Building fiscal capacity in developing countries: Evidence on the role of information technology

Merima Ali*, Abdulaziz B. Shifa†, Abebe Shimeles‡, Firew Woldeyes§

June 1, 2018

Abstract

Weak fiscal capacity for domestic resource mobilization is considered as one of the most important challenges in poor countries. Recently, many developing countries resorted to the application of information technology to consolidate tax mobilization; however, there is little systematic empirical evidence on the impact of such reforms. We attempt to narrow this gap by providing evidence from Ethiopia where there has been a recent surge in the adoption of electronic sales register machines (ESRMs). We use a unique administrative firm-level panel data covering all business taxpayers in Ethiopia and apply matching difference-in-difference method to account for possible bias that may arise due to selection. We find that adoption of ESRMs has a significant positive effect on tax payments and reported sales. Moreover, we find a positive effect on employment and no effect on net entry, suggesting that increased tax payments by registered taxpayers occurred without erosion of the tax base.

JEL Classification: H26, H32, O10, O55
Keywords: Developing economy; fiscal capacity; information technology; taxation.

1 Introduction

Economic development requires a state capable of mobilizing fiscal resources to finance the provision of essential public goods—a capacity that developing countries tend to lack. Weak fiscal capacity of states has thus received increased attention in the political economy of development.[1] Governments with the bare minimum of a tax administrative infrastructure, as is typical of developing countries, find it difficult to enforce tax compliance partly due to lack of reliable records on earnings by taxpayers. Thus, the potential that information technology (IT) afforded to gather and analyze large amounts of data on taxpayers at relatively minimal cost has caught the attention of tax authorities throughout developing countries. Tax reform efforts to enhance monitoring earnings and improve tax collection "in

*CHR Michelsen Institute and Syracuse University. Email: Merima.Ali@cmi.no
†Maxwell School, Syracuse University. Email: abshifa@maxwell.syr.edu
‡African Development Bank. Email: a.shimeles@afdb.org
§Ethiopian Development Research Institute. Email: w.firew@edri-eth.org

[1]See, for example, Bird, 1989; Tanzi and Zee, 2000; Acemoglu, 2005; Besley and Persson, 2010, 2011; Baskaran and Bigsten, 2013.

developing countries have generally centered on information technology" (Bird and Zolt, 2008, pp. 794). Nevertheless, there has been little, if any, systematic empirical evidence on the impact of those reforms. In this study, using administrative firm-level panel data on a large number of business taxpayers, we provide evidence on the impact of using the electronic sales register machines (henceforth, ESRMs) on tax revenues in the context of a developing country.

The focus of our study is a recent reform to expand the use of ESRMs in Ethiopia—a sub-Saharan African country with one of the lowest per capita incomes in the world and a minimal fiscal capacity. Starting in 2008, the Ethiopian Revenue and Customs Authority (ERCA) required several businesses to use ESRMs. The program has been rolled out over many rounds, and in 2014, over 60,000 firms adopted the new system to conduct business transactions. The machines register sales and print receipts. The transactions are then reported via a network to an ERCA server. Hence, once a firm starts using ESRMs, ERCA receives daily data on the firm's revenue. This provides ERCA with the ability to monitor reported revenues on a daily basis. With the traditional, paper-based receipts, this would have been prohibitively expensive and virtually impossible.

Even though ESRMs have the potential to provide more accurate transaction data to help minimize tax evasion, it is not clear whether developing countries can effectively harness ESRMs to generate higher tax revenues. First, developing countries may face technical and administrative challenges in implementing ESRMs. Operation of the machines requires a reliable supply of electricity and network infrastructure, as well as availability of technical capacity that can administer the network. In addition, processing of the massive data collected to aid decision-making would not be an easy fit, given the lack of requisite technical expertise and coordination failures among public agencies. Moreover, the machines do not enforce tax rules by themselves; they merely provide information on revenues. Whether the information is utilized to improve tax compliance depends on administrative and legal factors, as well as perception by businesses of the credibility of the threat posed by the new technology in terms of increasing oversight. For example, business owners may still evade taxes by paying bribes to tax officers, who would otherwise use the new set of data to track evasions. Thus, in settings where institutions are weak, as is typically the case in developing countries, the impact of gaining extra information on earnings may be minimal.

Second, even if ESRM adoption leads to increased enforcement among registered taxpayers, the overall effect on tax payments may go in either direction depending on how ESRMs affect the tax base. On the one hand, increasing taxes on final sales may lower demand, decrease production, and encourage exit out of (or discourage entry into) the formal sector. This effect implies that ESRMs can lead to erosion of the tax base. On the other hand, ESRM adoption may affect the decision of firms in a way that incentivizes them to operate at larger scales, hence expanding the tax base. For example, in the absence of ESRM adoption, evading taxes may require firms to operate on a smaller scale. This would be the case if, as firms become larger, they tend to leave easily detectable traces that make tax evasion harder, thereby encouraging the firms to remain small. However, if the adoption of ESRMs leads to more accurate revenue records irrespective of firm's scale of operation, firms may no longer have the incentive to perate in smaller scale in order to evade taxes. The adoption of ESRMs, by helping improve business records, may also have a direct effect on the firm's output and employment. This could happen if ESRMs help firms lower the cost of supervising their employees, making it easier for entrepreneurs to delegate tasks and expand their scale (Akcigit et al., 2016).

Hence, in order to provide a fuller picture of the effect of ESRMs on overall fiscal capacity, we empirically examine the impact of ESRMs on both tax payments by registered-taxpayers and proxies of the tax base. In order to account for possible bias that may arise due to selection into ESRM adoption, we use matching difference-in-difference (MDID) method. MDID has increasingly been used in the evaluation literature to address endogeneity concerns in studies using non-experimental data (Girma and Gorg, 2007; Becerril and Abdulai, 2010).

We find three major patterns in the data. First, firms report higher sales and pay more value-added tax (VAT) following ESRM adoption. Second, as a proxy for the effect of ESRM adoption on the tax base, we look at the effect on employment and rates of net entry into the formal sector. Rates of net entry is defined as the number of newly entered firms minus those that exited, as a share of the existing firms. We find that employment increases following ESRM adoption, suggesting that firms did not decrease their production in response to ESRM adoption. Furthermore, differences in rates of ESRM adoption across sectors or locations are not found to be associated with rates of net entry into the formal sector. In summary, the fact that reported sales and tax payments increased without lowering either employment or net entry suggests that ESRMs helped enhance overall fiscal capacity.

This paper contributes to the growing literature on the fiscal capacity of the state and tax compliance in developing countries. One of the important challenges for tax authorities in developing countries is the lack of accurate information on earnings (Jenkins and Kuo, 2000; Engel et al., 2001; Fisman and Wei, 2004; Boadway and Sato, 2009; Gordon and Li, 2009; Olken and Singhal, 2011). This motivated a number of recent studies that assessed alternative policy tools to provide tax authorities with more reliable information (see, e.g., Slemrod, 2008; Kumler et al., 2013; Naritomi, 2016; Carrillo et al., 2014; Pomeranz, 2015; Slemrod et al., 2017). Even though governments in many developing countries are expanding the adoption of electronic tax system to enhance their ability to gather, analyze, and monitor earnings information, we are not aware of any study examining the impact of electronic tax systems in the context of developing countries—a gap that our study attempts to narrow.

This paper is also related to the literature on the impact of IT on economic outcomes. These studies have mostly focused on the effect of IT on private sector productivity (Brynjolfsson and Hitt, 2000; Bresnahan et al., 2002; Stiroh, 2002). Despite the widespread adoption of IT in public service delivery, commonly known as "e-governance," assessment of the impact remains relatively unexplored (Garicano and Heaton, 2010). Two recent seminal contributions are Lewis-Faupel et al.

(2016) and Muralidharan et al. (2016), who studied the impact of IT use on public service delivery in the context of developing countries. Using evidence from India, Muralidharan et al. (2016) study the impact of using biometrically-authenticated payment systems on the effective delivery of targeted social transfer payments. Lewis-Faupel et al. (2016) documented the impact of electronic procurement on infrastructure provision in India and Indonesia. Our paper contributes to this strand of literature on IT and state capacity building in developing economies.

Research on the impact of tax reforms in developing countries is quite limited due to the lack of accurate data on tax payments. Our paper contributes to the few, but significant, advances that have recently been made in the use of administrative tax data from developing countries to study tax reforms (see, e.g., Kleven and Waseem, 2013; Best et al., 2015).

The paper is structured as follows. In Section 2, we discuss the institutional background of taxation in Ethiopia and describe the data. The empirical analysis proceeds in three steps. First, we report the preliminary results on the impact of ESRMs on reported sales, VAT, and employment (in Section 3). Then, we complement the preliminary results using evidence from matching diff-in-diff analysis (Section 4). Finally, we present results on the impact of ESRMs on net entry (Section 5). Concluding remarks follow in Section 6.

2 Background, data and outline of the empirical analysis

2.1 Background and Data

Our data set comes from Ethiopia—a country that was ravaged by a long civil war during the Cold War era and still remains one of the poorest countries in sub-Saharan Africa. In 2010, Ethiopia's GDP per capita was about 1,000 USD in current purchasing power parity. For comparison, this figure is only about one-third of the average in sub-Saharan Africa and less than one-thirtieth of the OECD average.[2]

The need for fiscal resources was no more apparent than in the lack of basic public infrastructure, such as roads that are needed to connect the markets across the country. However, as is the case with many developing countries, Ethiopia has a low level of fiscal capacity. The tax revenue, as a share of GDP, was about 12% during the decade 2001-2011. Ethiopia also relied heavily on taxes on international trade—a kind of tax that is relatively easy to enforce but tends to be more distortionary to the economy. More than 40% of Ethiopia's tax revenue came taxes from international trade—a very high ratio even by the standards of developing countries. About a third of the revenues came from income taxes.

Against this background, the government undertook two major reforms that are the focus of this study. The first one was the introduction of the VAT, which

[2]The per capita GDP for OECD, Sub-Saharan Africa and Ethiopia, respectively, are 34,483, 3,056 and 1041 (WDI online data bank, accessed on July 13, 2014).

is a main outcome variable in our paper. VAT was introduced in 2003 with the aim of broadening the domestic tax base and minimizing the dependence on trade taxes. VAT has become a significant source of government revenue contributing nearly one-fifth of domestic total tax revenue and half of indirect tax revenue. Since its introduction, the VAT rate has been set at 15%. The second reform was the adoption of ESRMs in 2008. By maintaining electronic record of business transactions, ESRMs are meant to minimize tax evasion by businesses.

Given the logistics challenges in implementing these reforms all at once, VAT registrations and ESRM adoptions were rolled out gradually. The rollout happened through a series of ad hoc directives issued by ERCA that made it mandatory for an increased number of firms to enroll in the reforms. The solid line in Figure 1 plots the number of VAT-registered taxpayers and ESRM users. VAT registration started with about 6,000 firms in 2003, gradually expanding, to reach over 140,000 firms by the end of 2014.[3] The adoption of ESRMs began with a few hundred firms in 2008. By 2014, about 73,000 taxpayers (out of over 140,000 VAT-registered taxpayers) adopted ESRMs[4].

Our data-set contains administrative tax records on the universe of VAT-registered firms in Ethiopia, both ESRM users and non-users (see Figure 1). The administrative records provide information on several factors such as sales, employment, VAT payments, location, types of business activity (sector), ownership structure, age and date of ESRM adoption. We have data covering the period from January 2003—the year of VAT introduction in Ethiopia—to the end of 2014. The data series are available on monthly frequency.

Table 1 presents some key characteristics of the firms in our sample. These firms are mostly businesses located on main streets across major cities in the country. Nearly half of the firms are from the capital city (Addis Ababa). Majority of the firms operate in the service sector. About 35% of the firms engage in retail, mostly representing shops on major city streets. About 20% of the firms are non-retail services, including services such as beauty salons, coffee shops, jewelry makers and so forth. Another 15% of the firms are from the construction sector, which has been booming in Ethiopia over the past years.

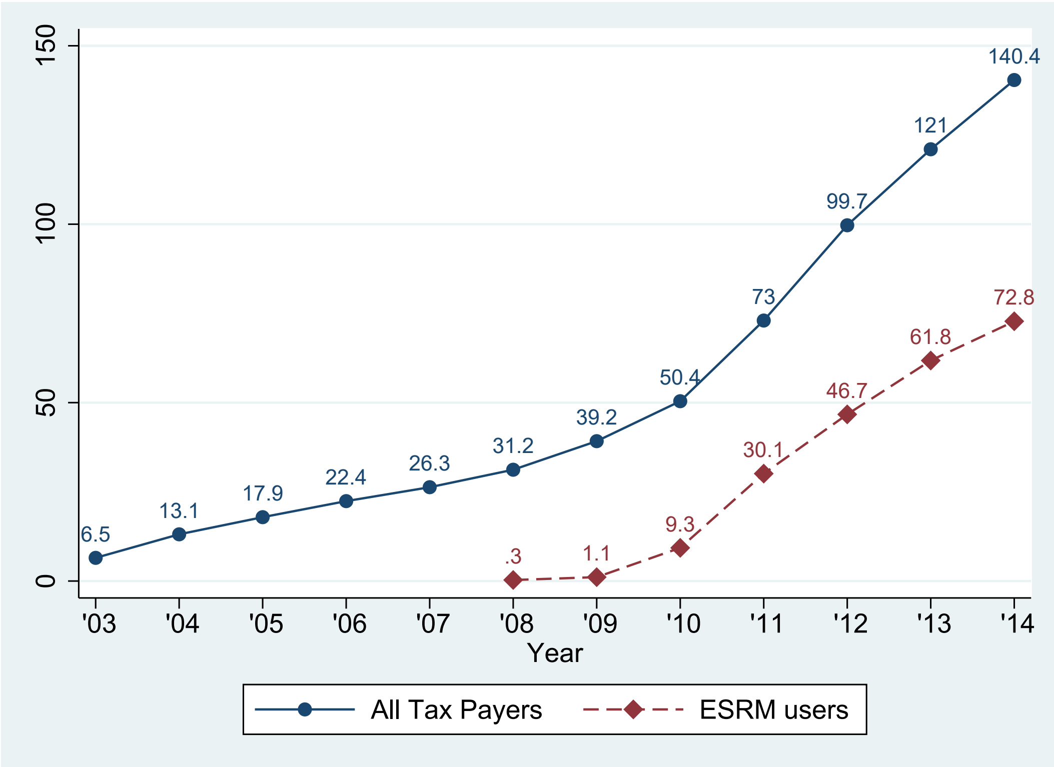
Most of the firms are family-run small-scale businesses that are typically single-outlet stores in cities, as indicated by a relatively high share (67.5%) of owner-operated firms, i.e. firms that did not report formally hiring outside labor. Of those who hired outside labor, the average employment stands at 54.6. However, the median employment is only 2.6 number of workers, implying that most of the firms are quite small.[5]

[3]The legislation for VAT registration exempted smaller firms. For a detailed theoretical discussion on the optimal VAT threshold, see Keen and Mintz (2004).

[4]As has been the case for VAT registration, smaller firms were not required to adopt ESRMs due to cost. According to our conversations with ERCA officials, the machines typically cost 5,000 to 13,000 Birr (about 250 to 650 USD in current market exchange rate), a significant expense for many businesses in Ethiopia. Once ERCA decides that a firm should use ESRM, the machines are installed at the firm's sales outlets/stores. This is done in the presence of IT technicians from ERCA who assess whether the installations satisfy the technical requirements and standards set by ERCA.

[5]This situation of firm distribution where the economy is dominated mostly by small firms

Figure 1: Number of VAT payers and ESRM adopters ('000).



On average, annual sales and VAT collections are 884,411 Birr and 112,223 Birr, respectively.[6] By the end of 2014, nearly half of the firms adopted ESRMs. We provide detailed comparisons between firms that adopted ESRMs and those that did not. These comparisons are reported in Tables A1–A4. We discus these comparisons in our matching analysis (Section 4).

2.2 Outline of empirical analysis

As discussed in the introduction, our focus is on the impact of ESRMs on overall tax capacity. ESRMs may affect tax capacity in two possible ways. First, by altering revenue data available to the tax authority, ESRMs may affect compliance behavior among registered taxpayers. This would be the case, for example, if ESRMs increase the cost of tax evasion (by increasing the likelihood of detection), hence improve compliance. Second, ESRMs may affect the tax base depending on how firms respond to ESRM adoption. For instance, if ESRMs lead to increased tax compliance, firms may respond by exiting from the formal sector (to shun ESRM adoption). The increase in tax payments may increase overall cost of production and hence induce firms to lower their output. As a result, ESRMs

is consistent with the broad pattern in developing countries (Hsieh and Olken, 2014). The high standard deviation in sales, VAT and employment in table one is also due to the large share of small firms in the economy

[6] With the nominal exchange rate of about 20 Birr/USD, these amounts correspond to 44,225 and 5,611 USD, respectively. The numbers are not adjusted for inflation.

Table 1: Descriptive statistics

	Mean	Std.	Median
Addis Ababa (dummy for located in Addis Ababa)	0.48	0.50	—
Sectors:			
Retail	0.35	0.48	—
Non-retail services	0.20	0.40	—
Construction	0.14	0.35	—
Others	0.31	0.46	—
Owner-operated	0.68	0.49	—
Annual sales ('000 Birr)	884.41	3,467.37	69.52
Annual VAT ('000 Birr)	112.22	459.10	7.92
Employment	54.64	2286.39	2.67
Adopted ESRM by 2014 (share)	0.48	0.50	—

The table presents descriptive statistics for 153,825 firms in our sample. We report the share of firms in Addis Ababa, the share of firms in each listed sectors, the share of firms operated by their owners (i.e. firs that hire no outside labor), mean annual sale, mean annual VAT, employment among non-owner operated firms and share of firms that adopted ESRMs by the end of 2014.

may lead to erosion of the tax base. In order to assess the effect of ESRMs on overall tax capacity, one thus has to look at the effects both on tax payments by registered taxpayers and on the tax base. To evaluate the former effect, we estimate the impact on VAT payments and reported sales. We examine the latter effect (i.e., on the tax base) by estimating the impacts on employment and net entry.

The empirical analyses on sales, VAT, and employment are undertaken at the firm level. The analyses on net entry are carried out at sectoral and regional levels, where we look at the association between variations in the rate of ESRM adoption and net entry across sectors and locations.

Presentation of the empirical results proceeds in three stages. In the next two sections, we present results on the impact of ESRMs on reported sales, VAT and employment (from firm-level analyses). We begin by looking at the preliminary evidences in Section 3. We then present further evidence using matching methods to address endogeneity concerns (in Section 4). Finally, we report results on the effect of ESRMs on firm net entry (in Section 5).

3 Impact of ESRMs on reported sales, VAT, and employment: Preliminary evidence

To provide preliminary estimates of the impact of ESRMs on reported sales, VAT and employment, we consider the regression equation:

$$y_{j,t} = \beta \times ESRM_{j,t} + \mu_j + \psi_t + \epsilon_{j,t} \qquad (1)$$

where $y_{j,t}$ is one of the three outcome variables in period t by firm j. μ_j and ψ_t are firm and time fixed-effects, respectively. $\epsilon_{j,t}$ is the error term. $ESRM_{j,t}$ is an indicator variable that equals 1 for the periods after ESRM adoption, and 0, otherwise. Our coefficient of interest is β. It is meant to capture the change in the outcome variables following ESRM adoption.

We aggregate the series into half-yearly frequency, so each period (denoted by t) represents six months. The reason for choosing half-yearly frequency is twofold. First, the half-yearly series is an intermediate option in the trade-off between minimizing noise from using a lower frequency (e.g., a year) and capturing the dynamics by using a higher frequency (e.g., a month). Second, as we shall see in Section 4, the number of firms that adopt ESRMs in a given month or quarter would be too few to undertake the matching analysis.

The inclusion of time-fixed effects in the specification (1) helps account for possible correlation between expansion of ESRM adoption and aggregate variables that may affect firm's revenue, such as economic growth, government spending, and inflation. Controlling for time-fixed effects is feasible due to the gradual implementation of the program through several rounds (as discussed in Section 2).

The firm-fixed effects, on the other hand, help address bias that may arise due to potentially systematic and time-invariant differences between firms that use ESRMs and those that do not. The identification in the above specification relies on variations within the firm as opposed to a cross-sectional comparison between groups of firms that used the ESRM and those that did not.

Table 2 reports estimates of the effects from the regression specification (1). Distributions of sales, VAT, and employment are not normal, due to both the presence of observations with zero values and some outliers on the right tail. In order to account for this, we use the log transformations (adding 1) as dependent variables (columns [1] through [3]). This transformation helps minimize the problem of outliers and enables us to use all observations.[7] Alternatively, we consider, as the dependent variable, a dummy indicating whether the firm reported positive values (columns [4] through [6]). Robust standard errors clustered at firm level are in parentheses.

Table 2: Results from fixed-effects panel regressions

Dependent variable					
$\log(1+sales)$	$\log(1+VAT)$	$\log(1+Emp.)$	$(Sales>0)$	$(VAT>0)$	$(Emp.>0)$
[1]	[2]	[3]	[4]	[5]	[6]
2.27	1.89	0.21	0.16	0.15	0.10
(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
Obs. $= 810,707$					
Firms $= 153,825$					

Notes: This table reports estimated impact of ESRM from a fixed-effects panel regressions (as specified by Equation 1). The outcome variables are reported sales, VAT and employment (in log scales as well as binary indicators for whether a firm reported positive values for the outcomes). Robust standard errors clustered by firms are in parentheses.

[7] Moreover, the effect of ESRMs is likely to depend on some base values due to factors like inflation and firm size, making log-transformed dependent variables more appropriate.

The first column presents the estimated impact of ESRMs on reported sales. We see that reported sales increase significantly following ESRM adoption. Even if ESRMs increase reported sales, VAT payments may or may not increase, depending on whether firms increase reported cost of inputs and lower their tax liability (Slemrod et al., 2017). The second column reports the effect on VAT, which shows that VAT payments also increased significantly following ESRM adoption.

As briefly discussed in Section 1, ESRMs may affect not only tax reporting but also actual output. On the one hand, if ESRMs increase effective taxes on final sales through stronger enforcement, firms could decrease production due to increased production cost as a result of the extra tax burden. On the other hand, one could also imagine plausible scenarios where ESRM adoption could incentivize firms to operate at larger scales, thereby expanding output. By providing more accurate revenue records to the revenue authority, adoption of ESRMs could minimize asymmetry between small and large firms with respect to their ability to hide revenues, and, hence reduce firms' incentive to operate at smaller scales with the intent of evading taxes. This in effect can increase firms' output and expand the tax base.[8] By helping improve business records, ESRMs may also have a direct positive effect on the firms' output. This could happen if, for example, ESRMs help firms lower the cost of supervising their employees, making it easier for entrepreneurs to delegate tasks and expand their scale (Akcigit et al., 2016).

The estimated changes in reported sales and VAT do not distinguish the changes between reported and actual output. Therefore, they are not satisfactorily informative about the effect of ESRMs on actual output. In fact, one cannot rule out the possibility that actual output may decrease while reported sales and VAT payments increase. This could happen if firms pay more taxes (due to improved compliance) even if their output decreases. Output is not directly observed in our dataset. However, we have data on employment, and therefore, we use it as an alternative dependent variable to examine the effect of ESRMs on output and firm size. This result is reported in the third column. We see that employment also increased significantly following ESRM adoption.

Columns [4]–[6] show that the results point to similar patterns when we consider the dummy indicators for whether the outcome variables have positive values. The likelihood that one observes positive values for reported sales, VAT payments and employment increases significantly following the adoption of ESRMs.

4 Impact on sales, VAT, and employment: Evidence from MDID analysis

Even though the estimates from fixed-effects regressions in Table 2 provide suggestive evidence on the impact of ESRMs, they could be biased if selection into ESRM

[8]Operating at larger scales may make harder to evade taxes since expanded scales need more reliance on formal/contractual relations due to, for example, the need to fill jobs with employees outside one's personal network (such professional managers, instead of family members). By leaving traces of legally verifiable records, the contractual relationships may in turn make it more difficult to hide business transactions from the tax authority.

adoption is associated with other time-varying factors that would have occurred in the absence of ESRM adoption. This would be the case, for example, if ERCA selected firms into ESRM adoption when firms acquired some productivity gains that are unobservable in the data (e.g. product innovation). In such a case, one cannot fully attribute estimates from the specification to ESRMs, since the outcome variables are likely to change even in the absence of ESRM adoption. To mitigate this concern, we now report results using the matching difference-in-difference (MDID) approach, which has increasingly been employed by the evaluation literature to address endogeneity concerns in studies using non-experimental data. A useful aspect of MDID, as discussed ahead, is that it combines the desirable features of both matching and difference-in-difference methods (Blundell and Dias, 2000).

4.1 Econometric framework

The aim of matching is to pair each firm that adopted ESRMs with those that did not, so that the non-adopters can be used as a counterfactual for the adopters in examining the impact of ESRMs on adopters. Compared to the approach in Section 3 where all of the non-adopters are included in the control group, matching is desirable since it minimizes the likelihood of bias if the non-adopters are considerably different from the adopters.

Once observations from treatment and control groups are matched (based on pre-treatment characteristics), standard matching methods use observations in the matched sample to estimate the difference in (weighted) mean outcome levels between treatment and control groups. However, since we have longitudinal data, we estimate the difference in mean differences (instead of the difference in mean levels) by using the following diff-in-diff equation.

$$y_{i,t} = \theta_4 \ (Post_t * Treated_i) + \theta_3 \ Post_t + \theta_2 \ Treated_i + \theta_1 + \epsilon_{i,t} \qquad (2)$$

$Post_t$ is a dummy for the post-treatment period. $Treated_i$ is an indicator for whether the firm belongs to the treated group. θ_1 and $\theta_1 + \theta_2$ are group-specific means for the treated and comparison groups, respectively. The coefficient of interest is θ_4. It captures the difference in trends between the treatment and comparison groups.

This procedure of combining matching and diff-in-diff methods helps exploit the advantage of both methods. Whereas the standard matching estimator (of differences in levels) would require the strong assumption that, in absence of the treatment, levels of outcome variables should be the same across treatment and control groups in a matched sample, causal interpretation in the matching diff-in-diff requires a relatively weaker assumption by allowing for unobserved time-invariant differences between the two groups (Smith and Todd, 2005). Hence, the combination of matching and diff-in-diff "has the potential to improve the quality of non-experimental evaluation results significantly" (Blundell and Dias, 2000, p. 438).

As shall be described in Section 4.3, we construct four matched samples. We then run the diff-in-diff equation (2) on each of the four samples. In the first sample, the treated groups adopted ESRMs in the first half of 2011 (2011:1). We then examine how the trends for these firms compare (before and after this adoption period) with the trends for firms that did not yet adopt ESRMs. In comparing the trends, we consider a window of two periods before and after ESRM adoption, so t = 2010:1, 2010:2, 2011:1, 2011:2, 2012:1. These periods represent the pre-adoption time (2010:1, 2010:2), the adoption period (2011:1) and the post-adoption period (2011:2, 2012:1). We construct the control group by matching the treated firms with those that had not adopted ESRMs by the end of our comparison period (2012:1).

In the second sample, the treated group consists of firms that adopted ESRMs in 2011:2. Correspondingly, we examine the trends during the periods t = 2010:2, 2011:1, 2011:2, 2012:1, 2012:2. The set of control firms in this sample are selected from firms that had not adopted ESRMs by 2012:2. In the remaining two samples, the treated firms consist of firms that adopted ESRMs in 2012:1 and 2012:2. The control groups are selected from the firms that had not adopted ESRMs during the corresponding comparison periods.

In constructing the matched samples, our focus on the periods 2011 and 2012 is due to sample size reasons. First, a relatively large number of firms that newly adopted ESRMs during 2011 and 2012 (see Figure 1). Second, there were relatively many firms that had not yet adopted ESRMs during those periods. Thus, these periods provide us with reasonable numbers of both treated and control groups to construct the matched samples.

4.2 Matching variables

We use several variables to match the treatment and control group. All of the variables are sourced from the administrative tax records. The matching covariates consist of two broad sets of pretreatment variables. The first set of covariates include several time-invariant firm characteristics, and, hence are meant to address the concern that treatment and control groups may have differential trends due to some persistent differences between them. These covariates include: a location indicator for whether the firm is in Addis Ababa; two indicators for sector, one for retail and another one for other services (the third group is manufacturing); an indicator for whether the firm is a sole proprietorship (as opposed to limited liability); and two dummies indicating the firm's size category as registered by ERCA (small, medium, and large).

The second set of variables are intended to match the control and treatment groups with respect to time-varying characteristics and, hence, are meant to address the concern that the estimated effects may be confounded by differential time-trends between the treatment and comparison groups due to temporary shocks that affect the two groups differently. To capture firms' pretreatment dynamics in sales, tax payments, and production, we include the first and second lags of the three outcome variables (sales, VAT, and employment). To account for shocks at sector and local levels, we include the lags of average sales at district and sector levels.

4.3 Matched samples and balancing tests

Choice of the matched sample is crucial, since it is likely to affect the estimated effects. In constructing the matched sample, we follow three broad sets of "best-practice" guidelines that the matching literature has identified (Caliendo and Kopeinig, 2008; Imbens, 2015; King et al., 2017). First, differences in the outcome variables between treated and control groups should not influence the choice of the matched sample. That is, the outcome variable should not be included as part of the matching covariates. This is meant to avoid bias owing to selectively picking a matched sample that supports one's favored hypothesis. The second guideline relates to the variance–imbalance trade-off. Balance between the treated and comparison groups is achieved by dropping observations until the treated and comparison units are reasonably similar in the remaining sample (i.e. matched sample). This process of pruning observations to arrive at the matched sample inherently involves a trade-off between size of the matched sample and the level of balance—dropping more observations to achieve better balance leads to fewer observations (i.e., higher variance). Thus, in choosing the matched sample, one should aim to optimize the variance-imbalance trade-off; that is, one should maximize the size of the matched sample for any given level of imbalance or minimize the imbalance for any given level of size (King et al., 2017). Third, given the variety of available matching approaches that one can choose from, it is important to ensure that the results are not driven by restrictively selecting matching algorithms among the available options (Imbens, 2015). One thus has to verify robustness of results to reasonable changes in the choice of matching algorithms to generate the matched sample. Following these guidelines, we construct the matched samples by selecting observations based on only the matching variables (but not the outcome variables). In order to examine the sensitivity of results, we report estimates using alternative approaches to constructing the matched samples. We also put particular attention on the variance-imbalance trade-off.

Since the treatment and control groups are matched typically across several characteristics, matching methods use a single distance metric in order to reduce the dimensionality problem. Let vector $\mathbf{x}_i$ denote the covariates of matching characteristics for firm i. The distance metric between two firms is some function of the covariate values for the two firms, $d_{i,j} = d(\mathbf{x}_i, \mathbf{x}_j)$. Hence the metric is meant to contain information on all the characteristics and serve as a measure of (dis)similarity between the two firms with respect to the matching characteristics. Observations in the treatment group are matched with their nearest neighbors in the control group, where neighborhood proximity is defined by the distance metric.

The most commonly used measures are computed using propensity scores and Mahalanobis distance. In propensity score matching, the probability of receiving the treatment (ESRM adoption), as a function of the matching characteristics, is estimated for each firm using a probit or logit model. Then, distance between any pair of firms i and j is defined as the absolute difference in the predicted probabilities ($\hat{P}$) for the pairs, $d_{i,j} = |\hat{P}(\mathbf{x}_i) - \hat{P}(\mathbf{x}_j)|$ (Rosenbaum and Rubin, 1983, 1985; Dehejia and Wahba, 2002). In Mahalanobis matching, the distance metric is defined as $d_{i,j} = (\mathbf{z}'_{i,j}\mathbf{V}\mathbf{z}_{i,j})^{1/2}$, where $\mathbf{z}_{i,j} \equiv \mathbf{x}_i - \mathbf{x}_j$ and $\mathbf{V}$ is the sample covariance matrix for the covariates. Notice that this distance metric would be equivalent to Euclidean distance if one replaces $\mathbf{V}$ with an identity matrix. Thus, Mahalanobis distance can be interpreted as Euclidean distance between normalized values of covariates (where the covariates are normalized using the covariance matrix).

We begin with a relatively straightforward matching, where each treated unit is matched with its nearest neighbor. Neighborhood distance among observations is defined based on the similarity of the matching covariates, as measured by Mahalanobis distance. Table 3 reports the level of balance between treatment and comparison groups—as measured by standardized differences—for each of the matching covariates.

Standardized differences are commonly used to examine the similarity of distribution of matching variables between treatment and comparison groups. First described in Rosenbaum and Rubin (1985), mean standardized bias (in percentages) for a covariate X between the treated and comparison units are given by:

$$MSB_{before}(X) = 100 \times \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{V_T(X)+V_C(X)}{2}}} \qquad MSB_{after}(X) = 100 \times \frac{\bar{X}_{TM} - \bar{x}_{CM}}{\sqrt{\frac{V_T(X)+V_C(X)}{2}}}$$

where $MSB_{before}(X)$ denotes the mean standardized bias (MSB) between the treated and comparison groups in the full sample and $MSB_{after}(X)$ denotes MSB in the matched sample. $\bar{X}_T$ and $\bar{X}_C$ are the sample means for the full treatment and comparison groups; $\bar{X}_{TM}$ and $\bar{X}_{CM}$ are the sample means for the treatment and comparison groups in the matched sample, and $V_T(X)$ and $V_C(X)$ are the sample variances for the full treatment and comparison groups. There is no universally accepted cutoff value for MSB to decide whether the matching is satisfactory. However, some authors suggest that MSB of 10% or 5% could roughly be considered as an indicator for negligible imbalance (Rosenbaum and Rubin, 1985; Caliendo and Kopeinig, 2008). Thus, in constructing the matched sample, we prune observations until—rather conservatively—the $MSBs$ are below 5% for each of the matching covariates.

Table 3 presents the balance statistics for each of the four matched samples.[9] The first three columns report balance statistics for the group in which treatment firms adopted ESRMs in the first half of 2011. We report $MSBs$ for the matched sample, as well the percent reduction in $MSBs$ (columns labeled "MSB" and "%Δ", respectively). We see that the matched samples have balanced the treatment and comparison groups quite successfully. In each of the four groups, the biases between the treatment and comparison groups more or less disappear in the matched samples. In almost every case, more than 90% of the biases in the unmatched sample have been removed in the matched samples.

Whereas the $MSBs$ compare each variable separately for the treated and comparison units, we also report *joint* statistics, where all the matching variables are taken together in comparing the treated and comparison groups (see Panel B of Table 3). These statistics are based on probit regressions where an indicator dummy for treatment status is regressed on the matching variables. The idea is that if the

[9]See Section 4.1 for a description of each matched sample.

Table 3: Mean standardized biases (MSB) and bias reduction (Δ) due to matching.

	Treatment group adopted ESRMs during:							
	2011:1		*2011:2*		*2012:1*		*2012:2*	
	MSB [1]	% Δ [2]	MSB [3]	% Δ [4]	MSB [5]	% Δ [6]	MSB [7]	% Δ [8]
Panel A: MSBs for individual matching variables								
$\log(1 + sales_{t-1})$	-0.5	99.5	-0.1	99.9	0	100	-0.3	99.6
$\log(1 + sales_{t-2})$	-0.4	99.5	-1.4	97.7	-0.1	99.8	-1.9	96.1
$sales_{t-1} > 0$	0.1	99.9	0	100	0	100	-0.2	99.7
$sales_{t-2} > 0$	0.1	99.9	-0.3	99.6	0	100	-1.6	96.8
$\log(1 + emp_{t-1})$	-0.1	93	0.1	93.8	0.1	98	1.2	93.1
$\log(1 + emp_{t-2})$	-0.8	87.8	0	99.7	0.1	97.3	0.8	93.9
$emp_{t-1} > 0$	0.4	91.1	0	100	0	100	0.3	99.1
$emp_{t-2} > 0$	-0.3	93.3	0	100	0	100	0.3	98.7
age	-0.7	97.4	1.2	96.1	1.5	98.2	1.6	97.2
age^2	0.1	99.3	1.1	95.7	1.4	96.8	1.6	95.5
Large	0	100	0.1	99.9	0	100	1.2	74.3
Small	0	100	0	100	0	100	0.2	58.4
LimitedLiability	-0.1	99.9	-2.4	94.2	-2.4	94.2	1.1	97.3
Retail	0.4	99.4	-0.2	99.7	-0.2	99.7	1.3	96.9
Service	0	100	0	100	0	100	0.1	99.7
$SectorSales_{t-1}$	1.5	94.5	2.8	93.3	2.8	93.3	2.2	95.4
$DestrictSales_{t-1}$	3.1	97	3.3	96.3	3.3	96.3	3.2	93.7
Addis Ababa	0	100	0.7	99	0.7	99	1.8	84.6
Panel B: Joint statistics								
Pseudo R-squared	0.001	—	0.002	—	0.001	—	0.001	—
Log-likelihood ratio (P-value)	0.428	—	0.888	—	0.985	—	0.999	—

Notes: This table reports balance statistics for matching covariates in four matched samples. The columns labeled as *MSB* report Mean Standard Biases among the matched firms. Columns that are labeled as %Δ report the bias reduction due to matching. The matched samples are categorized by the period of adoption among the treated firms. In the first two columns, the treatment groups adopted ESRMs in 2011:1. In the next 6 columns, the treatment groups adopted ESRMs in 2011:2, 2012:1 and 2012:2.

treated and comparison groups are similar (with respect to the regressors), then the regressors should provide little power to predict the likelihood of receiving treatment. The first row in Panel B reports pseudo R-squared values for each of the four samples. We see that R-squared values are virtually zero for all of the four matched samples. The second row of Panel B reports p-values for joint significance of the regressors in the probit regressions. These p-values also show that in the matched samples, one cannot reject the null that the regressors are jointly insignificant, providing no indication that differences between the treatment and control groups predict the likelihood of receiving treatment.

The mean comparisons for each covariates before and after the match between treated and control groups is further reported in the appendix (Tables A1–A4). The tables confirm that in the matched sample, the treated and control firms do not display significant differences.

Table 4 reports the diff-in-diff estimates for the four matched samples (i.e., the coefficient β in Equation (2)). In the top three rows, the dependent variables are sales, VAT, and employment (in log scales), respectively. The three bottom rows report the estimated coefficient where the dependent variables are indicators for whether the firm reported positive values for sales, VAT, and employment. The observed patterns affirm the earlier results from the fixed-effects panel regression reported in Table 2. Reported sales, VAT payments, and employment all increase following ESRM adoption. The matching estimates are generally larger than the estimates from the panel regressions, providing no indication whether the positive effects of ESRM adoption estimated from the panel regression are driven by selection.

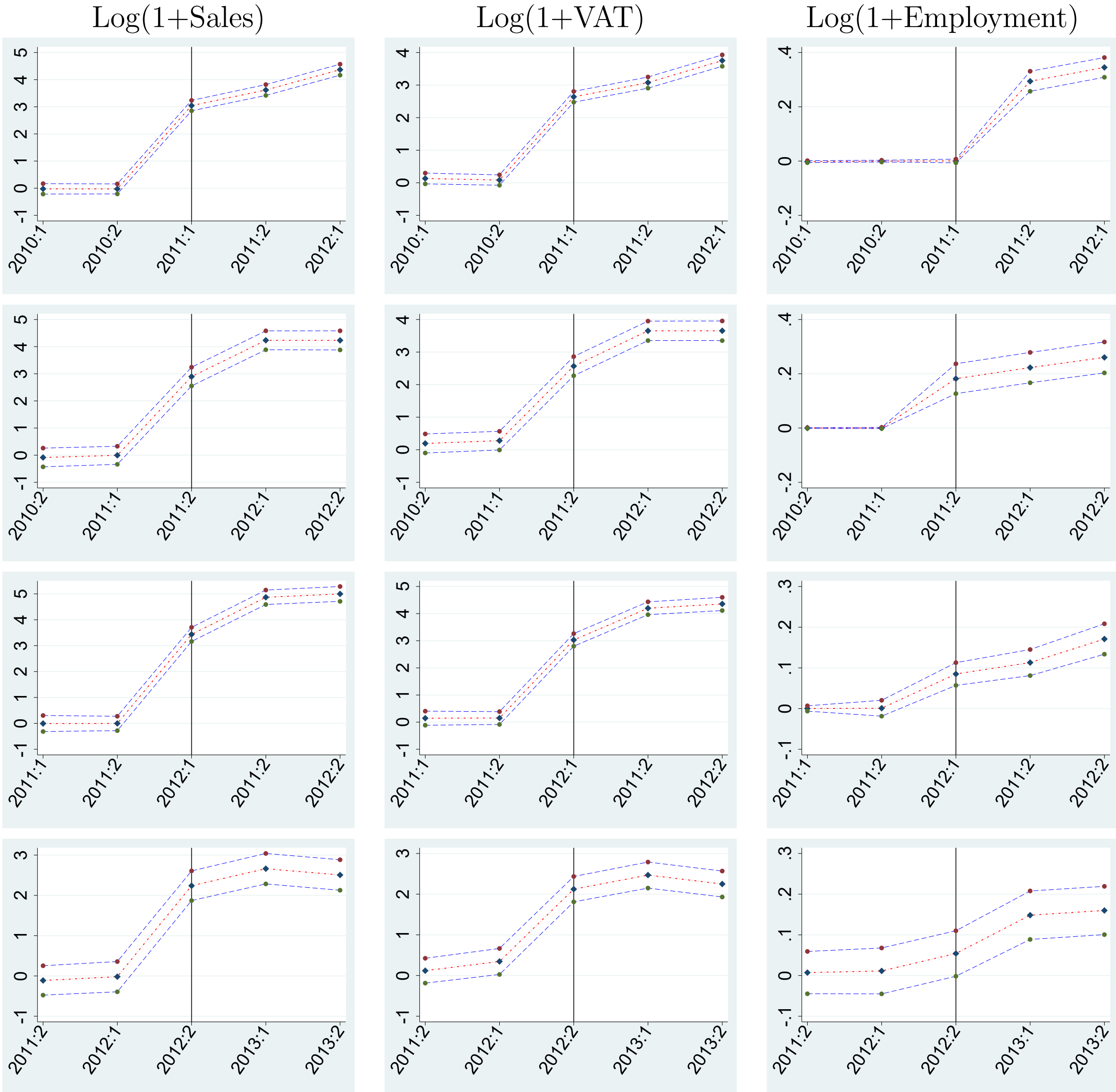
The visual display trend differences in Figure (2) present a relatively transparent look at the data. Dotted lines represent the 95% confidence intervals. For each of the four matched samples (corresponding to each row), we plot the trends for the three outcome variables (all in log scales). The period of adoption in each group is indicated by the vertical line, that is, the treated groups adopted ESRMs in the period marked by vertical lines. These plots mimic the patterns reported in Tables 3 and 4—that while there is no significant difference between treatment and comparison groups prior to ESRM adoption, the treatment groups display significantly larger values for each of the outcome variables.

We have undertaken three sets of robustness checks. First, we assess robustness of the results to the choice of alternative matching algorithms. Second, we examine whether the results were driven by external effects, which are particularly important in assessing the effect of ESRMs on overall outcomes (such as total tax revenue). Finally, we carry out placebo analyses examining the trend difference between treated and comparison groups during the prematching period (instead of the postmatching period). We will discuss each of them in the next sub-section.

4.4 Robustness analysis

We have implemented several matching strategies to examine sensitivity of the results to the choice of matching algorithms. Instead of using Mahalonibis distance, we conduct nearest-neighbor matching, using differences in propensity scores as

Figure 2: Trend differences (with 95% CI) betwen treated and comparison units.



Notes: The figures show the differences in reported sales, VAT and employment between treated and comparison groups for four matched samples (corresponding to each row). The vertical lines show the period of ESRM adoption for the treated group.

Table 4: Matching Diff-in-Diff results using nearest neighbor Mahalanobis matching

	Treatment group adopted ESRMs during:			
	2011:1	*2011:2*	*2012:1*	*2012:2*
	[1]	[2]	[3]	[4]
Dependent variable:				
Sales	3.71	3.83	4.45	2.50
	(0.23)	(0.24)	(0.20)	(0.17)
VAT	3.05	3.05	3.72	2.02
	(0.20)	(0.20)	(0.17)	(0.14)
Employment	0.22	0.22	0.12	0.11
	(0.04)	(0.05)	(0.02)	(0.02)
$\mathbb{1}_{Sales>0}$	0.29	0.30	0.37	0.20
	(0.02)	(0.02)	(0.02)	(0.01)
$\mathbb{1}_{VAT>0}$	0.29	0.29	0.37	0.19
	(0.02)	(0.02)	(0.02)	(0.01)
$\mathbb{1}_{Employment>0}$	0.10	0.13	0.08	0.06
	(0.02)	(0.02)	(0.01)	(0.01)
Observations	76,940	25,380	30,800	22,150
Firms	16,188	5,446	6,844	5,082

Notes: This table reports estimated impact of ESRM adoption using the specification given by Equation 2. The first column lists the dependent variables. The estimate are provided for four matched samples (corresponding to each columns [1] through [4]). In column [1], the treated group consists of firms that adopted ESRMs during the first half of 2011. The comparison group consists of firms that either never adopted ESRMS or adopted after a year (after 2012:2). Similarly, in columns [2], [3] and [4], the treated group adopted ESRMs during 2011:2, 2012:1 and 2012:2, respectively. The comparison group in each matched sample consist of either firms that never adopted ESRMs or firms that adopted ESRMs a year after the treated firms in the respective matched sample. Robust standard errors clustered by firms are in parentheses.

the distance metric. The results are reported in Table 5. We see that the propensity score matching tends to perform poorly in the variance–imbalance trade-off, in the sense that the numbers of firms in the matched samples are generally smaller than those in Mahalnobis matching.[10] However, the results remain the same.

Table 5: Matching Diff-in-Diff results using nearest neighbor propensity score matching

	Treatment group adopted ESRMs during:			
	2011:1 [1]	*2011:2* [2]	*2012:1* [3]	*2012:2* [4]
Dependent variable:				
Sales	3.65	3.76	4.14	2.03
	(0.24)	(0.22)	(0.18)	(0.17)
VAT	2.99	2.85	3.42	1.71
	(0.20)	(0.18)	(0.15)	(0.14)
Employment	0.14	0.19	0.13	0.12
	(0.08)	(0.03)	(0.02)	(0.02)
$\mathbb{1}_{Sales>0}$	0.30	0.30	0.34	0.16
	(0.02)	(0.02)	(0.02)	(0.01)
$\mathbb{1}_{VAT>0}$	0.30	0.27	0.34	0.16
	(0.02)	(0.02)	(0.02)	(0.01)
$\mathbb{1}_{Employment>0}$	0.10	0.12	0.09	0.06
	(0.02)	(0.01)	(0.01)	(0.01)
Observations	44,920	12,500	18,980	17,040
Firms	5,756	2,486	5,232	4,626

Notes: This table reports estimated impact of ESRM adoption using the specification given by Equation 2. The first column lists the dependent variables. The estimate are provided for four matched samples (corresponding to each columns [1] through [4]). In column [1], the treated group consists of firms that adopted ESRMs during the first half of 2011. The comparison group consists of firms that either never adopted ESRMS or adopted after a year (after 2012:2). Similarly, in columns [2], [3] and [4], the treated group adopted ESRMs during 2011:2, 2012:1 and 2012:2, respectively. The comparison group in each matched sample consist of either firms that never adopted ESRMs or firms that adopted ESRMs a year after the treated firms in the respective matched sample. Robust standard errors clustered by firms are in parentheses.

As noted by Dehejia and Wahba (2002) and several others, caliper matching tends to perform better in terms of the trade-off between balance and size. This is due to the fact that caliper matching—as opposed to nearest-neighbor matching, which includes only the nearest observation(s) from the control units—uses all comparison units within a predefined radius, and hence, includes "as many comparison units as available within the calipers, allowing for the use of extra (fewer) units when good matches are (not) available" (Dehejia and Wahba, 2002 p. 153–154). Thus, we run the analyses using caliper matching by implementing

[10]The relatively poor performance of propensity score matching with regard to the variance–imbalance trade-off is also noted by King et al. (2017).

the algorithm that Lechner et al. (2011) proposed.[11] The results are reported in Table 6. We see that the results remain similar.

Table 6: Matching Diff-in-Diff results using radius matching

	Treatment group adopted ESRMs during:			
	2011:1	*2011:2*	*2012:1*	*2012:2*
	[1]	[2]	[3]	[4]
Dependent variable:				
Sales	3.93	3.86	4.56	2.64
	(0.22)	(0.23)	(0.16)	(0.14)
VAT	3.21	3.11	3.79	2.12
	(0.19)	(0.19)	(0.14)	(0.12)
Employment	0.23	0.25	0.14	0.12
	(0.03)	(0.04)	(0.02)	(0.02)
$\mathbb{1}_{Sales>0}$	0.31	0.31	0.37	0.20
	(0.02)	(0.02)	(0.01)	(0.01)
$\mathbb{1}_{VAT>0}$Post	0.30	0.30	0.37	0.19
	(0.02)	(0.02)	(0.01)	(0.01)
$\mathbb{1}_{Employment>0}$Post	0.11	0.14	0.09	0.07
	(0.01)	(0.02)	(0.01)	(0.01)
Observations	106,910	72,925	107,560	119,175
Firms	22,943	15,806	23,780	25,679

Notes: This table reports estimated impact of ESRM adoption using the specification given by Equation 2. The first column lists the dependent variables. The estimate are provided for four matched samples (corresponding to each columns [1] through [4]). In column [1], the treated group consists of firms that adopted ESRMs during the first half of 2011. The comparison group consists of firms that either never adopted ESRMS or adopted after a year (after 2012:2). Similarly, in columns [2], [3] and [4], the treated group adopted ESRMs during 2011:2, 2012:1 and 2012:2, respectively. The comparison group in each matched sample consist of either firms that never adopted ESRMs or firms that adopted ESRMs a year after the treated firms in the respective matched sample. Robust standard errors clustered by firms are in parentheses.

We have also checked robustness of the results to adjusting the balancing criteria by which we prune the sample. Instead of using 5% as a cutoff MSB value to prune observations, we used 3% and 10% as cutoff MSB values. These changes are not found to alter the results—reported sales, VAT payments, and employment show a significant increase following ESRM adoption.

[11]The caliper-matching algorithm developed by Lechner et al. (2011) involves four major steps. First, probability of treatment (propensity score) is estimated using a standard probit model. Second, for each treatment unit, Mahalanobis distances between the treatment unit and comparison units are computed (over a subset of the matching covariates and the estimated propensity scores). Third, each treatment unit is matched with a set of comparison units that lie within a given distance (or radius) from the treatment unit. Finally, control units are weighted, based on their similarity (as defined by the Mahalanobis distance) with the treatment units with which they are matched.

Our next sets of robustness checks relate to the external effects, which may arise if the adoption of ESRMs by some firms affects outcomes for non-adopters. Such external effects may occur particularly among firms within the same locality or sector. For example, if ESRM adopters become more competitive (vis-à-vis non-adopters), non-adopters may lose some of their market shares; hence, sales, VAT payments, and employment by non-adopters may decrease. In this this case, the observed difference between adopters and non-adopters is due to not only increases by adopters, but also decreases by non-adopters. Hence, coefficients from the matching regressions may over-estimate the effect. On the contrary, the coefficients could under-estimate the effect if ESRM adopters become less competitive and lose some of their market share.

In order to address this concern, we reran the analysis, after excluding (from the comparison group) observations located in districts with relatively high levels of adoption rates. The assumption is that external effects are likely to matter in areas with high adoption rates. Hence, if the results are driven by external effects, they should not hold when the comparison group includes only the observations that are located in districts with low adoption rates (as they are unlikely to have been affected by the externalities). We first ranked districts according to the share of firms that adopted ESRMs in each district (as of the period in which the treatment firms adopted ESRMs). We then dropped all untreated observations located in districts where the adoption rate is above the median district and reran the analysis. Table 7 reports the results. The matched sample size is now smaller as we have fewer firms to match. However, the results remain the same.

One may, perhaps, be concerned that the external effects could work through sectors rather than locations. Thus, we also ranked sectors according to rate of ESRM adoption (i.e., the fraction of firms in the sector that adopted ESRMs). We then reran the analysis after excluding observations (from the comparison group) where the adoption rate for the firm's sector is above that of the median sector. We found similar results.

Finally, we have undertaken placebo analyses to assess whether the difference between the treated and comparison groups would occur in periods that are presumably unaffected by ESRM adoption. Recall that in the previously described matching analysis, we included—among others—the first and second lags of the outcome variables in the matching covariates. Thus, the matching analysis involved selecting treated and comparison firms based on their similarity during the *matching periods* (i.e., the two periods preceding ESRM adoption), and then comparing trend differences between the groups during the *postmatching periods*. The assumption is that, in absence of the treatment, the treated and comparison groups would have displayed similar trends outside the matching period. One would then expect the two groups to display similar trends during the *prematching periods* (as opposed to the postmatching periods), since ESRMs had not yet taken effect. The presence of significant differences during the prematching periods would cast doubt on the assumption that, in the absence of ESRM adoption, the two groups would have similar trends outside the matching periods.

Panel (A) of Figure (3) presents the trend differences in sales during the match-

Table 7: Accounting for external effects: Results from Matching Diff-in-Diff when comparison units are drown from districts with low adoption rates

	Treatment group adopted ESRMs during:			
	2011:1	*2011:2*	*2012:1*	*2012:2*
	[1]	[2]	[3]	[4]
Dependent variable:				
Sales	3.68	3.00	3.29	1.91
	(0.38)	(0.27)	(0.21)	(0.20)
VAT	3.02	2.14	2.83	1.63
	(0.32)	(0.22)	(0.18)	(0.17)
Employment	0.34	0.33	0.16	0.08
	(0.02)	(0.03)	(0.01)	(0.02)
$\mathbb{1}_{Sales>0}$	0.30	0.24	0.27	0.15
	(0.03)	(0.02)	(0.02)	(0.02)
$\mathbb{1}_{VAT>0}$	0.29	0.21	0.27	0.15
	(0.03)	(0.02)	(0.02)	(0.02)
$\mathbb{1}_{Employment>0}$	0.16	0.18	0.10	0.04
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	51,680	17,070	31,220	18,780
Firms	11,736	4,868	6,718	5,082

Notes: This table reports estimated impact of ESRM adoption using the specification given by Equation 2. The first column lists the dependent variables. The estimate are provided for four matched samples (corresponding to each columns [1] through [4]). In column [1], the treated group consists of firms that adopted ESRMs during the first half of 2011. The comparison group consists of firms that either never adopted ESRMS or adopted after a year (after 2012:2). Similarly, in columns [2], [3] and [4], the treated group adopted ESRMs during 2011:2, 2012:1 and 2012:2, respectively. The comparison group in each matched sample consist of either firms that never adopted ESRMs or firms that adopted ESRMs a year after the treated firms in the respective matched sample. Robust standard errors clustered by firms are in parentheses.

ing and postmatching periods, that is, it visually displays the impact of ESRM adoption estimated from the matching diff-in-diff specification (see Table 4). Panel (B) shows the placebo comparison where we plot trend differences during the matching and prematching periods. Each row represents one of the four matched samples. Whereas the first column shows the trends as we move from the matching period towards the postmatching period, the second column displays the trends as we move (back in time) from the matching period towards the prematching period. The vertical lines in the first column represent the actual treatment period (beginning of postmatching period), while the vertical lines in the second column show the placebo period (beginning of prematching period). The vertical axes within each row have the same scale so that displayed differences between the placebo and treatment effects are directly comparable.

These figures show that the increase in reported sales occurs only when we move from the matching period to the postmatching period. The changes, as we move from the matching period to the prematching periods (back in time), are relatively small and statistically insignificant, providing no indication that the treated comparison groups would diverge in the absence of ESRM adoption. We have done similar comparisons for VAT payments and employment, finding similar results—that significant differences arise only in the postmatching period (see Figure .4 and Figure .5 in the Appendix).

5 ESRMs and net entry

As discussed in Section 3, the potential contribution of ESRMs to build fiscal capacity may be undermined if firms lower their output in response to increases in the cost of production (due to extra tax liability), thereby leading to erosion of the tax base. Although the positive effect of ESRMs on employment suggests that this may not be the case, ESRMs may still affect the tax base if they affect the decision of firms to enter to (or exit from) the formal sector. For example, if ESRMs make it more difficult for firms to evade taxes, existing firms may respond by leaving the formal sector and operating informally where they can shun the adoption of ESRMs. Similarly, the threat of ESRM adoption may discourage potential new entrants into the formal sector.

Admittedly, identifying the causal effect of ESRM adoption on net entry (entry minus exit) is difficult. We nonetheless present the available preliminary evidence on the association between ESRM adoption and net entry. We consider the following regression:

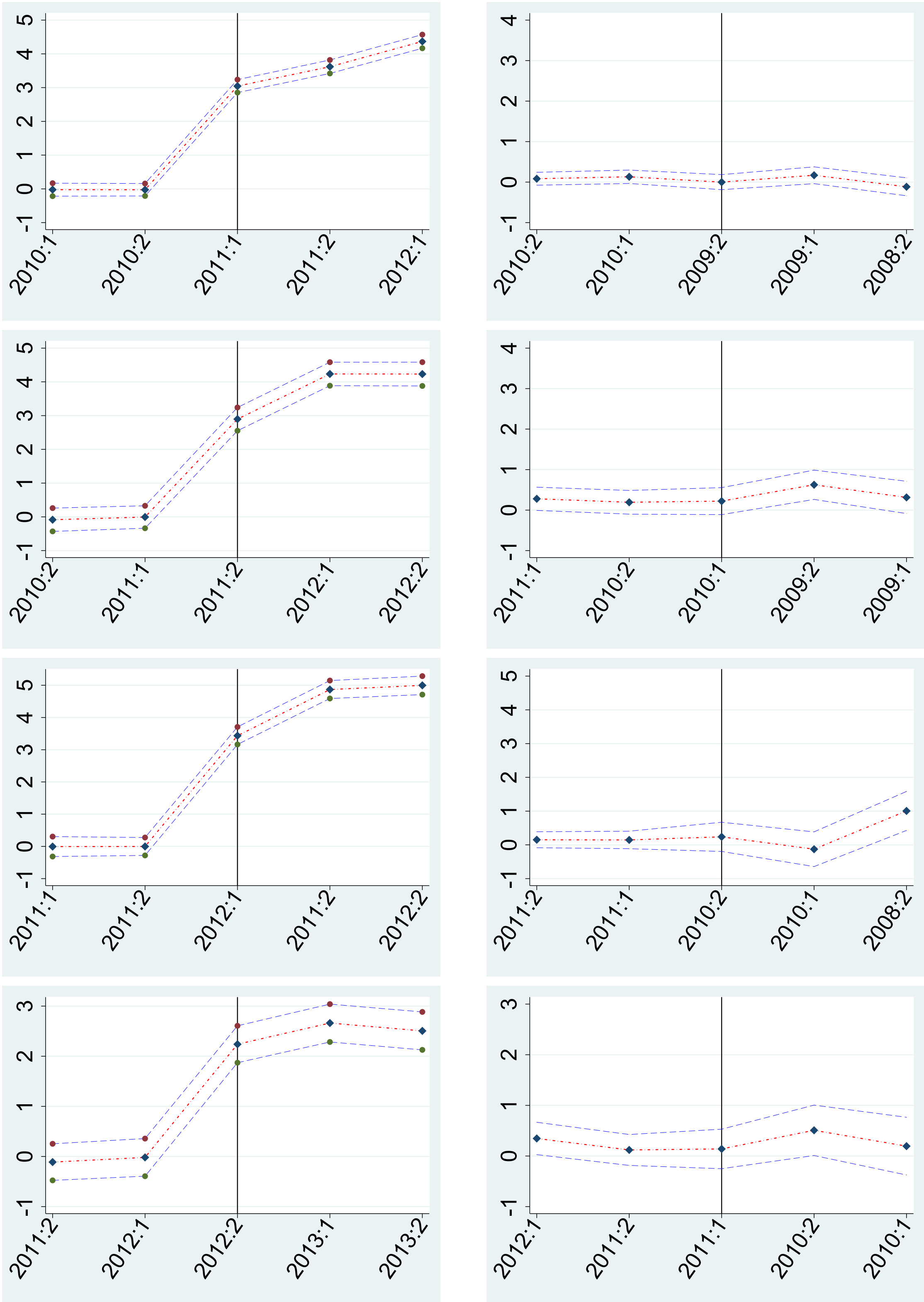
$$NetEntry_{d,t} = \psi \times AdoptionRate_{d,t-1} + \gamma_t + \omega_d + \epsilon_{d,t}$$

The outcome variable is the rate of net entry in district d, period t. The rate of net entry is defined as the number of newly entered firms minus those that exited (as a share of the total number of exiting firms). $AdoptionRate_{d,t-1}$ is the share of firms who adopted ESRMs in district d as of period $t-1$. γ_t and ω_n are time and district fixed effects. ψ is the coefficient of interest. A negative value of ψ

Figure 3: Trend differences of mean log sales during the matching, postmatching and prematching periods.

Panel (A): Treatment effect—trend during matching and post matching period

Panel (B): Placebo effect—trend during matching and prematching period



Notes: The figures show the differences in reported sales between treated and comparison groups for four matched samples (corresponding to each row). The vertical lines show the period of ESRM adoption for the treated group.

would suggest that districts with aggressive expansions of ESRMs are associated with lower rates of net entry.

Table 8: Net entry and the rate of ESRM adoption

	Dependent variable:			
	$NetEntry_{d,t}$		$NetEntry_{s,t}$	
	[1]	[2]	[3]	[4]
Right-hand-side variable:				
$AdoptionRate_{d,t-1}$	−0.09			
	(0.07)			
$AdoptionRate_{d,t-2}$		0.02		
		(0.08)		
$AdoptionRate_{s,t-1}$			−0.07	
			(0.07)	
$AdoptionRate_{s,t-2}$				−0.06
				(0.04)
Observations	2,491	2,429	2,259	2,235
Districts (or sectors)	245	242	197	197

Notes: Columns [1] and [2] show the correlations between the current rate of net entry into a district ($NetEntry_{d,t}$) and the lags of ESRM adoption rate in the district ($AdoptionRate_{d,t-1}$ and $AdoptionRate_{d,t-2}$). Columns [3] and [4] show show the correlation between the current rate of net entry into sectors s ($NetEntry_{s,t}$) and the lags of ESRM adoption rate in the sector ($AdoptionRate_{s,t-1}$ and $AdoptionRate_{s,t-2}$). Robust standard errors clustered by district (columns [1] and [2]) or sectors (columns [3] and [4]) are in parentheses.

The first column of Table 8 reports the estimate for ψ. There is no significant relationship between rate of ESRM adoption and net entry rate. In the second column, we include the second lag of ESRM adoption rate within the district ($AdoptionRate_{d,t-2}$) as the right-hand side variable. The result remains the same. In the third and fourth columns, we look at the correlation between the rate of net entry into a sector ($NetEntry_{s,t}$) and the lags of ESRM adoption rate by firms within that sector ($AdoptionRate_{s,t-1}$ in the third column and $AdoptionRate_{s,t-2}$ in the fourth column). These results also show no significant association between the rate of ESRM adoption and net entry across sectors.

6 Conclusion

Limited fiscal capacity of states has received increased attention as an important constraint to economic development. Having an effective tax system requires a vast administrative infrastructure capable of gathering, analyzing, and monitoring earnings information of a large number of taxpayers—a capacity that many developing countries tend to lack. Thus, the advent of electronic systems has attracted governments in many developing countries as a relatively cheap alternative for monitoring earnings information and improving fiscal capacity. In this study, using data from Ethiopia, we document the first empirical evidence on one such policy experiments.

We find that tax payments by firms increase in the aftermath of the ESRM adoption. We also find that the increased tax payments by registered taxpayers do not appear to have led to erosion of the tax base. By and large, these empirical patterns point to a possible positive contribution of the IT revolution to fiscal capacity in developing countries.

References

Acemoglu, Daron, "Politics and economics in weak and strong states," *Journal of Monetary Economics*, October 2005, *52* (7), 1199–1226.

Akcigit, Ufuk, Harun Alp, and Michael Peters, "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries," Working Paper 21905, National Bureau of Economic Research January 2016.

Baskaran, Thushyanthan and Arne Bigsten, "Fiscal Capacity and the Quality of Government in Sub-Saharan Africa," *World Development*, 2013, *45* (C), 92–107.

Becerril, Javier and Awudu Abdulai, "The Impact of Improved Maize Varieties on Poverty in Mexico: A Propensity Score-Matching Approach," *World Development*, 2010, *38* (7), 1024 – 1035.

Besley, Timothy and Torsten Persson, "State Capacity, Conflict, and Development," *Econometrica*, 2010, *78* (1), 1–34.

_ and _ , *Pillars of Prosperity: The Political Economics of Development Clusters*, Princeton University Press, 2011.

Best, Michael, Anne Brockmeyer, Henrik Kleven, and Johannes Spinnewijn, "Production vs Revenue Efficiency With Limited Tax Capacity: Theory and Evidence From Pakistan," *Journal of Political Economy*, 2015, (forthcoming).

Bird, R. M., "The Administrative Dimension of Tax Reform in Developing Countries," in Malcolm Gillis, ed., *Lessons from Tax Reform in Developing Countries*, Duke University Press Durham 1989, pp. 315–46.

Bird, Richard M. and Eric M. Zolt, "Technology and Taxation in Developing Countries: From Hand to Mouse," *National Tax Journal*, December 2008, *61* (4), 791–821.

Blundell, Richard and Monica Costa Dias, "Evaluation Methods for Non-Experimental Data," *Fiscal Studies*, 2000, *21* (4), 427–468.

Boadway, Robin and Motohiro Sato, "Optimal Tax Design and Enforcement with an Informal Sector," *American Economic Journal: Economic Policy*, 2009, *1* (1), 1–27.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, 2002, *117* (1), pp. 339–376.

Brynjolfsson, Erik and Lorin M. Hitt, "Beyond Computation: Information Technology, Organizational Transformation and Business Performance," *The Journal of Economic Perspectives*, 2000, *14* (4), pp. 23–48.

Caliendo, Marco and Sabine Kopeinig, "SOME PRACTICAL GUIDANCE FOR THE IMPLEMENTATION OF PROPENSITY SCORE MATCHING," *Journal of Economic Surveys*, 2008, *22* (1), 31–72.

Carrillo, Pau, Dina Pomeranz, and Monica Singha, "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," Technical Report 15-026, Harvard Business School Working Paper 2014.

Dehejia, Rajeev H. and Sadek Wahba, "Propensity Score-Matching Methods For Nonexperimental Causal Studies," *The Review of Economics and Statistics*, February 2002, *84* (1), 151–161.

Engel, Eduardo M. R. A., Alexander Galetovic, and Claudio E. Raddatz, "A Note on Enforcement Spending and VAT Revenues," *Review of Economics and Statistics*, 2001, *83* (2).

Fisman, Raymond and Shang-Jin Wei, "Tax Rates and Tax Evasion: Evidence from "Missing Imports" in China," *Journal of Political Economy*, 2004, *112* (2), pp. 471–496.

Garicano, Luis and Paul Heaton, "Information Technology, Organization, and Productivity in the Public Sector: Evidence from Police Departments," *Journal of Labor Economics*, 2010, *28* (1), pp. 167–201.

Girma, Sourafel and Holger Gorg, "Evaluating the foreign ownership wage premium using a difference-in-differences matching approach," *Journal of International Economics*, May 2007, *72* (1), 97–112.

Gordon, Roger and Wei Li, "Tax structures in developing countries: Many puzzles and a possible explanation," *Journal of Public Economics*, 2009, *93* (7-8), 855 – 866.

Hsieh, Chang-Tai and Benjamin A. Olken, "The Missing "Missing Middle"," *Journal of Economic Perspectives*, Summer 2014, *28* (3), 89–108.

Imbens, Guido W., "Matching Methods in Practice: Three Examples," *Journal of Human Resources*, 2015, *50* (2), 373–419.

Jenkins, Glenn P. and Chun-Yan Kuo, "A VAT Revenue Simulation Model for Tax Reform in Developing Countries," *World Development*, April 2000, *28* (4), 763–774.

Keen, Michael and Jack Mintz, "The optimal threshold for a value-added tax," *Journal of Public Economics*, March 2004, *88* (3-4), 559–576.

King, Gary, Christopher Lucas, and Richard A. Nielsen, "The Balance-Sample Size Frontier in Matching Methods for Causal Inference," *American Journal of Political Science*, 2017, *61* (2), 473–489.

Kleven, Henrik J. and Mazhar Waseem, "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *The Quarterly Journal of Economics*, 2013, *128* (2), 669–723.

Kumler, Todd, Eric Verhoogen, and Judith A. Frías, "Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico," Working Paper 19385, National Bureau of Economic Research August 2013.

Lechner, Michael, Ruth Miquel, and Conny Wunsch, "LONG-RUN EFFECTS OF PUBLIC SECTOR SPONSORED TRAINING IN WEST GERMANY," *Journal of the European Economic Association*, 2011, *9* (4), 742–784.

Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A. Olken, and Rohini Pande, "Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia," *American Economic Journal: Economic Policy*, August 2016, *8* (3), 258–83.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar, "Building State Capacity: Evidence from Biometric Smartcards in India," *American Economic Review*, October 2016, *106* (10), 2895–2929.

Naritomi, Joana, "Consumers as tax auditors," 2016.

Olken, Benjamin A. and Monica Singhal, "Informal Taxation," *American Economic Journal: Applied Economics*, October 2011, *3* (4), 1–28.

Pomeranz, Dina, "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, 2015, *105* (8), 2539–69.

Rosenbaum, Paul R. and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 1983, *70* (1), 41–55.

_ **and** _ , "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, 1985, *39* (1), 33–38.

Slemrod, Joel, "Does It Matter Who Writes the Check to the Government? The Economics of Tax Remittance," *National Tax Journal*, June 2008, *61* (2), 251–75.

_ **, Brett Collins, Jeffrey L. Hoopes, Daniel Reck, and Michael Sebastiani**, "Does credit-card information reporting improve small-business tax compliance?," *Journal of Public Economics*, 2017, *149*, 1 – 19.

Smith, Jeffrey A. and Petra E. Todd, "Does matching overcome LaLonde's critique of nonexperimental estimators?," *Journal of Econometrics*, 2005, *125* (1-2), 305 – 353. Experimental and non-experimental evaluation of economic policy and models.

Stiroh, Kevin J., "Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say?," *The American Economic Review*, 2002, *92* (5), pp. 1559–1576.

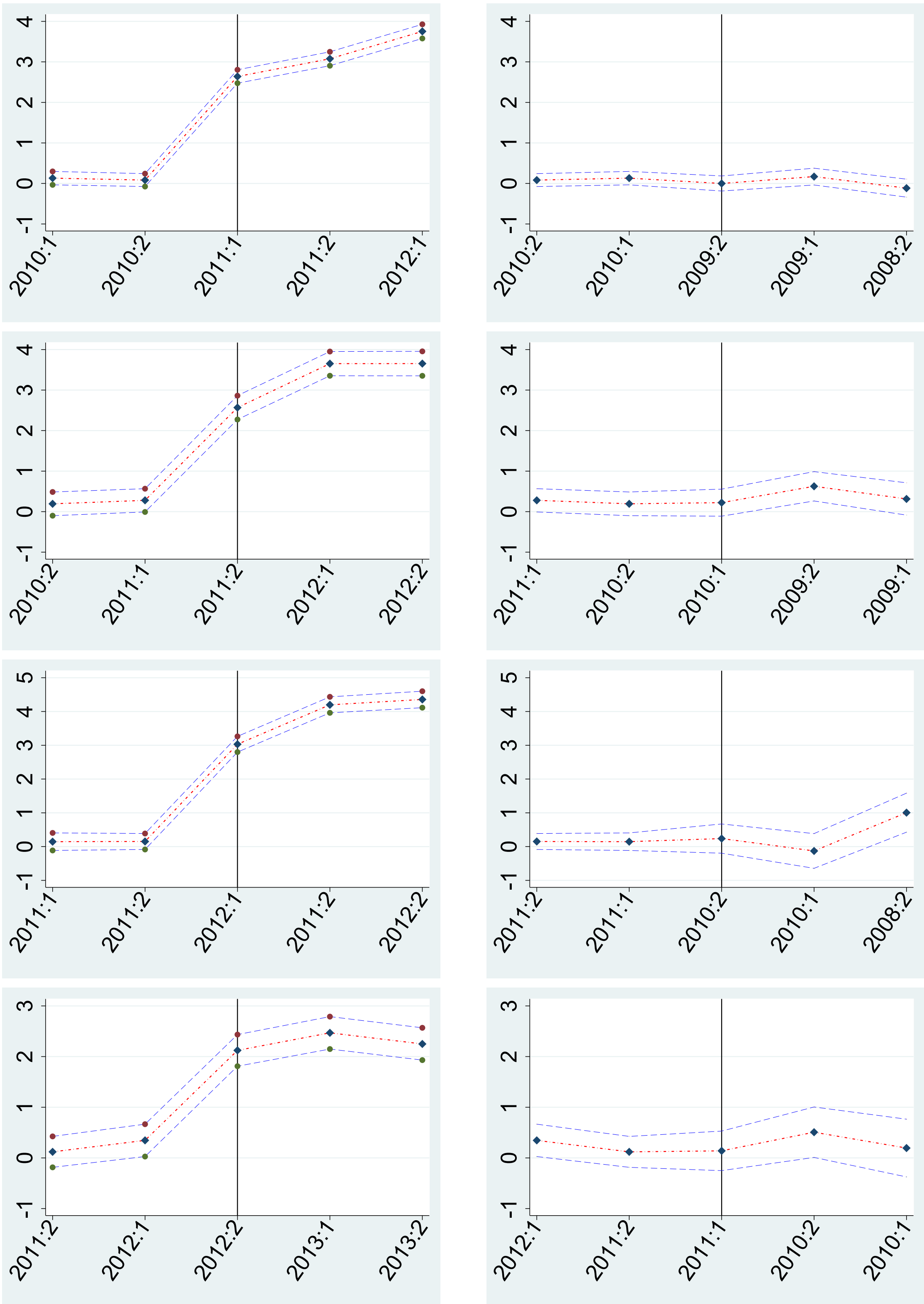
Tanzi, Vito and Howell H. Zee, "Tax Policy for Emerging Markets: Developing Countries," *National Tax Journal*, 2000, *53* (n. 2), 299–322.

Appendix

Figure .4: Trend differences in mean log VAT during the matching, postmatching and prematching periods.

Panel (A): Treatment effect—trend during matching and post matching period

Panel (B): Placebo effect—trend during matching and prematching period

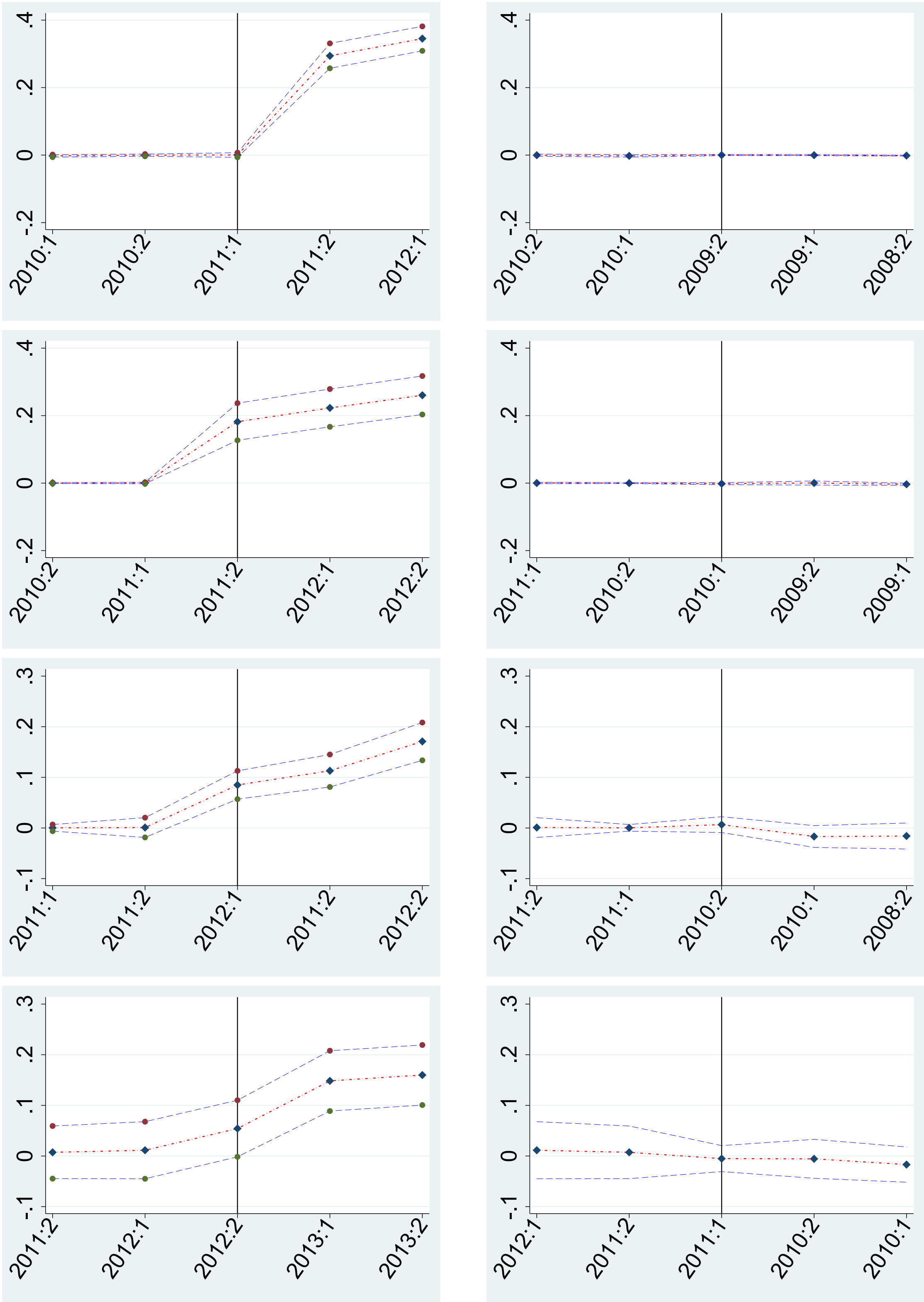


Notes: The figures show the differences in VAT between treated and comparison groups for four matched samples (corresponding to each row). The vertical lines show the period of ESRM adoption for the treated group.

Figure .5: Trend differences in log employment during the matching, postmatching and prematching periods.

Panel (A): Treatment effect—trend during matching and post matching period

Panel (B): Placebo effect—trend during matching and prematching period



Notes: The figures show the differences in employment between treated and comparison groups for four matched samples (corresponding to each row). The vertical lines show the period of ESRM adoption for the treated group.

Table A1: Balance statistics before and after matching—treated firms adopted ESRMs in 2011:1

Variable [1]	Un(matched) [2]	Treated [3]	Control [4]	MSB(%) [5]	t-value [6]
$\log(1+sales_{t-1})$	U	9.282	3.621	97.6	71.5
	M	9.217	9.243	−0.5	−0.3
$\log(1+sales_{t-2})$	U	8.636	4.092	75.3	55.3
	M	8.652	8.676	−0.4	−0.3
$sales_{t-1}>0$	U	0.733	0.289	99.1	72.3
	M	0.727	0.727	0.1	0.1
$sales_{t-2}>0$	U	0.683	0.327	76.3	55.8
	M	0.684	0.683	0.1	0.1
$\log(1+emp_{t-1})$	U	0.023	0.027	−1.1	−0.8
	M	0.005	0.005	−0.1	−0.1
$\log(1+emp_{t-2})$	U	0.008	0.025	−6.7	−4.4
	M	0.004	0.006	−0.8	−1.1
$emp_{t-1}>0$	U	0.013	0.009	4.2	3.2
	M	0.003	0.002	0.4	0.5
$emp_{t-2}>0$	U	0.004	0.008	−4.9	−3.4
	M	0.002	0.002	−0.3	−0.3
age	U	6.551	7.865	−28.5	−21
	M	6.489	6.523	−0.7	−0.5
age^2	U	65.003	82.177	−21.7	−16
	M	63.67	63.554	0.1	0.1
Large	U	0.926	0.571	89.8	60.4
	M	0.932	0.932	0	0
Small	U	0.063	0.068	−2.2	−1.6
	M	0.058	0.058	0	0
LimitedLiability	U	0.894	0.615	68.6	47
	M	0.895	0.896	−0.1	−0.1
Retail	U	0.492	0.182	69.5	53.2
	M	0.496	0.494	0.4	0.2
Service	U	0.138	0.063	25	19.4
	M	0.132	0.132	0	0
$SectorSales_{t-1}$	U	6.412	6.073	27.3	20
	M	6.408	6.389	1.5	1
$DestrictSales_{t-1}$	U	7.83	5.328	103.6	81.9
	M	7.731	7.656	3.1	1.6
Addis Ababa	U	0.986	0.543	122.6	78.9
	M	0.987	0.987	0	0

Notes: This table reports balance statistics before and after matching. Column [2] describes whether each row represents an unmatched (U) or matched (M) sample. Columns [3] and [4] report covariate means for treated and control groups, respectively. Mean Standard Biases are reported in Column [4]. The last column reports t-values for the mean differences between treated and control groups.

Table A2: Balance statistics before and after matching—treated firms adopted ESRMs in 2011:2

Variable [1]	Un(matched) [2]	Treated [3]	Control [4]	MSB(%) [5]	t-value [6]
$\log(1 + sales_{t-1})$	U	8.732	3.327	91.6	44.6
	M	8.694	8.699	−0.1	0
$\log(1 + sales_{t-2})$	U	7.164	3.475	61.6	30.5
	M	7.201	7.286	−1.4	−0.5
$sales_{t-1} > 0$	U	0.692	0.258	96.4	47
	M	0.689	0.689	0	0
$sales_{t-2} > 0$	U	0.583	0.28	64.2	31.8
	M	0.586	0.587	−0.3	−0.1
$\log(1 + emp_{t-1})$	U	0.027	0.03	−1.1	−0.5
	M	0.002	0.001	0.1	0.2
$\log(1 + emp_{t-2})$	U	0.021	0.029	−2.5	−1.1
	M	0.001	0.001	0	0
$emp_{t-1} > 0$	U	0.017	0.011	5.3	2.8
	M	0.002	0.002	0	0
$emp_{t-2} > 0$	U	0.01	0.01	−0.1	−0.1
	M	0	0	0	0
age	U	6.663	8.217	−32.2	−15.4
	M	6.594	6.534	1.2	0.5
age^2	U	67.985	90.633	−26.1	−12.3
	M	66.612	65.631	1.1	0.4
Large	U	0.918	0.575	85.9	34.8
	M	0.919	0.918	0.1	0.1
Small	U	0.053	0.066	−5.4	−2.5
	M	0.053	0.053	0	0
LimitedLiability	U	0.799	0.617	41	18.3
	M	0.801	0.811	−2.4	−1
Retail	U	0.418	0.12	71.2	40.4
	M	0.424	0.424	−0.2	−0.1
Service	U	0.174	0.048	40.8	24.7
	M	0.17	0.17	0	0
$SectorSales_{t-1}$	U	6.755	6.255	42.1	19.8
	M	6.747	6.713	2.8	1.1
$DestrictSales_{t-1}$	U	7.803	5.571	89.7	50.9
	M	7.734	7.652	3.3	1
Addis Ababa	U	0.86	0.531	76.6	32.7
	M	0.859	0.856	0.7	0.3

Notes: This table reports balance statistics before and after matching. Column [2] describes whether each row represents an unmatched (U) or matched (M) sample. Columns [3] and [4] report covariate means for treated and control groups, respectively. Mean Standard Biases are reported in Column [4]. The last column reports t-values for the mean differences between treated and control groups.

Table A3: Balance statistics before and after matching—treated firms adopted ESRMs in 2012:1

Variable [1]	Un(matched) [2]	Treated [3]	Control [4]	MSB(%) [5]	t-value [6]
$\log(1 + sales_{t-1})$	U	8.379	2.923	99.6	53.6
	M	8.238	8.241	0	0
$\log(1 + sales_{t-2})$	U	5.827	3.446	39.9	21.8
	M	5.801	5.806	−0.1	0
$sales_{t-1} > 0$	U	0.72	0.234	111.4	60.5
	M	0.712	0.712	0	0
$sales_{t-2} > 0$	U	0.482	0.275	43.8	24.3
	M	0.481	0.481	0	0
$\log(1 + emp_{t-1})$	U	0.149	0.106	7.1	3.7
	M	0.074	0.073	0.1	0.1
$\log(1 + emp_{t-2})$	U	0.023	0.034	−3.4	−1.6
	M	0.008	0.007	0.1	0.1
$emp_{t-1} > 0$	U	0.082	0.042	16.7	10
	M	0.045	0.045	0	0
$emp_{t-2} > 0$	U	0.013	0.013	−0.1	−0.1
	M	0.005	0.005	0	0
age	U	4.007	8.18	−84.2	−42.6
	M	3.76	3.686	1.5	0.7
age^2	U	37.059	95.059	−44.7	−22.5
	M	30.643	28.779	1.4	1.1
Large	U	0.807	0.583	50.1	24.8
	M	0.81	0.81	0	0
Small	U	0.042	0.065	−9.9	−4.9
	M	0.038	0.038	0	0
LimitedLiability	U	0.799	0.617	41	18.3
	M	0.801	0.811	−2.4	−1
Retail	U	0.418	0.12	71.2	40.4
	M	0.424	0.424	−0.2	−0.1
Service	U	0.174	0.048	40.8	24.7
	M	0.17	0.17	0	0
$SectorSales_{t-1}$	U	6.755	6.255	42.1	19.8
	M	6.747	6.713	2.8	1.1
$DestrictSales_{t-1}$	U	7.803	5.571	89.7	50.9
	M	7.734	7.652	3.3	1
Addis Ababa	U	0.86	0.531	76.6	32.7
	M	0.859	0.856	0.7	0.3

Notes: This table reports balance statistics before and after matching. Column [2] describes whether each row represents an unmatched (U) or matched (M) sample. Columns [3] and [4] report covariate means for treated and control groups, respectively. Mean Standard Biases are reported in Column [4]. The last column reports t-values for the mean differences between treated and control groups.

Table A4: Balance statistics before and after matching—treated firms adopted ESRMs in 2012:2

Variable [1]	Un(matched) [2]	Treated [3]	Control [4]	MSB(%) [5]	t-value [6]
$\log(1 + sales_{t-1})$	U	7.628	3.403	69.3	32.6
	M	7.615	7.633	−0.3	−0.1
$\log(1 + sales_{t-2})$	U	5.867	2.998	49.4	23.7
	M	5.871	5.983	−1.9	−0.6
$sales_{t-1} > 0$	U	0.603	0.264	72.8	34.3
	M	0.602	0.603	−0.2	−0.1
$sales_{t-2} > 0$	U	0.487	0.243	52.3	25.2
	M	0.487	0.495	−1.6	−0.5
$\log(1 + emp_{t-1})$	U	0.397	0.231	17.9	8.6
	M	0.388	0.376	1.2	0.4
$\log(1 + emp_{t-2})$	U	0.316	0.197	13.9	6.6
	M	0.31	0.303	0.8	0.3
$emp_{t-1} > 0$	U	0.19	0.091	29	15.1
	M	0.189	0.188	0.3	0.1
$emp_{t-2} > 0$	U	0.149	0.079	22.3	11.4
	M	0.149	0.148	0.3	0.1
age	U	5.057	8.037	−55.3	−23.6
	M	5.031	4.946	1.6	0.6
age^2	U	50.571	97.632	−35.5	−13.7
	M	49.668	47.566	1.6	0.8
Large	U	0.577	0.554	4.6	2.1
	M	0.577	0.571	1.2	0.4
Small	U	0.059	0.06	−0.5	−0.2
	M	0.059	0.058	0.2	0.1
LimitedLiability	U	0.788	0.603	41	17.3
	M	0.787	0.782	1.1	0.4
Retail	U	0.273	0.11	42.2	22.4
	M	0.273	0.268	1.3	0.4
Service	U	0.193	0.05	45	27.2
	M	0.192	0.192	0.1	0
$SectorSales_{t-1}$	U	7.111	6.143	47.9	21.1
	M	7.11	7.065	2.2	0.8
$DestrictSales_{t-1}$	U	6.763	5.565	50	22.1
	M	6.756	6.68	3.2	1.1
Addis Ababa	U	0.371	0.43	−12	−5.3
	M	0.37	0.361	1.8	0.6

Notes: This table reports balance statistics before and after matching. Column [2] describes whether each row represents an unmatched (U) or matched (M) sample. Columns [3] and [4] report covariate means for treated and control groups, respectively. Mean Standard Biases are reported in Column [4]. The last column reports t-values for the mean differences between treated and control groups.