifying invoice mills connected to exporters than to others either because the tax administration is more concerned about the overclaim of refund than the underpayment of VAT on domestic transactions or because of any data limitations. Measurement error introduced by such selective blacklisting of firms means that I underestimate the VAT loss caused by invoice mills. Specifically, invoice mills issuing invoices solely to non-export firms escape blacklisting and thus do not feature in my empirical framework.

## VIII Conclusion

Value-added tax has seen remarkable expansion in recent decades. Its popularity as a tax instrument in large part rests on its superior enforcement properties relative to other taxes. By creating (1) matchable 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 is argued to facilitate tax compliance. Any revenue-efficiency gains accruing from these mechanisms, however, need to be

balanced against additional evasion opportunities a destination-based VAT creates in export refunds. This paper estimates the extent of VAT noncompliance and mechanisms driving it in a representative emerging economy. For this purpose, it leverages a novel tax reform that cut the VAT rate applicable on five major industries of the country from 15% to 0%. With the reduction of the rate to zero, any incentive to mis-report approach zero, implying that by comparing the changes in firm behavior from pre- to post-reform periods one can infer the amount of evasion as it existed in the treated industries at the baseline.

The total revenue loss caused by noncompliance in a destination-based VAT is a sum of the refund overclaimed on exports and the tax underpaid on domestic B2C transactions. I find that at the baseline the refund was overclaimed by nearly 22% and domestic B2C sales were underreported by 43.5%. Together, this caused a total VAT revenue loss of PKR 38 billion in 2004, which is nearly 13% of the total VAT collected in the country in that year. The evaded amount translates into 11.5% (77% of the statutory rate of 15%) of the *true* aggregate B2C sales reported in the treated industries in the year. I also explore the role of invoice mills in facilitating this noncompliance. Invoice mills are firms that engage in no real business activity and exist solely to trade in VAT invoices. I find that the amount of excess refund claimed by exporters nearly equals the amount of input tax involved on invoices issued to them by invoice mills. This shows almost all of revenue loss on export refunds takes place through the medium of invoice mills.

## References

- ABADIE, ALBERTO, ATHEY, SUSAN, IMBENS, GUIDO W, & WOOLDRIDGE, JEFFREY. 2017. *When should you adjust standard errors for clustering?* Tech. rept. National Bureau of Economic Research.
- AGRAWAL, DAVID R, & ZIMMERMANN, LAURA. 2019. Production and evasion responses with limited state capacity. International Growth Centre Working Paper.
- ALMUNIA, MIGUEL, HJORT, JONAS, KNEBELMANN, JUSTINE, & TIAN, LIN. 2019. *Strategic or Confused Firms? Evidence from "Missing" Transactions in Uganda*. Mimeo.

- ARTAVANIS, NIKOLAOS, MORSE, ADAIR, & TSOUTSOURA, MARGARITA. 2016. Measuring Income Tax Evasion Using Bank Credit: Evidence from Greece. *The Quarterly Journal of Economics*, **131**(2), 739–798.
- ATKINSON, ANTHONY BARNES, & STIGLITZ, JOSEPH E. 1976. The design of tax structure: direct versus indirect taxation. *Journal of public Economics*, **6**(1-2), 55–75.
- BACHAS, PIERRE, & SOTO, MAURICIO. 2019. Not (ch) your average tax system: corporate taxation under weak enforcement. World Bank Policy Research Working Paper.
- BERTRAND, MARIANNE, DUFLO, ESTHER, & MULLAINATHAN, SENDHIL. 2004. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, **119**(1), 249–275.
- BROCKMEYER, ANNE, SMITH, SPENCER, HERNANDEZ, MARCO, & KETTLE, STEWART. 2019. Casting a Wider Tax Net: Experimental Evidence from Costa Rica. *American Economic Journal: Economic Policy*, **11**(3), 55–87.
- CARRILLO, PAUL, POMERANZ, DINA, & SINGHAL, MONICA. 2017. Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. *American Economic Journal: Applied Economics*, **9**(2), 144–64.
- EBRILL, LIAM P, KEEN, MICHAEL, BODIN, JEAN-PAUL, SUMMERS, VICTORIA P, *et al.* 2001. *The modern VAT*. International Monetary Fund.
- EISSA, NADA, & ZEITLIN, ANDREW. 2014. Using mobile technologies to increase VAT compliance in Rwanda.
- FBR. 2005. CBR Quarterly Review. 5(1). Federal Board of Revenue, Pakistan, <https://tinyurl.com/y4tuxp3q>.
- FELDMAN, NAOMI E, & SLEMROD, JOEL. 2007. Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns. *The Economic Journal*, **117**(518), 327–352.
- GADENNE, LUCIE, NANDI, TUSHAR K., & RATHELOT, ROLAND. 2019. *Taxation and Supplier Networks: Evidence from India*. Mimeo.
- GORODNICHENKO, YURIY, MARTINEZ-VAZQUEZ, JORGE, & PETER, KLARA SABIRANOVA. 2009. Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effects in Russia. *Journal of political economy*, **117**(3), 504–554.

- IMBENS, GUIDO W, & RUBIN, DONALD B. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- INTERNATIONAL TAX DIALOGUE. 2013. International Tax Dialogue, Key Issues and Debates in VAT, SME Taxation and the Tax Treatment of the Financial Sector. *International Tax Dialogue*.
- KEEN, MICHAEL. 2007. VAT attacks! *International Tax and Public Finance*, **14**(4), 365–381.
- KEEN, MICHAEL, & SMITH, STEPHEN. 2006. VAT fraud and evasion: What do we know and what can be done? *National Tax Journal*, 861–887.
- KLEVEN, HENRIK J., KNUDSEN, MARTIN B., KREINER, CLAUS THUSTRUP, PEDERSEN, SØREN, & SAEZ, EMMANUEL. 2011. Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, **79**(3), 651–692.
- NARITOMI, JOANA. 2019. Consumers as tax auditors. *American Economic Review*, **109**(9), 3031–72.
- OECD. 2017. Revenue Statistics, Organization for Economic Cooperation and Development. URL, <https://goo.gl/WC1M1L>.
- PISSARIDES, CHRISTOPHER A., & WEBER, GUGLIELMO. 1989. An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, **39**(1), 17 – 32.
- 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, REHMAN, OBEID UR, & WASEEM, MAZHAR. 2019. Pecuniary and non-pecuniary motivations for tax compliance: Evidence from Pakistan.
- WASEEM, MAZHAR. 2018a. Taxes, Informality and Income Shifting: Evidence from a Recent Pakistani Tax Reform. *Journal of Public Economics*, **157**, 41–77.

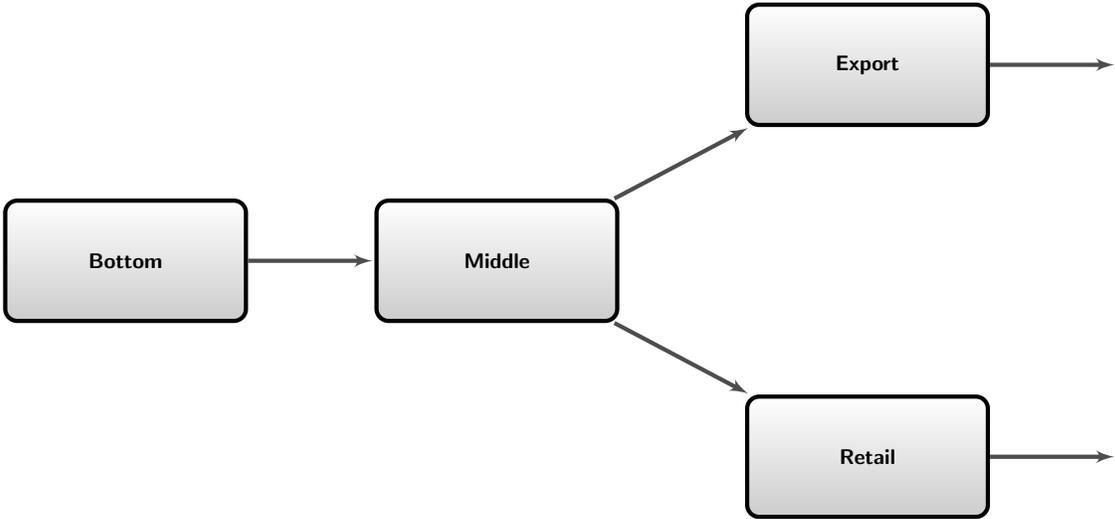
WASEEM, MAZHAR. 2018b. Using Movement of Exemption Cutoff to Estimate Tax Evasion: Evidence from Pakistan. Working Paper, University of Manchester.

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

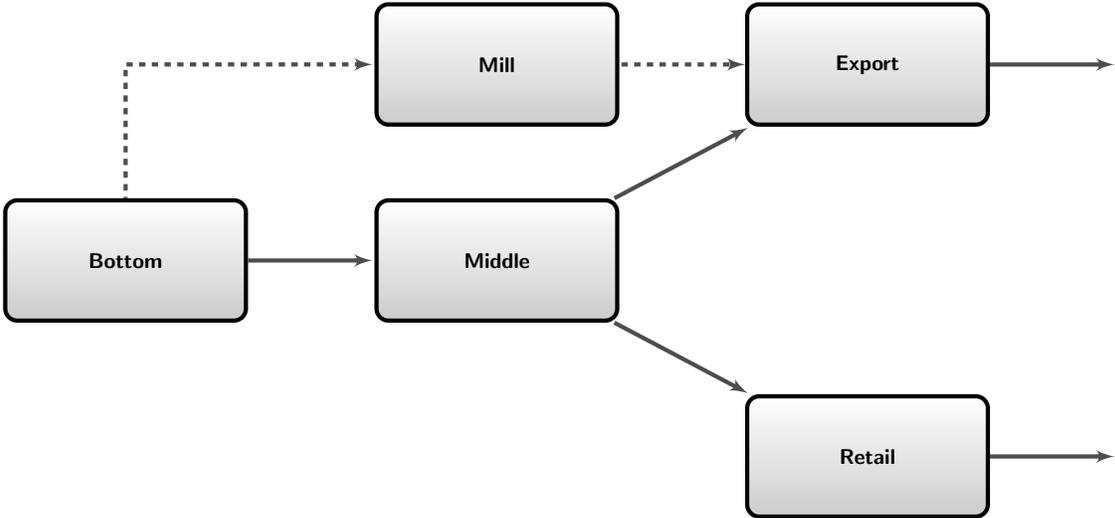
WASEEM, MAZHAR. 2019b. Information, Asymmetric Incentives, or Withholding: Understanding the Self-Enforcement of Value-added Tax. Working Paper, University of Manchester.

FIGURE I: INPUT OUTPUT LINKAGES

**A: Real Supply Chain**

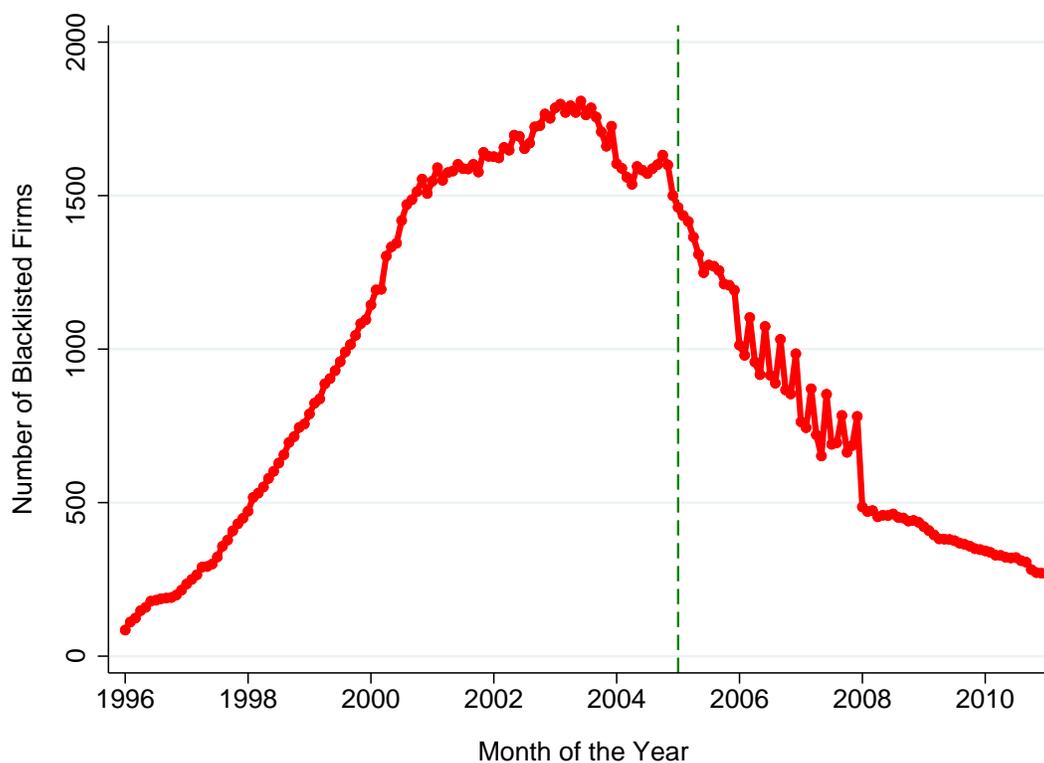


**B: Supply Chain With an Invoice Mill**



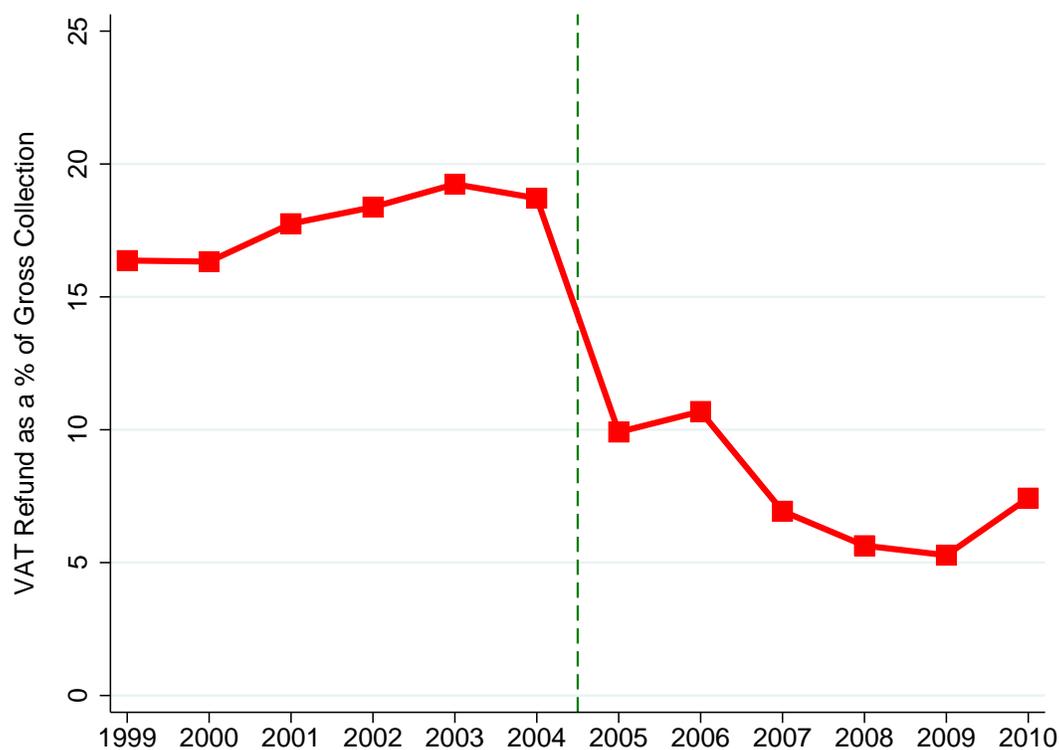
**Notes:** This figure displays a typical supply chain comprising three production stages. The top panel displays the *real* input output linkages between firms, showing the path the consumption good travels before it is exported or consumed in the domestic market. The bottom panel displays the role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

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



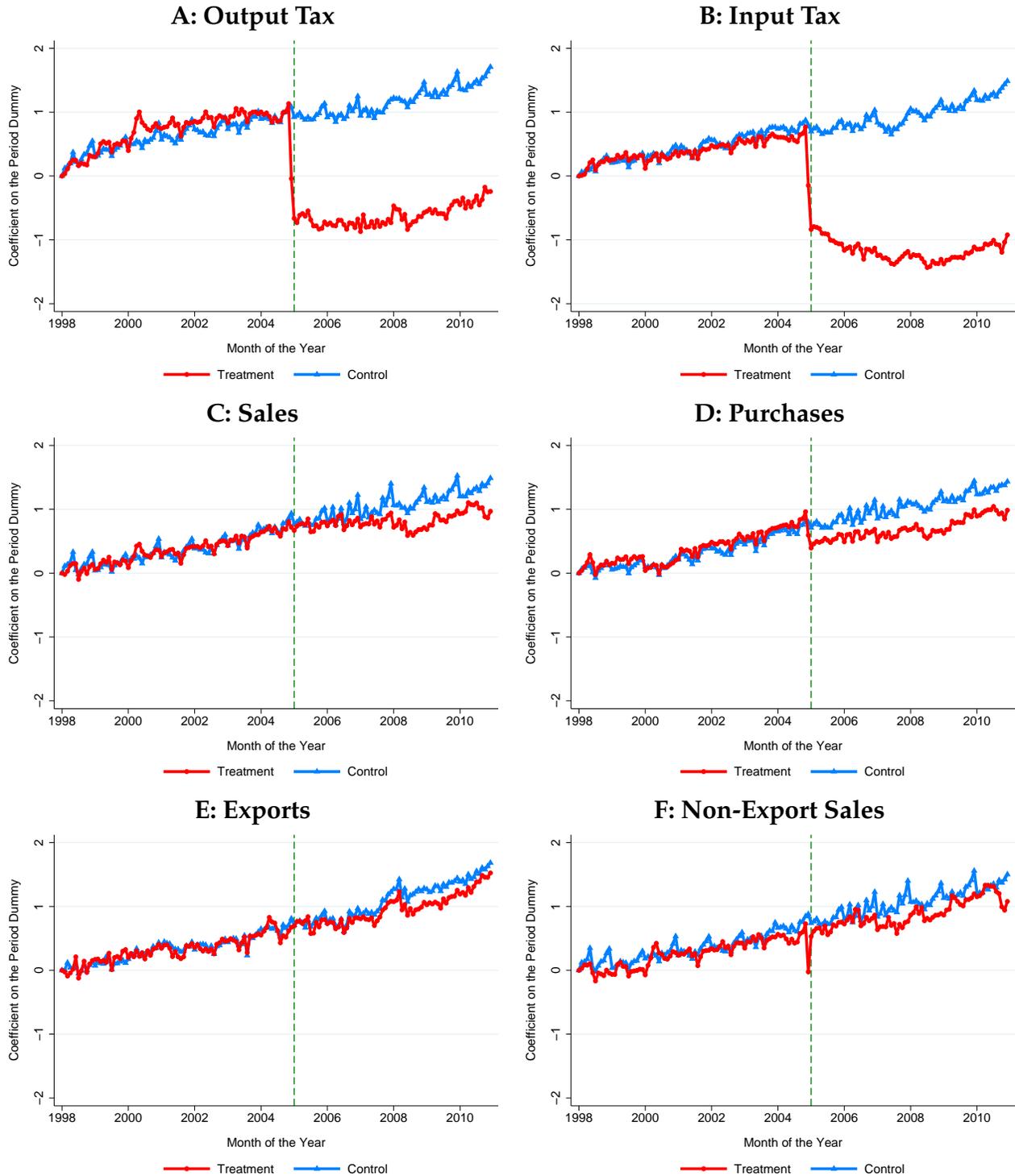
**Notes:** The figure shows the emergence, growth, and subsequent decline of invoice mills in the Pakistani setting. The sample begins from 1996, when a *broad-based* VAT with coverage extended to almost entire manufacturing and exports stages began in the country. Each marker in the curve denotes the number of blacklisted firms that file a return in the given month. As I note in section III.C, the FBR acquired the legal power to declare a firm blacklist from July 2003. Most of the returns included in this plot were therefore filed at the time the firm was still not blacklist. The year  $t$  in the horizontal axis denotes the month July of year  $t$ . The dashed, vertical line represents the time from which the zero-rating reform became applicable.

FIGURE III: REFUND AS A PROPORTION OF GROSS COLLECTION



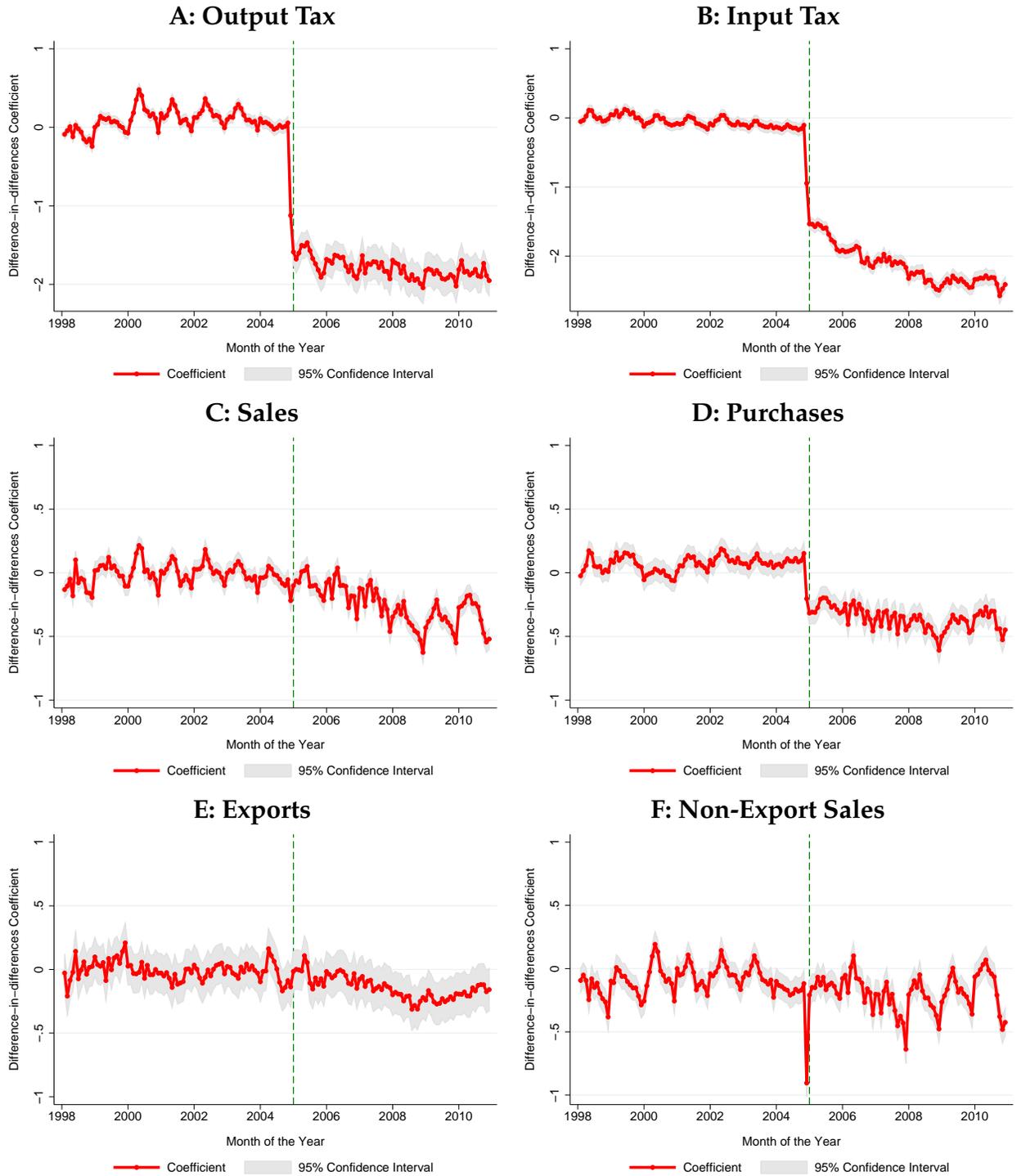
**Notes:** The figure shows the effects of the zero-rating reform on the VAT refund paid in Pakistan. Each marker in the plot denotes the aggregate VAT refund paid by the FBR to firms in *all* industries as a percentage of the gross VAT collection in that year. The data used for this plot are publicly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

FIGURE IV: FIRM BEHAVIOR TO THE TAX CUT



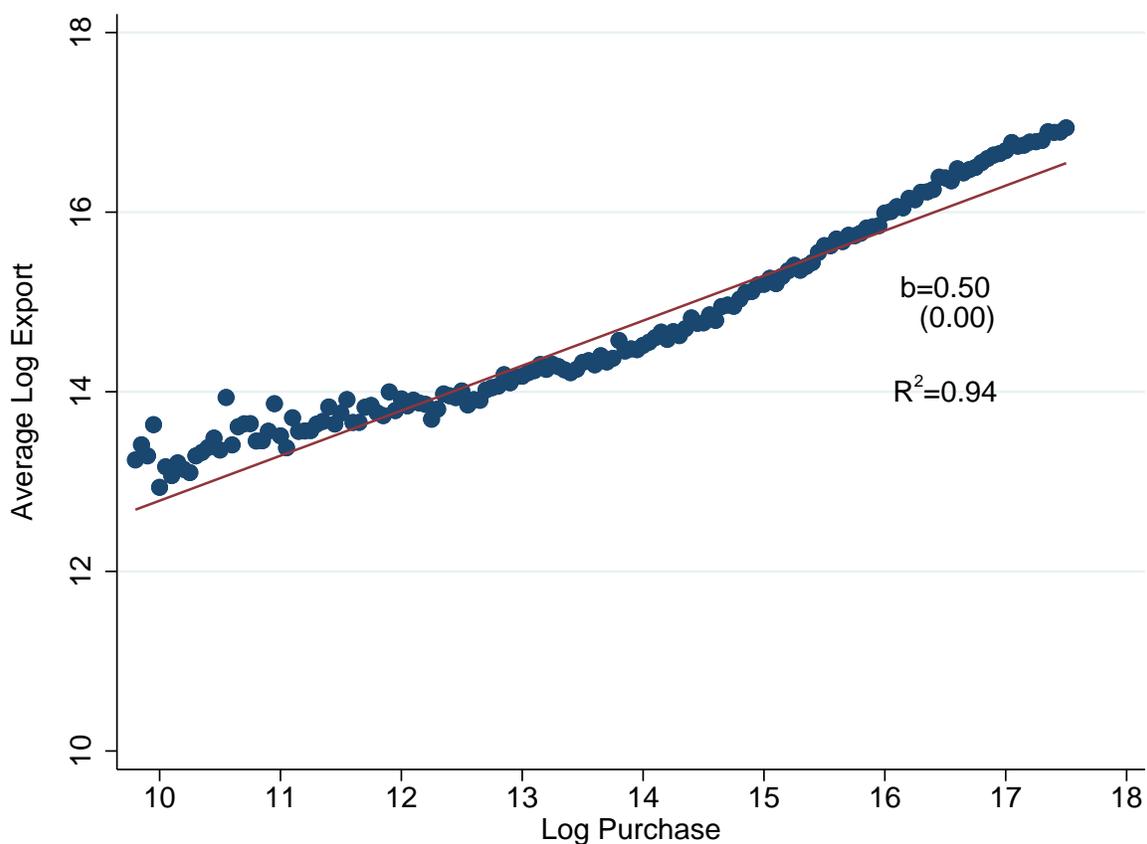
**Notes:** The figure compares the evolution of six VAT outcomes from the tax year 1998 to 2010 across the treatment and control groups. Treatment and control groups consist respectively of firms in the zero-rated and non-zero-rated industries. To construct these charts, I regress the log of the outcome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

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



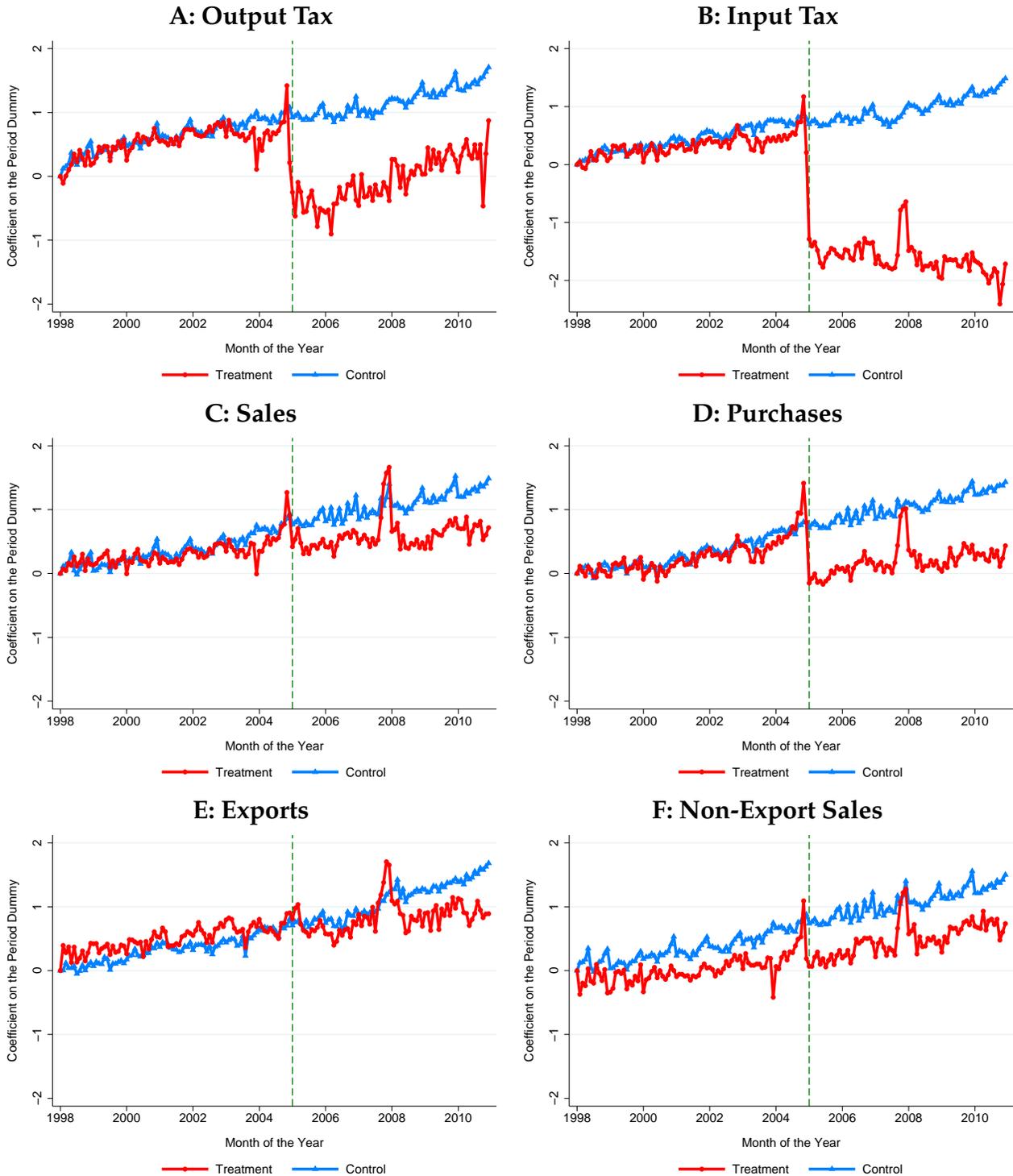
**Notes:** The figure shows the difference-in-differences version of the plots in Figure IV. To construct these charts, I regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, and month $\times$ treat dummies, dropping the dummies for July 1998. I then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

FIGURE VI: EXPORTS AS A FUNCTION OF PURCHASES



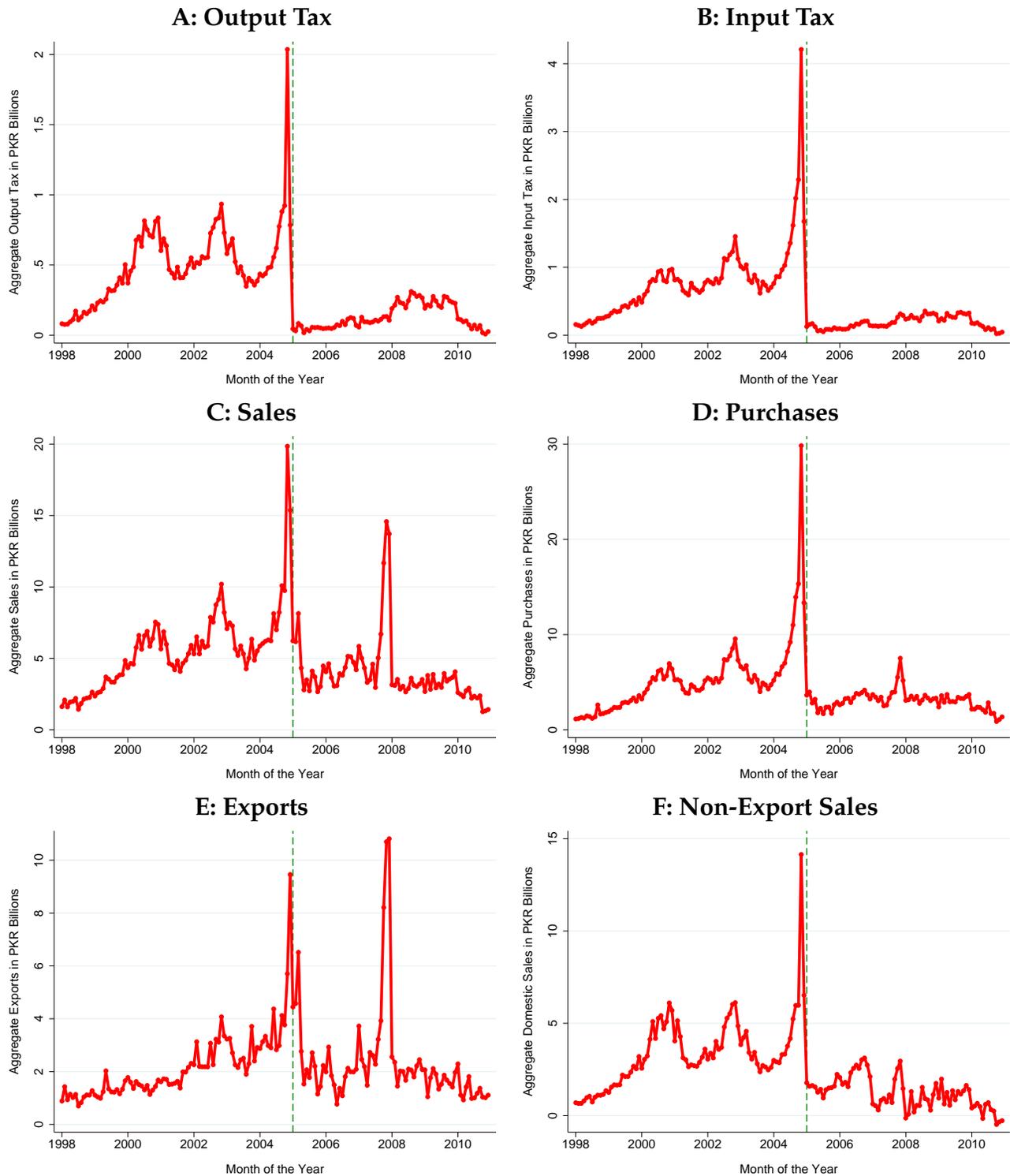
**Notes:** The figure explores the relationship between purchases of intermediates and exports reported by firms of the treated industries in the baseline years (1998-2004). The blue curve shows a binned scatter plot for the data. I group firms into bins on the basis of log of purchases reported by them.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

FIGURE VII: FIRM BEHAVIOR TO THE TAX CUT – BLACKLISTED FIRMS



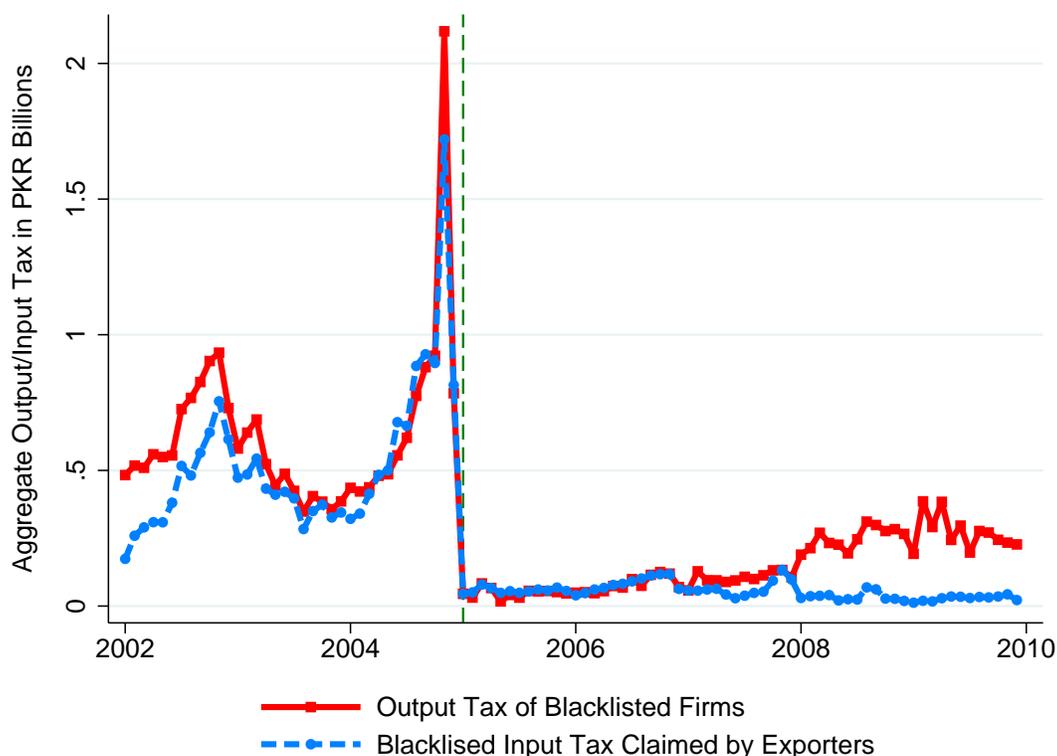
**Notes:** The figure compares the evolution of six VAT outcomes from the tax year 1998 to 2010 across the treatment and control groups. The treatment group here consists of blacklisted and suspended firms in the zero-rated industries only. The control group, as earlier, comprises all firms of the non-zero-rated 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 VIII: AGGREGATE VALUES OF VAT OUTCOMES – BLACKLISTED FIRMS



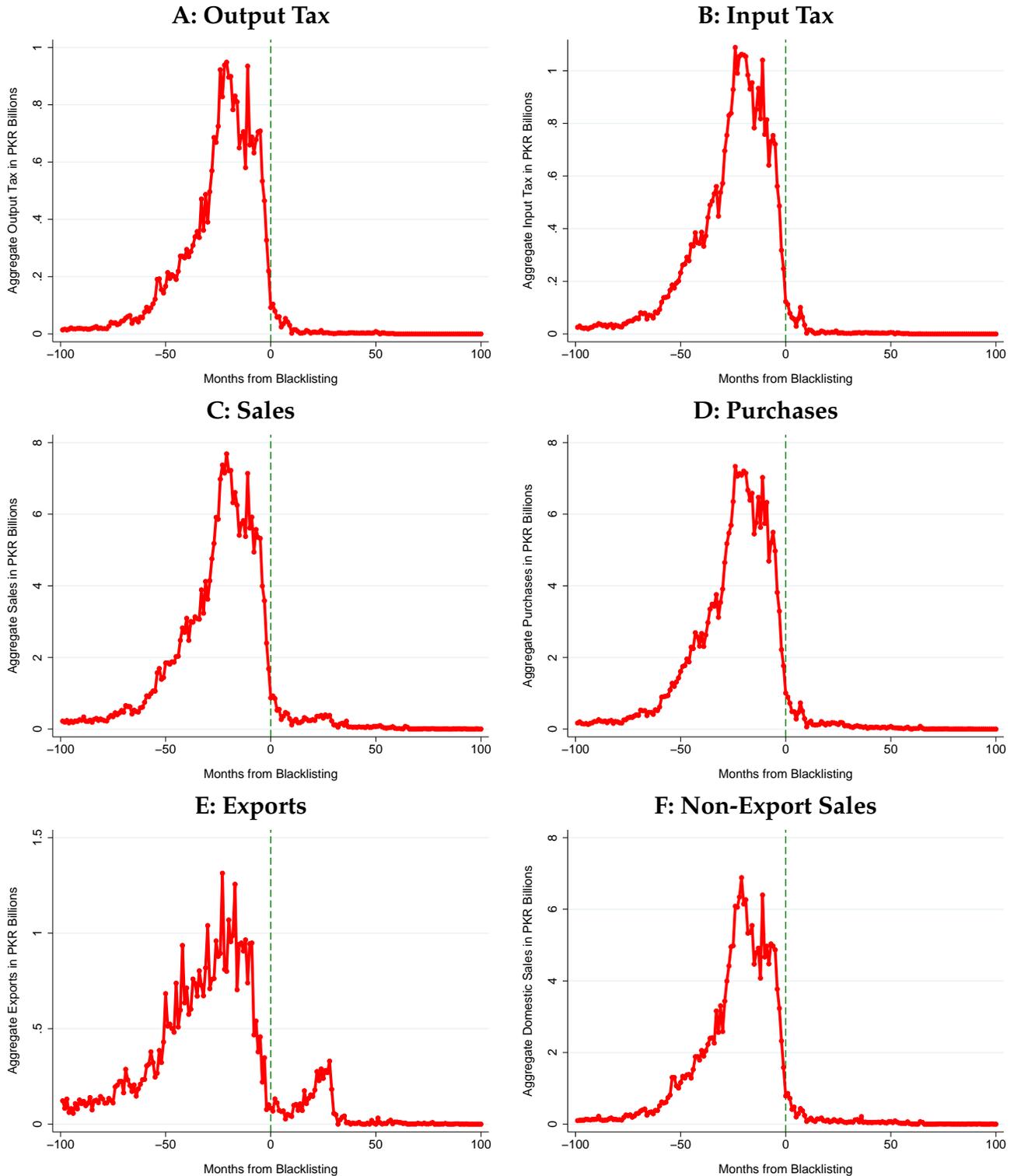
**Notes:** The figure illustrates how VAT outcomes evolve around the time of the zero-rating reform. The sample contains blacklisted and suspended firms operating in the zero-rated industries. Each panel of the figure shows the aggregate value of the outcome indicated in the title of the panel for the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

FIGURE IX: INVOICE MILLS AND VAT REFUND



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

FIGURE X: AGGREGATE VALUES OF VAT OUTCOMES – BLACKLISTED FIRMS



**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 blacklisted or its registration was suspended. Each panel of the figure shows the aggregate value of the outcome in 100 months prior to and 100 months after the event of blacklisting. To deal with outliers, I exclude ten observations with the highest values of the given outcome in the entire sample. For example, for constructing Panel A, I sort all firm-month observations on the basis of Output Tax in a descending order and exclude the top-ten observations.

TABLE I: INPUT OUTPUT LINKAGES UNDER THREE STYLIZED EXAMPLES

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

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

TABLE II: SUMMARY STATISTICS

	2003		2004	
	Treatment	Control	Treatment	Control
	(1)	(2)	(3)	(4)
1. # Observations	172,321	743,281	163,327	709,560
2. Output Tax	0.514 (0.013)	0.602 (0.028)	0.479 (0.012)	0.850 (0.045)
3. Input Tax	0.686 (0.014)	0.431 (0.017)	0.707 (0.013)	0.639 (0.031)
4. Sales	6.919 (0.895)	3.782 (0.122)	7.776 (0.116)	5.623 (0.202)
5. Purchases	4.600 (0.084)	2.678 (0.096)	6.158 (0.116)	4.223 (0.175)
6. Exports	2.614 (0.054)	0.311 (0.010)	4.207 (0.065)	0.503 (0.016)
7. Non-Export Sales	4.306 (0.893)	3.471 (0.118)	3.569 (0.082)	5.120 (0.194)
8. Major City	0.486 (0.001)	0.563 (0.001)	0.507 (0.001)	0.576 (0.001)
9. # Years Registered	3.695 (0.009)	3.592 (0.004)	3.238 (0.010)	3.108 (0.004)
10. # Years Active	8.083 (0.009)	8.608 (0.004)	8.401 (0.008)	8.883 (0.004)
11. Some Export	0.472 (0.001)	0.135 (0.000)	0.545 (0.001)	0.152 (0.000)
12. Some Import	0.416 (0.001)	0.369 (0.001)	0.476 (0.001)	0.415 (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 groups comprises all other firms. The first row of the table compares the number of firm-month observations for the two groups in the two prereform years. Subsequent rows compare the mean of eleven VAT outcomes and firm characteristics across the two groups. Major City denotes that the firm is registered in Karachi or Lahore, the two major cities of Pakistan. The variable # Years Registered reports the number of years up to 2003 since the firm's registration; # Years Active reports the number of years the firm remained active, filing its VAT return. Standard errors of the mean are in parenthesis.

TABLE III: FIRM BEHAVIOR TO THE TAX CUT

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

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

TABLE IV: FIRM BEHAVIOR TO THE TAX CUT – DYNAMICS

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

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

TABLE V: FIRM BEHAVIOR TO THE TAX CUT – TEXTILE VS. OTHERS

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

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

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

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

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

TABLE VII: CALCULATING VAT EVASION

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

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

TABLE VIII: FIRM BEHAVIOR TO THE TAX CUT – BLACKLISTED VS. OTHERS

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

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

TABLE IX: SUMMARY STATISTICS – BLACKLISTED FIRMS

	Blacklisted (1)	Suspended (2)	Others (3)
<u>A: VAT Outcomes (PKR Millions)</u>			
1. Sales	4.609	5.465	2.183
2. Purchases	7.043	5.781	1.875
3. Exports	0.649	1.187	0.388
4. Non-Export 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 Non-Export 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
<u>B: Firm Characteristics</u>			
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.

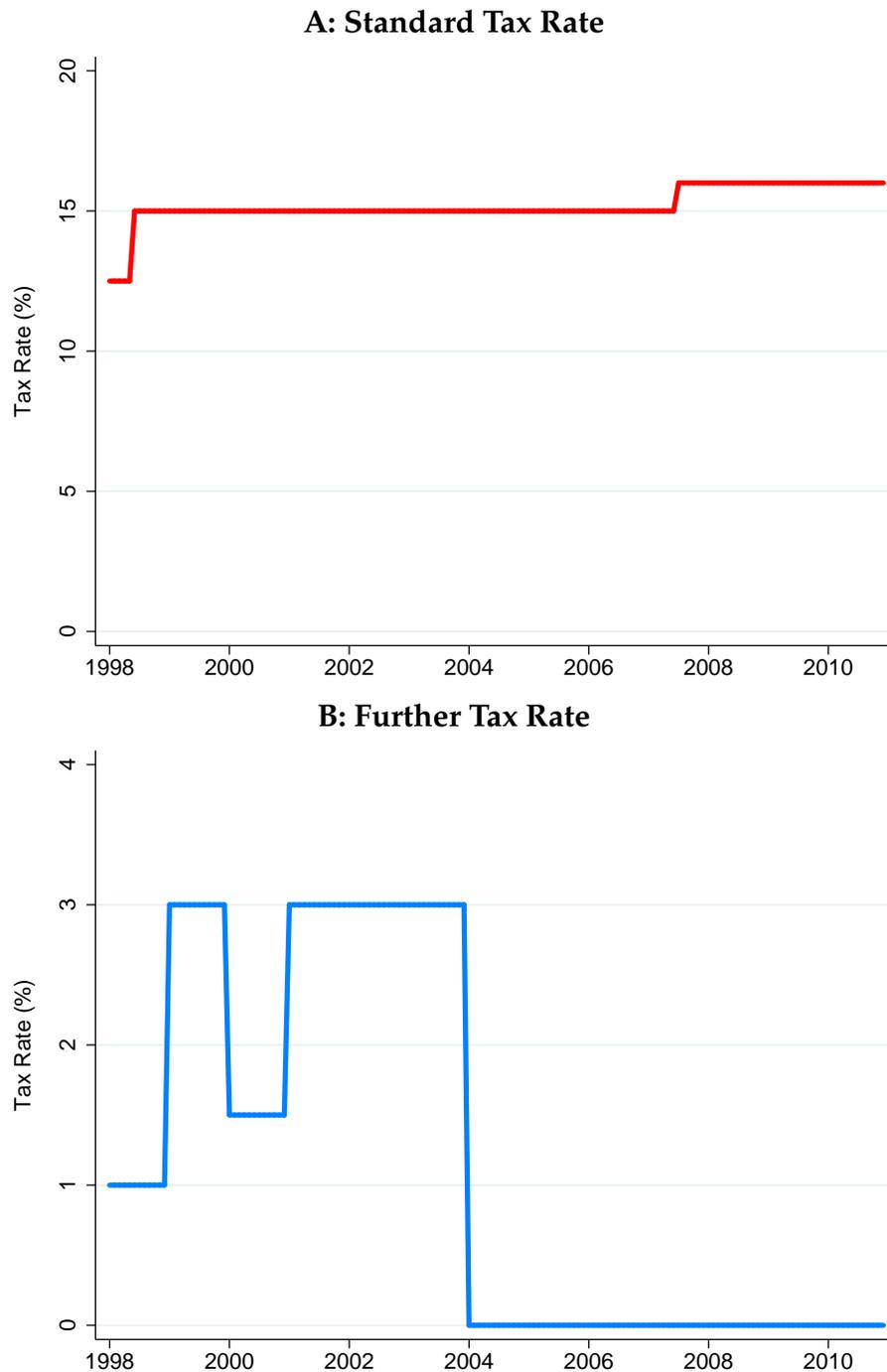
## A Online Appendix

### A.1 Definition of Variables

- (i) **Output Tax.** The value of VAT charged on sales made by a registered firm in a given tax period (month). It equals  $\tau \cdot \hat{s}_{it}$ , where  $\tau$  is the applicable VAT rate and  $\hat{s}_{it}$  is the value of sales reported by firm  $i$  in period  $t$ .
- (ii) **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  $\tau \cdot \hat{c}_{it}$ , where  $\tau$  is the applicable VAT rate and  $\hat{c}_{it}$  is the value of purchases of intermediates claimed by firm  $i$  in period  $t$ .
- (iii) **Sales.** The value of all goods and services supplied by a firm in a given tax period (month).
- (iv) **Purchases.** The value of all taxable intermediates acquired by a firm in a given tax period (month).
- (v) **Exports.** The value of all goods and services exported by a firm in a given tax period (month).
- (vi) **Non-Export Sales.** The value of all goods and services supplied by a firm minus the value of all goods and services exported by a firm in a given tax period (month).
- (vii) **Manufacturer.** A firm whose principal business activity is the manufacture of goods. Manufacturing is the process whereby a firm converts inputs into a distinct article capable of being put to use differently than inputs and includes any process incidental or ancillary to it.
- (viii) **Wholesaler.** Wholesaler' includes a dealer and means any person who carries on, whether regularly or otherwise, the business of buying and selling goods by wholesale or of supplying or distributing goods, directly or indirectly, by wholesale for cash or deferred payment or for commission or other valuable consideration or stores such goods belonging to others as an agent for the purpose of sale; and includes a person supplying taxable goods to a person who deducts income tax at source under the Income Tax Ordinance, 2001.

- (ix) **Retailer.** A person, supplying goods to general public for the purpose of consumption.
- (x) **Industry.** The Pakistani tax administration uses 4-digit Harmonized Commodity Description and Coding System (HS code) to classify firms into industry. The code, used by customs administrations throughout the world, divides all goods and services into 99 chapters (the first two digits in the code) and 21 sections. The sections broadly correspond to major industries in the country. I take the section a firm falls in as its industry. Table shows the sections, HS code, and description of these industries.
- (xi) **Major City** A firm registered either in Karachi or Lahore, the two largest cities in Pakistan on the basis of both population and GDP.
- (xii) **Initial Capital.** The value of initial capital of the firm, as reported by it at the time of registration for VAT.

FIGURE A.I: VAT RATES



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