The High 5

Building Fiscal Capacity:
The Role of ICT

Merima Ali, Abdulaziz B. Shifa, Abebe Shimeles and Firew Woldeyes

AFRICAN DEVELOPMENT BANK GROUP

Working Paper Nº 290

Abstract

Limited fiscal capacity poses a significant challenge in developing countries. To mitigate this challenge, the adoption of electronic tax systems has been at the forefront of tax reforms by many developing countries; however, there is little systematic empirical evidence on the impact of such reforms. We attempt to narrow this gap by documenting evidence from Ethiopia where there has been a recent surge in the use of electronic sales registry machines (ESRMs). Using administrative data covering all business taxpayers, we find that ESRM use resulted in significant increases in reported sales and tax payments. 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.

This paper is the product of the Vice-Presidency for Economic Governance and Knowledge Management. It is part of a larger effort by the African Development Bank to promote knowledge and learning, share ideas, provide open access to its research, and make a contribution to development policy. The papers featured in the Working Paper Series (WPS) are those considered to have a bearing on the mission of AfDB, its strategic objectives of Inclusive and Green Growth, and its High-5 priority areas—to Power Africa, Feed Africa, Industrialize Africa, Integrate Africa and Improve Living Conditions of Africans. The authors may be contacted at workingpaper@afdb.org.

Building Fiscal Capacity: The Role of ICT[1]

Merima Ali, Abdulaziz B. Shifa, Abebe Shimeles, Firew Woldeyes

[1]Merima Ali: CHR Michelsen Institute and Syracuse University. Email: Merima.Ali@cmi.no; Abdulaziz B. Shifa: Maxwell School, Syracuse University. Email:abshifa@maxwell.syr.edu. Abebe Shimeles: Africa Development Bank. Email: a.shimeles@afdb.org Firew Woldeyes: Ethiopian Development Research Institute, Email: w.firew@edri-eth.org.

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.[1] Weak fiscal capacity of states has thus received increased attention in the political economics of development.[2] 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 a relatively minimal cost has caught the attention of tax authorities throughout developing countries, and tax reform efforts to enhance monitoring earnings and improve tax collection 'in developing countries have generally centered on information technology' (Bird and Zolt, 2008). 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 world's lowest per capita incomes 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. The machines register sales and print out 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 earnings/transaction data and help minimize tax evasion, it is not obvious whether developing countries can effectively harness ESRMs to generate higher tax revenues. First, developing countries may face technological and administrative challenges in implementing ESRMs. Operation of the machines requires a fairly reliable provision of electricity and network infrastructure as well as availability relatively skilled workers that can administer the network. This could pose a challenge to developing countries due to lack of technical expertise and co-ordination 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/legal factors. For example, business owners may still evade taxes by paying more 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 extra earnings information may be minimal.

---

[1]For example, in 2006, the average GDP share of government revenue in low-income countries was 12.1%; however, for high-income OECD countries, the figure stood at 25.2% – twice the amount we observe in low-income countries. Source: World Development Indicators online data-base, accessed on June 28, 2014. The definition of high-income OECD and low-income countries follows the World Bank categorization.

[2]See, for example, Bird, 1980, Tanzi and Zee, 2000, Acemoglu (2005) Besley and Persson, 2010 and Besley and Persson (2011)

Second, even if ESRMs lead 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 operate at a smaller scale. This would be the case if, as a firm becomes larger, it tends to leave more traces of financial records and hiding the firm's revenues from the tax authority becomes more difficult, encouraging the firm to remain small. However, if the adoption of ESRMs leads to more accurate revenue records irrespective of the firm's scale of operation, lowering a firm's size will no longer be attractive as a tool to evade taxes. This may cause employment levels to increase following ESRM adoption. 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 both on tax payments by registered-taxpayers and proxies of the tax base.

We find three major patterns in the data. First, reported sales and tax payments by registered businesses increase substantially following ESRM adoption. Second, as a proxy for the effect of ESRM use on actual output (as opposed to reported sales), we look at the effect on employment by firms. We find that employment also increases following ESRM adoption, suggesting that firms appear to have increased their production in response to ESRM adoption. Third, we find that differences in rates of ESRM adoption across sectors or locations do not appear to affect rates of net entry into the formal sector. Hence, the fact that reported sales and tax payments increased without lowering neither employment nor 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 (Engel et al., 2001; Fisman and Wei, 2004; Olken and Singhal, 2011; Gordon and Li, 2009; Boadway and Sato, 2009). This motivated a number of recent studies that assess alternative policy tools to provide tax authorities with more reliable information. Generally, the studies examine the impact of third-party information to verify the accuracy of earnings reported by the taxpayer and minimize tax evasion (see, e.g., Carrillo et al., 2014; Pomeranz, 2015; Slemrod, 2008; Kumler et al., 2013; Naritomi, 2013). Even though governments in many developing countries are using the electronic tax system to enhance their ability to gather, analyze and monitor earnings information, we are not aware of any study examining the these policies – a gap that our study attempts to narrow.

Our 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 (Bresnahan et al., 2002; Stiroh, 2002; Brynjolfsson and Hitt, 2000). 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 exceptions are Lewis-Faupel et al. (2016) and Muralidharan et al. (2016), who study 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) document 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 will follow in Section 6.

2 Background, Data and Outline of Empirical Analysis

2.1 Background to ESRM adoption in Ethiopia

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 a third of the average in Sub-Saharan Africa and less than one-thirtieth of the OECD average.[3]

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. As a result, nearly ——% of government spending during the decade 2001-2011 came from non-tax sources such as international aid and loans. Ethiopia also relied heavily on taxes on international trade – a kind of tax that is relatively easy to enforce but that tends to be more distortionary to the economy. More than 40% of its tax revenue came taxes on international trade – a very high ratio even by the standards of developing countries. About a third of the revenues come from income taxes.

Against this background, the government undertook two major reforms that are the focus of this study. The first one is introduction of the VAT, which is one of the outcome variables in this analysis. VAT was introduced in 2003 with the aim of broadening the domestic tax base and minimize the dependence on trade taxes. VAT has now become a

---

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

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 is the adoption of ESRMs in 2008. By maintaining electronic record of business transactions, ESRMs are meant to minimize tax evasion by businesses.

The adoption of ESRMs and/or VAT registration is not mandatory for all firms. Whether a firm is required to register for VAT and/or adopt ESRMs is decided based on a broad set of criteria (such as size, location and sector) that ERCA outlines. Given the challenges in implementing these reforms all at once, the criteria have been revised gradually over several rounds, with each round expanding the set of firms that must register for VAT and/or adopt ESRM. Consequently, the number of both VAT-registered businesses and ESRM users has been increasing.

The solid line in Figure 1 plots the number of VAT-registered taxpayers during the years 2003-2013. The implementation of VAT started with about 6,000 firms in 2003 and gradually expanded, reaching about 97,000 firms by the end of 2014.[4] The implementation of ESRMs started with a few hundred firms in 2008 and gradually expanded. By 2014, about 50,000 taxpayers (out of over 97 thousands taxpayers) used ESRMs (see Fig. 1)[5].

---

[4]The legislation to implement the VAT imposed relatively stricter compliance requirements. For example, VAT registered business are required to use either ESRMs or paper receipts that are supplied by ERCA. This naturally implies a higher compliance cost both for firms to adhere to the requirements and for ERCA to enforce those requirements. As a result, the law excluded smaller firms whose turnover is not deemed to be large enough to justify the compliance cost to register for VAT. For a detailed theoretical discussion on the optimal VAT threshold, see Keen and Mintz (2004).

[5]As has been the case for the VAT registration, smaller firms have been excluded from ESRM use due to cost. According to our conversations with ERCA officials, the machines typically cost between 5,000 to 13,000 Ethiopian Birr (about 250 to 650 USD in current market exchange rate), a significant sum 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/standards set by ERCA.

Figure 1: Number of VAT payers and ESRM users ('000).
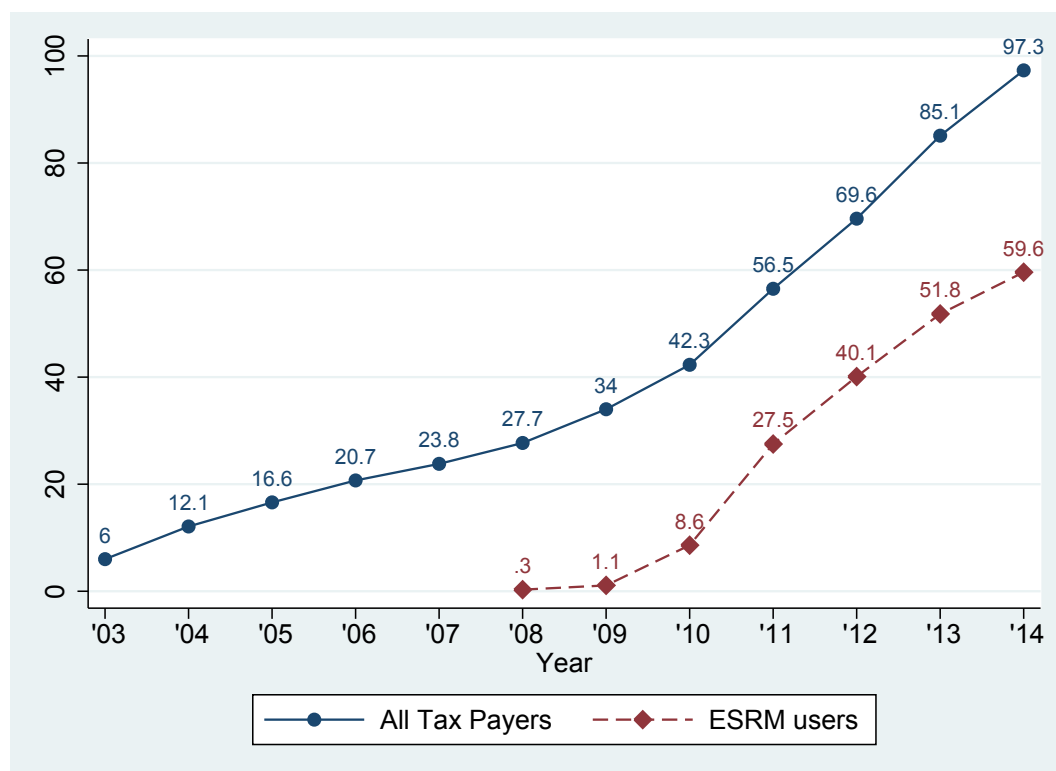
2.2 Descriptive Statistics

Our data-set contains administrative tax records on the entire set of VAT-registered firms in ERCA's database, covering the period from January 2003—the year of VAT introduction in Ethiopia—to the end of 2014. The administrative records provide information on several factors such as sales, employment, VAT payments, location, types of business activity (sector), ownership structure, age, date of ESRM adoption and category of the firm in the official tax classifications. The data series are available on monthly frequency. In our benchmark analysis, we aggregate the series into half-yearly frequency. 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 top panel in Table 1 presents moments of the three key outcome variables in our analysis: sales, value added tax and employment. There are 152,353 number of firms and about 1.1 million observations. The average half-yearly sales and VAT are about 3 million and 327 thousands Birr, respectively[6]. At about 11% of reported sales, the average VAT is well below the official VAT rate of 15% on final sales. This disparity could happen due to the various tax exemption that firms may get. The average number of workers per firm stands at 94.4. In nearly half of the cases, we observe zero values for sales and VAT. Similarly, only in 17% of the cases that we see positive values for employment, which is

---

[6]The numbers are not adjusted for inflation.

perhaps not surprising given that most of the firms are family-run small businesses.

The middle and lower panels present moments of the outcome variables for two sub-samples that differ by adoption status. In the middle (lower) panel, we have observations drawn from ESRM users (non-users). ESRM users tend to have higher means for all of reported sales, VAT payments and number of employees. ESRM users also tend to report positive values for all of the three outcomes more frequently than non-users.

2.3 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), and 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 may lead to erosion of the tax base. In order to asses 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 assess the effects on tax payments by registered taxpayers, we estimate the impact on VAT payments and reported sales. In order to examine whether ESRMs affect the tax base, we look at the impacts on employment and net entry.

The empirical analyses on sales, VAT and employment are undertaken at firm level. The analyses on net entry is carried out at sector and region 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 will then present further evidence using using matching methods to address endogeneity concerns (in Section 4). Finally, we will report results on the effect of ESRMs on firm entry/exit (in Section 5).

3 Impact of ESRMs on Reported Sales, VAT and Employment: Preliminary Evidence

3.1 Econometric framework

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

$$y_{j,t} = \beta \times ESRM_{j,t} + \mu_j + \psi_t + \varepsilon_{j,t} \tag{1}$$

Table 1: Descriptive statistics

Sample:	Sales	VAT	Employment	$\mathbb{1}_{Sales>0}$	$\mathbb{1}_{VAT>0}$	$\mathbb{1}_{Employment>0}$	Observations	Firms
Sample:								
All firms	2,983.5	326.8	94.4	0.50	0.48	0.17	1,128,648	152,353
	(644529.7)	(72402.2)	(153.88)	(0.50)	(0.49)	(0.38)		
Adopted ESRMs	4,856.4	464.0	153.9	0.70	0.70	0.36	409,358	74,054
	(717,484)	(17767.3)	(4802.3)	0.45	(0.46)	(0.48)		
Not adopted ESRMs	1,917.6	248.7	60.5	0.39	0.36	0.07	719,290	133,065
	(599,051.8)	89,698.9	3450.9	0.49	0.48	0.25		

The table presents means and standard deviations (in parentheses) of reported sales (*Sales*), VAT payments (VAT) and employment. *Sales* and VAT are in thousands Birr. The variables $\mathbb{1}_{Sales>0}$, $\mathbb{1}_{VAT>0}$ and $\mathbb{1}_{Employment>0}$ are indicators for whether the firm has non-zero values for *Sales*, VAT and employment. The statistics are reported for the whole sample as well as two sub-samples that differ by status of ESRM adoption.

$y_{j,t}$ 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. $\varepsilon_{j,t}$ is the error term. $ESRM_{j,t}$ is an indicator variable that equals one for the periods after ESRM use, and zero otherwise. Our coefficient of interest is β. It is meant to capture the change in the outcome variables following ESRM use.

3.2   Identification check

The important identification assumption in the above regression framework is that timing of ESRM adoption should not be correlated with differential trends in the outcome variables that would have occurred in the absence of ESRM adoption. Even though we will address the issue of endogeneity more robustly using matching analysis in Section 4, we now present some preliminary evidence suggesting that the timing of ESRM adoption does not appear to be correlated with differential trends in a way that may bias estimates from Eq. 1 above.

One can plausibly imagine three sets of possibilities that may lead to violation of this assumption. First, there may be a possible correlation between expansion of ESRM adoption and other macro variables that may affect firm revenue. Aggregate economic trends such as economic growth, government spending and inflation may affect both the timing of the government's action on ESRM use and the firms' revenue. For example, one may worry that the government may expand the adoption of ESRMs when its financing needs increase, say, due to increased government expenditure. However, increases in government expenditure may as well affect firm revenues through changes in aggregate demand, causing a spurious correlation between ESRM adoption and sales. The time-fixed effects included in the specification above address this kind of concern. The inclusion of time fixed effects is feasible thanks to the gradual implementation of the program through several rounds (as discussed in Section 2).

The second possible source of bias relates to potentially systematic and time-invariant differences between firms that use ESRMs and those that do not. For example, if ESRM users tend to engage in sectors where firms are larger, they may report a higher level of outcomes simply because of their size (as opposed to the effect of ESRM use). This potential problem is addressed in the above specification due to the inclusion of firm fixed effects. The identification in 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.

A third source of bias arises if selection into ESRM use 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 selects firms into ESRM adoption when firms acquire some productivity gains (such as product innovation) that are not observable to us in the data. In such a case, one cannot fully attribute estimates from the above specification to ESRMs since the outcome variables are likely to change even in the absence of ESRM adoption.

As a first step to examine this concern, we started out with interviewing ERCA officials and reviewing official documents to asses the context in which ESRMs were rolled out. Our preliminary reading of the context does not suggest that such a correlation between ESRM adoption and differential time trends is a significant factor to drive the empirical patters. First of all, adoption of ESRMs was mandatory where ERCA took the the

9

decision on which firms should adopt. Thus, the issue of self-selection by firms is less of a concern.

Second, the expansion of ESRMs was a significant logistical challenge for ERCA, and the decisions to role out ESRMs were primarily driven by logistical concerns rather than targeting firms with the highest potential for growth at the time of ESRM adoption. As a result, ERCA resorted to a set of ad hoc criteria. In the appendix (Table A1), we report a summary of directives that were issued by ERCA outlining (in a very general terms) the type of firms that should adopt ESRMs. Local tax offices select firms following the broad criteria outlined in those directives. Typically, factors like sector and location were used to determine which firms should adopt ESRMs. For example, in one of the rounds, firms operating on the main business streets of Addis Ababa (the capital city) were ordered to adopt. In another round, supermarkets, restaurants, jewelry stores and hotels operating in Addis Ababa were required to adopt. The use of these kind of very rough criteria is not particularly surprising in the context of a developing country where tax authorities have limited information on taxpayers. Effectively, ERCA resorted to easily observable criteria such as firm location and sector, which mostly reflect firm attributes that are unlikely to fluctuate over time and should be captured by the firm-fixed effects.

Figure 2: Mean of log sales by period of ESRM adoption.
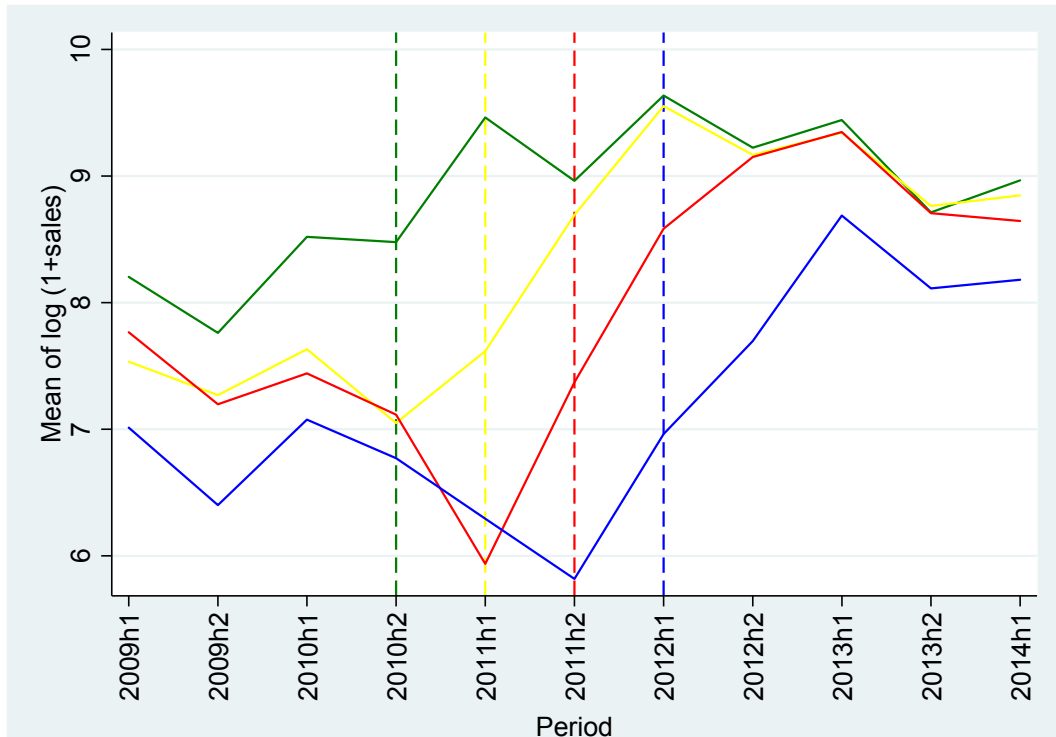
A visual inspection of the trends in sales seem to affirm that ERCA did not select ESRM adopters in response to systematic changes in trends. Figure 2 display the cross-sectional averages of log sales for four groups of firms around the time of ESRM adoption. The solid green line plots the trend for those firms that adopted ESRMs in 2010:2 (i.e. the second half of 2010). The yellow, red and blue lines plot the trends for firms that adopted

10

ESRMs in 2011:1, 2011:2 and 2012:1, respectively. Each dashed vertical broken line indicates the period of ESRM adoption for the group whose trend plot has the corresponding color. First, the trends appear fairly parallel prior to the first instance of ESRM adoption among the firms (2010:1). Second, this pattern of parallel trend appears to hold until a group adopts ESRMs and breaks-off from the trend. These patterns do not suggest that adoption of ESRMs is preceded by systematic changes in trends.

We have also tested whether past growth in sales is correlated with the probability of ESRM adoption. We found the correlation to be virtually zero (less than 10^{-6}) and statistically insignificant. We have done similar tests for VAT and employment and found very small and statistically insignificant correlations between lag of growth in VAT/employment and the likelihood of ESRM adoption.

3.3 Results

Table 2 reports estimates of Eq. 1. Distributions of sales, VAT and employment are not normal due to a large number of observations with zero values and some outliers on the right tail (see Table 1). 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, for each period, whether the firm reported positive values (columns (4) through (6)). Robust standard errors are clustered at firm level are in parenthesis.

Table 2: Results from fixed-effects panel regressions

| | Dependent variable | | | | | |
	Sales	VAT	Employment	$(Sales_{j,t} > 0)$	$(VAT_{j,t} > 0)$	$(Employment_{j,t} > 0)$
$ESRM_{j,t}$	2.27***	1.89***	0.37***	0.16***	0.15***	0.11***
	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	820731	820731	820710	820731	820731	820731
Firms	152353	152353	152353	152353	152353	152353

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The first column presents the estimated impact of ESRMs on reported sales. This effect is interesting because ESRMs are primarily meant to improve the accuracy of reported sales. The second column reports the effect on VAT. We see that both reported sales and VAT have 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, increasing taxes on final sales may lower demand and decrease production, eroding the tax base. On the other hand, one could also imagine plausible scenarios where ESRM adoption may incentivize firms to operate at larger scales, hence expanding output and the tax base. For example, consider a scenario where tax evasion is relatively easier for firms operating at smaller scales because are large scale

---

[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.

operations require more reliance on formal/contractual relations. Leaving more traces of legally financial records (as the firm's size increase) may in turn make it more difficult to hide revenues from the tax authority. As a result, firms may choose to remain small. Now imagine further that by providing more accurate revenue records irrespective of the firm's scale of operation, the adoption of ESRMs eliminates this asymmetry in the ability of large and small firms to hider revenues. Then, with the adoption of ESRMs, lowering a firm's size may no longer be an effective strategy to evade taxes, reducing the firm's incentive to lower production as a way of evading taxes. This effect suggests the adoption of ESRMs may increase output.

ESRMs, by helping improve business records, may also have a direct positive effect on the firm's output. This could happen, for example, 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).

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 where ESRMs increase reported sales, decrease production and the increase from reporting dominates the decrease in production. Actual output is not directly observed in our dataset. However, we have data on employment and we therefor 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 has 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.

Taken together, these empirical results suggest that – following ESRM adoption – reported sales and VAT payments increased without undermining the tax base, as indicated the increase in employment.

4   Impact on Reported Sales, VAT and Employment: Evidence Using Matching Diff-in-Diff Analysis

Even though the above identification checks do not indicate evidence of differential trends between ESRM users and non-users in the lead up to ESRM adoption, they do not necessarily guarantee that the comparison group constitutes a valid counterfactual for causal inference. Thus, 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 below, 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 has adopted ESRMs with a firm that has not so that the non-adopter can be used as a counterfactual for the adopter 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 firms in the control group are considerably different from the adopters.

Since the treatment and control groups are matched across several pre-adoption characteristics, matching methods involve using 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 them 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, Mahalonibs distance can be interpreted as Euclidean distance between normalized values of covariates (where the covariates are normalized using the covariance matrix).

Once observations from treatment and control groups are matched (based on some matching criteria employing either of the distance metrics), 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 rather estimate the difference in mean differences (instead of the difference in mean levels) by using the following diff-in-diff equation.

$$y_{j,t} = \beta \times Post_t \times Treated_j + \gamma \times Post_t + \alpha_{treated} \times Treated_j + \alpha_{control} + \varepsilon_{j,t} \qquad (2)$$

$Post_t$ is a dummy for the post-treatment period. $Treated$ is an indicator for whether the firm is treated. the treatment indicator. $\alpha_{control}$ and $\alpha_{control} + \alpha_{treated}$ are group-specific means for the control and treatment groups, respectively. The coefficient of interest is β—it captures the difference in trends (instead of levels) 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 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).

We report estimates of Eq. 2 for four matched samples, where the samples are categorized based on the period of adoption for the treated firms. In the first matched sample, the treated group includes firms that adopted ESRMs in 2011:1. The periods of adoption in the other three matched samples are 2011:2, 2012:1 and 2012:2. We focus on the periods during 2011 and 2012 because relatively large number of firms adopted ESRMs during those years (see Figure 1). While estimating Eq. 2, we consider a window of two periods (one year) before and after the period of adoption. For example, in the first matched sample, we include observations from 2010:1 through 2012:1. $\hat{\beta}$ for this sample captures the break in the adopters' trend in 2011:1 (the period of adoption for the treated group). The control group for this sample consist of observations drawn from the set of firms that either never adopted or adopted only after 2012:2. Similarly, the control groups in the other three matched samples consist of observations selected from firms that either never adopted or adopted only three periods after the period of adoption for the respective sample.

4.2 Matching variables

We use several variables to match the treatment and comparison groups—a total of 18 variables are included in the benchmark estimates. All of the variables are sourced from the administrative tax records. Besides providing a relatively accurate information on taxpayers, the use of administrative records is potentially advantageous in that they constitute ERCA's information set on taxpayers, hence they are likely to be relevant for selection. Remember that ESRM adoption was not voluntary—ERCA decides on which firms should adopt ESRMs. To the extent that firms were selected for ESRM adoption based on their characteristics observable to ERCA, the administrative records, which constitute ERCA's information set on firms, are likely to be relevant for selection. Thus, matching on variables from administrative records is beneficial in order to match the treatment and control groups with respect to firm characteristics that are important for selection.

The matching covariates consist of two broad set of pre-treatment variables. The first set of covariates include several time-invariant firm characteristics, 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 which one of the three statutory size categories that the firm belongs to (small, medium and large).

The second set of variables are intended to match the control and treatment groups with respect to time-varying characteristics, 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' pre-treatment 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

Within the matched sample, the distribution of matching variables should be similar between treatment and control groups. This is achieved by selectively dropping observations until the treatment and comparison groups are balanced (with respect to the matching variables) within the remaining (i.e. matched) sample.

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., 2016). First, differences in the outcome variables between treated and control groups should not influence 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 that 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, i.e. one should maximize size of the matched sample for any given level of imbalance or minimize imbalance for any given level of size (King et al., 2016). 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 algorithm 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 assess the sensitivity of results, we report estimates using alternative approaches to construct the matched samples. We also put particular attention on the variance-imbalance trade-off.

For the sake of transparency, 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 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 the MSBs are below 5% for each of the matching covariates.

Table 3: Mean standardized biases and bias reduction due to matching.

| | 2011:1 | | | 2011:2 | | | 2012:1 | | | 2012:2 | | |
	U	M	% Δ	U	M	% Δ	U	M	% Δ	U	M	% Δ
Panel A: MSBs for individual matching variables												
$Sales_{t-1}$	92.8	-0.2	99.8	86.4	-0.3	99.6	97.6	0.3	99.7	65.3	-0.1	99.8
$Sales_{t-2}$	71.5	-0.1	99.8	58.3	-0.8	98.7	39.5	0.1	99.7	47.4	-1.8	96.2
$Sales_{t-1} > 0$	94.9	0.0	100.0	91.5	0.0	100.0	109.0	0	99.7	68.4	-0.1	99.9
$Sales_{t-2} > 0$	73.2	0.0	100.0	61.1	0.0	100.0	43.5	0.0	100.0	49.9	-1.5	97.0
$Employment_{t-1}$	-0.9	0.0	100.0	-1.1	0.0	98.2	9.0	0.2	97.2	20.4	0.9	95.6
$Employment_{t-2}$	-7.5	0.0	100.0	-4.3	0.0	100.0	-4.3	-0.0	99.4	15.8	0.4	97.5
$Employment_{t-1} > 0$	2.5	0.0	100.0	3.3	0.0	100.0	15.0	0.0	100.0	27.0	0.3	98.8
$Employment_{t-2} > 0$	-5.9	0.0	100.0	-1.6	0.0	100.0	-2.3	0.0	100.0	23.7	0.4	98.5
Age	-29.3	-0.8	97.4	-31.8	0.2	99.2	-82.0	0.5	99.4	-52.9	0.8	98.5
Age^2	-22.8	-0.5	97.7	-26.4	0.4	98.7	-43.7	0.7	98.3	-32.9	1.4	95.7
$CategoryA$	94.0	0.0	100.0	85.8	0.0	100.0	50.8	0.0	100.0	5.1	0.9	83.0
$CategoryB$	-1.4	0.0	100.0	-3.8	0.0	100.0	-9.3	0.0	100.0	0.5	0.2	64.2
$LimitedLiability$	69.6	0.0	100.0	42.4	0.0	100.0	64.6	0.0	100.0	40.1	1.4	96.9
$Retail$	69.3	0.0	100.0	69.6	0.0	100.0	72.2	0.0	100	40.4	0.3	99.2
$Service$	25.0	0.0	100.0	40.1	0.0	100.0	47.4	0.0	100.0	42.5	0.1	99.7
$SectorSales_{t-1}$	29.5	1.0	96.8	44.5	2.7	94.0	97.1	2.1	97.8	46.4	2.3	95.1
$RegionSales_{t-1}$	129.8	0.0	100.0	86.8	0.3	99.7	75.7	3.6	95.2	48.3	3.1	93.5
$AddisAbaba$	129.8	0.0	100.0	78.9	0.0	100.0	18.7	0.0	100.0	-9.8	2.1	80.1
Panel B: Joint statistics												
Pseudo R-squared	0.514	0.000	—	0.356	0.001	—	0.333	0.001	—	0.156	0.001	—
Log-likelihood ratio (P-value)	0.00	1.00	—	0.00	1.00	—	0.00	1.0	100.0	0.00	1.00	—

Treatment group adopted ESRMs during:

In Panel A of Table 3, we report the MSBs for four groups of firms—categorized according to the period in which the treatment firms adopted ESRMs (see Section 4.1). The first three columns report balance statistics for the group whose treatment firms adopted ESRMs in the first half of 2011. For each group, we report MSBs both for the unmatched sample and matched sample (columns labeled "U" and "M", respectively). We also report the percent reduction in MSBs (columns labeled "%Δ").

We see that the matched samples have balanced the treatment and comparison groups quite successfully. In each of the four groups, the MSBs are relatively large for the unmatched sample. However, these 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 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 for each of the eight 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.

Table 4 reports the diff-in-diff estimates for the four matched samples (i.e. the coefficient β in Eq. 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 the matching estimates are generally larger than the estimates from the panel regressions, providing no indication that the positive effects of ESRM adoption estimated from the panel regression are driven by selection.

A visual display of the trend differences (along with the 95% confidence intervals) are presented in Figure 3. For each of the four matched sample (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, i.e. the treated groups adopted ESRMs right after 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 have significantly larger values for each of the outcome variables.

We have undertaken two sets of robustness checks. First, we assess whether the results are robust to reasonable changes in the matching implementation. Second, we examine whether the results are driven by external effects, which are particularly important in assessing the effect of ESRMs on overall outcomes (such as total tax revenue). We will

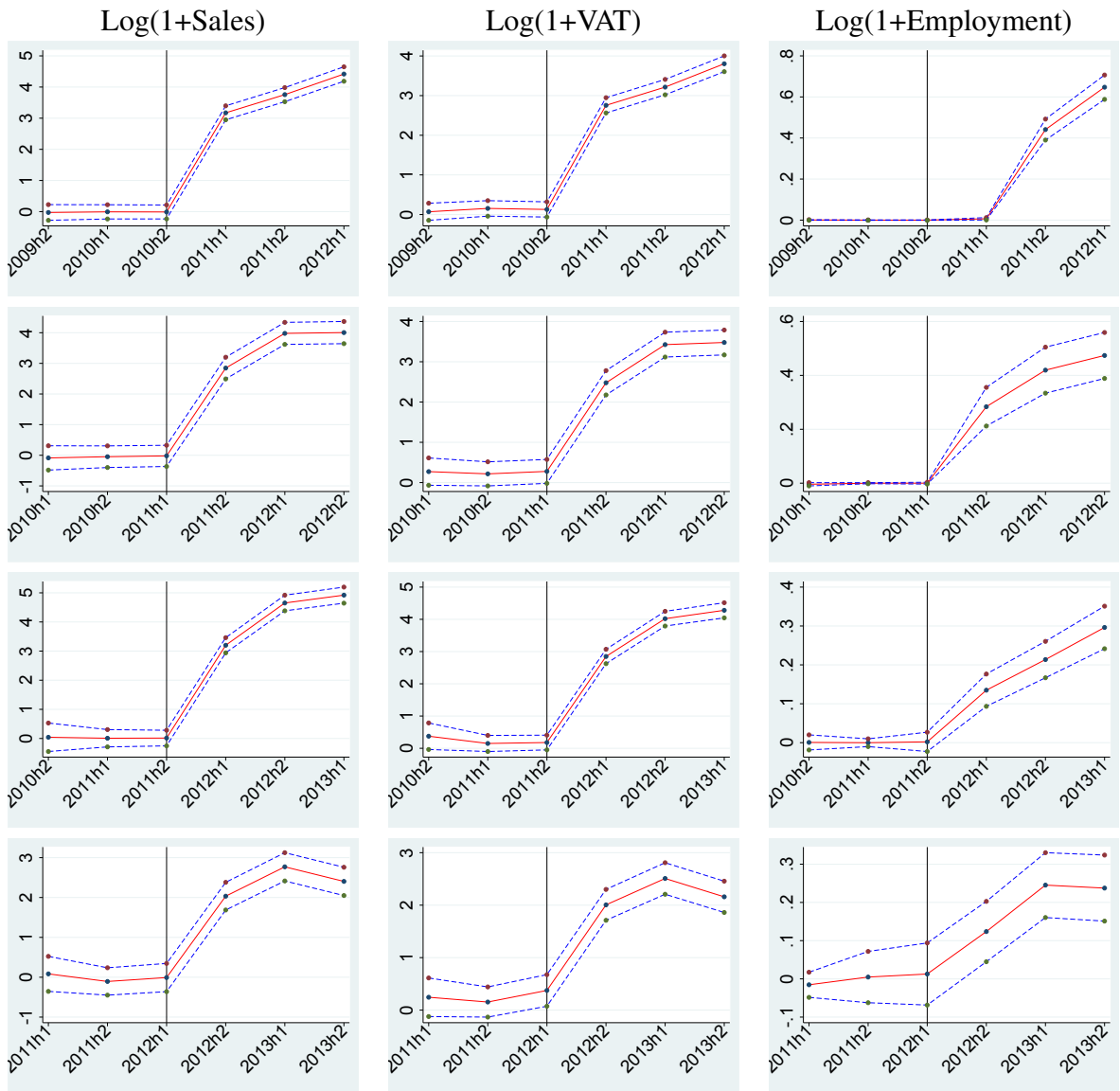Figure 3: Trend differences (with 96% CI) betwen treated and comparison units.

Table 4: Impact of ESRMs: Regression results from matched observations

| Dependent variable: | Treatment group adopted ESRMs during: | | | |
	2011:1	*2011:2*	*2012:1*	*2012:2*
Sales	3.79***	3.64***	4.25***	2.46***
	(0.21)	(0.22)	(0.21)	(0.16)
VAT	3.12***	2.88***	3.55***	1.96***
	(0.18)	(0.19)	(0.17)	(0.13)
Employmnet	0.36***	0.39***	0.21***	0.19***
	(0.05)	(0.06)	(0.02)	(0.03)
$\mathbb{1}_{Sales>0}$	0.30***	0.29***	0.35***	0.19***
	(0.02)	(0.02)	(0.02)	(0.01)
$\mathbb{1}_{VAT>0}$	0.30***	0.28***	0.35***	0.18***
	(0.02)	(0.02)	(0.02)	(0.01)
$\mathbb{1}_{Employment>0}$	0.11***	0.13***	0.09***	0.06***
	(0.01)	(0.02)	(0.01)	(0.01)
Observations	58680	24340	33590	25410
Firms	7890	3749	5073	4448

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

discuss each of them in the next sub-section.

4.4   Robustness Analysis

We have implemented several matching strategies to assess sensitivity of the results to the choice of matching algorithms. Instead of using Mahalonibis distance, we implemented nearest-neighbor matching using differences in propensity scores as the distance metric. The propensity score matching tends to perform poorly in the variance-imbalance trade-off in the sense that the remaining number of observations in the matched sample is generally smaller than those in Mahalnobis matching.[8]

We have also implemented caliper matching as proposed in Lechner et al. (2011). Caliper matching tends to deliver the best trade-off between balance and size. This is perhaps 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, 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. 153–4).'

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. In all of the alternative matching implementations, we found that the results remain the same—that reported sales, VAT

---

[8]The relatively poor performance of propensity score matching with regard to the balance-variance trade-ff is also noted by King et al. (2016).

payments and employment increase following ESRM adoption.

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, and 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. The coefficients will under-estimate the effect if ESRM adopters become less competitive and lose some of their market share.

In order to address this concern, we rerun the analysis where we exclude the comparison observations located in districts with relatively high level of adoption rates. The assumption is that external effects are likely to matter in areas with high adoption rates. Then, if the results are driven by external effects, they should not hold when the comparison group includes observations that are located in districts with low adoption rate (as they are unlikely to have been affected by the externality). 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.

One may be concerned that the external effects may 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 excluded untreated units from sectors where ESRM adoption rate is above that of the median sector.

In all of these robustness exercises, we found that the results remain the same.

5 Impact of ESRMs on Net Entry

As discussed above, the potential contribution of ESRMs to build fiscal capacity may be undermined if firms lower their output in response to increases in cost of production (due to extra tax liability), thereby leading to erosion of the tax base. However, the positive effect of ESRMs on employment suggests that this may not be the case. Perhaps an equally important channel whereby ESRMs may affect the tax base is through their effect on 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 operate informally where they can shun the adoption of ESRMs. Similarly, the threat of ESRM adoption may discourage potential new entrants to 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:

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

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 firms who newly entered minus of those who 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 n as of period $t-1$. γ_t and ω_n are time and district fixed-effects. γ is the coefficient of interest. A negative value of γ would suggest that districts with aggressive expansion of ESRMs are associated with lower rate of net entry.

Table 5: fised business (columns 1-3) and fised location (columns 3-6) share and Net entry

| | Dependent variable: | | | |
	$NetEntry_{d,t}$		$NetEntry_{s,t}$	
$AdoptionRate_{d,t-1}$	−0.08			
	(0.068)			
$AdoptionRate_{d,t-2}$		0.04		
		(0.08)		
$AdoptionRate_{s,t-1}$			−0.07	
			(0.06)	
$AdoptionRate_{s,t-2}$				−0.06
				(0.04)
Observations	2509	2447	2259	2235
Groups (districts or sectors)	245	242	197	197

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The first column of Table 5 reports the estimate for γ. We see that 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 column, 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 the sector ($AdoptionRate_{s,t-1}$ in the third column and $AdoptionRate_{s,t-2}$). We see that there is 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 earning information on 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, we document the first empirical evidence on one such policy experiments using data from Ethiopia.

We find that tax payments by firms increase in the aftermath of the ESRM use. 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. However, this

conclusion comes with an important caveat. We estimated the effect on firms that are already registered as taxpayers. This list of businesses does not cover a large number of micro and small enterprises in the country that operate informally and that are not registered as taxpayers. Thus, assessing effectiveness of such programs in the informal segment of developing countries warrants future research.

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.

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 1980, 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 Non-experimental 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.

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.

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

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 Nielsen, "The Balance-Sample Size Frontier in Matching Methods for Causal Inference," *American Journal of Political Science*, 2016, *In Press.*

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," unpublished 2013.

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.

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 A   Appendix

Table A1: Criteria for rolling out ESRM use – summary of directives issued by ERCA

Date when directive was issued	Types of firms required to start using ESRMs by the directive
–	VAT-registered hotels, super-markets, cafeterias, pastries and beverage groceries
July 7, 2009	All large tax payers except government-owned firms, banks, insurance companies, transport companies
October 11, 2009	Jewelry stores
December 4, 2009	Category A or B hotels, super-markets, cafeterias, pastries and beverage groceries, butcheries*
February 5, 2010	Category A or B furniture stores, building materials stores, auto spare part stores, electronics and computer stores, leather suppliers and electrical appliances stores
March 30, 2010	All VAT-registered businesses operating in Addis Ababa except insurance companies, banks, transport companies and government enterprises.
December 22, 2010	All VAT-registered businesses operating in Addis Ababa
April 21, 2011	All businesses operating in major malls in Addis Ababa
July 30, 2011	All category A and B businesses operating in Addis Ababa except transport companies

 Source: ERCA. This table presents a summary of directives issued by ERCA during July 2009 to July 2011. The second column presents the type of firms that are required to start using ESRMs by the directives. The table is a translation of the original document in Amharic (translated by the authors).
 * Firms are categorized into A or B based on their size. Category A firms are relatively larger than B.