The London School of Economics and Political Science

Essays on Firms in Developing Countries

Tsogsag Nyamdavaa

A thesis submitted to the Department of Economics of the London School of Economics and Political Science for the degree of Doctor of Philosophy. London, July 2021

To my family.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 40,000 words.

Disclaimer

Chapters one and two use administrative data provided by the General Department of Taxation of Mongolia. The views expressed in these chapters are solely those of the authors and do not necessarily reflect the views of the data providers.

Acknowledgements

This work wouldn't have been possible without the help of many great people. I would like to express my deep gratitude to my supervisors Rachel Ngai, Guy Michaels and Joana Naritomi for all their guidance and support throughout these years. I thank Guy for his unwavering high standards that sharpened my thinking and motivated me to work harder. I am very thankful for his frequent in-depth feedbacks that helped me to be more consistent and diligent toward my research. I thank Rachel for her never-ending support and encouragement from the onset of this journey. She is an exceptional role model and has helped me to become a better researcher and person from in different aspects. Our meeting in her office to discuss not only my research but also to talk about motherhood and family life has helped me stay on track and get things in perspective. I am indebted to Joana for always being supportive and willing to dedicate time to hearing my research ideas and even my doubts. It has always been intellectually stimulating and inspiring to interact with her. I thank her for reminding me to be confident and bold.

I am also grateful to Francesco Caselli, Ethan Ilzetzki, Xavier Jaravel, Per Krusell, Camille Landais, Daniel Reck, and Johannes Spinnewijn and many others at CFM for their advice and comments during meetings and seminars at LSE. I appreciate the help of the staff at LSE, especially Mark Wilbor and Coralie Simmons for their administrative support. I would like to thank the General Department of Taxation of Mongolia for their cooperation and for providing access to their administrative datasets.

I am very blessed to make so many great friends from this journey and it is definitely one of the highlights of my time at LSE: Andrea Alati, Miguel Bandeira, Fabio Bertolotti, Adrien Bussy, Thomas Drechsel, Andreas Ek, Martina Fazio, Dana Kassem, Sevim Kosem, Kieu-Trang Nguyen, Arthur Seibold. Torsten Figueiredo Walter deserves special thanks for many discussions that led to much improvement of this work.

I wouldn't be able to go far along the PhD process without the constant help of my family. My deepest gratitude is to my parents for their unconditional love and support. I thank them for teaching me the value of knowledge and providing me with endless opportunities to pursue my dreams from my childhood. They are the biggest role models in life and without them I wouldn't be where I am today. One of the precious gifts of my life has arrived during this challenging journey — my lovely son. All the worries and problems would go away at the sight of him. He is the most gentle human being and I thank him for his non-demanding nature that made parenthood life easier in the last few years of my

PhD. Last but by no means least, I thank my one and only, my beloved husband for always believing in me and encouraging me since our U-Tokyo days. I thank him for his patience and for comforting me during my difficult times. He is the biggest source of inspiration and I am so blessed to have him in my life.

Abstract

This thesis examines firms in developing countries. The first chapter provides evidence on the effects of consumer monitoring on tax evasion behaviour of firms along the supply chain. I study a Mongolian government program, which incentivises consumers to report their purchases. First, I estimate the effect of the program on corporate income tax (CIT) and value-added tax (VAT), by comparing retailers who are directly affected, and wholesalers, who are only indirectly affected. I find that retailers increase their reported sales, but partly offset this by artificially inflating their costs on CIT returns. As a result, retailers' CIT liabilities increase by 11%. In comparison, their VAT liabilities increase by 31% because VAT is less prone to such cost manipulation. Second, I find that the program also increases the VAT liabilities of upstream firms by about 15% when they are more likely to sell to (monitored) retailers, compared to the upstream firms that sell to firms that are not directly monitored. The program does not, however, affect the upstream firms' reported CIT liabilities. My findings highlight the enforcement advantage of VAT compared to CIT and that consumer monitoring enhances the self-enforcing mechanism in VAT along the supply chain. The second chapter characterises tax-evading firms using the same program in Mongolia. In particular, I study the firms that reported an abnormally large growth in their sales in the year that the program was launched, which suggests that those firms had previously been evading more taxes. I find that tax evasion was particularly prevalent among smaller firms, and conditional on firm size it was more common among older firms. My findings also suggest that tax evasion was more common in the capital city. The third chapter studies how resources should be allocated across firms. In particular, I theoretically investigate a trade-off between static and dynamic optimality conditions in terms of resource allocation across firms in the presence of learning-by-doing (LBD) effect. As in the standard misallocation literature, the static efficiency requires firms to have the same marginal revenue products (MRP) within each sector. In contrast, in the long run, I show that it is optimal to have some degree of dispersion in the MRP when there is an endogenous productivity growth through LBD mechanism. Then I compare

the implications from the dynamic and the static models quantitatively by using firm-level panel data from Indonesia. First, I show that firms' productivity process exhibits LBD mechanism. Namely, small and younger firms have lower productivity level but have higher productivity growth compared to larger and older firms. Second, I simulate both models and find that the dynamic model increases the aggregate TFP more in the long run than the static optimality conditions where I remove all dispersion in MRP.

Contents

List of Figures

Chapter : Resource Misallocation and Learning-by-Doing

List of Tables

Chapter 1

Does Consumer Monitoring Reduce Corporate Tax Evasion Along the Supply Chain? Evidence from Mongolia

This paper tracks the effects of consumer monitoring on firms' tax evasion along the supply chain. To do so, I study a Mongolian government program, which incentivises consumers to report their purchases. First, I estimate the effect of the program on corporate income tax (CIT) and value-added tax (VAT), by comparing retailers who are directly affected, and wholesalers, who are only indirectly affected. I find that retailers increase their reported sales, but partly offset this by artificially inflating their costs on CIT returns. As a result, retailers' CIT liabilities increase by 11%. In comparison, their VAT liabilities increase by 31% because VAT is less prone to such cost manipulation. Second, I find that the program also increases the VAT liabilities of upstream firms by about 15% when they are more likely to sell to (monitored) retailers, compared to the upstream firms that sell to firms that are not directly monitored. The program does not, however, affect the upstream firms' reported CIT liabilities. My findings highlight the enforcement advantage of VAT compared to CIT and that consumer monitoring enhances the self-enforcing mechanism in VAT along the supply chain.

1.1 Introduction

Power to tax lies at the heart of state development, and state capacity, in turn, is an important factor for economic development. However, it is well known that developing countries tax very little. Specifically, the tax-to-GDP ratio is positively correlated with countries' level of development (Besley and Persson, 2014). Therefore, it is crucial to study tax enforcement and explore the ways to strengthen it in developing countries. Firms play a crucial role in taxation in all modern tax systems. They remit the majority of tax revenues to the government, either with regard to their own tax liabilities or through the withholding of taxes of employees or other businesses (Kopczuk and Slemrod, 2006; Slemrod and Valayudhan, 2017).[1] Hence, much of the recent literature on tax enforcement and development has focused on firms.

A growing body of literature has documented that third-party information reporting in the form of consumer monitoring, whistle-blowers and paper trails could enhance tax enforcement because tax authorities can use it to verify firms' tax reporting (for example Naritomi, 2019; Pomeranz, 2015; Carrillo et al., 2017; Slemrod et al. 2017). In particular, it is well-known that VAT has a self-enforcing mechanism that creates a paper trail on transactions between firms, which makes it harder to hide the business-to-business (B2B) transactions from the authorities. The reason is that each B2B transaction is reported twice, once by the seller and once by the buyer, which enables the authorities to cross-check the information and detect any misreporting by the firms. However, this built-in enforcement mechanism breaks down at the end of the supply chain as final consumers do not typically report their purchases. This creates an opportunity for the firms at the end of the supply chain to under-report sales and potentially collude with upstream firms to evade tax.

This paper tracks the effects of using final consumers as "firms' sales auditors" on firms' tax reporting behaviour along the supply chain. To do so, I exploit an anti-tax evasion program, called E-receipt program, implemented by the Mongolian government in 2016, where consumers are incentivised to report their purchase.[2] Using rich administrative tax data on firms' tax returns as well as their trade network that span the period between 2014 and 2018, I study the changes in tax liabilities of firms along the supply chain. Specifically, I focus on firms' VAT and CIT liabilities because they are the main taxes that firms remit

[1] To be specific, firms remit 85% of total tax revenue in OECD countries and India (Slemrod and Valayudhan, 2017).

[2] Similar consumer monitoring programs were implemented in many other countries such as Brazil, Taiwan, Portugal and Slovakia. See Section 1.2.2 for details of the program.

in most countries.[3] More importantly, comparing these taxes in one setting highlights the importance of third-party information. For CIT, it is known that third-party information on firms' sales does not necessarily lead to more tax payments even though it increases firms' reported sales. This is because firms take advantage of the fact that costs are less verifiable for tax authorities and offset the effect on their CIT liabilities by reporting higher costs. For VAT, as discussed above9, firms' reported costs on VAT returns are constrained by the declarations of suppliers. Therefore, it is harder to manipulate costs on VAT returns.

I start by studying the direct effect of the program on the firms at the end of the supply chain — retailers — because their sales are directly monitored by final consumers. To identify the effects on retailers' CIT and VAT reporting behaviour I use variation in treatment intensity. I use a difference-in-difference (DiD) estimation method, where I take retailers as a treatment group and wholesalers as a control group.[4] I find that the program increases retailers' reported sales on CIT returns by 20%. However, retailers increase their reported costs by 23% in response to the program, leading to only 11% increase in their CIT obligations. Using tax audit data, I find that some of the increase in reported costs is due to increased misreporting. On the other hand, I find stronger effects on VAT reporting. Retailers' reported sales on VAT return increase by 42%. Even though they report higher input costs their VAT liabilities increase by 31%. There are two potential reasons why I find a larger increase in retailers' VAT liabilities than CIT liabilities. First, the composition of the firms used for CIT and VAT analysis is different. All firms submit CIT returns in Mongolia, but only large firms are liable for VAT. In other words, firms below the VAT threshold file only CIT returns, but larger firms submit both CIT and VAT returns. Therefore, to directly compare the CIT and VAT response, I restrict my sample to the large VAT-liable firms. I still find larger effect on retailers' VAT than CIT: their VAT and CIT liabilities increase by 25% and 17%, respectively. The second reason is the fact that it is relatively difficult for firms to over-report their costs on VAT returns because they could be cross-checked with reported sales of their trading partners. Consistent with this, the audit data do not produce any sign of increased cost over-reporting on VAT returns, unlike CIT.

It is important to note that any increase in reported costs of a VAT-liable retailer must

[3] They constitute more than 40% of the total tax revenue all around the world as shown in Figure A1. The share is higher for low and middle-income countries. In particular, CIT and VAT together made up 47% of total tax revenue in Mongolia in 2016. In comparison, payroll tax accounted for 15% of the total tax revenue in Mongolia.

[4] This estimation method is commonly used to study the effects of similar policies in the literature. For example, Naritomi, 2019 uses the same DiD estimation strategy to estimate the direct effect of the "NFP" program in Brazil, which employs consumers as the third-party reporters.

be associated with an increase in its upstream firms' sales because of the self-enforcing mechanism in VAT. This can happen if the retailer had been colluding with its upstream firms and hiding (some of) the B2B transactions before the intervention. In other words, any increase in reported input costs of retailers on VAT returns implies collusion along the supply chain. Clearly such collusion is beneficial for the upstream firms because it reduces their reported sales and tax liabilities. However it is not straightforward to see why retailers have an incentive to collude and underreport their costs, but there could be a number of reasons. For example, it would look suspicious to the tax authorities if a retailer declares purchasing costs of a good but does not report the sales. By hiding both purchases and sales of the good, the retailer can keep all the profits to themselves without paying any tax. Also the retailer would appear smaller on tax returns and hence could stay off the radar of the tax authorities.[5] Once final consumers start reporting their purchases to the tax authorities the retailers would want to increase their reported costs and thus stop colluding with upstream firms. This leads to an increase in reported sales of the upstream firms and potentially their tax liabilities.

Therefore, next, I explore the indirect effects of the program up the supply chain. I utilise my firm network data that cover periods before and after the intervention. I define upstream firms as firms that have ever supplied a retailer before the intervention, and rank them in terms of their share of pre-intervention sales to retailers. Then I use the DiD estimation approach to estimate the effects on upstream firms, where the treatment group is the firms, whose average pre-intervention share of sales to retailers is above the median, and the firms below the median are categorised as the control group. I find no significant effect on CIT liabilities of upstream firms with above-median sales to retailers compared to the below-median firms. In contrast, their VAT liabilities are estimated to increase by at least 15%. As a robustness check, I run transaction-level DiD within each upstream firm, where I compare its sales to retailers to its sales to non-retail firms. I find that upstream firms' sales to retailers increase by at least 22% in contrast to their sales to other firms. These results suggest that the E-receipt program does not only affect the firms at the end of the supply chains, but also has a positive indirect effect on their suppliers.

It is worth noting that both analyses of the direct effect on retailers and the indirect

[5] Another example is that the retailer could be offered a discount by the upstream firms, the sellers. If retailers do not report their purchases on their tax returns, the sellers would not have to pay tax on those sales and transfer some of the gains to the buyer. Therefore, it can be profitable for both the seller and the buyer to hide their transactions. Alternatively, the retailers could be involved in some underground activities, such as selling alcohol without a license, hence have an incentive to hide both sales and costs from tax authorities.

effect on upstream firms underestimate the true effects of the E-receipt program. The direct effect analysis uses wholesalers as a control group for retailers. The underlying assumption for this strategy is that wholesalers would have behaved similarly to retailers in the absence of the intervention (parallel trend assumption) and that wholesalers are not affected by the program. The data indicate a reasonable parallel trend in the sales of retailers and wholesalers before the intervention, which validates the parallel trend assumption. However, the wholesalers are likely to be affected by the program both directly and indirectly. Wholesalers are likely to be directly affected because they could sell to final consumers. Also, not surprisingly, wholesalers are classified as upstream firms, and I find substantial spillover effect on the upstream firms. Therefore, the estimated effects are a lower bound of the true direct effects on retailers. To investigate the extent of underestimation, I run another version of DiD regression, in which I use the wholesalers that never sell to any retailers as a control group. I identify such wholesalers using the firm network data. The results suggest a substantial underestimation for both CIT and VAT analysis. Similarly, the indirect analysis lead to an underestimation because the upstream firms in the control group sells to retailers to some extent and thus affected by the program. Acknowledging these limitations of the analysis indicates that the overall impact of the E-receipt program on tax revenue is even larger.

Lastly, I do a simple cost-benefit analysis of the program. To implement the E-receipt program the Mongolian government promises 20% of the VAT to the consumers as well as it holds monthly lottery events. Moreover, it bears some administrative costs such as expenses associated with installing IT systems and wage costs of the IT engineers. Considering these costs, I find that a 30.4% increase in VAT payments is needed for the program to break even.[6] As discussed above, I find that retailers' VAT liabilities increase by 31%, which is just enough to cover the costs of the program. Hence if one focuses only on the direct effects on retailers' VAT liabilities, as in the previous literature, then the program would appear not being able to increase the total tax revenue for the government. Once we account for the spillover effect on upstream firms' VAT liabilities and other tax

[6] My calculation does not include compliance costs for consumers and firms. The compliance cost for consumers is negligible because it is very easy to report their purchases to the government. See Section 1.2.2 for details of the program. Any compliance costs for firms should be reflected in their CIT liabilities and I find that retailers' CIT liabilities increase by 11%. In this sense, firms' compliance costs are accounted for in my analysis. Furthermore, there are other intangible aspects of the program. Firstly, the program could be changing certain societal norms that may have long-lasting effects even after the program ends. These changes include people getting used to asking for receipts, an increase in tax awareness, greater attention to the public expenditure, and more demand for efficient public spending and so on. On the other hand, the program increases firms' tax burden of the firms which could also increase the efficiency costs of the CIT and VAT. Moreover, I do not study any changes in tax incidence or transfer of the tax burden. Even though these are interesting and important aspects of tax enforcement they are beyond the scope of this project.

bases such as CIT, then it is clear that the program leads to larger tax revenue in Mongolia.

Related literature. This paper adopts a holistic approach to study the effects of a consumer monitoring program, and contributes to the literature on the role of third-party reporting in tax enforcement in two important ways. First, I show that it is crucial to include the final consumers into VAT reporting, as this ensures better enforcement along the whole supply chain. To date, the literature has studied the effects of consumer monitoring only on firms at the end of the supply chain (Naritomi, 2019). I extend this further and document that consumer monitoring does not only affect the downstream firms but also upstream firms along the VAT chain indirectly. This chain effect in VAT has been studied in the literature both theoretically (De Paula and Scheinkman, 2010) and empirically (Pomeranz, 2015). Specifically, Pomeranz, 2015, finds that increased tax enforcement can have spillover effects on the targeted firms' trading partners. However, its analysis focuses on the firms suspected of tax evasion, and the data collection process potentially entails some attrition and selection bias concerns. I, on the other hand, study the entire population of firms in the trade sector and their network using official administrative tax data. Second, I reconcile the different effects of third-party information on CIT and VAT in the literature by studying these taxes together. In particular, the literature has found firms' limited ability to adjust their reported costs for VAT (Naritomi, 2019), but close to full adjustment of costs for CIT (Carrillo et al. 2017; Slemrod et al. 2017). Also, for CIT, firms' reported costs are found to be much more elastic than their reported sales (Bachas and Soto, 2019). I reconcile these different findings by studying both CIT and VAT in one setting in the context of consumer monitoring. I find that the built-in enforcement mechanism in VAT is the driving force of larger effects on firms' VAT liabilities. On the other hand, for CIT, I discover that firms are substituting away from under-reporting sales to over-reporting costs when there is an improvement in sales enforcement. To the best of my knowledge, this paper is the first to offer direct evidence that firms respond to improved sales enforcement by increasing cost misreporting on CIT returns.

The remainder of this paper is structured as follows: Section 1.2 provides a background on the Mongolian tax system and explains the relevant datasets and their summary statistics. It also describes the policy intervention — the E-receipt program. Section 1.3 describes the empirical strategy and presents the results. Section 1.4 shows a simple cost-benefit analysis of the program, and Section 2.4 concludes.

1.2 Institutional Background and Policy Intervention

This paper utilises a nationwide anti-tax evasion program in Mongolia to study the effects of consumer monitoring on tax evasion behaviour of firms along the supply chain. In this section, first, I briefly describe the institutional background and tax system in Mongolia. Then I explain the anti-tax evasion program. Lastly, I discuss the datasets and provide summary statistics.

1.2.1 Mongolian Economy and Tax System

Mongolia is a lower-middle-income country and its GDP per capita (PPP) was around \$12,200 in 2018.[7] In this sense, the country's level of development is similar to Sri Lanka. However, Mongolia is often compared to Kyrgyzstan because both countries are land-locked, rich in mineral resources, both were under the influence of the Soviet Union and have a small population, even though Kyrgyzstan has a lower GDP than Mongolia.

Tax evasion is an indispensable part of the shadow economy, whose measures could indicate the extent of tax evasion in the economy.[8] The size of the shadow economy in Mongolia between 1999 and 2006 was estimated to be 18% of its GDP, while the average share of the informal economy for other 88 developing countries the same year was 35% (Schneider et al., 2010). Therefore, Mongolia is not considered to have a relatively high share of activities in its unofficial economy.

In this paper, I focus on two taxes — CIT and VAT, which are the main taxes firms remit in most countries. In particular, together they made up 47% of the total tax revenue in Mongolia in 2016. For CIT, there is no threshold for eligibility, and hence all firms submit CIT returns. Hence my CIT dataset contains information on the universe of formal firms. However, not all firms submit VAT returns. There is a VAT threshold in Mongolia, whereby firms with sales above the threshold have to register as VAT-liable firms.[9] On average, 30% of the firms in the CIT data are VAT-liable each year. Once a firm becomes a VAT-liable, it has to submit VAT returns on top of the CIT returns to the tax authorities. Having firms filing both CIT and VAT returns enables me to compare the

[7] Worldbank databank website, United Nations "World Economic Situation and Prospects 2018"

[8] One of the broadest definitions of the shadow economy is "those economic activities and the income derived from them that circumvent or otherwise avoid government regulation, taxation or observation" as defined in Schneider, 2012.

[9] If a firm is caught not having registered for VAT even though its annual sales are above the threshold there will be penalties. Also, it has to pay the owed VAT for the period it would have been a VAT-liable firm.

tax reporting behaviour of firms of these taxes.

Interestingly, there does not seem any systematic cross-checking between the two tax returns even though both could be submitted by the same firms. Data show a large discrepancy between, for example, the reported total sales on the two tax returns for VAT-liable firms.[10] This is possibly due to the fact that the submission frequency as well as the way firms report their sales, costs and tax liabilities on CIT and VAT returns differ. Specifically, CIT returns are submitted quarterly, and values such as sales and costs are reported in cumulative terms. That is, in quarter one firms report sales and costs applicable for only quarter one, but in quarter two firms report the sum of quarter one and two. In quarter four, firms report their annual revenue and costs. In contrast, VAT returns are submitted monthly and reported values, such as sales and costs, corresponding to the respective month. Therefore, firms could take advantage of the fact that it is not straightforward to compare CIT and VAT returns for tax authorities and respond differently to changes in tax enforcement.

The tax base for CIT is profit, which is the difference between revenue and total costs.[11] On the other hand, the VAT base is the value-added of firms, which is equal to total sales minus the cost of input purchases.[12] In practice, VAT-liable firms collect VAT on their sales (from the final consumers and other firms) and subtract the value of VAT that they pay on their intermediate purchase and transfer the difference to the government. To prove the collected VAT as well as the VAT payment on their purchases, firms submit VAT invoices, which contain information such as the tax IDs of the trading partners, both upstream (suppliers) and downstream (buyers) firms, and the relevant transaction values. Therefore, each B2B transaction ends up being reported twice, once by the seller and once again by the buyer. This is called the VAT credit-invoice scheme that enables the tax authorities to verify firms' self-reported values on VAT returns by cross-checking. My data suggest reasonable cross-checking between reported values within VAT reporting, unlike the comparison between CIT and VAT as mentioned before.

Moreover, both CIT and VAT rates are 10% and stayed the same throughout the period of my analysis.[13] In comparison, the world average rate for CIT and VAT in 2017 was

[10] More than 15% of the firms report different total sales on CIT and VAT returns, where the difference accounts for 10% of the sales declared on the CIT returns.

[11] There are some restrictions on the deductible costs. For example, firms are not allowed to deduct costs associated with paying fines, penalties, VAT and city tax payments.

[12] The difference between total costs for CIT and input purchasing costs for VAT is that total costs contain not only input purchasing costs but also wage costs, administrative costs and other costs.

[13] Actually there are two rates for CIT in Mongolia: 10% if the annual revenue is below 3 billion MNT ($\approx$ 1,150,000 USD), and 25% if the annual revenue is above 3 billion MNT. I assume the CIT rate is 10% for the sake of simplicity since most of the firms in the sample have an annual revenue below 3 billion

25% and 16%, respectively.[14]

Lastly, I use tax audit data from operational tax audits. Each year the tax authorities calculate firms' tax evasion risk score using their internal and external (third-party) information. Based on these scores, they choose which firms to audit. There are also non-routine tax audits at the requests of third parties such as courts, the police or other types of whistle-blowers. Subsequently, on average, 10% of the firms are audited each year. Generally, the tax audits examine the last five years of tax returns and other financial documents and check for any irregularities and inconsistencies.[15] If any tax incompliance is found, the firm is urged to pay the corresponding tax duties and fines. More importantly, any discovered misreporting of sales and costs is aggregated to the annual value for each type of tax return and recorded in the audit reports. Therefore, the tax audit data show if a firm was found misreporting on its CIT and/or VAT returns, and if so, how much is the under-report sales and/or over-reported costs is for each audited year. It is worth noting that audit data are at an annual level, unlike the CIT and VAT returns data.

1.2.2 E-receipt Program

The Mongolian government introduced an anti-tax evasion program, called E-receipt program, in January 2016.[16] The purpose of the program is to use final consumers as informants about firms' sales to disincentivise the firms from hiding their revenue. The program incentivises consumers to report their receipts of purchase in two ways:

- Consumers receive 20% of the VAT that they paid on their purchase if they register the receipt. The tax rebate is transferred to the consumers' bank account annually in January the following year.

- The registered receipt automatically turns into a lottery ticket regardless of the face value. The tax authorities hold a lottery event every month. The prize amount varies month to month and ranges from 20 million MNT (≈7700 USD), which is equal to the current VAT threshold in Mongolia of 500 million MNT (≈190000 USD).

MNT: around 1.5% of the firms have annual sales of more than 3 billion MNT each year.

[14] Sources: Tax foundation webpage — https://taxfoundation.org/publications/corporate-tax-rates-around-the-world/; and IMF Tax Policy Assessment Framework (TPAF) — https://www.imf.org/external/np/fad/tpaf/pages/vat.htm. Both accessed on 31 October 2020.

[15] The audit coverage period can be less than five years if the firm is established or was audited within the last five years.

[16] E-receipt program was put in place at the start of 2016 but the tax authorities already started publicizing it in late 2015. Even though it is possible, in principle, firms started reacting to the announcement by the end of 2015, as I show later that most of the effects began to appear after 2016.

By law, all firms have to participate and issue E-receipts whenever they sell to final consumers. Once an E-receipt is issued by a firm the transaction information is semi-automatically sent to the tax authorities and the sales value has to be accounted for on the firm's tax returns, whether the consumer submits it or not.[17] E-receipts have to satisfy some requirements: they have to contain a unique, system-generated 35-digit code, a 10-digit lottery code and a QR code in addition to sales details such as the item's face value, item details, prices, date, and the tax ID of the firm. Therefore, to be able to issue E-receipts, firms need to update or buy a new registry system and POS machines (receipt printing machines) that connect to the system via the Internet. Because of these fixed costs, there is a gradual enrollment of firms as shown in Figure A2.[18] If a firm is found not to issue E-receipts or if it refuses to issue E-receipts, consumers can report it to the tax authorities and the firm will be required to pay a penalty and could potentially face a tax audit.

The role of consumers is to make sure firms issue E-receipts and send them to the E-receipt system. It is easy for consumers to enrol in the program: they simply sign up to the E-receipt system via the website or the free mobile application by entering their details and bank account information.[19] Once the account is set up, the consumer can register receipts at any time using the E-receipt website or the mobile application as long as they have access to the Internet.[20]

Tax authorities hold a lottery event every month during which they choose the winners from that month's pooled E-receipts. An E-receipt has to be reported by both the seller firm and the consumer before the lottery event to be a valid receipt for the monthly lottery event.[21] The lottery event takes place in the middle of the month — around the 15th or 16th of each month — a few days after the VAT return submission deadline, which is the 10th of each month. This is to make sure that the consumers and firms submit their receipts before the VAT return submission date. If consumers submit their receipts after the lottery event, then the receipts will not be included in any future lottery event but

[17]E-receipts are automatically sent to the E-receipt system if the POS machine that issues the receipt is connected to the Internet. Firms can delay this information transmission for at most three working days.

[18]It is said that chain supermarkets or large retailers and wholesalers already had relatively modern registry systems even before the E-receipt program. Therefore, it is sufficient for them to simply update their system, which is cheaper. If it is not possible or suitable, large firms are in the financial position to invest in a registry system. On the other hand, for small firms it could be a considerable burden to buy a new registry system. To decrease the costs for small firms, it is made possible to print the system-generated receipts from the web-browser or to send them via email. Hence, for small firms with a small consumer base, it is sufficient to have a computer with Internet access and a printer.

[19]To be able to receive the VAT refunds and lottery prizes, consumers need to enter their full name, email address, phone number and government-issued ID number.

[20]Internet coverage is relatively advanced in Mongolia, and there are many places with free Wi-Fi, especially in the capital city, Ulaanbaatar.

[21]In the first three months the authorities held lottery events twice each month to attract more people.

they will still be eligible for the VAT rebate at the end of the year.

This lottery scheme is adopted to minimise the possibility of collusion between consumers and firms. However, there is a risk that firms may offer discounts to consumers to persuade them to collude and hide transactions from the authorities. For example, firms can collude with consumers by offering them a 10% discount, which is the VAT rate, if they do not ask for E-receipts. From the consumers' point of view, they need to choose between the firm's offer of a 10% discount now, and the government's offer of a 2% VAT rebate next year plus their luck in the lottery. If the consumers are myopic and/or do not believe that they have a high chance of winning the lottery then they might choose the firm's offer. This will attenuate the effects of the E-receipt program.

Lastly, it is noteworthy that the VAT threshold increased five-fold from 10 million MNT ($\approx$ 3800 USD) to 50 million MNT ($\approx$ 19000 USD) in January 2016, at the same time as the E-receipt program was initiated. This shift in the VAT threshold could, in general, affect the estimation of the effects of the E-receipt program, especially in the case of CIT reporting.[22] But, as I show below, the results survive qualitatively even if I restrict my samples to the firms who have always been VAT-liable suggesting that the VAT threshold shift does not drive the results.

1.2.3 Datasets

This project uses four (unbalanced) panel datasets, which are CIT and VAT returns data, VAT invoice data and operational tax audit data. All of them span the period between 2014 and 2018. Since the E-receipt program started in January 2016, the datasets cover two years before the intervention and three years after the program was initiated.

As discussed before, the CIT data contain information on the universe of formal firms. The main type of information I use from CIT returns is the firms' the reported total sales, total costs and CIT liabilities. Similarly, I use the information on the reported total sales, purchasing costs and VAT liabilities from VAT returns, but only for VAT-liable firms. In the main analysis, I focus on firms with strictly positive tax liabilities and summary statistics of CIT and VAT data are presented in Table 1.1.[23] The CIT sample contains

[22]It is known that firms bunch below the VAT threshold by misreporting their sales and/or restricting their production or activity. When the threshold increases firms could stop bunching and report larger revenue sums if misreporting had existed previously. If this is mainly true for retailers then it could result in over-estimation.

[23]The main results survive qualitatively if I use the entire sample that contains observations with non-positive tax liabilities as shown in Appendix A.15.
Also, as mentioned before, on CIT returns firms report their sales, costs and tax liabilities in cumulative

25,000 firms, of which 6,500 are retailers and 18,500 are wholesalers. For VAT data, there are 14,900 VAT-liable firms, of which 3,200 are retailers and 11,700 are wholesalers.[24] As expected, wholesalers are larger in size than retailers and have larger tax bills. Also, VAT-liable firms report larger sales and costs.

Table 1.1: Summary statistics - CIT & VAT data

(a) CIT returns				(b) VAT returns			
	mean	sd	count		mean	sd	count
Retailers				*Retailers*			
Sales	20.68	74.12	71,454	Sales	61.52	241.22	32,816
Expenses	19.24	70.47	71,454	Purchases	43.01	180.30	32,816
CIT liab	0.10	0.48	71,454	VAT liab	1.48	5.79	32,816
Wholesalers				*Wholesalers*			
Sales	64.05	137.86	130,684	Sales	121.26	318.76	101,136
Expenses	58.88	131.93	130,684	Purchases	69.84	220.36	101,136
CIT liab	0.40	0.94	130,684	VAT liab	3.96	8.66	101,136

Note: Table 1.1a presents descriptive statistics of the main variables from CIT returns. Sales, Expenses and CIT liab. are the quarterly gross reported sales, purchases and CIT liabilities of firms. Table 1.1b presents descriptive statistics of the main variables from VAT returns. Sales, Expenses and VAT liab. are the quarterly gross reported sales, purchases and VAT liabilities of VAT-liable firms. All nominal values are in thousand USD (1 MNT = 2600 USD).

Moreover, the VAT invoice data provide information on transactions between all VAT-liable seller-buyer pairs, which allows me to study the spillover effect of the E-receipt program on the upstream firms of retailers. I define the upstream firms as the firms that have ever sold to any retailer before the intervention. A total of 4,600 upstream firms are identified, most of which belong to trade (wholesale or retail), manufacturing and professional activities such as consulting sectors as shown in Table 1.2.

Table 1.3 presents descriptive statistics of transaction values from the VAT invoice data. In particular, I sum the upstream firms' sales to retailers within each quarter and summarise it in the part *Sales to retailers*, and similarly their total quarterly sales to non-trade sector firms such as hotels and schools as shown in *Sales to other firms*. On average, upstream firms sell twice as much (in terms of value of transaction) to non-retail firms as

terms. For example, in quarter two firms report the sum of quarter one and two, and in quarter four firms report their annual sales, costs and tax liability. Therefore, to calculate the quarterly revenue and costs, I subtract the previous quarter's value from the current quarter unless it is quarter one. Because I take the difference the quarterly sales and costs could result in a negative due to reporting or data quality issues. I drop such cases in the main analysis.

[24] It might seem unusual to have more wholesalers than retailers. My definition of retailer and wholesaler is based on firms' 4-digit industry classification code (ISIC Rev.2), which is reported on CIT returns. Even if some retailers are mistakenly classified as wholesalers, this will lead to an underestimation of the effect of the program.

Table 1.2: Industries of upstream firms

	Frequency	Percentage	Cum.Percentage
Administrative activities	56	1.22	1.22
Agriculture	46	1.00	2.23
Arts	9	0.20	2.42
Construction	171	3.74	6.16
Education	12	0.26	6.42
Electricity	64	1.40	7.82
Finance	35	0.76	8.58
Health	10	0.22	8.80
Hotel	91	1.99	10.79
IT	141	3.08	13.87
Manufacturing	386	8.43	22.30
Mining	31	0.68	22.98
Other services	35	0.76	23.74
Professional activities	311	6.79	30.54
Public administration	197	4.30	34.84
Real estate	32	0.70	35.54
Trade	2,855	62.36	97.90
Transportation	59	1.29	99.19
Water supply	37	0.81	100.00
Total	4,578	100.00	

the transaction values are more than twice as the value of the transaction to retailers.

Table 1.3: Summary statistics - VAT invoice data

	mean	sd	count
Sales to retailers			
Trans. Value	20.98	91.96	48,202
Sales to other firms			
Trans. Value	46.90	146.03	71,505

Note: Table 1.3 presents descriptive statistics of transaction values from VAT invoice data. *Sales to retailers* represents the average upstream firm's quarterly gross sales to retailers (summed over all retailers) and *Sales to other firms* shows its gross sales to non-retail and non-wholesale firms. All nominal values are in thousand USD (1 MNT = 2600 USD).

Lastly, I use tax audit data that come from operational tax audits. As mentioned before, tax audits usually cover the last five years of tax returns and other financial documents. Therefore, I use the firms audited in 2017 or after so that the audited period covers both pre- and post-intervention periods.[25] The audit data contain information on the year of audit, whether any misreporting was discovered on their CIT and/or VAT returns, and if so, the value of the under-reported sales and/or over-reported costs for each audited year.

[25] I drop firms that are audited before 2016 because the audited period will be between 2011 and 2015, which does not cover the post-intervention period.

CIT and VAT audit data are summarised in Table 1.4, which report (annual) values of under-reported sales and their share in firms' true sales for each type of tax return. The true sales are calculated as the sum of reported sales and hidden sales. They also provide summary statistics of (annual) values of over-reported costs, and their share in the true costs in each type of tax return. The true costs equal the difference between reported costs and the over-reported costs.

In particular, in CIT audit data, there are in total 4,000 firms audited between 2017 and 2018, of which 1,300 are retailers and 2,700 are wholesalers. The data are unbalanced, therefore, there are 960 retailers and 1,856 wholesalers in a year. Retailers under-report 2.33% (over-report 1.12%) of the total sales (costs). Wholesalers misreport 2.72% (1.76%) of the total sales (costs). For VAT audit data, there are fewer firms as expected: 1,060 VAT-liable retailers and 2,308 VAT-liable wholesalers in total. On average, 755 retailers and 1,566 wholesalers are audited in a year. VAT-liable firms are more likely to under-report their sales and less likely to over-report their costs on VAT returns compared to CIT data. In particular, VAT-liable retailers misreport 4.24% (0.63%) of the total sales (costs) and VAT-liable wholesalers under-report 4.28% (over-report 2.08%) of the total sales (costs).[26]

[26]Firm composition in CIT and VAT data is different because CIT audit data contain not only VAT-liable firms but also non-VAT-liable firms. Therefore, I compare CIT and VAT audit data for VAT-liable firms only and summary statistics are reported in Table A1. It shows that VAT-liable firms are more likely to under-report their sales and less likely to over-report their costs on VAT returns compared to CIT returns.

Table 1.4: Summary statistics - Audit data

(a) CIT returns

	mean	sd	count
Retailers			
Under.rep.sales ($)	0.72	6.13	4,819
share (%)	2.33	12.00	4,819
Over.rep.costs ($)	0.96	11.90	4,819
share (%)	1.12	5.09	4,819
Wholesalers			
Under.rep.sales ($)	3.73	30.34	9,281
share (%)	2.72	15.25	9,281
Over.rep.costs ($)	4.80	27.12	9,281
share (%)	1.76	6.96	9,281

(b) VAT returns

	mean	sd	count
Retailers			
Under.rep.sales ($)	1.81	15.60	3,775
share (%)	4.24	21.93	3,775
Over.rep.costs ($)	0.64	8.96	3,775
share (%)	0.63	4.78	3,775
Wholesalers			
Under.rep.sales ($)	4.32	29.53	7,831
share (%)	4.28	24.75	7,831
Over.rep.costs ($)	4.12	46.91	7,831
share (%)	2.08	9.30	7,831

Note: Table 1.4a and 1.4b present summary statistics of CIT and VAT audit data, respectively. Specifically, it summarises (annual) values of under-reported sales and their share in firms' true sales on each type of tax returns. The true sales are calculated as the sum of reported sales and hidden sales. Similarly, it provides summary statistics of (annual) values of over-reported costs, and their share in true costs in each type of tax return. The true costs are calculated as the difference between reported costs and the over-reported costs. All nominal values are in thousand USD (1 MNT = 2600 USD).

1.3 Empirical Analysis

This section empirically studies the effects of the E-receipt program on the tax evasion behaviour of firms along the supply chain. The purpose of the program is to use consumers as third-party reporters to reduce firms' sales misreporting. Therefore, firms at the end of supply chains — retailers — are directly affected by the program. I call the effects of the program on retailers as the "direct effect" and analyse it in Subsection 1.3.1. Next, I examine the spillover effects up the supply chain in Subsection 1.3.2, which I call the "indirect effect" of the program. In particular, I study the changes in tax liabilities of retailers' upstream firms.

1.3.1 Direct Effects — Retailers

To identify the direct effect of the program on retailers, I use the difference-in-difference (DiD) estimation approach, where I take retailers as a treatment group and wholesalers as a control group. Wholesalers are considered to be a reasonable control group for retailers because they both belong to the trade sector, and are likely to be affected by the same macro shocks.[27] However, one can think of a few caveats with this approach. First, my

[27] This identification strategy is commonly found in the literature. For example, Naritomi, 2019 adopts this strategy to study the effects of a similar consumer monitoring intervention in Brazil.

analysis is restricted to the trade sector only. More importantly, this estimation approach underestimates the true direct effect of the program due to two reasons. First, wholesalers could be directly treated by the E-receipt program if they sell to final consumers. Second, as I discuss later, there could be a spillover effect of the E-receipt program on the wholesalers via retailers. To investigate this further, I run another version of DiD regression, in which I use the wholesalers that never sell to any retailers as a control group. I identify such wholesalers using the firm network data from the VAT invoices. The results suggest substantial underestimation. Therefore, it is important to acknowledge these caveats of the identification strategy.

I am interested in estimating the effects of the program on retailers' CIT and VAT reporting behaviour. Below, I analyse them separately because of the following three reasons: First, the firms that submit the CIT & VAT returns are different. CIT data include all the firms whereas VAT data include only VAT-liable firms. Second, even though VAT-liable firms fill out both CIT and VAT returns, the reported values such as total sales and total costs do not necessarily match one-to-one between the two tax returns. This is because no systematic cross-checking is done by the authorities between the information on CIT and VAT returns.[28] Third, a more critical difference between CIT & VAT is the credit-invoice scheme inherent in VAT, which makes sure that VAT-liable trading partners monitor one another. As I explain later, this difference plays a vital role when interpreting the results. I start from the CIT data first because they cover all formal firms. Then I move on to VAT data and discuss the role of the credit-invoice design.

1.3.1.1 CIT

I start by showing that wholesalers are a valid control group, i.e., there is no pre-trend in reported sales before the intervention. I do this in two ways. First, I make a sector-level comparison between the retail and wholesale sector. Specifically, I aggregate sales of retailers each period and standardise it by dividing the sums by the pre-intervention mean value of the sums.[29] I do the same for wholesalers and plot them over time in panel (a) in Figure 1.1. As we can see from the plot, there is no pre-trend before the policy change, but total sales of retailers start to increase more compared to wholesalers in 2016. The gap between sales of retailers and wholesalers starts widening over time, and I attribute this divergence to the E-receipt program under the assumption that wholesalers are a valid

[28] It is said that if the tax officers manually cross-check the tax returns, then any unusually large gap would be noticeable. In that case, they would contact the firm and ask them to justify such disparities.

[29] The reason I divide the sums by the pre-intervention average sales is to make the visual comparison easier because wholesalers are larger in general.

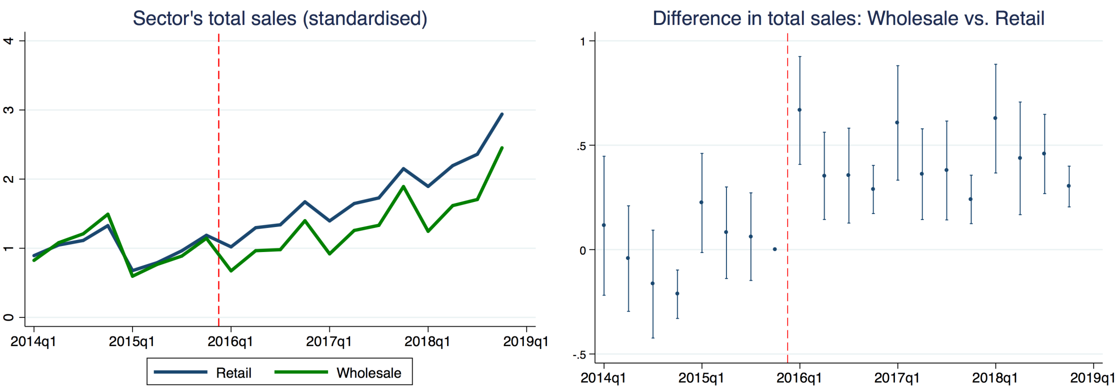
control group.[30]

Next, to establish the parallel trend I do a firm-level analysis, where I run the following flexible DiD regression:

$$ln(Y_{its}) = \gamma_i + \delta Quarter_t + \sum_{t=-8}^{11} \beta^t (Treat_{is} \cdot Quarter_t) + u_{its} \tag{1.1}$$

where subscripts i, t, s represent firm, quarter and 4-digit ISIC Rev.2 industry code respectively. $Treat_{is}$ equals one if firm i is a retailer, otherwise zero. The left had side variable $ln(Y_{its})$ is the log of quarterly revenue of the firms i in sector s in period t. In this regression I include firm fixed effect γ_i and quarter fixed effect $Quarter_t$. Therefore, my coefficients of interest are βs. I cluster the error terms by using 4-digit industry code. The estimated βs are plotted in panel (b) in Figure 1.1 which prove that there is no pre-trend. In particular, the confidence intervals always include zero before 2016, and the βs after the intervention are positive and significantly different from zero. This means that the wholesale sector is a valid control for retailers, and that E-receipt program significantly increased retailers' reported sales relative to wholesalers.

Figure 1.1: Pre-trend in CIT data



(a) Sectors' standardised total sales

(b) Coefficients from the flexible DiD regression

Note: Panel (a) displays the changes in the sales of retail and wholesale sectors reported on CIT returns. Each line is the total sales reported by all firms aggregated by retail or wholesale sectors scaled by the pre-intervention average quarterly sales each sector group. The graph plots the raw sales. Thus there are spikes in quarter four each year due to seasonality. The vertical dashed red line represents the start of the E-receipt program, which is January 1, 2016. Panel (b) plots the coefficients from firm-level regression, expressed in equation 1.1, using CIT data.

To see the effect of the E-receipt program on other variables such as CIT liabilities reported

[30]The graph shows spikes in quarter four each year because it plots raw aggregate sales. In Appendix A.5, I report a version of the graph where I plot aggregate sales of each sector after controlling for quarter-of-year FEs in Figure A4a. It corrects for the seasonality and still confirms the pre-trend assumption between retail and wholesale sectors.

on CIT returns, I run the following simple DiD regression:

$$ln(Y_{its}) = \gamma_i + \delta Post_t + \beta Treat_{is} \cdot Post_t + u_{its} \tag{1.2}$$

where subscripts i, t, s represent firm, quarter and 4-digit ISIC Rev.2 industry code respectively. $Treat_{is}$ equals one if firm i is a retailer, otherwise zero. Similarly, $Post_t$ equals one if the quarter falls after January 2016 and zero otherwise. The left-hand side variable $ln(Y_{its})$ is the variable of interest such as a log of quarterly revenue, costs or tax liabilities of the firms. Since I take the log of the dependent variable, the firms with zero tax liabilities (firms with zero or negative profits) drop out of the sample. Only the firms with strictly positive profits are included in the analysis. I include firm fixed effect γ_i in the regressions and cluster the error terms by using 4-digit industry code. β represents the average percentage increase in reported sales, costs and liabilities of retailers in the 3-year time period after the intervention compared to wholesalers.

Table 1.5: Direct effects - CIT returns

	(1) Sales	(2) Costs	(3) CIT
DD coef	0.198***	0.226***	0.114**
	(0.0697)	(0.0801)	(0.0536)
Firm FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry
Observations	202,138	202,138	202,138
Adjusted R^2	0.76	0.74	0.61

Note: This table displays the results from the regression equation 1.2. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. The dependent variables are a log of firms' reported quarterly sales, costs or tax liabilities on CIT returns. Only the firms with strictly positive profits are included in the analysis because I take a log of the dependent variable. That is the firms with zero tax liabilities (firms with zero or negative profits) drop out of the sample. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results are presented in Table 1.5.[31] All regressions are weighted by firms' average quarterly sales before the intervention.[32] The dependent variables are a log of firms' reported quarterly total sales, costs or tax liabilities on CIT returns. Column 1 shows that the E-receipt program induced retailers to report 20% higher sales relative to wholesalers. However, in column 2, retailers reported an increase in costs by 22.6%. This increase in costs partially offsets the effect on CIT liabilities, and thus CIT liabilities increase by

[31] Parallel trend in costs and CIT liabilities are shown in Figure A6 in Appendix A.6.

[32] Table A2 in Appendix A.7 shows results from unweighted regressions, which are consistent with Table 1.5.

11.4% in column 3.

As mentioned before, using wholesalers as a control group leads to an underestimation. This is because the wholesalers could be directly treated by the E-receipt program if they also sell to final consumers. Moreover, as I discuss later, there could be a spillover effect of the E-receipt program on the wholesalers via retailers. To investigate the extent of the underestimation, I run another version of DiD regression, in which I use the wholesalers that never sell to any retailers as a control group. I identify such wholesalers using the firm network data from the VAT invoices. The results are reported in Table A3. The estimated coefficients of sales and costs are above 60% suggesting a substantial underestimation.

These changes occur on retailers' CIT returns because consumers started monitoring them. To see if the program induced any real response, the number of workers and the total value of wages are analysed. The available payroll data cover the period between quarter one in 2015 to quarter three in 2018 only. Therefore, in Table 1.6, I report not only the changes in retailers' reported wages and workers, but also the regression results using the main variables (sales, costs, CIT) for this period. The first three columns confirm that the increase in costs offsets the increase in sales and thus leaving no significant increase in CIT liabilities. The last two columns show that retailers do not report a larger number of workers and wages after the intervention compared to wholesalers. This suggests that the increase in reported sales and costs is due to a reporting effect, and there is no actual increase in production.

Next, I focus on the increase in the reported costs. It has been documented in the literature that firms and individuals increase their reported costs on CIT returns in response to increased third-party information on firms' sales (Slemrod et al. 2017; Carrillo et al. 2017). Specifically, they tend to adjust costs that are more difficult to verify such as "other administrative costs". I study this in Table 1.7, where I decompose the increase in total costs into changes into its components: production, administrative and other costs. In particular, production costs contain material input costs, transportation, packaging and shipment costs, insurance costs and labour costs that are associated with production procedures. Administrative costs consist of marketing costs, travel expenses, labour costs of administrative staff etc. Other costs include non-operating costs such as interest payments, costs from currency exchange and other one-off or unusual costs. On average, production, administrative and other costs make 70%, 28% and 2% of the total costs, respectively. The last three columns in Table 1.7 show that the increase in total costs is mainly driven by an increase in production and administrative costs. The coefficient on

Table 1.6: No real response by retailers

	Main variables			Real response	
	(1)	(2)	(3)	(4)	(5)
	Sales	Costs	CIT	Wages	Workers
DD coef	0.139***	0.151***	0.0181	-0.00455	0.0157
	(0.0482)	(0.0518)	(0.0275)	(0.0266)	(0.0354)
Firm FE	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry
Observations	79,610	79,610	79,610	79,610	79,610
Adjusted R^2	0.83	0.82	0.72	0.92	0.91

Note: This table displays the results from the regression equation 1.2. The first three columns take a log of quarterly sales, costs or tax liabilities as dependent variables. Only the firms with strictly positive profits are included in the analysis because I take a log of the dependent variable. That is the firms with zero tax liabilities (firms with zero or negative profits) drop out of the sample. The payroll data covers Q1 in 2015 to Q3 in 2018 only. Therefore, less observation compared to Table 1.5. The dependent variables in the last two columns are log of total wages and number workers. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

other costs in column 4 is insignificant even though it is positive.[33]

Table 1.7: Decomposition of total costs reported on CIT returns

	(1)	(2)	(3)	(4)
	Total costs	Production	Admin	Other
DD coef	0.226***	0.233***	0.131***	0.213
	(0.0801)	(0.0698)	(0.0391)	(0.223)
Firm FE	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry
Observations	202,138	144,922	171,558	29,262
Adjusted R^2	0.74	0.70	0.78	0.40

Note: This table decomposes the total costs in column 1 into production, administrative and other costs, which are reported in columns 2-4. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

These results are slightly different from the findings in the existing literature mentioned above, namely, that so-called "hard to verify" other costs do not play a role in explaining total costs. Nevertheless, this does not rule out the possibility of firms artificially inflating

[33]Some of the firms do not classify the costs accurately and pool all their costs into one category such as production costs or administrative costs. This is the case for 30% of the sample. I do the same regression analysis by dropping those firms. The results are reported in Table A5 and they are consistent with the outcomes in Table 1.5.

their production and administrative costs to offset the effect of larger sales. In fact, an increase in the reported total costs could be either legitimate or illegitimate, or both. The increase is legitimate if the retailers start declaring the costs that are associated with the final sales that are disclosed by the consumers. This means the retailers used to hide both sales and costs associated with goods or services sold to the final consumers before the intervention. It is clear that retailers would have an incentive to hide their sales. However it is not straightforward to see why retailers have an incentive to hide and underreport their costs. But, if a retailer declares the purchasing costs of a good but does not report the sales, it would look suspicious to the tax authorities. Therefore, retailers are willing to underreport their costs as long as they can hide the corresponding final sales to consumers. Hiding both sales and costs associated with trading goods and services is beneficial to retailers. This is because they can keep the profits to themselves without paying any tax.[34] Moreover, there could be other reasons why retailers might want to suppress their reported costs. For example, they could be offered a discount by the upstream firms, the sellers. If retailers do not report their purchase on their tax returns, the sellers would not have to pay tax on those sales and transfer some of the gains to the buyer. Therefore, it can be profitable for both the seller and the buyer to hide their transactions. Alternatively, the retailers could be involved in some underground/illegal activities, selling alcohol without a license, hence hide both sales and costs from tax authorities. In all these cases, once the E-receipt program forces retailers to report their final sales, they would have an incentive to declare the previously hidden costs. Hence the increase in reported costs is legitimate.

On the other hand, the increase in reported costs is illegitimate if the retailers artificially inflate their costs to decrease the CIT liabilities. Since the E-receipt program makes it harder for firms to hide their sales they might want to substitute away from under-reporting their sales to over-reporting costs to keep their CIT liabilities small. This is feasible because the E-receipt program monitors only the sales of the firms, and not costs. Also, firms' reported costs on CIT returns are less verifiable for the tax authorities than sales.

Having a legitimate or illegitimate increase in reported costs has very different implications on the effectiveness of the E-receipt program to fight with tax evasion and increase tax revenue for the government. A genuine increase in reported costs result in the intervention (at least partially) successfully decreasing the size of the shadow economy even though

[34] Figure A3 in Appendix A.4 illustrates this using an example. A retailer buys a good from a wholesaler at a price 5 and sells it to a consumer at a price 8, generating a profit of 3. If the retailer declares both purchase and sales of the good, it has to pay at least the associated income tax. Therefore the gain of trading the good for the retailer is less than 3. If the retailer hides both its purchase and sales, then gain is 3.

the tax liability does not increase much. On the other hand, if firms increase their costs artificially, then the program is failing in its fight with tax evasion. It is easier for firms to misreport their sales and costs on CIT returns since there is no credit-invoice scheme as in VAT. Therefore, we cannot directly tell that the increase in retailers' reported costs is legitimate and should be associated with an increase in sales of upstream firms. Therefore, I investigate this further by using the audit data to shed light on the changes in retailers' misreporting behaviour.

To see the effect of the E-receipt program on the misreporting behaviour of retailers on their CIT returns I use firms' misreported sales and costs that are discovered during the operational audits. The audited data are summarised in Table 1.4.[35] Also, I calculate shares of misreported sales and costs in firms' true sales and costs. I define the true sales as the sum of misreported sales and reported sales, and true costs as the difference between reported costs and misreported costs. Here I implicitly assume that firms always want to under-report their sales and over-report their costs.[36] Then I run the simple DiD regression expressed in equation 1.2 and the results are presented in Table 1.8. The first (last) four columns in Table 1.8 analyse the misreporting of sales (costs). The dependent variables in columns 1 and 2 are the log of reported sales and calculated true sales. In column 3, I use the share of misreported sales as a dependent variable, which is the ratio between discovered hidden sales and true sales. Column 4 uses a log of the misreported sales. Similarly, in columns 5 and 6, I use a log of reported costs and calculated true costs as right-hand side variables. Columns 7 and 8 use the share and the (log of) value of misreported costs. I calculate the share of the misreported costs by dividing the value of misreported costs by true costs.

Columns 1 to 4 suggest that audited retailers misreport their sales less in the period after the intervention was initiated. As we can see from columns 1 and 2, retailers' reported sales increase more than true sales.[37] This is because retailers' tendency to under-report sales decrease after the intervention. Column 3 shows that retailers' share of misreported sales decreases by 0.7%. Column 4 shows that the value of hidden sales of retailers decreases by 1.5%, even though the coefficient is not significantly estimated.[38] On the other hand,

[35] As mentioned before, the audited firms are not chosen randomly. However, as long as the criteria to choose firms have not changed during the sample period the DiD estimation approach should estimate the effect of the program on audited firms. In contrast, the external validity of the estimated effect of the program is still questionable if the audited firms are systematically different from the general population of the firms.

[36] However, there are some evidence that firms' misreporting behaviour is not always optimal. For example, Almunia et al., 2019 shows that 29% of firms misreport own sales and purchases such that their tax liabilities increase.

[37] However, the estimated coefficients are not statistically different from one another.

[38] This lack of statistical significance is potentially due to the smaller sample size used in Column 4.

Table 1.8: Misreporting on CIT returns (annual values)

	Sales				Costs			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reported	True	Misreport (%)	Misreport ($)	Reported	True	Misreport (%)	Misreport ($)
DD coeff	0.193***	0.181***	-0.705*	-0.0151	0.199***	0.183***	0.376	0.362**
	(0.0385)	(0.0399)	(0.377)	(0.249)	(0.0397)	(0.0415)	(0.268)	(0.137)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Observations	14,100	14,100	14,100	2,246	14,100	14,100	14,100	3,091
Adjusted R^2	0.74	0.74	0.22	0.68	0.79	0.78	0.23	0.74

Note: This table displays the results from the regression equation 1.2 using CIT audit data. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. The first (last) four columns analyses misreporting of sales (costs). The dependent variables in columns 1 and 2 are the log of reported sales and calculated true sales. I calculate true sales by adding misreported sales to reported annual sales. In column 3, I use the share of misreported sales as a dependent variable, which is the ratio between discovered hidden sales and true sales. Column 4 uses a log of the misreported sales. Similarly, in column 5 and 6, I use a log of reported costs and calculated true costs as right-hand side variables. True costs are calculated by subtracting misreported costs from reported annual costs. Column 7 and 8 use the share and the (log of) value of misreported costs. I calculate the share of the misreported costs by dividing the value of misreported costs by true costs. All regressions are weighted by firms' average annual sales reported on CIT returns before the intervention. Time and firm fixed effects are included in all regressions. Standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

columns 5 to 8 imply that retailers misreported their costs more after the intervention. In particular, reported costs increase by 20% and calculated true costs increase by 18%.[39] Column 7 indicates that the share of over-reported costs increases by 0.4% even though the estimated coefficient is not statistically significant. The value of misreported costs increases by 36% as shown in column 8.[40]

These results support the hypothesis that retailers are substituting away from under-reporting sales to over-reporting costs on their CIT returns. In particular, audit data shows that retailers artificially increase their reported costs to decrease their tax liabilities since it is harder for them to hide sales due to the intervention. The estimated coefficients in columns 5 and 6 imply that at least 2% of the increase in reported costs is due to illegitimate cost over-reporting.[41] To the best of my knowledge, this is the first time

Sample size decreases in Column 4 because the dependent variable is a log of the misreported sales, and not all audited firms got caught misreporting their sales during the tax audits.

[39] However, the estimated coefficients are not statistically different from one another.

[40] Note that the number of observations used in columns 4 and 8 is smaller than the other columns. This is because some firms do not misreport sales and/or costs in some years and are dropped out because I take a log of the dependent variable. I use several other measures of misreported values sales and costs in Table A6. In particular, I use a log of one plus the misreported values and a dummy for positive misreported sales and costs. The results qualitatively confirm the fact that retailers misreport their sales less but are more likely to over-report their costs after the intervention.

[41] It should be noted that this 2% is potentially an underestimate of the actual increase in retailers' artificial cost over-reporting. It is possible that tax audits do not reveal all misreporting. For example, tax auditors cannot identify cost misreporting due to their lack of knowledge about the business. Or tax auditors could be influenced by corruptions and collude with the audited retailers and hide their misreporting. In these cases, the calculated true costs could still be an overestimate of firms true costs.

direct evidence is provided showing that firms are over-reporting their costs more on CIT returns in response to increased enforcement on firms' sales.

1.3.1.2 VAT

In this subsection, I analyse the VAT data employing the same research strategy used in the previous subsection 1.3.1.1. First, I establish that VAT-liable wholesalers are a valid control group for VAT-liable retailers. Panel (a) in Figure 1.2 compares the total sales of VAT-liable retailers to VAT-liable wholesalers.[42]. The graph shows that the sales of VAT-liable retailers and wholesalers move roughly parallel to each other before the policy change, validating the no pre-trend assumption. Total sales of VAT-liable retailers start to increase more than VAT-liable wholesalers around the start of the E-receipt program.[43] The gap between the sales of retailers and wholesalers widens over time.

Panel (b) in Figure 1.2 plots the estimated βs from equation 1.1, where the dependent variable is the log of the quarterly reported sales on firms' VAT returns. The graph shows that the estimated coefficients before 2016 are not significantly different from zero, implying that the VAT-wholesalers are a valid control for VAT-retailers. The estimated βs increase and are significantly different from zero after the introduction of the E-receipt program, implying the program's differential effect on retailers' reported sales.

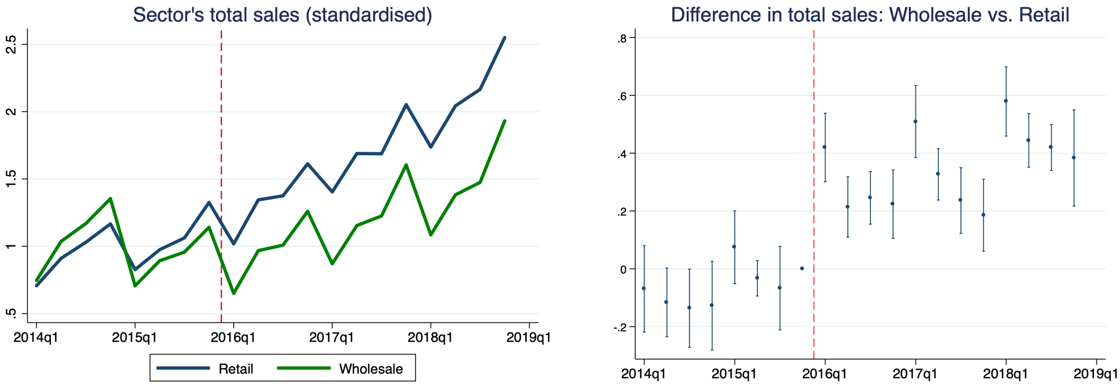
To see the effect on other variables such as reported costs and VAT liabilities, I run the simple DiD regression expressed in equation 1.2 using the VAT data. For dependent variables I use the log of quarterly total sales, purchasing costs and VAT liabilities of the firms.The results from the weighted regressions, weighted by the mean pre-intervention sales of firms, are presented in Table 1.9. Column 1 shows that VAT-liable retailers' reported sales increase by 42%. Even though the reported purchases increase by 38% as reported in column 2, they do not cancel out the effect on the final VAT liabilities of the

Moreover, since the operational audit data is subject to selection issues in terms of which firms get audited, these results need to be interpreted with caution.

[42] I aggregate the sales of retailers each period and standardise it by dividing the sums by the pre-intervention mean value of the sum. I do the same thing for wholesalers and plot them over time in panel (a) in Figure 1.2. Since they are raw quarterly sales the lines show spikes in quarter four each year. To control for this seasonality I regress the aggregate sales on quarter-of-year FEs and analyse the residuals. The residuals are plotted in Figure A4b in Appendix A.5, and they show a similar pattern as in Figure 1.2 confirming the parallel trend assumption.

[43] It might seem that the divergence between the sales of retail and wholesale sectors appear in quarter one in 2015 already. One potential reason for this is the changes in the number of VAT-liable retailers compared to the number of VAT-liable wholesalers as depicted in Figure A5a in Appendix A.5. This suggests that the slight increase in the retail sales in quarter one in 2015 is partially due to adjustments at the extensive margin. A more noticeable gap between retail and wholesale sales emerges at the start of the E-receipt program in quarter one in 2016. Also, as I discuss next, analysis of firm-level sales in Panel (b) Figure 1.2 confirms this.

Figure 1.2: Pre-trend in VAT data



(a) Sectors' standardised total sales (b) Coefficients from the flexible DiD regression

Note: Panel (a) displays the changes in the sales of retail and wholesale sectors reported on VAT returns. Each line is the raw sales reported by all firms aggregated by retail or wholesale sectors scaled by the pre-intervention average quarterly sales each sector group. The graph plots the raw sales. Thus there are spikes in the last quarter each year due to seasonality. The vertical dashed red line represents the start of the E-receipt program, which is January 1, 2016. Panel (b) plots the coefficients from equation 1.1 estimated from firm-level regression using firms' reported sales on VAT returns.

retailers. VAT liabilities increased substantially, by 31%, in column 3.[44]

Table 1.9: Direct effects - VAT returns

	(1) Sales	(2) Purchase	(3) VAT
DD coef	0.416***	0.378***	0.312**
	(0.0380)	(0.0579)	(0.124)
Firm FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry
Observations	130,311	130,311	130,311
Adjusted R^2	0.74	0.73	0.62

Note: This table displays the results from the regression equation 1.2 using the VAT data. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. The dependent variables are a log of firms' reported quarterly total sales, purchasing costs or VAT liabilities. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In comparison to the CIT analysis reported in Table 1.5 there is a larger effect on retailers' reported sales, purchasing costs and VAT-liabilities on VAT returns. This could be due to the following two reasons. First, the samples used in the CIT versus the VAT analysis

[44] As before, to investigate the extent of underestimation, I run a DiD regression, in which I use the wholesalers that never sell to any retailers as a control group. I identify such wholesalers using the firm network data from the VAT invoices. The results are reported in Table A4. The estimated coefficients of sales, costs and VAT liabilities are above 64% suggesting a substantial underestimation.

consist of different firms. The VAT sample consists of only VAT-liable firms while the CIT sample includes both non-VAT-liable and VAT-liable firms. Second, there could be a genuinely differential effect on values reported on CIT and VAT returns. To disentangle these effects, I use the same sample of firms (only VAT-liable) for both CIT and VAT analysis.

The results are reported in Table 1.10.[45] The first three columns take a log of quarterly sales, total costs and CIT liabilities reported on CIT returns as dependent variables. Similarly, I use a log of quarterly sales, purchasing costs or VAT reported on VAT returns as dependent variables for the last three columns. Comparing the first three columns to the last three, I still find a larger effect on the reported sales and tax liabilities reported on VAT returns compared to CIT returns. Specifically, the reported sales on VAT returns increase more compared to CIT returns, as in columns 1 and 4. This might seem surprising because the reported sales on VAT and CIT returns should be equal to each other. However, as mentioned before, firms could be taking advantage of the fact that it is not so straightforward to compare CIT and VAT returns for tax authorities, and thus they may manipulate their reported values. The comparison of reported quarterly total sales on CIT and VAT returns are reported in Table A8 and it confirms the large discrepancy between the two reported sales.[46]

Unlike reported sales, I cannot directly compare the total costs reported on CIT returns and purchasing costs on VAT due to their different definitions. Total costs include all types of costs such as purchasing, labour and administrative costs. Unfortunately, on CIT returns, total costs are decomposed to only production, administrative and other costs.[47] Nevertheless, as shown in columns 2 and 5, the total costs reported on CIT returns increase more compared to the purchasing costs on VAT returns .[48] Moreover, VAT-liable retailers' VAT liabilities increase more (by 25%) compared to their CIT liabilities, which increase

[45] The sample size is 81,000, which is smaller than sample used in VAT analysis in Table 1.9. This is because I match the VAT sample to the CIT sample, where all observations have positive VAT and CIT liabilities. An observation is dropped if, for example, a firm has positive CIT liabilities but non-positive VAT liabilities or vice versa.

[46] There are some legitimate reasons for why reported sales on CIT and VAT returns could differ. For example, there are different accounting rules such as revenue recognition rules for CIT and VAT. Unfortunately, the available data are not sufficient to separate how much of the discrepancy is due to these legitimate rules.

[47] The production costs contain not only purchasing costs but also labour, transportation and insurance costs that are associated with production procedures.

[48] I study the increase in reported purchasing costs on VAT returns further. I decompose the increase in purchases into its components: total input costs are split into the deductible and non-deductible input costs. Summary statistics are presented in Table A10 where it can be seen that deductible costs make 99% of the total input costs which equals the total purchasing costs for 90% of the sample. Nonetheless, I analyse the each component using the equation 1.2 and the results are presented in Table A9. The results show that both deductible and non-deductible costs increase. It is worth noting that non-deductible input costs increase more compared to deductible costs even though it does not affect firms' VAT liabilities.

by 17%.

Table 1.10: CIT and VAT comparison

	CIT returns			VAT returns		
	(1)	(2)	(3)	(4)	(5)	(6)
	Sales	Total costs	CIT	Sales	Purchase	VAT
DD coef	0.206***	0.256***	0.171***	0.254***	0.237***	0.249**
	(0.0548)	(0.0639)	(0.0368)	(0.0626)	(0.0810)	(0.0977)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
Observations	81,027	81,027	81,027	81,027	81,027	81,027
Adjusted R^2	0.77	0.76	0.59	0.77	0.75	0.68

Note: This table is to compare the effect of the E-receipt program on values reported on CIT and VAT returns. The first three columns take a log of quarterly sales, total costs or CIT reported on CIT returns as dependent variables. The last three columns use a log of quarterly sales, purchasing costs or VAT reported on VAT returns as dependent variables. The sample consists of only VAT-liable firms. The regression is specified in equation 1.2. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. Time and firm fixed effects are included in all regressions. All regressions are weighted using pre-intervention average sales. Standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

One potential explanation for the smaller increase in reported costs but larger increase in reported tax liabilities on VAT returns compared to CIT returns in Table 1.9 is the credit-invoice scheme inherent in VAT reporting. The credit-invoice scheme makes it harder for VAT-liable firms to misreport their costs (and B2B sales) on VAT returns because the values can be cross-checked with the declarations of firms' trading partners. That is, VAT-liable firms do not over-report their input costs on VAT returns, unlike CIT returns, to the extent that they offset the effect of higher reported sales.[49]To test this further I turn to the VAT audit data.

To examine the changes in misreporting behaviour of VAT-liable firms on their VAT returns I run the regression in equation 1.2. Table 1.11 presents the results and the first (last) four columns analyse the misreporting of sales (costs). The dependent variables in columns 1 and 2 are the log of reported sales and calculated true sales, respectively. I calculate true sales by adding misreported sales to reported annual sales. In column 3, I use the share of misreported sales as a dependent variable, which is the ratio between discovered hidden sales and true sales. Column 4 uses a log of the misreported sales. Similarly, in columns 5 and 6, I use a log of reported costs and calculated true costs as right-hand side variables. True costs are calculated by subtracting misreported costs from reported annual costs. Columns 7 and 8 use the share and (log of) value of misreported

[49]Of course, it is still possible that VAT firms can misreport their B2B transactions if the partner, especially the buyer, is a non-VAT firms. I discuss this further in section 1.3.2.

Table 1.11: Misreporting on VAT returns (annual values)

	Sales				Costs			
	(1) Reported	(2) True	(3) Misreport (%)	(4) Misreport ($)	(5) Reported	(6) True	(7) Misreport (%)	(8) Misreport ($)
DD coeff	0.324***	0.307***	-1.513**	-0.231	0.277***	0.289***	-0.700***	-0.180
	(0.0379)	(0.0365)	(0.632)	(0.149)	(0.0435)	(0.0433)	(0.250)	(0.505)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Observations	11,606	11,606	11,606	2,174	11,606	11,606	11,606	1,983
Adjusted R^2	0.78	0.78	0.22	0.62	0.78	0.77	0.19	0.62

Note: This table displays the results from the regression equation 1.2 using VAT audit data. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. The first (last) four columns analyses misreporting of sales (costs). The dependent variables in columns 1 and 2 are the log of reported sales and calculated true sales. I calculate true sales by adding misreported sales to reported annual sales. In column 3, I use the share of misreported sales as a dependent variable, which is the ratio between discovered hidden sales and true sales. Column 4 uses a log of the misreported sales. Similarly, in column 5 and 6, I use a log of reported costs and calculated true costs as right-hand side variables. True costs are calculated by subtracting misreported costs from reported annual costs. Column 7 and 8 use share and (log of) value of misreported costs. I calculate the share of the misreported costs by dividing the value of misreported costs by true costs. All regressions are weighted by firms' average annual sales reported on VAT returns before the intervention. Time and firm fixed effects are included in all regressions. Standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

costs. I calculate the share of the misreported costs by dividing the value of misreported costs by true costs.

Similar to the CIT audit data analysis in Table 1.8, columns 1 to 4 in Table 1.11 suggest that audited retailers misreport their sales less after the intervention. In particular, column 3 shows that retailers' share of misreported sales decreases by 1.5%. Column 4 shows that the value of hidden sales of retailers decrease by 23%, even though the coefficient is not significantly estimated. Columns 5 to 8 imply that retailers misreported their costs less on their VAT returns after the intervention, unlike the case of CIT. Particularly, column 7 indicates that the share of over-reported costs decreases by 0.7%. The value of misreported costs increases by 18%, even though insignificantly estimated, as shown in column 8.[50] These results suggest that the combination of consumer reporting and the credit-invoice scheme in VAT makes it harder for VAT-liable firms to misreport not only sales but also their purchasing costs.

A summary of the direct effect analysis

[50] Note that the number of observations used in columns 4 and 8 is smaller than the other columns. This is because some firms do not misreport sales and/or costs in some years and are dropped out because I take a log of the dependent variable. I use several other measures of misreported values sales and costs in Table A7. In particular, I use a log of one plus the misreported values and a dummy for positive misreported sales and costs. The results qualitatively confirm the fact that retailers misreport both their sales and costs on VAT returns after the intervention.

On CIT returns, I find that retailers declare 20% higher sales after the implementation of the E-receipt program, and this is due to the pure reporting effect. I do not find any real effect on retailers' production, which is proxied by their number of workers. However, this increase in sales does not directly translate into larger CIT liabilities because of higher reported costs. I document that the increase in total reported costs is mainly due to changes in the production and administrative costs of retailers. Moreover, using CIT audit data I find that some part of the rise (at least 2%) in reported costs is due to misreporting. This is because of the combination of firms' incentive to decrease their tax liabilities and the inability of tax authorities to verify the reported costs. To the best of my knowledge, this paper is the first to document the fact that firms respond to improved sales enforcement by increasing cost misreporting on CIT returns.

On the other hand, on VAT returns, VAT-liable retailers' reported sales, costs and VAT liabilities all increase more than 30%. More importantly, the increase in input costs does not offset the increase in sales, and thus VAT liabilities increase by 31%. As discussed above, one of the reasons for the E-receipt program having a substantial effect on VAT liabilities compared to CIT liabilities is the existence of the credit-invoice scheme in VAT reporting.

One thing worth analysing further is the increase in reported input costs of the VAT liabilities. Because of the credit-invoice scheme, the increase in input purchase should also mean an increase in the sales of the upstream firms — the suppliers to the VAT-retailers. This can happen if the upstream firms and retailers were colluding and hiding their trade from the authorities before the intervention. The E-receipt program forces retailers to report their sales truthfully. This, in turn, will induce an incentive for retailers to increase the reported costs; hence collusion with the upstream firms may break. If this hypothesis is true, then it means that consumer monitoring — the E-receipt program — affects not only the firms at the end of the supply chain but also the upstream firms. Therefore, the whole supply chain may well be affected by the E-receipt program. I analyse this in the next section.

Lastly, it is worth noting that the above results are the lower bounds of the effects of the E-receipt program. This is because I exploit the variation in the intensity of treatment to estimate the direct effect of the program on retailers. In particular, I compare retailers' tax reporting behaviour to that of wholesalers. The implicit assumption for this identification strategy is that the wholesalers are not affected by the program. However, in reality, some wholesalers may sell to final consumers and be affected by the program directly. Also, as briefly explained above, the wholesalers could be treated by the program indirectly. Therefore, my analysis in this subsection underestimates the true direct effects of the intervention on retailers.

1.3.2 Indirect Effects — Upstream Firms of Retailers

This subsection explores if the E-receipt program has any effect on the upstream firms of the retailers. To identify the upstream firms, I use VAT invoice data where I observe the VAT-liable buyer-seller pairs and their volume of transactions at a quarterly frequency. I define the upstream firms as the firms that have ever sold to any retailer before the intervention. A total of 4,600 upstream firms are identified and most of them belong to trade (wholesale or retail), manufacturing and professional activities such as consulting sectors as shown in Table 1.2. Using these upstream firms, I estimate the spillover effect in two ways, which are transaction-level within an upstream firm analysis and firm-level between upstream firm analysis.

I start with the transaction-level within an upstream firm analysis. I adopt a DiD estimation approach where I take the upstream firms' sales to retailers as a treatment group, and their sales to buyers in non-trade sectors as a control group.[51] In essence, I compare the change in sales to the retail sector to the change in sales to other sectors within each upstream firm. Sales to retailers is the treatment group because the retailer is directly monitored by the consumers. For each seller I calculate the total sales to the retail sector by aggregating the volume of transaction over retail buyers each quarter. Similarly, I compute the quarterly total sales to buyers in other sectors.[52] Using these transaction values I run the following DiD regression:

$$ln(T_{itr}) = \gamma_i + \sigma Treat_{ir} + \delta Post_t + \beta Treat_{ir} \cdot Post_t + u_{itr} \tag{1.3}$$

where subscripts i and t correspond to (upstream) firm and quarter as before. The subscript r represents the industry of the buyers (retail vs non-retail). Variable $Treat_{ir}$ equals one if firm i sells to retail sector r, otherwise zero.[53] $Post_t$ equals one if the quarter falls

[51] The non-trade buyers can be, for example, manufacturing, consulting, or construction firms and are non-retail and non-wholesale buyers.

[52] Summary statistics of the share of sales to each group of buyers are shown in Table 1.3.

[53] The main difference between this equation 1.3 and 1.2 is the variable $Treat$. In equation 1.2 the variable $Treat_{is}$ is at firm level and equals one if a firm belongs to retail sector, zero otherwise. In equation 1.3 the variable $Treat_{ir}$ is at transaction level and equals one if the buyer is a retailer, zero otherwise.

after January 2016 and zero otherwise. The left-hand side variable $ln(T_{itr})$ is the log of firm i's total sales to sector r in month t. I include firm fixed effect γ_i and the error terms are clustered at sellers' 4-digit industry level.

It is important to recall that, most of the upstream firms belong to the trade sector, either retailers or wholesalers, as shown in Table 1.2. And we know that retailer and wholesalers could be directly affected by the E-receipt program as discussed in the previous section. Including them in the analysis of indirect effects could contaminate the estimation of the indirect effect of the intervention. Therefore, I run several regressions for robustness by including and excluding them from the sample.

Table 1.12: Indirect effects — Transaction-level DiD

	log(Transaction value)		
	(1)	(2)	(3)
	All sellers	Non-retail	Non-trade
DD coef	0.224**	0.335***	0.397**
	(0.113)	(0.0885)	(0.182)
Buyer Ind.FE	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry
Observations	119,053	105,776	44,956
Adjusted R^2	0.63	0.63	0.63

Note: This table displays the results from the regression equation 1.3. The variable DD coef is defined as the interaction between a dummy for time period and a dummy that equals one if a buyer's sector is retail, zero otherwise. The dependent variables is a log of firm i's sales to sector r in quarter t. The first column uses all upstream firms regardless of their industry. In columns 2 I drop upstream firms that are retailer sector. The last column excludes both retailers and wholesalers from the analysis. All regressions are weighted by suppliers' average quarterly total sales before the intervention. Time and supplier fixed effects are included in all regressions. Standard errors are clustered at 4-digit industry level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I expect the coefficient on the cross term to be positive if there is an indirect effect on the upstream firms. Table 1.12 presents the results. The first column uses all upstream firms regardless of their industry. In columns 2, I drop the upstream firms that are in the retail sector from the sample. The last column excludes both retailers and wholesalers from the analysis. All regressions are weighted by suppliers' average quarterly total sales before the intervention.[54]

Table 1.12 shows that there is a positive effect on upstream firms' sales to retailers compared to their sales to non-trade buyers. The effect increases as I exclude retailers and

[54]Table A11 presents results from unweighted regressions. Estimated coefficients are positive even though they are not significant.

wholesalers from the analysis. Specifically, in column 1, sales to retailers increase by 22.4% when I use all upstream firms. The estimated coefficient increases to 33.5% when I drop retail upstream firms from the sample. In column 3, where I keep only non-trade sellers, the estimated coefficient is 40%. This means non-trade upstream firms sales to retailers increase more than their sales to other firms. The main idea behind this result is the change in retailers' incentive to collude with upstream firms due to consumer monitoring. The intuition behind these results is the change in retailers' incentive to collude with upstream firms. Because of the consumer monitoring retailers are forced to disclose their previously hidden sales. This will induce retailers to break the collusion and report larger costs to decrease their tax liabilities.

Next, I estimate the firm-level indirect effect on upstream firms. To do so, I rank the upstream firms and divide them into two groups based on their share of sales to retailers before the intervention. The firms, whose share of sales to retailers is above the median are classified as a treatment group and firms below the median are used as a control group.

Since the analysis is at the firm-level, I can examine whether the parallel-trend assumption holds. As before, for each quarter, I aggregate the reported sales on CIT returns of the upstream firms in the treatment group and standardise it by dividing the sums by the pre-intervention mean value of the sums. I do the same for the firms in the control group and plot them over time in panel (a) in Figure 1.3. Similarly, panel (b) plots the aggregate sales reported on VAT returns for each group. As we can see from the plots, there is no pre-trend before the policy change, but total sales of the treatment group start to increase more compared to the control group in 2016. The gap between them starts widening over time, and I attribute this divergence to the E-receipt program.[55]

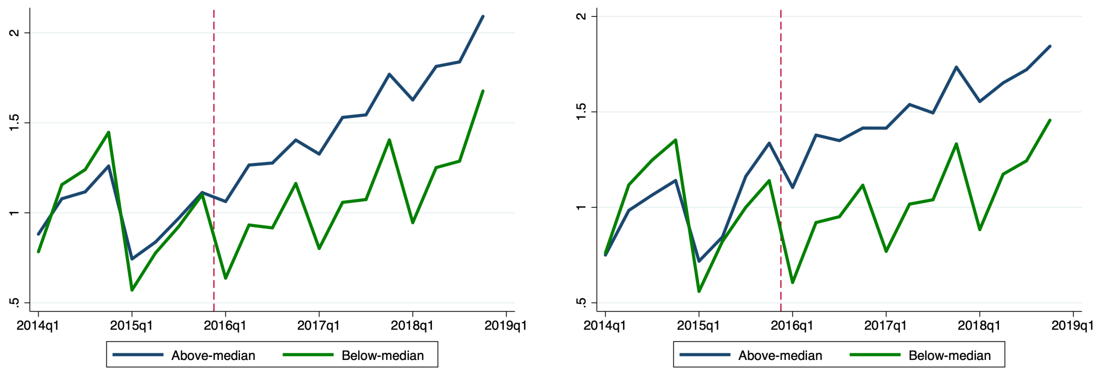
Then, I run firm-level DiD regressions specified in equation 1.4 to examine the changes in reported sales further.

$$ln(Y_{itR}) = \gamma_i + \sigma Treat_{iR} + \delta Post_t + \beta Treat_{iR} \cdot Post_t + u_{itR} \tag{1.4}$$

where subscripts i and t correspond to firm and quarter as before. The variable $Treat_{iR}$ takes one if firm i is above the median in terms of its volume of sales to retailers pre-intervention, zero otherwise. $Post_t$ equals one if the quarter falls after January 2016 and zero otherwise. The left-hand side variable $ln(Y_{it})$ is the log of firm i's quarterly total

[55]Figure A9 plots the same graphs using CIT and VAT liabilities and it confirms the parallel-trend assumption as well.

Figure 1.3: Pre-trend in upstream firms' sales



(a) Sales on CIT returns

(b) Sales on VAT returns

Note: Panel a (b) displays the changes in the total sales of the upstream firms in treatment and control groups reported on CIT (VAT) returns. Each line is the sum of sales reported by firms in the treatment or control groups scaled by the pre-intervention average quarterly sales of each group. The graph plots the raw sales. Thus there are spikes in quarter four each year due to seasonality. The vertical dashed red line represents the start of the E-receipt program, which is January 1, 2016.

sales in month t. I include firm fixed effect γ_i and the error terms are clustered at sellers' 4-digit industry level.

Table 1.13: Indirect effects — Firm-level DiD — Sales

	Sales on CIT returns			Sales on VAT returns		
	(1)	(2)	(3)	(4)	(5)	(6)
	All sellers	Non-retail	Non-trade	All sellers	Non-retail	Non-trade
DD coeff	0.258***	0.265***	0.257***	0.298***	0.315***	0.351***
	(0.0431)	(0.0399)	(0.0956)	(0.0616)	(0.0559)	(0.134)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Weight	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
Observations	69,314	61,658	26,349	69,314	61,658	26,349
Adjusted R^2	0.64	0.63	0.65	0.64	0.63	0.64

Note: This table displays the results from the regression equation 1.4. The variable DD coef is defined as the interaction between a period (pre- and post-intervention) dummy and a dummy variable, which equals one if firm i is above the median in terms of its volume of sales to retailers pre-intervention, zero otherwise. The dependent variables in the firms three columns are a log of firm i's quarterly total sales reported on CIT returns. The last three columns use reported sales on VAT returns as a dependent variable. Columns 1 and 4 include all upstream firms in the analysis regardless of their industry. In columns 2 and 5, I drop upstream firms that are retailer sector. Columns 3 and 5 exclude both retailers and wholesalers from the analysis. All regressions are weighted by suppliers' average quarterly total sales before the intervention. Period and supplier fixed effects are included in all regressions. Standard errors are clustered at 4-digit industry level.

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

Table 1.13 presents the results. The first (last) three columns correspond to changes in sales reported on CIT (VAT) returns. Columns 1 and 4 include all upstream firms in the analysis regardless of their industry. In columns 2 and 5, I drop upstream firms that are in the retail sector. Columns 3 and 5 exclude both retailers and wholesalers from the analysis. All regressions are weighted by suppliers' average quarterly total sales before the intervention. Columns 1, 2, 3 show that reported sales of upstream firms with above-median sales to retailers increase by 26% compared to those who sell less to retailers. In contrast, there is a larger effect on reported sales on VAT. Specifically, upstream firms' reported sales on VAT returns increase by at least 30%. However, it is documented that this increase in reported sales does not necessarily lead to larger tax liabilities, especially for CIT. To see this, I study the changes in the upstream firms' tax liabilities. Table 1.14 shows the results. As shown in columns 1-3, there is no significant effect on CIT liabilities. By contrast, VAT liabilities increase at least by 15% in columns 4-6. It is worth noting that these estimates are the lower bound of the true indirect effect since the control group is affected by the program to some degree. This is because sales of firms with lower rank are expected to increase to some degree since they also sell to retailers.

Table 1.14: Indirect effects — Firm-level DiD — Tax liabilities

	CIT liabilities			VAT liabilities		
	(1) All sellers	(2) Non-retail	(3) Non-trade	(4) All sellers	(5) Non-retail	(6) Non-trade
DD coeff	0.0270	0.0172	-0.0232	0.179***	0.151***	0.1771***
	(0.0395)	(0.0424)	(0.0910)	(0.0315)	(0.0312)	(0.0448)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Weight	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
Observations	46,727	40,718	16,986	51,741	46,375	20,611
Adjusted R^2	0.54	0.53	0.54	0.67	0.65	0.68

Note: This table displays the results from the regression equation 1.4. The variable DD coef is defined as the interaction between a period (pre- and post-intervention) dummy and a dummy variable, which equals one if firm i is above the median in terms of its volume of sales to retailers pre-intervention, zero otherwise. The dependent variables in the firms three columns are a log of firm i's quarterly CIT liabilities. The last three columns use reported VAT liabilities as a dependent variable. Columns 1 and 4 include all upstream firms in the analysis regardless of their industry. In columns 2 and 5, I drop upstream firms that are retailer sector. Columns 3 and 5 exclude both retailers and wholesalers from the analysis. All regressions are weighted by suppliers' average quarterly total sales before the intervention. Period and supplier fixed effects are included in all regressions. Standard errors are clustered at 4-digit industry level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

These results suggest that the E-receipt program has a positive effect on upstream firms along the VAT chain for the following reasons: first, retailers are forced to report their sales truthfully and increase their reported sales because of consumer monitoring. To decrease VAT liabilities they increase their reported purchasing costs. This, in turn, is likely to result in less collusion between retailers and their upstream firms and reveal previously hidden transactions between them. In other words, because of the credit-invoice mechanism in VAT, the increase in retailers' reported purchasing costs has to be associated with a rise in upstream firms' sales and VAT liabilities. Hence, consumer monitoring does not only affect the firms at the end of the supply chain, the retailers, as documented in the literature; rather, its effects propagate up the VAT chain. Therefore, the total impact on the economy is larger than previously thought.

1.4 Cost-benefit Analysis

In this section, I show a simple cost-benefit analysis of the E-receipt program. To implement the program, the government had to bear some costs and it is still not clear whether the program leads to an increase in tax revenue. First of all, it promises consumers 20% of the VAT paid on their purchases. Second, a lottery event is held every month and the lottery prizes get transferred to the winners' bank account every month. The total money spent on lottery prizes corresponds to 13.7% of the total VAT rebate costs. Lastly, there are other costs associated with developing an IT system, preparing the infrastructure of the E-receipt database, wage salaries of the IT workers, etc. These administrative costs account for 2.9% of the total VAT rebate costs. These cost estimates correspond to the Mongolian economy as a whole. Unfortunately, the portion of the costs that correspond to the trade sector is unknown. Therefore, I assume the same pattern holds for the trade sector.

To see whether the program pays off, I calculate the percentage increase in VAT revenue to break even. I define $VAT\ Rev_0$ as the VAT revenue of the trade sector in the absence of the E-receipt program. VAT revenue after implementing the program is denoted by $VAT\ Rev_1$.

$$
\begin{aligned}
&VAT\ Rev_0 = VAT\ Rev_1 * (1 - \underbrace{0.2}_{\text{rebate}} - (\underbrace{0.137}_{\text{lottery}} + \underbrace{0.029}_{\text{admin}}) * 0.2) \\
&\frac{VAT\ Rev_1}{VAT\ Rev_0} - 1 = 0.304
\end{aligned}
$$

The above calculation shows that a 30.4% increase in VAT payment will generate the same VAT revenue for the government net of the costs. As we have seen in Section 1.3.1, retailers' VAT liability increased by 31.2%, which is just enough to break even. Therefore, the previous literature significantly underestimates the effects of the consumer monitoring program since it does not consider the spillover effect on retailers' CIT liabilities as well as the indirect effect of the program on the upstream firms' VAT liabilities. If we include them in the calculation, the program is successful for increasing the government's tax revenue.

It is worth noting that I did not include firms' and consumers' compliance costs into the calculation. As discussed before, the compliance cost for consumers is negligible because it is possible to register a receipt as long as consumers have a cell phone. In contrast, there is a higher cost for firms since some firms have to update or buy a new registry system. Unfortunately estimates of such costs do not exist.[56] Furthermore, there are other intangible aspects of the program in terms of both benefits and costs. As for the former, the program may change societal norms that have long-lasting effects even after the program ends. These changes include people getting used to asking for receipts, an increase in tax awareness, greater attention to the public expenditure and demand for more efficient public spending and so on. On the other hand, the program increases the tax burden of the firms and thus could increase the efficiency costs of the CIT and VAT. Moreover, I do not study any changes in tax incidence or transfer of the tax burden. Even though these are interesting and important aspects of tax enforcement they are beyond the scope of this project.

1.5 Conclusion

This paper studies the role of consumer monitoring on firms' tax reporting behaviour along the supply chain. To do so, I exploit rich administrative tax data and an anti-tax evasion program implemented by the Mongolian government that incentivises consumers to report their transactions.

I start by studying the effect of the program on tax reporting behaviour of firms at the end of the supply chain — retailers. Retailers mainly sell to final consumers, and thus they are directly affected by the program. I document that consumer monitoring increases retailers' reported sales on their CIT returns by 20%. However, the effect of larger sales is partially offset by over-reporting costs. I confirm this by using tax audit data that suggest a large increase in reported costs on CIT returns is partly explained by cost misreporting. In other words, because of the consumer monitoring firms find it harder to misreport

[56] However, such costs should be reflected in firms' CIT liabilities and I find that retailers' CIT liabilities increase by 11%. In that sense, compliance costs for firms are reflected in my analysis.

their sales and thus substitute away from under-reporting sales to over-reporting costs to decrease their CIT liabilities. Thus, I find retailers' CIT liabilities increase by 11%. On the other hand, I find a stronger effect on retailers' VAT liabilities, which increase by 31%. This is because retailers' reported costs on VAT returns are constrained by the declarations of suppliers hence they are not freely adjusted. These results suggest that different opportunities for cost adjustment faced by firms in CIT and VAT ultimately lead to the different effects of consumer monitoring on their CIT and VAT liabilities.

Next, I examine how the effects of consumer monitoring propagate through the firm network. Because of the self-enforcing mechanism in B2B trade in VAT, any increase in reported input costs should be associated with an increase in upstream firms' sales. Accordingly, I find that upstream firms that sell to retailers increase their VAT liabilities by 17%. In contrast, I do not find any significant effect on their CIT liabilities. These results highlight the enforcement advantage of VAT compared to CIT and suggest that consumer monitoring enhances the self-enforcement mechanism in VAT. At the same time, it also highlights the fact that the credit-invoice system in B2B trade is not a silver bullet. This is because the self-enforcing mechanism breaks down at the end of the supply chain since consumers do not usually report their purchase. This creates opportunities for firms to evade VAT along the supply chain by, for example, colluding with one another. Therefore, it is important to include the final consumers into VAT reporting and thus ensuring better enforcement along the whole supply chain.

Taking together the effects of consumer monitoring on downstream and upstream firms, the economy-wide impact of the policy is larger than previously found in the literature.

References

Miguel Almunia, François Gerard, Jonas Hjort, Justine Knebelmann, Dorothy Nakyambadde, Claude Raisaro, and Lin Tian. An Analysis of Discrepancies in Tax Declarations Submitted Under Value-Added Tax in Uganda. *IGC report*, (February):36, 2017.

Miguel Almunia, Jonas Hjort, Justine Knebelmann, and Lin Tian. Strategic or Confused Firms? Evidence from " Missing " Transactions in Uganda. 2019.

Miguel Almunia and David Lopez-Rodriguez. Under the radar: The effects of monitoring firms on tax compliance. *American Economic Journal: Economic Policy*, 10(1):1–38, 2018.

Annette Alstadsæter, Niels Johannesen, and Gabriel Zucman. Tax evasion and inequality. *American Economic Review*, 109(6):2073–2103, 2019.

Marcelo Arbex and Enlinson Mattos. Optimal sales tax rebates and tax enforcement consumers. *Oxford Economic Papers*, 67(2):479–493, 2015.

Pierre Bachas and Mauricio Soto. Not(ch) Your Average Tax System : Corporate Taxation Under Weak Enforcement. pages 1–56, 2015.

M. Chatib Basri, Mayara Felix, Rema Hanna, and Benjamin Olken. Tax Administration vs. Tax Rates: Evidence from Corporate Taxation in Indonesia. 2020.

Youssef Benzarti, Dorian Carloni, and France Youssef Benzarti. Who Really Benefits from Consumption Tax Cuts? Evidence from a Large VAT Reform in. 2017.

Youssef Benzarti, Dorian Carloni, Jarkko Harju, and Tuomas Kosonen. What Goes Up May Not Come Down: Asymmetric Incidence of Value-Added Taxes. *National Bureau of Economic Research*, 2017.

Timothy Besley and Torsten Persson. *Pillars of Prosperity*. 2011.

Timothy Besley and Torsten Persson. Taxation and Development. *Handbook of Public Economics*, 5(December 2011):51–110, 2013.

Timothy Besley and Torsten Persson. Why do developing countries tax so little? *Journal of Economic Perspectives*, 28(4):99–120, 2014.

Paul Carrillo, Dina Pomeranz, and Monica Singha. Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*, 9(2):144–164, apr 2017.

Áureo de Paula and José A. Scheinkman. Value - Added Taxes, Chain Effects, and Informality. *American Economic Journal: Macroeconomics*, 2(October):195–221, 2010.

Abel Embaye. Underground economy estimates for non-OECD countries using currency demand method , 1984-2005. *MPRA Paper No. 20308*, 2007.

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

Jarkko Harju, Tuomas Matikka, and Timo Rauhanen. The Effect of VAT Threshold on the Behavior of Small Businesses: Evidence and Implications. Technical report, 2015.

Jarkko Harju, Tuomas Matikka, and Timo Rauhanen. Compliance costs vs. tax incentives: Why do entrepreneurs respond to size-based regulations? *Journal of Public Economics*, 173:139–164, 2019.

Jensen. Employment structure and the rise of the modern tax system. *NBER Working Paper*, 2019.

Andrew Johns and Joel Slemrod. The distribution of income tax noncompliance. *National Tax Journal*, 63(3):397–418, 2010.

Michael Keen. What Do (and Don't) We Know about the Value Added Tax ? A Review of Richard M. Bird and Pierre-Pascal Gendron's The VAT in Developing and Transitional Countries. *Journal of Economics Literature*, 47(1):159–170, 2009.

Michael Keen. The Anatomy of the VAT; by Michael Keen; IMF Working Paper 13/111; May 1, 2013. Technical report, 2013.

Michael Keen and Ben Lockwood. IMF Working Paper- The Value-Added Tax: Its Causes and Consequences. pages 1–36, 2007.

Michael Keen and Joel Slemrod. Optimal tax administration. *Journal of Public Economics*, 152:133–142, 2017.

Henrik Jacobsen Kleven, Martin Knudsen, Claus Thustrup Kreiner, Soren Pedersen, and Emmanuel Saez. Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, 79(3):651–692, 2011.

Henrik Jacobsen Kleven, Claus Thustrup Kreiner, and Emmanuel Saez. Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries. *Economica*, 83(330):219–246, 2016.

Wojciech Kopczuk and Joel Slemrod. Putting Firms into Optimal Tax Theory. *AEA PAPERS AND PROCEEDINGS*, 2006.

Adriana Veronica Litra. The Underground Economy in Romania. *Ovidius University Annals, Economic Sciences Series*, XVIII(2):173–176, 2018.

Joana Naritomi. Consumers as Tax Auditors. *American Economic Review*, 109(9):3031–3072, 2019.

ONS. Multi-factor Productivity, Indicative Estimates to 2012. (January):1–20, 2014.

Dina Pomeranz. No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105(8):2539–2569, 2015.

Dina Pomeranz and Joseacute Vila-Belda. Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities. *Annual Review of Economics*, 11:755–781, 2019.

Friedrich Schneider. The shadow economy and tax evasion: What do we (not) know? *CESifo Forum*, 13(2):3–12, 2012.

Friedrich Schneider, Andreas Buehn, and Claudio E. Montenegro. Shadow Economies All Over the World: New Estimates for 162 Countries from 1999 to 2007. *Policy Research Working Paper 5356, World Bank*, 2010.

Joel Slemrod. Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, 21(1):25–48, 2007.

Joel Slemrod. Does it matter who writes the check to the government? The economics of tax remittance. *National Tax Journal*, 61(2):251–275, 2008.

Joel Slemrod. Tax compliance and enforcement†. *Journal of Economic Literature*, 57(4):904–954, 2019.

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

Joel Slemrod and Shlomo Yitzhaki. Tax avoidance, evasion, and administration. *Handbook of Public Economics*, 3(2):201–202, 2002.

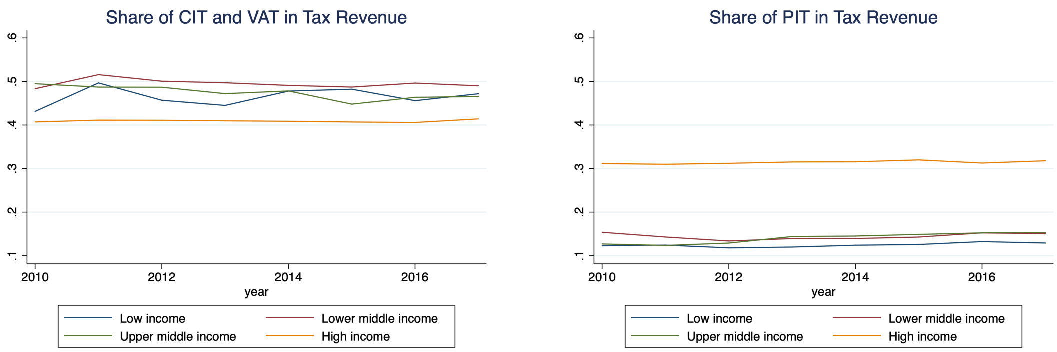
Joel B. Slemrod and Tejaswi Velayudhan. Do Firms Remit at Least 85 Percent of Tax Everywhere? *SSRN Electronic Journal*, 2017.

A Appendix

A.1 Share of CIT and VAT in Tax Revenue Across Countries

Figure A1 depicts the composition of total tax revenue across a group of countries: low, lower-middle, upper-middle and high-income countries. The sample data consist of 115 countries, of which 42 are high-income, 38 are upper-middle, 20 are lower-middle and 15 are low-income.[57]

Figure A1: Composition of tax revenue



(a) Share of CIT and VAT

(b) Share of PIT

Subfigure A1a shows that CIT and VAT constitute around 40% of total tax revenue in high-income countries. It is slightly higher — 47% — for low and middle-income countries. On the other hand, Subfigure A1b shows that personal income tax (PIT) make 30% of tax revenue for developed countries. The share of PIT for low and middle-income countries is around 15% of tax revenue, which is much lower compared to that of high-income countries. Therefore, CIT and VAT together make the largest share of tax revenue, especially in low and middle-income countries.

[57]Data sources are IMF Macroeconomic and financial data (https://data.imf.org/) and WorldBank open data (https://data.worldbank.org/)

A.2 Misreporting on CIT vs VAT Returns for VAT-liable Firms

Audited VAT-liable firms and their misreporting behaviour is summarised in Table A1. In particular, summary statistics of misreported values on the CIT audit data is reported in Table A1a and VAT audit data in Table A1b. They show that VAT-liable firms are more likely to under-report their sales and less likely to over-report their costs on VAT returns compared to CIT returns. A plausible explanation for this observation is the existence of the credit-invoice scheme in VAT. For VAT, firms reported purchasing costs are constrained by suppliers' declaration and hence it is harder to over-report costs on VAT returns.

Table A1: Summary statistics - Audit data for VAT-liable firms

(a) CIT returns

	mean	sd	count
Retailers			
Under.rep.sales ($)	0.70	5.25	3,723
share (%)	1.60	8.48	3,723
Over.rep.costs ($)	1.15	13.36	3,723
share (%)	0.88	4.30	3,723
Wholesalers			
Under.rep.sales ($)	4.04	31.32	7,636
share (%)	2.18	13.18	7,636
Over.rep.costs ($)	5.29	28.76	7,636
share (%)	1.63	6.54	7,636

(b) VAT returns

	mean	sd	count
Retailers			
Under.rep.sales ($)	1.80	15.66	3,738
share (%)	3.98	20.62	3,738
Over.rep.costs ($)	0.64	9.00	3,738
share (%)	0.63	4.79	3,738
Wholesalers			
Under.rep.sales ($)	4.33	29.73	7,621
share (%)	4.06	23.99	7,621
Over.rep.costs ($)	3.78	40.29	7,621
share (%)	2.04	9.23	7,621

Note: Table A1a and A1b present summary statistics of CIT and VAT audit data, respectively. Specifically, it summarises (annual) values of under-reported sales and their share in firms' true sales on each type of tax returns. The true sales are calculated as the sum of reported sales and hidden sales. Similarly, it provides summary statistics of (annual) values of over-reported costs, and their share in true costs in each type of tax return. The true costs are calculated as the difference between reported costs and the over-reported costs. All nominal values are in thousand USD (1 MNT = 2600 USD).

A.3 Firms Enrollment in E-receipt Program

Figure A2: Share of retailers issuing E-receipts

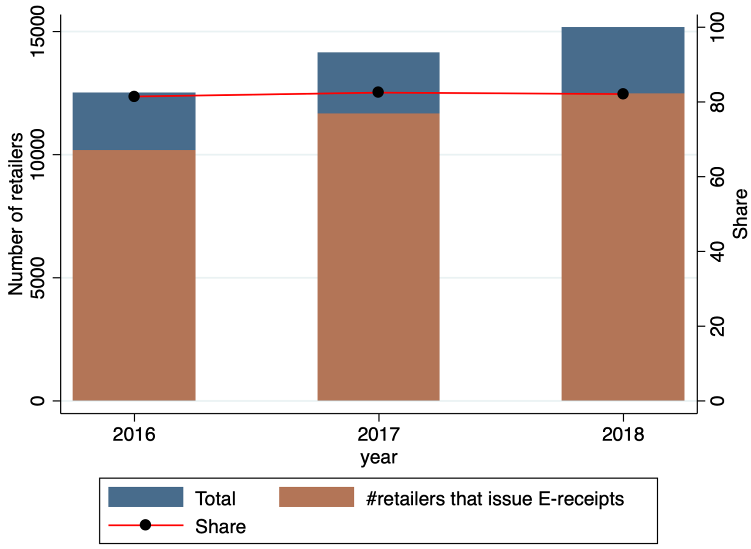


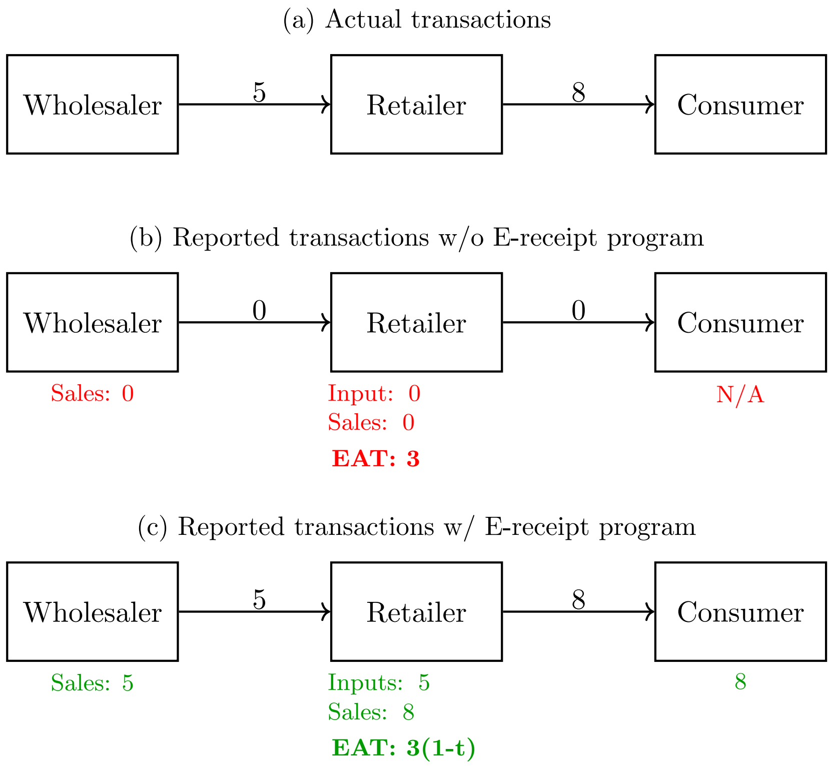
Figure A2 shows the total number of retailers that submit corporate income tax as well as the number and share of retailers that issue E-receipts. It illustrates the gradual enrollment of the retailers: the share of retailers that were enrolled in the E-receipt program was 51% in 2016 and increased to 56% in 2018.

A.4 Retailers' incentive to collude with upstream firms

In this section I explain why the effects of the E-receipt program propagate up the supply chain. To do so, I use a simple illustration of a supply chain, where I assume a retailer buys a good from a wholesaler at a price 5, and sells it to a consumer at a price 8 as illustrated in Figure A3a.

In Figure A3b I show how firms report these transactions in the absence of the E-receipt program. Since the consumer does not report their purchase to tax authorities the retailer can hide its final sales of 8. However, if the retailer declares the associated purchasing costs of 5 but not sales, it might send a red signal to the authorities. Therefore, the retailer potentially has an incentive to collude with the wholesaler and hide its purchasing costs of 5. Such misreporting of sales and costs allows the retailer to obtain the profits of 3 without paying any taxes. On the other hand, it is profitable for the wholesaler to hide its sales of 5, which leads to less tax liabilities. Also, the wholesaler may collude with its upstream firms/suppliers. Hence such collusion can happen along the whole supply chain.

Figure A3: Collusion along supply chain



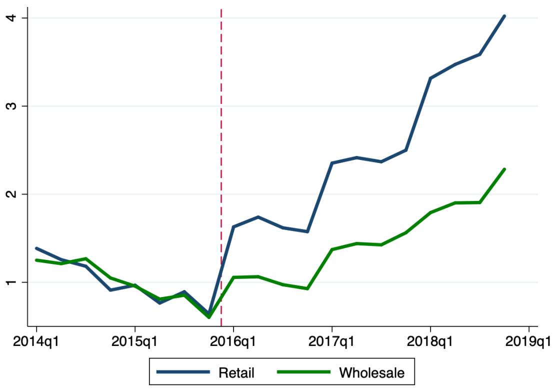
One E-receipt program is in place, the consumer start reporting the purchase of 8, this forces the retailer to declare its final sales. This, in turn, leads to a break of the collusion between the retailer and wholesaler as the retailer now has an incentive to declare the purchasing costs of 5. The sales and costs reported by the firms are illustrated in Figure A3c. In this case, the retailer has to pay taxes, thus its earnings after tax is $3(1-t)$.

Moreover, there could be other reasons why retailers might want to collude with their upstream firms. For example, upstream firms could offer a discount if they agree to hide their trade. If retailers do not report their purchase on their tax returns, the sellers would not have to pay tax on those sales and transfer some of the gains to the retailers. Therefore, it can be profitable for both upstream firms and retailers to hide their transactions. Alternatively, the firms could be involved in some underground/illegal activities, selling alcohol without a license, hence have an incentive to collude and hide their transactions from tax authorities.

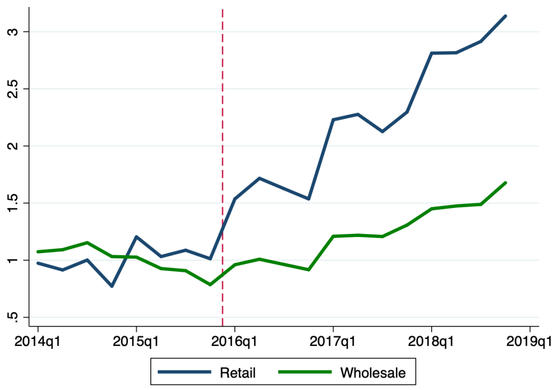
A.5 Pre-trend in reported sales after quarter fixed effects

In the main text I plot sector-level sales of wholesalers and retailers using firms' CIT (VAT) returns in Figure 1.1 (Figure 1.2) to see if the wholesalers are a valid control group. However, since they are the aggregates of raw reported sales of each firm, they exhibit spikes in quarter four each year. To control for this seasonality I regress the aggregate sales on quarter-of-year FEs and analyse the residuals. Figure A4a and Figure A4b show the CIT and VAT residuals, respectively. They are consistent with the no pre-trend assumption in the reported sales on CIT and VAT returns.

Figure A4: Pre-trend in sector-level sales, correcting for seasonality



(a) Sales reported on CIT returns

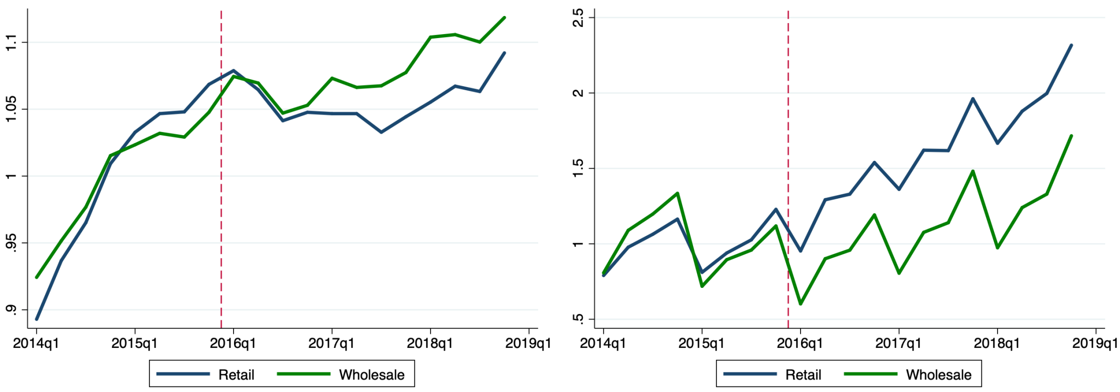


(b) Sales reported on VAT returns

Note: Panel (a) displays the changes in the sector-level sales of retailers and wholesalers reported on CIT returns after controlling for quarter-of-year fixed effect. In other words, each line plots the residuals after regressing industry-level sales of retailers and wholesalers on quarter fixed effects. Similarly, Panel (b) displays the residuals using reported sales on VAT returns by retailers and wholesalers. The vertical dashed red line represents the start of the E-receipt program, which is January 1, 2016.

One might think that the divergence between retailers and wholesalers' reported sales appear in quarter one in 2015. This is especially visible for reported sales on VAT returns in Figure 1.2 and Figure A4b. However, as depicted in Figure A5a, number of VAT-liable retailers increase more compared to number of VAT-liable wholesalers in quarter one in 2015. This suggests that the slight increase in the retail sales in quarter one in 2015 is partially due to adjustments at the extensive margin, not due to the changes in firm-level sales. Once I plot the average reported sales of each sector, which is the ratio between the aggregate sales divided by the number of firms, in Figure A5b, such early divergence is not as apparent as before in Figure 1.2.

Figure A5: Number of VAT-liable firms and average sales in each sector



(a) Number of VAT-liable firms

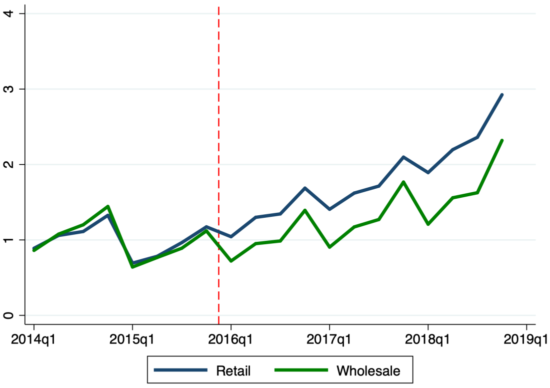
(b) Average sales reported on VAT returns

Note: Panel (a) displays the changes in the number of VAT-liable retailers and wholesalers, scaled by the pre-intervention average number of retailers and wholesalers. Panel (b) shows the average sales of VAT-liable retailers and wholesalers reported on VAT returns. Average sales in each sector are calculated by aggregating reported sales of the firms and dividing the sum by the number of firms. The average sales are also standardised by dividing by the mean pre-intervention average sales of each sector. The vertical dashed red line represents the start of the E-receipt program, which is January 1, 2016.

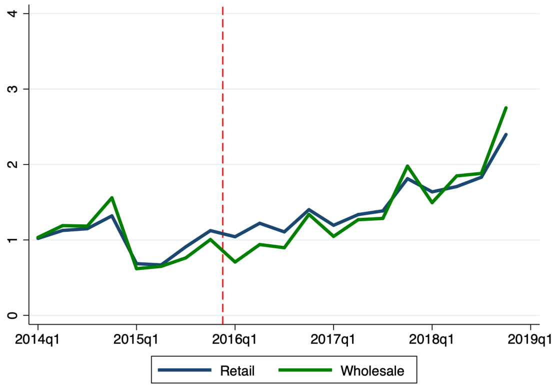
A.6 Pre-trend in reported costs and tax liabilities

In Figure 1.1 and 1.2 I show the parallel trend assumption holds in terms of firms' reported sales on CIT and VAT returns. In Figure A6 I do the same thing using firms' total costs and CIT liabilities reported on CIT returns in panels (a) and (b), and total purchasing costs and VAT liabilities from firms' VAT returns in panels (c) and (d). In particular, panel (a) plots the aggregate total costs reported on CIT returns of all retailers and wholesalers scaled by the pre-intervention average quarterly costs each sector group. Panel (b) shows the CIT liabilities of retailers and wholesalers standardised in the same way using pre-intervention average CIT liabilities. It shows that there is no pre-trend before January 2016, but total costs of retailers start to increase more compared to wholesalers after 2016. Panel (b) plots the standardised sector-level CIT liabilities of retail and wholesale sector. It also confirms the parallel trend assumption, but it exhibits only a short-lived larger effect on retailers' CIT liabilities. Panel (c) displays total purchasing costs of retailers and wholesalers divided by pre-2016 average sector-level purchasing costs, and confirms there is no-pre trend. Lastly, panel (d) shows sector-level VAT liabilities reported by retailers and wholesalers, and it confirms the parallel trend assumption as well.

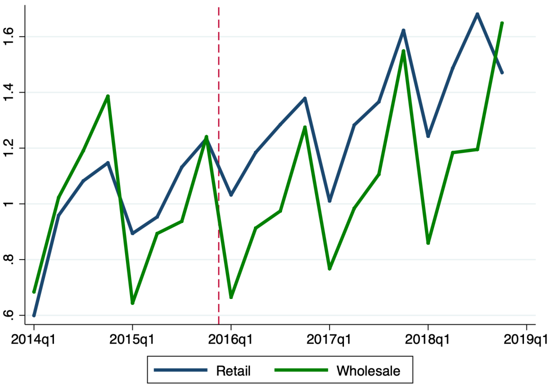
Figure A6: Pre-trend in CIT and VAT data



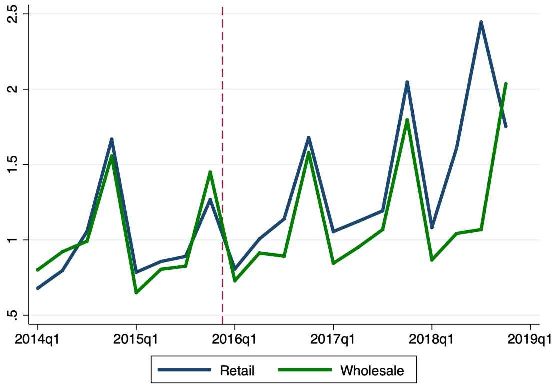
(a) Sectors' standardised total costs - CIT data



(b) Sectors' standardised CIT liabilities



(c) Sectors' standardised purchase - VAT data



(d) Sectors' standardised VAT liabilities

Note: In panel (a) and (b) I use data from firms' CIT returns. They display sector-level total costs and CIT liabilities reported by retailers and wholesalers on their CIT returns. In particular, panel (a) shows total costs reported by all firms aggregated by retail or wholesale sectors scaled by the pre-intervention average quarterly costs each sector group. Panel (b) shows the CIT liabilities of retailers and wholesalers standardised in the same way using pre-intervention average CIT liabilities. Panel (c) and (d) use data from firms' VAT returns. In panel (c) I plot standardised total purchasing costs reported by retailers and wholesalers on their VAT returns. Panel (d) shows the standardised VAT liabilities of retailers and wholesalers. The graphs plot the raw data, hence there are spikes in quarter four each year due to seasonality. The vertical dashed red line represents the start of the E-receipt program, which is January 1, 2016.

A.7 Complementary Tables — Unweighted regression

Table A2 presents results from unweighted regressions specified in equation 1.2. The first three columns show the coefficients from regressions that use a log of quarterly sales, costs or tax liabilities as dependent variables. In the last three columns I decompose the change in total costs into changes into its components: log of production, administrative and other costs are the dependent variables. The results are consistent with outcomes from weighted regressions shown in Table 1.5. From column 1 we can see that the E-receipt

Table A2: Direct effect - CIT returns (unweighted)

	Main variables			Cost decomposition		
	(1)	(2)	(3)	(4)	(5)	(6)
	Sales	Costs	CIT	Production	Admin	Other
DD coef	0.147***	0.178***	0.00894	0.216***	0.0403*	0.131
	(0.0473)	(0.0519)	(0.0185)	(0.0403)	(0.0210)	(0.104)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
Observations	202,138	202,138	202,138	144,922	171,558	29,262
Adjusted R^2	0.83	0.81	0.69	0.79	0.79	0.59

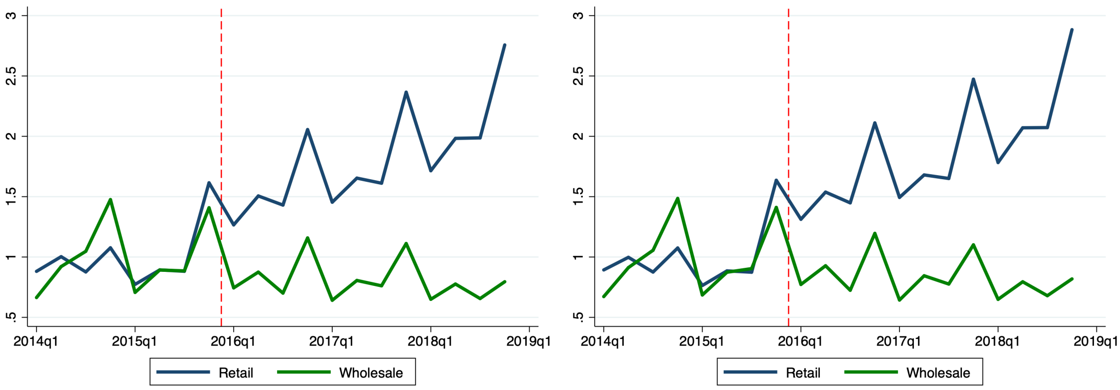
Note: This table displays the results from the regression equation 1.2. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. The first three columns take a log of quarterly sales, costs or tax liabilities as dependent variables. Only the firms with strictly positive profits are included in the analysis because I take a log of the dependent variable. That is the firms with zero tax liabilities (firms with zero or negative profits) drop out of the sample. The last three columns decompose the change in total costs into changes into its components: they take a log of production, administrative and other costs as dependent variables. Time and firm fixed effects are included in all regressions. Standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

program induced retailers to report 15% higher sales relative to wholesalers. However, in column 2, retailers' reported costs increased by 18%. This increase in costs offsets the effect on CIT liabilities, and there is no significant increase in CIT liabilities. The last three columns show that an increase in total costs is mainly driven by the rise in production and administrative costs. The coefficient on other costs is insignificant even though it is positive. This result is slightly different from the findings in the existing literature (for example Carrillo et al. 2017), where they document that firms in Ecuador tend to increase costs that are more difficult to verify such as "other administrative costs" in response to increased third-party information on sales.

A.8 Underestimation of Direct Effects

It is important to note that analyses of the direct effect on retailers underestimate the true effects of the E-receipt program. To identify the direct effect, I use wholesalers as a control group for retailers. The underlying assumption for this strategy is that wholesalers would have behaved similarly to retailers in the absence of the intervention (parallel trend assumption) and that wholesalers are not affected by the program. The data exhibit a reasonable parallel trend in the sales of retailers and wholesalers before the intervention, which validates the parallel trend assumption. However, the wholesalers are likely to be affected by the program both directly and indirectly. Wholesalers are likely to be directly

Figure A7: Underestimation: Pre-trend in CIT data



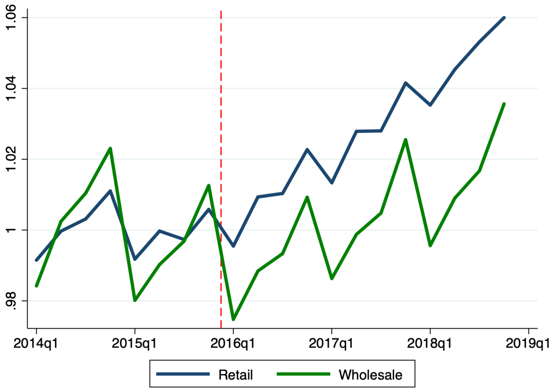
(a) Sectors' standardised total sales

(b) Sectors' standardised total costs

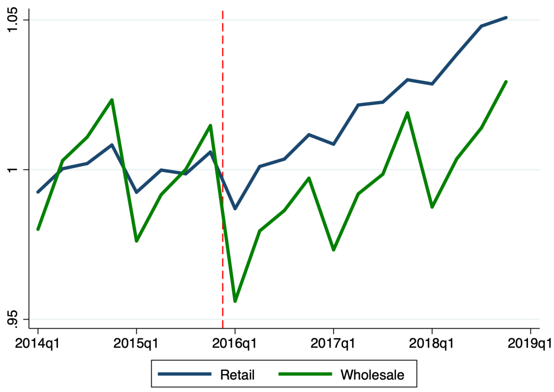
Note: Panel (a) and (b) display the changes in the total sales and costs of retail and wholesale sectors reported on CIT returns. Wholesale sector contains only the firms that never sell to any retailers between 2014 and 2018. Each line is the total sales reported by all firms aggregated by retail or wholesale sectors scaled by the pre-intervention average quarterly sales each sector group. The graph plots the raw sales. Thus there are spikes in quarter four each year due to seasonality. The vertical dashed red line represents the start of the E-receipt program, which is January 1, 2016.

affected because they could sell to final consumers. Also, not surprisingly, wholesalers are classified as upstream firms, and I find substantial spillover effect on the upstream firms in Section 1.3.2. Therefore, the estimated effects are a lower bound of the true direct effects on retailers. To investigate the extent of the underestimation, I change the control group to the wholesalers that never sell to any retailers. I identify such wholesalers using the firm network data from VAT invoice. To test the no pre-trend assumption I plot the industry level sales and costs from firms' CIT (VAT) returns in Figure A7 (Figure A8) and they show reasonable parallel trend.

Figure A8: Underestimation: Pre-trend in VAT data



(a) Sectors' standardised total sales



(b) Sectors' standardised total purchasing costs

Note: Panel (a) and (b) display the changes in the total sales and purchasing costs of retail and wholesale sectors reported on VAT returns. Wholesale sector contains only the firms that never sell to any retailers between 2014 and 2018. Each line is the total sales reported by all firms aggregated by retail or wholesale sectors scaled by the pre-intervention average quarterly sales each sector group. The graph plots the raw sales. Thus there are spikes in quarter four each year due to seasonality. The vertical dashed red line represents the start of the E-receipt program, which is January 1, 2016.

In Table A3 I report the regression results from the simple DiD specifications in equation 1.2. The results using CIT data are reported in. Similarly, Table A4 reports the results using VAT data.The estimated coefficients on sales and costs (as well as on VAT liabilities) are above 60% suggesting a substantial underestimation.

Table A3: Underestimation of the direct effect - CIT returns

	(1) Sales	(2) Costs	(3) CIT
DD coef	0.608***	0.755***	-0.0613
	(0.202)	(0.224)	(0.226)
Firm FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry
Observations	38,939	38,939	38,939
Adjusted R^2	0.87	0.86	0.88

Note: This table displays the results from the regression equation 1.2. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. Unlike the main specification in Table 1.5, I use the wholesalers that never sell to retailers as a control group, which are identified from the firm network data. The dependent variables are a log of firms' reported quarterly sales, costs or tax liabilities on CIT returns. Only the firms with strictly positive profits are included in the analysis because I take a log of the dependent variable. That is the firms with zero tax liabilities (firms with zero or negative profits) drop out of the sample. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Underestimation of the direct effect - VAT returns

	(1) Sales	(2) Purchase	(3) VAT (final)
DD coef	0.695***	0.696***	0.640***
	(0.0375)	(0.0597)	(0.125)
Firm FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry
Observations	76,973	76,973	76,973
Adjusted R^2	0.72	0.70	0.65

Note: This table displays the results from the regression equation 1.2 using the VAT data. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. Unlike the main specification in Table 1.9, I use the wholesalers that never sell to retailers as a control group, which are identified from the firm network data. The dependent variables are a log of firms' reported quarterly sales, costs or tax liabilities on CIT returns. Only the firms with strictly positive profits are included in the analysis because I take a log of the dependent variable. That is the firms with zero tax liabilities (firms with zero or negative profits) drop out of the sample. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.9 Cost Decomposition Sample

Some firms do not report the decomposition of the total costs accurately. For example, production costs are equal to the total costs and the other two components are zero. I do the same regression analysis by dropping those firms. The results are reported in Table A5 and they are consistent with outcomes in Table 1.5.

Table A5: Direct effect - CIT returns

	Main variables			Cost decomposition		
	(1) Sales	(2) Costs	(3) CIT	(4) Production	(5) Admin	(6) Other
DD coef	0.182***	0.195***	0.0986*	0.223***	0.118***	0.206
	(0.0646)	(0.0714)	(0.0567)	(0.0698)	(0.0351)	(0.230)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
Observations	139,257	139,257	139,257	125,803	129,517	27,494
Adjusted R^2	0.76	0.74	0.59	0.70	0.80	0.40

Note: This table displays the results from the regression equation 1.2 using the cost decomposition sample: I drop firms that do not report the decomposition of the total costs accurately. For example, I exclude the firms that report production costs equal to the total costs but the other two components are zero. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. The first three columns take a log of quarterly sales, costs or tax liabilities as dependent variables. Only the firms with strictly positive profits are included in the analysis because I take a log of the dependent variable. That is the firms with zero tax liabilities (firms with zero or negative profits) drop out of the sample. The last three columns decompose the change in total costs into changes into its components: they take a log of production, administrative and other costs as dependent variables. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.10 Robustness Checks for Changes in Retailers' Misreporting Behaviour

To study the changes in misreporting behaviour of retailers on CIT returns I use a log of the original value of misreported sales and costs in Table 1.8. Since some of the audited firms do not misreport their sales and/or costs in some years and such observations are dropped out of my sample because I take log. To avoid this I use several other measures of misreported values sales and costs in Table A6. In particular, I use a log of one plus the misreported values and a dummy for positive misreported sales and costs. The results qualitatively confirm the fact that retailers misreport their sales less but more likely to over-report their costs after the intervention as in Table 1.8. Table A7 presents the estimated coefficients from the same analysis using VAT audit data.

Table A6: Other measures for misreported sales and costs on CIT returns

	Misreported Sales ($)			Misreported Costs ($)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Original	log(x+1)	Dummy	Original	log(x+1)	Dummy
DD coeff	-0.0151	-0.0328	-0.00531	0.362**	1.053***	0.0582***
	(0.249)	(0.275)	(0.0181)	(0.137)	(0.296)	(0.0192)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
Observations	2,246	14,100	14,100	3,091	14,100	14,100
Adjusted R^2	0.68	0.14	0.14	0.74	0.30	0.29

Note: This table CIT audit data and uses different measures for values of misreported sales and costs. In particular, for columns 1-3 (4-6), I use a log of the original value misreported sales (costs), a log of one plus the misreported sales (costs), and a dummy for positive misreported sales (costs), respectively. Since some firms do not misreport sales and costs (zero value), the number of observations is smaller in columns 1 and 4. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. All regressions are weighted by firms' average annual sales reported on CIT returns before the intervention. Time and firm fixed effects are included in all regressions as expressed in equation 1.2. Standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Other measures for misreported sales and costs on VAT returns

	Misreported Sales ($)			Misreported Costs ($)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Original	log(x+1)	Dummy	Original	log(x+1)	Dummy
DD coeff	-0.231	-0.362*	-0.0206	-0.180	-1.073***	-0.0671***
	(0.149)	(0.201)	(0.0139)	(0.505)	(0.278)	(0.0180)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
Observations	2,174	11,606	11,606	1,983	11,606	11,606
Adjusted R^2	0.62	0.14	0.14	0.62	0.30	0.31

Note: This table VAT audit data and uses different measures for values of misreported sales and costs. In particular, for columns 1-3 (4-6), I use a log of the original value misreported sales (costs), a log of one plus the misreported sales (costs), and a dummy for positive misreported sales (costs), respectively. Since some firms do not misreport sales and costs (zero value), the number of observations is smaller in columns 1 and 4. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. All regressions are weighted by firms' average annual sales reported on CIT returns before the intervention. Time and firm fixed effects are included in all regressions as expressed in equation 1.2. Standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Comparison of sales reported on CIT and VAT returns

	min	mean	med	max	sd	count
Sales on CIT return	0	84	18	689	153	82,023
Sales on VAT return	0	124	18	2,442	331	82,023
Difference (CIT-VAT)	-2,430	-41	0	674	213	82,023
Share of diff in CIT sales	-1,324,575	-58	0	100	4,946	82,023

Note: This table presents summary statistics of total reported sales on CIT and VAT returns as well as their difference. All nominal values are in thousand USD (1 MNT = 2600 USD).

A.11 Comparison of Sales Reported on CIT and VAT returns

In section 1.3.1.2 I document that retailers' reported sales on VAT returns respond more compared to CIT returns. One potential explanation is that firms could be taking advantage of the fact that it is not straightforward to compare CIT and VAT returns for tax authorities and manipulate their reported values. Specifically, values reported on both tax returns are cross-checked manually by tax officers and it is not done for all firms. And it is not straightforward to compare because VAT returns are submitted monthly and values corresponding to the respective month are reported. In contrast, CIT returns are sent quarterly and values are in cumulative values. Table A8 compares quarterly total reported sales (in thousand USD) on CIT and VAT returns and their difference. It shows that there is a large difference between the sales reported on CIT and VAT returns.

A.12 VAT Data — Decomposition of Purchasing Costs

In Table 1.9 I document that the purchasing costs of VAT-liable retailers increase by 39%. I decompose the increase in purchases into its components using the equation 1.2: total purchasing costs are split into the deductible and non-deductible input costs on VAT returns.[58] The results are presented in Table A9. It results show that both deductible and non-deductible costs increase. It is worth noticing that non-deductible input costs increase more compared to deductible costs even though it does not affect firms' VAT liabilities.

Table A9: Decompostion of purchasing costs reported on VAT returns

	Weighted			No weight		
	(1)	(2)	(3)	(4)	(5)	(6)
	Total	Deductible	Non-deductible	Total	Deductible	Non-deductible
DD coef	0.378***	0.296**	0.883***	0.303***	0.312***	1.010***
	(0.0579)	(0.128)	(0.247)	(0.0357)	(0.0387)	(0.198)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry	Industry
Observations	130,309	130,164	9,099	130,309	130,164	9,099
Adjusted R^2	0.73	0.74	0.55	0.70	0.70	0.60

Note: This table displays the results from regressions expressed in equation 1.2. The first three columns represent regressions weighted by firms' average quarterly sales before the intervention. The last three columns are for unweighted regressions. The dependent variables are a log of the total, deductible and non-deductible input costs. Variable DD coef is defined as the interaction between a dummy for retail sectors, and a dummy variable that equals one for the periods after January 2016, zero otherwise. Time period (before and after-intervention) and firm fixed effects are included in all regressions. Standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A10 presents summary statistics of quarterly total, deductible and non-deductible purchasing costs reported on VAT returns. It can be seen that deductible costs make 99% of the total input costs and it equals the total purchasing costs in 90% of the sample.

[58] Summary statistics are presented in Table A10 and it can be seen that deductible costs make 99% of the total input costs and it equals the total purchasing costs in the 90% of the sample.

Table A10: Summary statistics - Purchasing costs decomposition

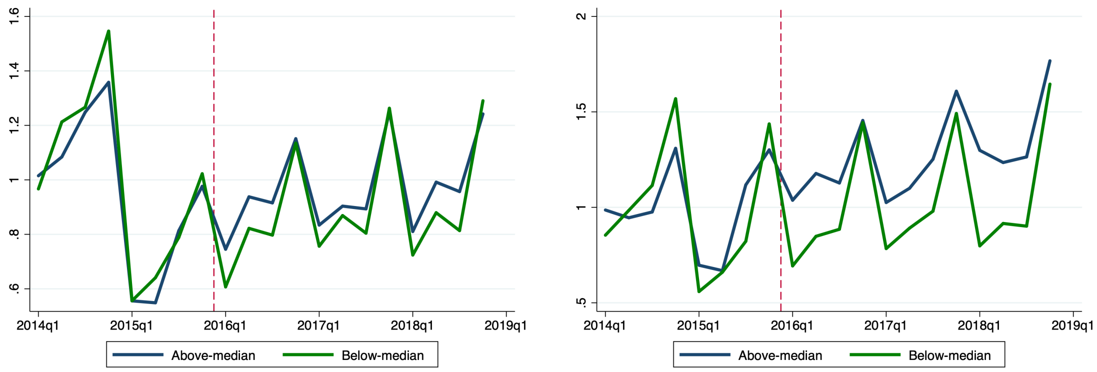
	min	mean	med	max	sd	count
Value						
Total purchase	0	57	6	13,064	237	132,586
Deductible	0	55	5	13,064	226	132,586
Non-deductible	0	2	0	8,259	55	132,586
	.					
Share (%)						
Deductible	0	98.9	100	100	8	132,586
Non-deductible	0	1.1	0	100	8	132,586

Note: Table A10 presents descriptive statistics of quarterly total, deductible and non-deductible purchasing costs reported on VAT returns. All nominal values are in thousand USD (1 MNT = 2600 USD).

A.13 Analysis of Indirect Effects — Pre-trend in Tax Liabilities

I examine whether the parallel-trend assumption holds for CIT and VAT liabilities of the upstream firms in Figure A9. In particular, for each quarter, I aggregate the reported CIT liabilities of the upstream firms in the treatment group and standardise it by dividing the sums by pre-intervention mean value of the sums. I do the same for the firms in the control group and plot them over time in panel (a) in Figure A9. Similarly, panel (b) plots the aggregate VAT liabilities for each group. As we can see from the plots, there is no pre-trend before the policy change, but total sales of the treatment group start to increase more compared to the control group in 2016. The gap between them starts widening over time, and I attribute this divergence to the E-receipt program.

Figure A9: Indirect effects — Pre-trend in tax liabilities



(a) CIT liabilities (b) VAT liabilities

Note: Panel a (b) displays the changes in the total CIT (VAT) liabilities of the upstream firms in treatment and control groups. Each line is the sum of sales reported by firms in the treatment or control groups scaled by the pre-intervention average quarterly tax liabilities of each group. The graph plots the raw sales. Thus there are spikes in quarter four each year due to seasonality. The vertical dashed red line represents the start of the E-receipt program, which is January 1, 2016.

A.14 Analysis of Indirect Effects — Transaction-level DiD (no weight)

Table A11 presents results from unweighted regressions specified in equation 1.3. The first column uses all upstream firms regardless of their industry. In columns 2 I drop upstream firms that are retailer sector. The last column excludes both retailers and wholesalers from the analysis. The estimated coefficients are positive even though they are not significant. Therefore, it also suggests that there is a positive effect on upstream firms' sales to retailers compared to their sales to non-trade buyers as discussed in section 1.3.2.

Table A11: Indirect effects — Transaction-level DiD (no weight)

	log(Transaction value)		
	(1)	(2)	(3)
	All sellers	Non-retail	Non-trade
DD coef	0.0178	0.0253	0.00825
	(0.0454)	(0.0305)	(0.0551)
Buyer Ind.FE	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes
Weight			
Cluster	Industry	Industry	Industry
Observations	119,053	105,776	44,956
Adjusted R^2	0.61	0.61	0.60

Note: This table displays the results from the regression equation 1.3. The variable DD coef is defined as the interaction between a dummy for time period and a dummy that equals one if a buyer's sector is retail, zero otherwise. The dependent variable is a log of upstream firms' quarterly sales to (retail vs non-retail) downstream firms. The first column uses all upstream firms regardless of their industry. In columns 2 I drop upstream firms that are retailer sector. The last column excludes both retailers and wholesalers from the analysis. All regressions are unweighted. Time and supplier fixed effects are included in all regressions. Standard errors are clustered at 4-digit industry level.

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

A.15 Robustness Checks Using Unbalanced Data

In the main analysis, I focus on firms with strictly positive profits and tax liabilities. This is because I take a log of the variables but the firms' profits and tax liabilities can be zero or even negative. In this subsection, I show that the main results survive qualitatively even if I include the firms with non-zero profits and tax liabilities in the analysis.

I start by analysing the CIT data as in the main text. Table A12 corresponds to the Table 1.5, and it is consistent the main result. If anything, it suggests that the effects of the E-receipt program on firms' reported sales, costs and CIT liabilities are stronger as the estimated coefficients are larger.

Next, I analyse if there the E-receipt program caused is any real response. That is, I examine if there are any differential changes in retailers' number of workers and the value of wages. Table A13 corresponds to the Table 1.6, and it confirms that there is no real response. I don't find any significant effect on retailers' workers and wages. In other words, any changes in retailers' reported sales, costs, and CIT liabilities are due to changes in reporting behaviour.

Table A12: Direct effects - CIT returns — Unbalanced data

	(1) Sales	(2) Costs	(3) CIT
DD coef	0.403***	0.344***	0.141
	(0.0331)	(0.0424)	(0.105)
Firm FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry
Observations	314,097	314,097	213,168
Adjusted R^2	0.51	0.65	0.50

Note: This table displays the results from the regression equation 1.2. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. The dependent variables are a log of firms' reported quarterly sales, costs or tax liabilities on CIT returns. That is the firms with zero tax liabilities (firms with zero or negative profits) drop out of the sample. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A13: No real response by retailers — Unbalanced data

	Main variables			Real response	
	(1) Sales	(2) Expenses	(3) CIT	(4) Wages	(5) Workers
DD coef	0.280***	0.228***	0.0763	0.0522	0.0715
	(0.0367)	(0.0284)	(0.0766)	(0.0587)	(0.0929)
Firm FE	Yes	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry	Industry
Observations	143,795	143,795	94,099	143,795	143,795
Adjusted R^2	0.57	0.74	0.53	0.89	0.88

Note: This table displays the results from the regression equation 1.2. The first three columns take a log of quarterly sales, costs or tax liabilities as dependent variables. The payroll data covers Q1 in 2015 to Q3 in 2018 only. Therefore, less observation compared to Table 1.5. The dependent variables in the last two columns are log of total wages and number workers. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Lastly, I turn to VAT data. Table A14 corresponds to the Table 1.9, and it is consistent the main result. If anything it implies even stronger effects of the E-receipt program on VAT-liable firms. In particular, it shows that VAT-liable retailers' reported sales increase by 45%. Even though the reported purchases increase by 39% as reported in column 2, they do not cancel out the effect on the final VAT liabilities of the retailers. VAT liabilities increased substantially, by 31%, in column 3.

Table A14: Direct effects - VAT returns — Unbalanced data

	(1) Sales	(2) Purchase	(3) VAT
DD coef	0.450***	0.386***	0.312**
	(0.0287)	(0.0523)	(0.124)
Firm FE	Yes	Yes	Yes
Cluster	Industry	Industry	Industry
Observations	183,793	183,793	130,316
Adjusted R^2	0.64	0.70	0.62

Note: This table displays the results from the regression equation 1.2 using the VAT data. The variable DD coef is defined as the interaction between a dummy for retail sectors and a dummy that equals 1 for the periods after January 2016. The dependent variables are a log of firms' reported quarterly total sales, purchasing costs or VAT liabilities. Time and firm fixed effects are included in all regressions. All regressions are weighted by firms' average quarterly sales before the intervention and standard errors are clustered at 4-digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 2

Characterising Tax-Evading Firms: Evidence from Mongolia

Understanding which firms evade taxes is important for developing economies who are trying to broaden their tax base and ensure equitable compliance. In this paper I characterise tax-evading firms using a Mongolian government program, which incentivises consumers to report their purchases. I find that as in other countries, this consumer monitoring increases firms' reported sales and tax liabilities. I then study the firms that reported an abnormally large growth in their sales in the year that the program was launched, which suggests that those firms had previously been evading more taxes. I find that tax evasion was particularly prevalent among smaller firms, and conditional on firm size it was more common among older firms. My findings also suggest that tax evasion was more common in the capital city.

2.1 Introduction

The state capacity of collecting taxes is one of the determinants of long-run economic growth. A well-designed tax system helps the government to raise public funds to spend on provision of public goods and services that are essential for economic development. (Barro, 1990; Besley and Persson, 2013; Acemoglu et al., 2018) However, in reality, tax structures are distorted by features such as tax evasion and avoidance, particularly in developing countries. Tax evasion and avoidance not only lead to less tax revenue for the government but also it causes inequality of tax burden across agents that induce misallocation of resources in the economy. Therefore, detecting tax evaders is a crucial step for tax authorities in developing economies to broaden their tax base and ensure equitable compliance.

Modern tax systems heavily rely on firms because firms remit a large fraction of the tax revenue either with regard to their own tax liabilities or through the withholding of taxes of employees or other businesses (Kopczuk and Slemrod, 2006; Slemrod and Valayudhan, 2017).[1] Therefore, identifying tax-evading firms is particularly useful for raising tax revenue.

In this paper I study the types of firms that are more likely to evade tax. To do so, I exploit a nationwide anti-tax evasion program, called E-receipt program, implemented by the Mongolian government in 2016. The program uses consumers as informants about firms' sales by incentivising them to report their purchase to the tax authorities.[2] In chapter 1 of this thesis I document that the E-receipt program informs us of the extent of pre-intervention tax evasion of firms. In particular, I find that the program induces retailers to report larger sales, at least by 20%. This suggests that the retailers would have hidden 20% of its sales from the tax authorities and evaded the corresponding tax in the absence of the E-receipt program. Therefore, to identify tax-evading firms I study the retailers that reported an abnormally large growth in their sales in the year that the program was launched. However, this line of reasoning depends on the effectiveness of the E-receipt program. Retailers are more likely to be classified as tax evaders if they participate in the E-receipt program diligently. On the other hand, the retailers that under-report their sales are not labelled as tax evaders if they don't comply with the program. Therefore, my measures underestimate the extent of tax evasion by retailers.

[1] To be specific, firms remit 85% of total tax revenue in OECD countries and India (Slemrod and Valayudhan, 2017).

[2] Similar policies were implemented in Brazil, Spain, Taiwan and other countries.

I begin by estimating the extent of tax evasion for each retailers using annual administrative tax data from CIT returns that spans the period 2013 and 2018. Specifically, I use the following three measures for the extent of retailers' tax evasion: the increase in retailers' sales growth above its firm-specific trend *in* 2016; above its firm-specific trend *after* 2016; and finally, the increase in retailers' sales growth after 2016 compared to that of wholesalers. The last one uses a difference-in-difference estimation approach, where I use retailers as the treatment group and wholesalers as the control group. Arguably, each of these three methods of estimating tax evasion degree has its caveats. The first one, which looks at a sudden increase in sales growth at the start of the E-receipt program underestimates firms' tax evasion extent because of the gradual enrollment of firms (and consumers). The number of retailers enrolled in the E-receipt program in each year is shown in Figure B1 and it displays a slightly slow take-up rate by retailers. Therefore, not all retailers who evade tax are affected by E-receipt program and increase their reported sales in 2016. The second approach fixes the issue of gradual enrollment by firms. However, it may classify the firms with accelerating growth in sales as tax evaders. As for the third approach, there are some potential violations to the parallel trend assumption between retailers and wholesalers as discussed in chapter 1. For example, some wholesalers could be treated by the E-receipt program because they sell directly to final consumers and violates assumptions of the control group. These potential violations imply that the third approach underestimates the extent of tax evasion. Nevertheless, I show that all three measures produce similar results.

I then associate the estimated tax evasion level of firms to their observable characteristics on tax returns such as firms' size, age, location and others. I document several patterns in the data. First, I find that firm size is negatively correlated with tax evasion. That is, small firms experience higher growth in their reported sales. There could be many explanations for this result. For example, firms with different sizes can have distinct tax evasion tendency because they face different audit probability, risk aversion, and subject to a different amount of third-party information such as credit card usage. Moreover, for small firms, the person making the evasion decision (e.g., whether or not to give a receipt) is also the residual claimant of the tax money saved, while for large firms, this decision is usually made by an employee who does not benefit directly. It is important to understand why firm size is negatively associated with tax evasion, but this is beyond the scope of this chapter.

Second, I find that older retailers evade tax more after controlling for firm size. Third, retailers in the capital city of Mongolia, called Ulaanbaatar, experience higher growth in sales and thus appear to evade tax more. One of the possible explanations is that firms in Ulaanbaatar are treated intensively (for example, due to better campaigning by tax authorities) and thus increase their reported sales more compared to firms in other regions. Fourth, firms that have ever issued E-receipts experience higher sales growth. This reassures that the changes in reported sales are driven by the E-receipt program. Other observable location-specific characteristics such as GDP per capita, population, number of firms in the province, the distance between Ulaanbaatar and the province in which the firm is located are not related to firms' tax evasion behaviour.

In the last part of my analysis I study how the estimated tax evasion of firms differ across different firm size groups. In particular, I associate the mean and standard deviation of firms' estimated tax evasion level vary across different decile of firm size distribution. I find that the mean is decreasing in percentiles, confirming the negative relation between firm size and tax evasion. This implies that there is vertical inequality across firms in terms of their tax. Moreover, I document that the dispersion of the estimated tax evasion of firms is negatively correlated with the size decile. For example, the lowest decile has the largest dispersion in tax evasion estimates across the firms in the decile. This suggests that there is horizontal inequality across firms in terms of their tax evasion level.

This paper closely relates to the literature on tax evasion of firms. In particular, several papers document the negative relationship between firm size and tax evasion (Pomeranz, 2015; Kleven et al., 2016; Naritomi, 2019). They suggest that larger firms evade tax less because they produce more third-party information such as paper trials. For example, Kleven et al., 2016 shows theoretically that it becomes harder for firms to evade (payroll) tax by colluding with employees as the number of employees increases. As for the empirical studies, Pomeranz, 2015, documents that a significant part of the higher evasion in smaller firms may be driven by a weaker paper trail as it is associated with the share of the sales to final consumers. Similarly, Naritomi, 2019, finds that the sheer size of a firm could deter under-reporting since the number of third-parties firms interact with can have a monitoring effect. This is because, for example, one of the consumers could become a whistle-blower and reveal firm's tax evasion. My results are consistent with such results. Moreover, I add to this literature by studying the relationship between tax evasion and other firm-level characteristics such as age and location.

The remainder of this paper is structured as follows. Section 2.2 explains the relevant datasets and summary statistics. Section 2.3 discusses the empirical analysis and results. I conclude in Section 2.4.

2.2 Data

I use administrative tax data from firms' CIT returns. The data is at an annual frequency and it covers the period between 2013 and 2018.[3] The main analysis uses a balanced sample, where I observe the firms throughout the sample period.[4] Since the main tax evasion variable is the growth in reported sales, I observe firms' growth rate in sales for five years between 2014 and 2018 in the final sample. 2013 data is used to calculate the sales growth of firms in 2014. Therefore the data covers two years before and three years after the E-receipt program was initiated. In this paper I mainly focus on retailers (I use wholesalers in one of the specifications as a control group). The final sample size is 18110, which consists of 3622 retailers. Their observable characteristics (mostly in year 2015) are summarised in Table 2.1. The sample mainly consists of small retailers as we can see from annual sales ($Sales_{15}$, $Sales_{pre}$) and the number of workers ($\#workers15$).[5] Retailers' age in 2015 varies between 3 and 24.[6] Table 2.1b summarises dummy variables and it shows that 26% of the retailers (952 out of 3,622) are located in the capital city and more than half of them are VAT-liable throughout the sample. Also, over 85% of the firms (3,107 out of 3,622) has ever issued a E-receipt.

Table 2.1: Firm characteristics

	min	mean	med	max	sd	count
Age_{15}	3	9	9	24	5	3,622
$Sales_{15}$	0	304	9	170,578	4,529	3,622
$Sales_{pre}$	0	331	8	211,143	5,204	3,622
$\#workers_{15}$	1	6	2	300	18	2,559
$Wages_{15}$	0	11	2	1,697	62	2,559

(a) Flow variables

	Dummy=1	Dummy=0	Total
Capital city	952	2,670	3,622
Always VAT firm	1,941	1,681	3,622
Ever VAT firm	2,412	1,210	3,622
E-receipt firm	3,107	515	3,622

(b) Dummy variables

Note: Table 2.1a (Table 2.1b) summarises firms' observable characteristics represented by flow (dummy) variables. Age_{15}, $Sales_{15}$, $\#workers_{15}$ and $Wages_{15}$ are firms' age, number of workers, annual wages and sales in year 2015. $Sales_{pre}$ is the mean sales before the intervention (year 2014 and 2015). Capital city equals one if a firm is located in the capital city Ulaanbaatar, zero otherwise. E-receipt firm=1 if a firm has ever issued E-receipt between 2016 and 2018. All nominal values are in units of thousand USD.

Definitions of tax evasion are based on how firms' sales evolve around the start of the

[3]This chapter uses annual data in contrast to chapter 1, where the data is at a quarterly level. This is because I am interested in sales growth of retailers. In Mongolia, firms report their sales and costs at cumulative terms on CIT returns. For example, in quarter 2, firms report the sum of sales in quarter 1 and 2. Similarly, they report their annual sales in quarter 4. Moreover, because firms can use year-end tax adjustment declarations, tax authorities (as well as firms) care more about what they report at the end of the year. Thus some firms skip some quarters without reporting their tax liabilities and submit their tax returns at the end of the year. Therefore, I use annual frequency to minimise the related noise in data.

[4]I do the same analysis using the unbalanced sample in Appendix B.2 and results are similar.

[5]The nominal values such as annual sales and (total) wages are in units of thousand USD in Table 2.1a. This is why the minimum value of annual sales and wages in the sample are shown as 0 in Table 2.1a (they are around $3).

[6]Since I use the balanced sample between 2013 and 2018, the minimum age for retailers in 2015 is 3.

Table 2.2: Firms' annual sales level and growth (by year)

	Sales (thousand USD)					
	min	mean	med	max	sd	count
2014	0.0004	359	7	251,708	5,918	3,622
2015	0.0004	304	9	170,578	4,529	3,622
2016	0.0001	311	12	173,159	4,327	3,622
2017	0.0004	368	14	216,005	5,111	3,622
2018	0.0000	486	17	320,580	7,099	3,622
Total	0.0000	365	12	320,580	5,491	18,110

(a) Flow variables

	Sales growth (%)					
	min	mean	med	max	sd	count
2014	-8.42	0.22	0.09	11.78	1.07	3,622
2015	-7.40	0.03	0.01	8.56	0.94	3,622
2016	-13.54	0.28	0.18	8.46	1.02	3,622
2017	-11.27	0.13	0.12	15.72	0.94	3,622
2018	-9.49	0.19	0.17	10.58	0.96	3,622
Total	-13.54	0.17	0.10	15.72	0.99	18,110

(b) Dummy variables

Note: Table 2.2a (Table 2.2b) summarises firms' annual sales level by year (sales growth rate). All nominal values are in thousand USD and growth rates are in percentages.

E-receipt program. I summarises firms' annual sales level and growth by year in Table 2.2. Table 2.2a shows summary statistics of sales level in thousand USD, Table 2.2b summarises sales growth in percentages. We can see that mean and median sales growth rates are the largest in 2016 (0.28% and 0.18% respectively).

2.3 Empirical analysis

2.3.1 Changes in retailers' reported sales

I use the timing of the E-receipt program to identify tax-evading retailers. In particular, I detect the retailers that reported an abnormally large growth in their sales in the year that the program was launched, which suggests that those firms had previously been evading more taxes.

To identify tax-evaders I run the following three regressions using the balanced sample consisting of retailers only.

1. Retailers that increase growth of sales ***in 2016 above trend***

$$\Delta ln(Y_{it}) = \gamma_i + \gamma_t + \beta_{16} \cdot D_{2016} \cdot Z_{it} + u_{it} \tag{2.1}$$

where i and t indicate firm and year, respectively. Y_{it} is annual sales and $\Delta ln(Y_{it})$ is sales growth rate of firm i in year t. D_{2016} is a dummy for year 2016 and Z_{it} represents observable firm characteristics of firm i in year t such as firm size, age and location. γ_i is the firm fixed effect and it controls for the firm-specific trend in sales growth. Similarly, γ_t is the year fixed effect and it controls for common trend across for all retailers each year. β_{16} is the coefficient of interest, which shows the correlation between firms' observable characteristics Z_{it} and the extent of tax evasion measured as sales growth in 2016 above its trend. I associate the sales growth of retailers in 2016 as measured by β_{16} as the degree of their tax evasion.[7]

2. Retailers that increase growth of sales ***after 2016 above trend***

$$\Delta ln(Y_{it}) = \gamma_i + Post_t + \beta_{post} \cdot Post_t \cdot Z_{it} + u_{it} \quad (2.2)$$

where $Post_t = 1$ if $t \geq 2016$, and equals zero otherwise. It controls for pre and post-intervention trend in firms' sales growth. β_{post} is the coefficient of interest, which shows the relation between firms' observable characteristics and the extent of their tax evasion measured as sales growth after 2016 above its trend. Thus, according to this specification I classify firms with higher growth rate after 2016 as tax evaders.

3. Retailers that increase growth of sales ***after 2016 compared to wholesalers***

$$\Delta ln(Y_{its}) = \gamma_i + Post_t + \beta \cdot Treat_s \cdot Post_t + \beta_{DiD} \cdot Treat_s \cdot Post_t \cdot Z_{it} + u_{its} \quad (2.3)$$

where s represents sector, which is either retail or wholesale. $Treat_s = 1$ if firm i is a retail firm, and zero otherwise. β_{DiD} is the coefficient of interest, which shows the relation between firms characteristics and the extent of firms' tax evasion measured as an increase in retailers' sales after 2016 compared to wholesalers.

It is important to acknowledge the potential caveats of these three specifications. The first one underestimates tax evasion because of the gradual enrollment of firms (and consumers) shown in Figure B1. The second approach corrects the issue of firms' gradual enrollment because it is using the changes in average growth rates pre- and post-intervention. However, it may classify the firms with accelerating growth in sales as tax evaders, hence it could result in overestimation of tax evasion. The third approach is subject to any potential violations to the parallel trend assumption between retailers and wholesalers discussed in chapter 1. For example, consumer reporting may affect wholesalers reported sales because some wholesalers sell to final consumers. Therefore, the third approach underestimates the extent of tax evasion of retailers.

Nevertheless, using the above three specifications I study the relationship between the extent of tax evasion measured as sales growth and firm characteristics summarised in Table 2.1. Specifically, I use firm age, different measures of firm size ($Sales_{15}$, $Sales_{pre}$,

[7]The identifying assumption is that firm-specific idiosyncratic shocks, u_{it}, are uncorrelated with observables such as firm size, age and location of the firms expressed by Z_{it}.

$\#workers_{15}$ and $Wages_{15}$), location, VAT status, and whether the firm issued any E-receipt between 2016 and 2018.

Table 2.3: Sales growth and firm characteristics

	$\Delta ln(Y_{it})$ — Sales growth in 2016					$\Delta ln(Y_{it})$ — Sales growth after 2016				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dummy*log($Sales_{pre}$)	-0.107***	-0.104***	-0.119***	-0.116***	-0.133***	-0.0902***	-0.0803***	-0.118***	-0.0907***	-0.130***
	(0.0105)	(0.0123)	(0.0165)	(0.0125)	(0.0166)	(0.0112)	(0.00987)	(0.0161)	(0.0101)	(0.0177)
Dummy*log(Age_{15})	0.120***				0.103**	0.204***				0.187***
	(0.0383)				(0.0436)	(0.0317)				(0.0293)
Dummy*Capital		0.130***			0.116***		0.0630***			0.0461**
		(0.0212)			(0.0217)		(0.0175)			(0.0208)
Dummy*VAT_{always}			0.113		0.0531			0.239***		0.185***
			(0.0718)		(0.0725)			(0.0475)		(0.0542)
Dummy*E-receipt				0.266***	0.250***				0.211***	0.149**
				(0.0663)	(0.0776)				(0.0490)	(0.0563)
Observations	18110	18110	18110	18110	18110	18110	18110	18110	18110	18110
Adjusted R^2	0.141	0.141	0.141	0.142	0.143	0.143	0.140	0.142	0.141	0.146

Note: Dependent variables are firms' annual growth in sales. The first five columns report results from the regression equations 2.1: Dummy variable interacted with firm characteristics variable is D_{2016}, which is a dummy for year 2016. The last five columns show results from equation 2.2: Dummy variable is $Post_t$, which equals one if the period is after 2016 and zero otherwise. Independent variable $Sales_{pre}$ is the average pre-intervention annual sales and Age_{15} is the number of years since the establishment date. Capital, VAT_{always} and E-receipt are dummy variables that indicate if the firm is located in the capital city, if the firm has been VAT-liable throughout the sample period, and the firm has ever issued any E-receipt, respectively. Regressions are weighted by sales at the beginning of the sample period, 2014. Standard errors are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I begin with the first two specifications in equations 2.1 and 2.2. The results are shown in Table 2.3. The first five columns report results from the regression equations 2.1: the dummy variable interacted with firm characteristics is D_{2016}, which equals one for year 2016. The last five columns show results from equation 2.2: dummy variable is $Post_t$, which equals one if the period is after 2016 and zero otherwise. Regressions are weighted by firms' sales at the beginning of the sample period, 2013. The first row shows that there is a negative relationship between sales growth and firm size measured as mean sales before the intervention. Specifically, the first 5 columns suggest that 1% increase in size leads to around 10 percentage points (p.p.) decrease in sales growth in 2016. Similarly, 1% increase in size results in 9 p.p. decrease in sales growth across the 3-year period following implementation the program as shown in the last 5 columns. As mentioned before I use a balanced sample. However, the fact that small firms have higher exit rate and higher growth could lead to overestimation of small firms' tax evasion. Therefore, I report results from unbalanced data in Table B1 and it gives similar results. Moreover, to check the robustness of the negative relation between size and tax evasion I use different measures of firm size in Table B2. I use sales in year 2014 and 2015 ($Sales_{14}$ and $Sales_{15}$), number of workers ($\#workers_{15}$) and total wages ($Wages_{15}$) at the end of 2015 as alternative size measures and they produce qualitatively the same results. These analyses suggest that smaller retailers experience higher growth in sales above its trend compared to larger retailers due to the E-receipt program. In other words, smaller retailers would have hidden more sales and evaded more taxes in the absent of E-receipt program.

Moreover, the second row in Table 2.3 also show that older firms evade tax more (after controlling for size). In particular, retailers' sales growth rate in 2016 increase by 10 p.p. as their age increase by 1%. Similarly, 1% increase in age lead to 20 p.p increase in sales growth in the three-year period after the E-receipt program was initiated. In addition, the third row suggests that firms in the capital have higher sales growth compared to firms in other regions. This relation survives even after controlling for Ulaanbaatar's distinctive and observable characteristics such as level of economic development, poverty and population size as shown in Table B3 and B4.[8] Last but not least, the last row reports that firms that issue E-receipt have higher growth in sales, which reassures the effect of the intervention.

Next, I divide the retailers into two groups: firms in Ulaanbaatar and firms in other regions. Then I run the regressions specified in equations 2.1 and 2.2 and results are reported in Table 2.4. The same pattern holds within as well as outside Ulaanbaatar: smaller and older retailers have higher sales growth.

Table 2.4: Firm size and tax evasion within and outside the capital city

	$\Delta ln(Y_{it})$ — Sales growth in 2016			$\Delta ln(Y_{it})$ — Sales growth after 2016		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy*log($Sales_{pre}$)	-0.127***	-0.140***	-0.139***	-0.128***	-0.131***	-0.132***
	(0.0261)	(0.0280)	(0.0282)	(0.0273)	(0.0267)	(0.0276)
Dummy*log(Age_{15})	-0.00948		-0.00794	0.186**		0.186**
	(0.139)		(0.138)	(0.0667)		(0.0653)
Dummy*E-receipt		0.345*	0.345*		0.233*	0.233*
		(0.194)	(0.195)		(0.123)	(0.119)
Observations	4760	4760	4760	4760	4760	4760
Adjusted R^2	0.136	0.137	0.137	0.138	0.137	0.139

(a) Within Ulaanbaatar

	$\Delta ln(Y_{it})$ — Sales growth in 2016			$\Delta ln(Y_{it})$ — Sales growth after 2016		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy*log($Sales_{pre}$)	-0.104***	-0.110***	-0.121***	-0.0826***	-0.0928**	-0.0951**
	(0.0331)	(0.0355)	(0.0372)	(0.0278)	(0.0336)	(0.0339)
Dummy*log(Age_{15})	0.143***		0.146***	0.0910**		0.0921**
	(0.0325)		(0.0316)	(0.0329)		(0.0330)
Dummy*E-receipt		0.234**	0.238**		0.182**	0.183**
		(0.0915)	(0.0962)		(0.0684)	(0.0691)
Observations	13350	13350	13350	13350	13350	13350
Adjusted R^2	0.145	0.145	0.146	0.145	0.145	0.146

(b) Outside Ulaanbaatar

Note: Dependent variables are firms' annual growth in sales. The first three columns report results from the regression equations 2.1: Dummy variable interacted with firm characteristics variable is D_{2016}, which is a dummy for year 2016. The last three columns show results from equation 2.2: Dummy variable is $Post_t$, which equals one if the period is after 2016 and zero otherwise. Independent variable $Sales_{pre}$ is the average pre-intervention annual sales and Age_{15} is the number of years since the establishment date. E-receipt is a dummy variable that equals one if a firm has ever issued any E-receipt. Regressions are weighted by sales at the beginning of the sample period, 2014. Standard errors are clustered at 4 digit industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, I turn to specification in equation 2.3, where I compare reported sales of retailers to wholesalers pre- and post-intervention using difference-in-difference (DiD) strategy. The results are reported in Table 2.5. The dependent variables in the first five columns are a

[8]There could be other unobservable characteristics of Ulaanbaatar that derives the relation. For example program campaign is more effective in Ulaanbaatar so that firms and consumers actively participate in the E-receipt program compared to those in other regions, resulting in higher sales growth for firms in the city.

log of sales and the last five columns use annual sales growth. The reason for using the log of sales is to make the analysis consistent and comparable with chapter 1. The first column shows that retailers' sales increase 29% compared to wholesalers after the policy intervention.[9] To see the heterogeneous effect of the intervention on firms with different characteristics I add terms where I interact firm characteristics with Treat*Post. Columns 2-6 suggest that smaller and older retailers' sales increase more compared to larger and younger retailers. From columns 8-12 we can see that smaller and older retailers have higher sales growth indicating larger tax evasion pre-intervention, which is consistent with the previous analysis in Table 2.3 and 2.4. Also it is verified that firms in Ulaanbaatar and that issue E-receipts experience higher growth in sales.

Table 2.5: Sales growth and firm characteristics — DiD

	$ln(Y_{it})$						$\Delta ln(Y_{it})$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Retail*Post	0.287***	1.842***	0.432***	1.847***	1.915***	1.992***	-0.0175	1.341***	-0.315***	1.112***	1.122***	1.151***
	(0.0693)	(0.512)	(0.0766)	(0.481)	(0.496)	(0.447)	(0.0203)	(0.166)	(0.0808)	(0.181)	(0.182)	(0.189)
Treat*Post*log($Sales_{pre}$)		-0.0898***		-0.0895***	-0.0934***	-0.127***		-0.0784***		-0.0902***	-0.0908***	-0.103***
		(0.0253)		(0.0272)	(0.0288)	(0.0280)		(0.00966)		(0.0112)	(0.0112)	(0.0113)
Treat*Post*log(Age_{15})			-0.0682*	-0.00500	-0.0274	-0.0247			0.140***	0.204***	0.200***	0.201***
			(0.0335)	(0.0399)	(0.0307)	(0.0310)			(0.0347)	(0.0317)	(0.0318)	(0.0305)
Treat*Post*Capital					0.179***	0.181***					0.0265	0.0271
					(0.0459)	(0.0473)					(0.0191)	(0.0183)
Treat*Post*E-receipt						0.572***						0.214***
						(0.0762)						(0.0501)
Observations	56895	56895	56895	56895	56895	56895	56895	56895	56895	56895	56895	56895
Adjusted R^2	0.845	0.845	0.845	0.845	0.845	0.846	0.105	0.106	0.105	0.106	0.106	0.106

Note: This table reports results from regressions specified in equation 2.3. The dependent variables in the first four columns are a log of sales and annual sales growth in the last four columns. Regressions are weighted by sales at the beginning of the sample period, 2013. Standard errors are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.3.2 Estimated tax evasion level of retailers

This subsection analyses dispersion in the extent of tax evasion across retailers and studies whether it is explained by retailers' observed characteristics. I start by estimating the degree of tax evasion for each retailer using the three different definitions of tax evasion specified in the previous subsection 2.3.1. Next, I associate the estimated tax evasion level to firm characteristics. Last, I analyse if the dispersion in estimated tax evasion level is explained by the observable characteristics.

[9]The coefficient is larger than what is found in chapter 1, which is around 20%. This difference is because I use balanced sample here in chapter 3 and unbalanced sample in chapter 1.

I use the following equations to estimate tax evasion level β_i for each retailer:

$$\Delta ln(Y_{it}) = \gamma_i + \gamma_t + D_{2016} \cdot \beta_{i,16} + u_{it} \tag{2.4}$$

$$\Delta ln(Y_{it}) = \gamma_i + Post_t + Post_t \cdot \beta_{i,post} + u_{it} \tag{2.5}$$

$$ln(Y_{its}) = \gamma_i + Post_t + Treat_s \cdot Post_t \cdot \beta_{i,DiD} + u_{its} \tag{2.6}$$

where subscripts i, t and s indicate firm, year and industry, respectively. Y_{it} is sales and $\Delta ln(Y_{it})$ is growth rate in sales of firm i in year t. D_{2016} is a dummy for year 2016, and $Post_t = 1$ if $t \geq 2016$, and equals zero otherwise. $Treat_s = 1$ if firm i is a retail firm, and zero otherwise. Equation 2.4 defines tax evaders as retailers with higher sales growth in 2016 above its trend. Similarly, equation 2.5 take retailers with high sales growth after 2016 as tax evaders. Equation 2.6 defines tax evaders as retailers with higher sales increase compared to wholesalers after the intervention. Equation 2.4 and 2.5 uses sample that consists of only retailers, while equation 2.6 uses sample of retailers and wholesalers. As before, I use balanced sample where I observe each firm throughout the period 2014 and 2018. $\beta_{i,16}$, $\beta_{i,post}$ and $\beta_{i,DiD}$ are the coefficient of interest, which represents firm i's tax evasion level according to each definitions of tax evaders.

I summarise the estimated coefficients, $\hat{\beta}_i$s in Table 2.6.[10] Table 2.6a shows that estimated $\hat{\beta}_i$s can be both positive and negative. Firms with positive $\hat{\beta}_i$s are classified as tax evaders since $\hat{\beta}_i$s are proxies for the extent of firms' tax evasion. Table 2.6b summarises the retailers where all three $\hat{\beta}_i$s are positive. Even though all three $\hat{\beta}_i$s aim to identify tax evaders, their definitions differ slightly. Therefore, I study their correlation in Table 2.7. The correlation coefficients suggest that they are positively correlated and the relation becomes stronger if I restrict the sample to retailers with only positive $\hat{\beta}_i$s as shown in Table 2.7b.

Table 2.6: Summary statistics of retailers' estimated tax evasion level, $\hat{\beta}_i$s

	min	mean	med	max	sd	count
$\hat{\beta}_{16}$	-16.53	0.06	-0.02	10.06	1.21	3,622
$\hat{\beta}_{post}$	-5.48	0.07	0.07	7.17	0.86	3,622
$\hat{\beta}_{DiD}$	-8.85	0.24	0.19	6.83	1.07	3,622

(a) All sample

	min	mean	med	max	sd	count
$\hat{\beta}_{16}$	0.00	0.98	0.63	10.06	1.22	1,076
$\hat{\beta}_{post}$	0.00	0.65	0.44	7.17	0.81	1,076
$\hat{\beta}_{DiD}$	0.00	1.02	0.79	6.83	0.90	1,076

(b) Positive $\hat{\beta}_i$s

Note: This table presents the summary statistics of retailers' $\hat{\beta}_i$s. In particular, the summary statistics in panel (a) contains all firms, and panel (b) focuses on the retailers with positive $\hat{\beta}_i$s.

[10] There is variation in the standard errors of the estimated coefficients. For simplicity, I focus on the estimated coefficients only.

Table 2.7: Correlation coefficients between $\hat{\beta}_i$s

	$\hat{\beta}_{16}$	$\hat{\beta}_{post}$	$\hat{\beta}_{DiD}$
$\hat{\beta}_{16}$	1		
$\hat{\beta}_{post}$	0.529***	1	
$\hat{\beta}_{DiD}$	0.519***	0.457***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(a) All sample

	$\hat{\beta}_{16}$	$\hat{\beta}_{post}$	$\hat{\beta}_{DiD}$
$\hat{\beta}_{16}$	1		
$\hat{\beta}_{post}$	0.814***	1	
$\hat{\beta}_{DiD}$	0.696***	0.594***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) Positive $\hat{\beta}_i$s

Note: This table presents the correlation coefficients of retailers' $\hat{\beta}_i$s. In particular, panel (a) shows the correlation coefficients of all retailers, and panel (b) focuses on the retailers with positive $\hat{\beta}_i$s.

Next, I associate retailers' tax evasion level, $\hat{\beta}_i$s, to their observable characteristics. In other words, I use $\hat{\beta}_i$s as dependent variables and run cross-sectional regression specified in equation 2.7.

$$\hat{\beta}_i = \alpha + \delta \cdot Z_i + u_i \tag{2.7}$$

where subscript i represents retailer. Z_{it} represents observable firm characteristics of retailer i such as firm size, age and location. The coefficient of interest is δ, which shows the relation between firm characteristics and the degree of tax evasion of retailers measured by $\hat{\beta}_i$.

Since the main dependent variables are estimated in equations 2.4, 2.5 and 2.6 and have their standard errors, I need to be careful when I interpret the results from equation 2.7.

First, I analyse the extensive margin — compare retailers with positive vs. negative $\hat{\beta}$s. I create dummy variables $\hat{\beta}_i^{dummy}$ that equals one if $\hat{\beta}_i$ is positive, zero otherwise. Then I regress $\hat{\beta}_i^{dummy}$ on firm characteristics and results are presented in Table 2.8. The dependent variables in the first six columns use the tax evasion definitions specified in equations 2.4, 2.5 and 2.6. The dependent variables in the last two columns are $\hat{\beta}_{All}^{dummy}$, which equals one if all three $\hat{\beta}_{16}$, $\hat{\beta}_{post}$ and $\hat{\beta}_{DiD}$ are positive at the same time, zero otherwise. The coefficients on the first row suggest that smaller retailers are more likely to have positive $\hat{\beta}$s and thus likely to be classified as tax evaders. The rows two and three imply that older retailers and retailers in Ulaanbaatar are more likely to be tax evaders, although the coefficients are not always significant.

Second, I analyse the intensive margin using the estimated $\hat{\beta}$ and the results are reported in Table 2.9. The results confirm that smaller and older retailers and retailers in Ulaanbaatar tend to evade tax more.

Table 2.8: Sales growth and firm characteristics — extensive margin

	$\hat{\beta}_{16}^{dummy}$		$\hat{\beta}_{post}^{dummy}$		$\hat{\beta}_{DiD}^{dummy}$		$\hat{\beta}_{All}^{dummy}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log(Sales_{pre})$	-0.0504***	-0.0516***	-0.0462***	-0.0464***	-0.0310***	-0.0317***	-0.0470***	-0.0488***
	(0.00514)	(0.00505)	(0.00806)	(0.00785)	(0.00724)	(0.00718)	(0.00566)	(0.00558)
$\log(Age_{15})$	0.0228	0.0181	0.107***	0.106***	-0.0199	-0.0227	0.0353**	0.0282*
	(0.0171)	(0.0174)	(0.0199)	(0.0206)	(0.0160)	(0.0165)	(0.0148)	(0.0160)
Capital		0.0365*		0.00621		0.0211		0.0547***
		(0.0180)		(0.0197)		(0.0177)		(0.0163)
Observations	3622	3622	3622	3622	3622	3622	3622	3622
Adjusted R^2	0.037	0.038	0.037	0.036	0.033	0.033	0.038	0.041

Note: Dependent variables are dummy variables $\hat{\beta}^{dummy}$ that equals one if $\hat{\beta}$ is positive, zero otherwise. Specifically, the first six columns use the tax evasion definitions specified in equations 2.4, 2.5 and 2.6. The dependent variables in the last two columns are $\hat{\beta}_{All}^{dummy}$, which equals one if all three $\hat{\beta}_{16}$, $\hat{\beta}_{post}$ and $\hat{\beta}_{DiD}$ are positive, zero otherwise. $Sales_{pre}$ is the average pre-intervention annual sales and Age_{15} is the number of years since the establishment date. Capital is a dummy variables that indicate if the firm is located in the capital city. Regressions are weighted by sales at the beginning of the sample period, 2014. Standard errors are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

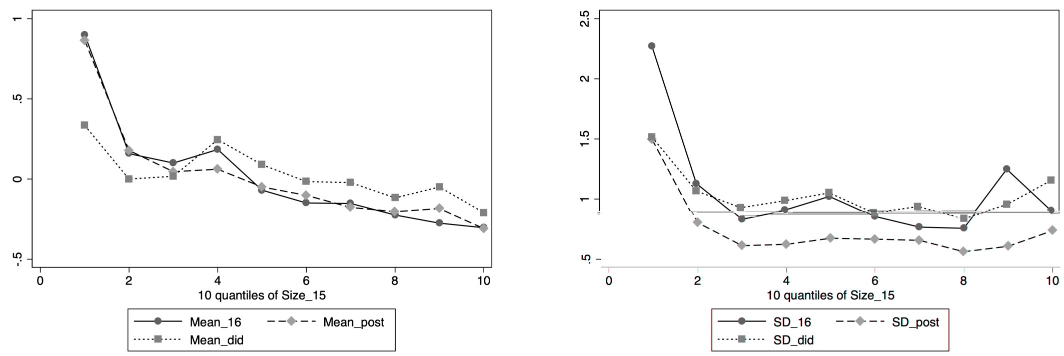
Table 2.9: Sales growth and firm characteristics — intensive margin

	$\hat{\beta}_{16}$		$\hat{\beta}_{post}$		$\hat{\beta}_{DiD}$	
	(1)	(2)	(3)	(4)	(5)	(6)
$\log(Sales_{pre})$	-0.131***	-0.135***	-0.113***	-0.114***	-0.0966***	-0.104***
	(0.0154)	(0.0145)	(0.0124)	(0.0120)	(0.0182)	(0.0184)
$\log(Age_{15})$	0.134***	0.117***	0.219***	0.213***	-0.00496	-0.0326
	(0.0311)	(0.0337)	(0.0287)	(0.0287)	(0.0314)	(0.0313)
Capital		0.132***		0.0461		0.212***
		(0.0331)		(0.0334)		(0.0488)
Observations	3622	3622	3622	3622	3622	3622
Adjusted R^2	0.044	0.046	0.073	0.073	0.042	0.050

Note: Dependent variables $\hat{\beta}_{16}$, $\hat{\beta}_{post}$ and $\hat{\beta}_{DiD}$ are determined in equations 2.4, 2.5 and 2.6, respectively. $Sales_{pre}$ is the average pre-intervention annual sales and Age_{15} is the number of years since the establishment date. Capital is a dummy variables that indicate if the firm is located in the capital city. Regressions are weighted by sales at the beginning of the sample period, 2014. Standard errors are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Last, I study the dispersion in $\hat{\beta}$ across different size groups of retailers. I calculate percentiles of size distribution of retailers using reported sales in 2015. Then calculate mean and standard deviation of $\hat{\beta}$ (after controlling for firm age and location) for each size percentile and plot them in Figure 2.1. We can see from Figure 2.1a that mean value of estimated $\hat{\beta}$ is decreasing in percentiles, confirming the negative relation between firm size and tax evasion. This implies that there is vertical inequality across firms in terms of their tax evasion level. Figure 2.1b plots standard deviation in the estimated $\hat{\beta}$ for each decile. It shows that dispersion in the estimated $\hat{\beta}$ decreases as the size decile increases. The lowest decile has the largest dispersion in $\hat{\beta}$ across firms in the decile. Moreover, each decile has positive standard deviation, which means that there is variation in tax evasion level across firms within each size groups. This suggests that there is horizontal inequality across firms in terms of their tax evasion level.

Figure 2.1: Mean and standard deviation of $\hat{\beta}$, by size percentiles



(a) Mean

(b) Standard deviation

2.4 Conclusion

This chapter utilises an interesting policy intervention in Mongolia, called E-receipt program, to study firms tax evasion behaviour. The E-receipt program gives incentives to final consumers to report their purchases and uses them as third-party reporters of firms sales. Therefore, this intervention reveals firms' hidden revenue and exposes associated tax evasion. In this paper, I use the timing of this policy intervention to identify tax-evading firms. Specifically, I classify firms with a notable increase in reported sales growth above their firm-specific trend in the year the E-receipt program was initiated as tax evaders. Then I associate firms' changes in sales growth to their observable characteristics. The empirical analysis shows that smaller firms and older firms evade tax more. I also find that firms in Ulaanbaatar, the capital city of Mongolia, evade more compared to firms in other regions. Moreover, firms that issue E-receipts experience higher sales growth, which reassures the effect of the E-receipt program.

References

Miguel Almunia and David Lopez-Rodriguez. Under the radar: The effects of monitoring firms on tax compliance. *American Economic Journal: Economic Policy*, 10(1):1–38, 2018.

Annette Alstadsæter, Niels Johannesen, and Gabriel Zucman. Tax evasion and inequality. *American Economic Review*, 109(6):2073–2103, 2019.

Montgomery B. Angell. Tax Evasion and Tax Avoidance. *Columbia Law Review*, 38(1):80, 1938.

R. J. Barro and X. Sala-I-Martin. Convergence. *Journal of Political Economy*, pages 223–251, 1992.

Robert J. Barro. Government Spending in a Simple Model of Endogeneous Growth. *Journal of Political Economy*, 98(5, Part 2):S103–S125, 1990.

Timothy Besley. Norms , Enforcement , and Tax Evasion. 2019.

Timothy Besley and Torsten Persson. Taxation and Development. *Handbook of Public Economics*, 5(December 2011):51–110, 2013.

Timothy Besley and Torsten Persson. Why do developing countries tax so little? *Journal of Economic Perspectives*, 28(4):99–120, 2014.

Travis Chow, Bin Ke, Hongqi Yuan, and Yao Zhang. What Types of Publicly Listed Firms Evade Taxes? Evidence from China. *SSRN Electronic Journal*, 2019.

Mark Dincecco and Gabriel Katz. State Capacity and Long-run Economic Performance. *Economic Journal*, 126(590):189–218, 2016.

Christoph Eichhorn. The Implications of Tax Evasion for Economic Growth. pages 1–20, 2004.

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

Michelle Hanlon, Lillian F. Mills, and Joel B. Slemrod. An Empirical Examination of Corporate Tax Noncompliance. *Cambridge: Cambridge University Press*, (1025), 2006.

Jarkko Harju, Tuomas Matikka, and Timo Rauhanen. Compliance costs vs. tax incentives: Why do entrepreneurs respond to size-based regulations? *Journal of Public Economics*, 173:139–164, 2019.

HM Revenue & Customs. *Tackling tax avoidance, evasion, and other forms of noncompliance*. Number March. 2019.

Internal Revenue Service. Tax Gap Estimates for Tax Years 2011–2013. (Rev.9-2019):2008–2010, 2019.

Anders Jensen. Employment structure and the rise of the modern tax system. *NBER WORKING PAPER SERIES*, 2019.

Andrew Johns and Joel Slemrod. The distribution of income tax noncompliance. *National Tax Journal*, 63(3):397–418, 2010.

Henrik Jacobsen Kleven, Martin Knudsen, Claus Thustrup Kreiner, Soren Pedersen, and Emmanuel Saez. Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, 79(3):651–692, 2011.

Henrik Jacobsen Kleven, Claus Thustrup Kreiner, and Emmanuel Saez. Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries. *Economica*, 83(330):219–246, 2016.

Wojciech Kopczuk and Joel Slemrod. Putting Firms into Optimal Tax Theory. *AEA PAPERS AND PROCEEDINGS*, 2006.

Adriana Veronica Litra. The Underground Economy in Romania. *Ovidius University Annals, Economic Sciences Series*, XVIII(2):173–176, 2018.

Li Liu, Ben Lockwood, Miguel Almunia, and Eddy Tam. VAT Notches, Voluntary Registration, and Bunching. *IMF Working Papers*, 19(205), 2019.

Joana Naritomi. Consumers as Tax Auditors. *American Economic Review*, 109(9):3031–3072, 2019.

Odd E. Nygård, Joel Slemrod, and Thor O. Thoresen. Distributional Implications of Joint Tax Evasion. *Economic Journal*, 129(620):1894–1923, 2019.

Dina Pomeranz. No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105(8):2539–2569, 2015.

Somchai Richupan. Measuring tax evasion. *Finance and Development*, 21(4):38–40, 1984.

Joel Slemrod. An Empirical Test for Tax Evasion. *The Review of Economics and Statistics*, 67(2):232–238, 1985.

Joel Slemrod. Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, 21(1):25–48, 2007.

Joel Slemrod. Tax compliance and enforcement†. *Journal of Economic Literature*, 57(4):904–954, 2019.

Joel Slemrod, Marsha Blumenthal, and Charles Christian. Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 79(3):455–483, 2001.

Joel Slemrod and Caroline Weber. Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, 19(1):25–53, 2012.

Joel Slemrod and Shlomo Yitzhaki. Tax avoidance, evasion, and administration. *Handbook of Public Economics*, 3(2):201–202, 2002.

Gabriel Tourek. Targeting in Tax Behavior : Evidence from Rwandan Firms. *Unpublished manuscript*, 2019.

Fangjun Wang, Shuolei Xu, Junqin Sun, and Charles P. Cullinan. Corporate Tax Avoidance: a Literature Review and Research Agenda. *Journal of Economic Surveys*, 0(0):1–19, 2019.

B Appendix

B.1 Firms enrollment in the E-receipt program

Figure B1: Number of retailers in E-receipt database

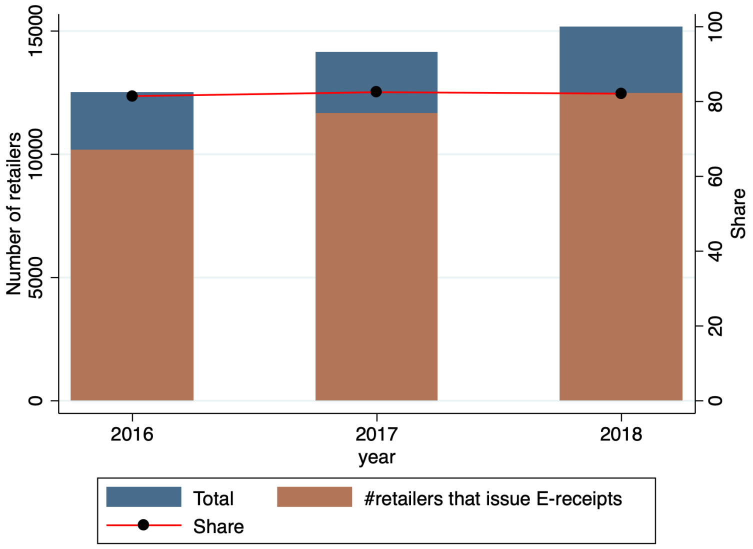


Figure B1 shows the total number of retailers that submit corporate income tax as well as the number and share of retailers that issue E-receipts. It illustrates a slight gradual enrollment of the retailers: the share of retailers were enrolled in the E-receipt program was 79% in 2016 and increased to 81% in 2018.

B.2 Robustness checks using unbalanced data

The main analysis uses a balanced data between 2014 and 2018. As a robustness check, I show the main results hold if I use unbalanced data. The sample size of the unbalanced data is 23,900 and it contains 6,800 retailers (balanced data size is 18,110 and 3,620 retailers). The results from the regressions specified in equations 2.1 and 2.2 using the unbalanced data are shown in Table B1 (similar to Table 2.3). Dependent variables are firms' annual growth in sales. The first four columns report results from the regression equations 2.1: Dummy variable interacted with firm characteristics variable is D_{2016}, which is a dummy for year 2016. The last four columns show results from equation 2.2: Dummy variable is $Post_t$, which equals one if the period is after 2016 and zero otherwise. All regressions are weighted by sales at the beginning of the sample period, 2014. Similar to the main analysis using balanced data, the results show that sales growth is negatively associated with firm size, positively with age. Moreover, retailers in the capital and retailers that issue E-receipts experience higher sales growth.

Table B1: Sales growth and firm characteristics using unbalanced data

	$\Delta ln(Y_{it})$ — Sales growth in 2016				$\Delta ln(Y_{it})$ — Sales growth after 2016			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dummy*log($Sales_{pre}$)	-0.107***	-0.0993***	-0.128***	-0.119***	-0.0923***	-0.0747***	-0.137***	-0.0913***
	(0.0121)	(0.0116)	(0.0202)	(0.0122)	(0.0111)	(0.00932)	(0.0146)	(0.00974)
Dummy*log(Age_{15})	0.174***				0.284***			
	(0.0456)				(0.0374)			
Dummy*Capital		0.162***				0.0574**		
		(0.0248)				(0.0259)		
Dummy*VAT_{always}			0.198**				0.379***	
			(0.0778)				(0.0424)	
Dummy*E-receipt				0.312***				0.244***
				(0.0659)				(0.0613)
Observations	23923	23923	23923	23923	23923	23923	23923	23923
Adjusted R^2	0.244	0.243	0.243	0.244	0.247	0.242	0.246	0.243

Note: Dependent variables are firms' annual growth in sales. The first four columns report results from the regression equations 2.1: Dummy variable interacted with firm characteristics variable is D_{2016}, which is a dummy for year 2016. The last four columns show results from equation 2.2: Dummy variable is $Post_t$, which equals one if the period is after 2016 and zero otherwise. Independent variable $Sales_{pre}$ is the average pre-intervention annual sales and Age_{15} is the number of years since the establishment date. Capital, VAT_{always} and E-receipt are dummy variables that indicate if the firm is located in the capital city, if the firm has been VAT-liable throughout the sample period, and the firm has ever issued any E-receipt, respectively. Regressions are weighted by sales at the beginning of the sample period, 2014. Standard errors are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.3 Other size measures

In the main analysis I use pre-intervention mean sales as a measure of firm size. In this section I use other size measures such as sales in year 2014 and 2015 ($Sales_{14}$) and $Sales_{15}$), number of workers ($\#workers_{15}$) and total wages ($Wages_{15}$) at the end of 2015. The results are shown in Table B2 and they confirm that there is a negative relationship between firm size and tax evasion.

B.4 Tax evasion and characteristics of Ulaanbaatar

Firms in Ulaanbaatar experience higher sales growth as shown in Table 2.3. This observation is not explained by observable characteristics of Ulaanbaatar. To see this, I add province-level economic development, poverty and population and firm densities as additional explanatory variables. The results are reported in Table 2.3 and they suggest a negative relationship between Capital dummy and sales growth survives. Moreover, I use the maximum level of night light of provinces as an alternative measure of economic development and results in Table B4 still suggest that firms in Ulaanbaatar have higher

Table B2: Sales growth and firm characteristics using different size measures

	$\Delta ln(Y_{it})$ — Sales growth in 2016			$\Delta ln(Y_{it})$ — Sales growth after 2016		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy*log($Sales_{14}$)	-0.0713***	-0.0686***	-0.0761***	-0.0461***	-0.0372***	-0.0427***
	(0.0100)	(0.0118)	(0.0120)	(0.00870)	(0.00864)	(0.00930)
Dummy*log(Age_{15})	0.0970**			0.174***		
	(0.0371)			(0.0324)		
Dummy*Capital		0.109***			0.0392**	
		(0.0212)			(0.0174)	
Dummy*E-receipt			0.191***			0.121**
			(0.0661)			(0.0563)
Observations	18110	18110	18110	18110	18110	18110
Adjusted R^2	0.138	0.138	0.138	0.138	0.136	0.136

(a) Sales in 2014

	$\Delta ln(Y_{it})$ — Sales growth in 2016			$\Delta ln(Y_{it})$ — Sales growth after 2016		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy*log($Sales_{15}$)	-0.172***	-0.168***	-0.185***	-0.150***	-0.140***	-0.155***
	(0.0139)	(0.0160)	(0.0160)	(0.0132)	(0.0123)	(0.0123)
Dummy*log(Age_{15})	0.167***			0.248***		
	(0.0415)			(0.0322)		
Dummy*Capital		0.166***			0.0968***	
		(0.0244)			(0.0224)	
Dummy*E-receipt			0.402***			0.337***
			(0.0693)			(0.0510)
Observations	18110	18110	18110	18110	18110	18110
Adjusted R^2	0.154	0.154	0.156	0.158	0.154	0.157

(b) Sales in 2015

	$\Delta ln(Y_{it})$ — Sales growth in 2016			$\Delta ln(Y_{it})$ — Sales growth after 2016		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy*log($\#workers_{15}$)	-0.0329	-0.0301	-0.0289	-0.0396**	-0.0201	-0.0215
	(0.0235)	(0.0243)	(0.0242)	(0.0141)	(0.0140)	(0.0136)
Dummy*log(Age_{15})	0.0527			0.196***		
	(0.0392)			(0.0344)		
Dummy*Capital		0.0463*			0.0446**	
		(0.0248)			(0.0198)	
Dummy*E-receipt			0.0584			0.125*
			(0.0752)			(0.0717)
Observations	12795	12795	12795	12795	12795	12795
Adjusted R^2	0.143	0.143	0.143	0.146	0.143	0.143

(c) Number of workers

	$\Delta ln(Y_{it})$ — Sales growth in 2016			$\Delta ln(Y_{it})$ — Sales growth after 2016		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy*log($Wages_{15}$)	-0.0319*	-0.0311	-0.0287	-0.0368***	-0.0236*	-0.0230*
	(0.0179)	(0.0202)	(0.0195)	(0.0112)	(0.0115)	(0.0113)
Dummy*log(Age_{15})	0.0558			0.198***		
	(0.0381)			(0.0349)		
Dummy*Capital		0.0567**			0.0540**	
		(0.0273)			(0.0196)	
Dummy*E-receipt			0.0606			0.128*
			(0.0789)			(0.0721)
Observations	12795	12795	12795	12795	12795	12795
Adjusted R^2	0.143	0.143	0.143	0.146	0.143	0.143

(d) Total wages

Note: Dependent variables are firms' annual growth in sales. The first three columns of each table report results from the regression equations 2.1: Dummy variable interacted with firm characteristics variable is D_{2016}, which is a dummy for year 2016. The last three columns show results from equation 2.2: Dummy variable is $Post_t$, which equals one if the period is after 2016 and zero otherwise. Regressions are weighted by sales at the beginning of the sample period, 2013. Standard errors are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

sales growth.

Table B3: Tax evasion and characteristics of Ulaanbaatar

	$\Delta ln(Y_{it})$ — Sales growth in 2016						$\Delta ln(Y_{it})$ — Sales growth after 2016					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dummy*log($Sales_{pre}$)	-0.110***	-0.109***	-0.109***	-0.109***	-0.109***	-0.109***	-0.0908***	-0.0905***	-0.0905***	-0.0904***	-0.0907***	-0.0897***
	(0.0107)	(0.0108)	(0.0108)	(0.0109)	(0.0110)	(0.0109)	(0.0112)	(0.0113)	(0.0113)	(0.0112)	(0.0113)	(0.0112)
Dummy*log(Age_{15})	0.106**	0.115**	0.116**	0.112**	0.111**	0.113**	0.200***	0.205***	0.203***	0.207***	0.207***	0.206***
	(0.0449)	(0.0482)	(0.0489)	(0.0467)	(0.0466)	(0.0473)	(0.0318)	(0.0294)	(0.0288)	(0.0320)	(0.0320)	(0.0312)
Dummy*Capital	0.110***	0.181***	0.197***	0.118***	0.110***	0.124***	0.0265	0.0600*	0.0490	0.121***	0.123***	0.0908**
	(0.0239)	(0.0554)	(0.0552)	(0.0225)	(0.0261)	(0.0365)	(0.0191)	(0.0338)	(0.0366)	(0.0215)	(0.0248)	(0.0371)
Dummy*GDPPC		-0.0986	-0.104	-0.192**	-0.194**	-0.203**		-0.0467	-0.0432	0.0433	0.0386	0.0265
		(0.0676)	(0.0652)	(0.0684)	(0.0721)	(0.0725)		(0.0458)	(0.0453)	(0.0512)	(0.0474)	(0.0600)
Dummy*Poverty			0.0540			0.0607			-0.0367			-0.00246
			(0.0489)			(0.0624)			(0.0575)			(0.0468)
Dummy*Pop dens.				0.0265**		-0.00986				-0.0255**		-0.0935
				(0.0109)		(0.0907)				(0.00970)		(0.0906)
Dummy*Firm dens.					0.0224**	0.0313					-0.0201**	0.0604
					(0.00986)	(0.0797)					(0.00802)	(0.0796)
Observations	18110	18110	18110	18110	18110	18110	18110	18110	18110	18110	18110	18110
Adjusted R^2	0.142	0.142	0.142	0.142	0.142	0.142	0.143	0.143	0.143	0.144	0.144	0.144

Note: Dependent variables are firms' annual growth in sales. The first six columns of each table report results from the regression equations 2.1: Dummy variable interacted with firm characteristics variable is D_{2016}, which is a dummy for year 2016. The last six columns show results from equation 2.2: Dummy variable is $Post_t$, which equals one if the period is after 2016 and zero otherwise. Regressions are weighted by sales at the beginning of the sample period, 2013. Standard errors are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B4: Ulaanbaatar — night light as a measure of economic development

	$\Delta ln(Y_{it})$ — Sales growth in 2016						$\Delta ln(Y_{it})$ — Sales growth after 2016					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dummy*log($Sales_{pre}$)	-0.110***	-0.111***	-0.111***	-0.112***	-0.112***	-0.112***	-0.0908***	-0.0910***	-0.0913***	-0.0897***	-0.0899***	-0.0895***
	(0.0107)	(0.0103)	(0.0103)	(0.0103)	(0.0103)	(0.0104)	(0.0112)	(0.0114)	(0.0114)	(0.0116)	(0.0116)	(0.0113)
Dummy*log(Age_{15})	0.106**	0.116**	0.116**	0.112**	0.112**	0.112**	0.200***	0.202***	0.203***	0.207***	0.207***	0.206***
	(0.0449)	(0.0473)	(0.0472)	(0.0465)	(0.0464)	(0.0461)	(0.0318)	(0.0317)	(0.0320)	(0.0322)	(0.0322)	(0.0322)
Dummy*Capital	0.110***	0.156***	0.156***	0.0885*	0.0766	0.0488	0.0265	0.0334	0.0331	0.127***	0.132***	0.0943**
	(0.0239)	(0.0315)	(0.0315)	(0.0468)	(0.0535)	(0.0824)	(0.0191)	(0.0393)	(0.0325)	(0.0282)	(0.0319)	(0.0444)
Dummy*NL_{max}		-0.131*	-0.135	-0.182*	-0.196*	-0.254		-0.0198	-0.0606	0.0510	0.0606	0.00940
		(0.0740)	(0.0929)	(0.105)	(0.110)	(0.154)		(0.0878)	(0.0823)	(0.0774)	(0.0843)	(0.0896)
Dummy*Poverty			-0.00567			-0.0436			-0.0632			0.00442
			(0.0691)			(0.0861)			(0.0626)			(0.0544)
Dummy*Pop dens.				0.0174		-0.0314				-0.0242**		-0.0978
				(0.0141)		(0.128)				(0.00917)		(0.0915)
Dummy*Firm dens.					0.0165	0.0474					-0.0204**	0.0660
					(0.0125)	(0.115)					(0.00850)	(0.0801)
Observations	18110	18110	18110	18110	18110	18110	18110	18110	18110	18110	18110	18110
Adjusted R^2	0.142	0.142	0.142	0.142	0.142	0.142	0.143	0.143	0.143	0.144	0.144	0.144

Note: Dependent variables are firms' annual growth in sales. The first six columns of each table report results from the regression equations 2.1: Dummy variable interacted with firm characteristics variable is D_{2016}, which is a dummy for year 2016. The last six columns show results from equation 2.2: Dummy variable is $Post_t$, which equals one if the period is after 2016 and zero otherwise. Regressions are weighted by sales at the beginning of the sample period, 2013. Standard errors are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Resource Misallocation and Learning-by-Doing

This paper investigates a trade-off between static and dynamic optimality conditions for resource allocation across firms in the presence of learning-by-doing (LBD). The standard static efficiency requires firms to have the same marginal revenue products (MRP) within each sector. In contrast, I show theoretically that dynamic efficiency condition implies dispersion in the MRP across firms when productivity growth is endogenous due to LBD. I then compare the implications of the dynamic and static models quantitatively using firm-level panel data from Indonesia. I show that firms' productivity growth is consistent with LBD, whereby small and younger firms have lower productivity, but higher productivity growth compared to larger and older firms. I simulate the dynamic model and find that aggregate productivity is higher in the long run when we allow for some dispersion in MRP.

3.1 Introduction

It is well documented that firms in narrowly defined industries vary a lot in terms of their size and productivity (for example Foster et al. (2001), Syverson (2010)). The way resources are allocated across those heterogeneous firms both at the extensive and intensive margins have an important effect on the aggregate productivity and growth (for example Hsieh and Klenow (2009), Restuccia and Rogerson (2017)). There is considerable consensus that, at the extensive margin, the lowest productive firms should exit the market (for example Aghion and Howitt (1992), Caballero et al. (2008) and Kwon et al. (2015)). As for the intensive margin, high-productive firms should use more resources and produce more than lower productive firms (Bailey et al. (1992), Foster et al. (2008), Syverson (2010), Foster et al. (2017), and Haltiwanger (2016)). These allocative efficiency conditions require equalised marginal revenue products (MRPs) across firms when there are diminishing marginal products. Consequently, it is thought that any observed dispersion in MRPs is a sign of misallocation of resources in the economy (for example Hsieh and Klenow (2009)).

This paper, however, challenges the idea that it is optimal to eliminate or allocate less resources to firms with currently low-productive firms. Some of the currently low-productive firms might have a capacity of reaching high productivity in the future, for example through learning-by-doing (LBD) mechanism. Therefore firms' future productivity prospects should be taken into consideration. Specifically, this paper studies the implications of internalising future endogeneous productivity path into firms optimization problem on allocation of resources across firms.

First, I study if firms' productivity process show any signs of LBD mechanism. To do so, I utilise the Indonesian manufacturing firm-level data and estimate several measures of firm-level revenue and physical productivity. The estimated productivity measures show that smaller and younger firms have lower productivity level but their productivity growth is higher than that of larger and older firms. This pattern holds even after controlling for survival bias of smaller and younger firms.[1] A possible explanation for this empirical pattern is that small and young firms improve their productivity as they accumulate experience by producing more, which is the LBD mechanism. More importantly, it suggests diminishing LBD effect on productivity — the incremental rise in future productivity decreases as the firm becomes larger and older.

[1]It is well documented that exit rates of small and young firms are higher than large and old firms.

Next, to highlight the importance of the inter-temporal link in productivity through LBD mechanism, I build a dynamic, partial equilibrium model. Specifically, I incorporate the LBD mechanism into a standard partial equilibrium model discussed in Hsieh and Klenow (2009), where monopolistically competitive firms maximise their period profits by taking their demand into account. The LBD effect is modelled in a way that future productivity of firms depends on the current productivity level as well as the current production. That is, firms can increase their future productivity by producing more but the increase is decreasing in current productivity level. Hence, firms maximise their present discounted future profits by internalising the inter-temporal link in productivity through the LBD mechanism. I show that the dynamic optimality conditions require dispersion in MRPs as long as firms are heterogeneous in terms of their current productivity level. The intuition is that low-productive firms produce more and have lower MRPs in order to take advantage of the future high productivity through the LBD mechanism. In comparison, in the absence of the LBD mechanism firms maximise only their period profits. The static optimality conditions require firms' MRPs to be equalised within an industry as in the standard misallocation literature such as Hsieh and Klenow (2009). That is, the currently high-productive firms use more resources and produce more than currently low-productive firms to reach the static efficiency. This result highlights that some dispersion in MRPs could be beneficial in the long run and it is not always the case that all observed dispersion in MRPs is due to distortions.

To see the implications in the long run I simulate both the static and dynamic models over time. I feed both models with the same arbitrary productivity distribution and compare their transitional dynamics and the speed of reaching the steady state. Both economies start with firms, who are heterogeneous in terms of their initial productivity level, and converges to the steady state, where all firms reach the highest possible productivity level in the productivity distribution. Even though firms in the repeated static model do not internalise the LBD effects their productivity in the next period increases because of the LBD effect. In contrast, in the dynamic model, lower productive firms produce more to utilise the LBD effects and thus their productivity increases faster and reach the steady state faster than the static model.

This paper relates to several strands of literature. First, it contributes to misallocation literature by showing a mechanism/channel that justifies some degree of dispersion in MRPs even when there is no distortion. There are other literature that has identified other possible reasons that lead to heterogeneous MRPs in the no-distortion economy. For example, overhead input requirements discussed in Bartelsman, Scarpetta and Haltiwanger

(2013), and quadratic preferences and linear demand schedules discussed in Melitz and Ottaviano (2008) lead to dispersion in MRPs. I add to this literature by showing that the LBD mechanism justifies a certain level of dispersion in MRPs as potentially productive small and young firms should have lower MPRs than older and larger firms.

Next, this paper relates to the productivity estimation literature. In particular, I estimate several measures of firm-level productivity and study their relationship. Estimating productivity at the firm level has always been a challenge because firm-level prices and quantities are not widely available. The most commonly used productivity measure is revenue-based productivity, $TFPR$, where industry-level prices are used to deflate firms' sales. Therefore, the estimated $TFPR$ reflects not only the physical productivity of firms, $TFPQ$, but also firm-level price heterogeneity and it is affected by the demand structure that firms face. Broadly, there are mainly two ways to estimate $TFPR$. The first approach is to assume Cobb-Douglas revenue function and use industry-level cost shares of input expenditures as input elasticities. The second approach is to estimate the elasticities using regression techniques.

A few papers that have access to firm-level prices calculate $TFPQ$. In particular, Foster et al. (2008), Foster et al. (2015) study commodity-like manufacturing sectors, such as ready-mixed concrete and raw cane sugar. In this paper, I estimate two revenue-based firm-level productivity estimates, $TFPR$s: one is based on Cobb-Douglas production function, and the other is based on production/productivity estimation method discussed in Wooldridge (2009). I also calculate three physical productivity measures using firm-level prices: two of them are the ratio of the $TFPR$s and prices, and the last one uses the structural assumptions used in Hsieh and Klenow (2009), such as Cobb-Douglas production function.I study the relationship between these productivity measures and find that all of them are positively correlated. This is consistent with Foster et al. (2017), where they find that $TFPR$s based on the Cobb-Douglas function assumption and regression approaches have similar properties and are highly correlated. Using the firm-level prices of firms in Indonesian manufacturing sector I estimate not only $TFPR$s but also $TFPQ$s. I find all productivity measures are positively correlated.

Lastly, this paper is related to a large body of literature on firm-level productivity. As discussed in the survey paper Syverson (2010), there are external and internal within-firm factors that influence firm-level productivity. External factors include, for example, trade liberalization (e.g. Pavcnik (2002), Eaton and Kortum (2002) and Melitz (2003)), deregulation (for example, Bridgman et al. (2009)) and FDI-driven productivity spillovers

(Keller and Yeaple (2009), Arnold and Javorcik (2009), Fernandes and Isgut (2015)). Especially, Arnold and Javorcik (2009) examine how FDI affect plant-level productivity using the Indonesian Census of Manufacturing between 1983 and 2001. Their results suggest that FDI leads to significant productivity improvements and its effects are observed in the consequent periods.

On the other hand, the internal within-firm elements that affect firm-level productivity include managerial practice (e.g., Bloom and Van Reenen (2014)) and IT and R&D (for example, van Ark et al. (2008), Doraszelski and Jaumandreu (2009), Aw et al. (2011)). LBD mechanism, which is the main mechanism discussed in this paper, is considered one of such internal factors. I document that the productivity process of firms in Indonesia exhibits a LBD mechanism. Specifically, using the productivity estimates mentioned above I find that young and small firms have lower productivity level but the productivity increases as the firms become larger and older. This is consistent with the traditional LBD literature such as Wright (1936), Arrow (1962) and David (1973) document empirically that firms' unit cost (productivity) is negatively associated with cumulative output. Also, there is a large body of evidence that firm size increases with age and this pattern is attributed to learning in young firms (Dunne et al. (1989), Baldwin et al. (2000)). However, more recent literature, such as Baily et al. (1992), Bartelsman and Dhrymes (1998), study firms in the U.S. and find that younger firms have higher productivity, which differs from my finding. One possible reason is that these studies focus on the U.S, a developed country, while I use Indonesian manufacturing firms. Moreover, Jensen et al. (2001) finds that younger firms adopt more modern and productive technology than the average incumbents, but also, older firms are more down on the learning curve. Jensen et al. (2001) claims that these two channels have similar effects, hence young and old firms do not differ much in terms of their productivity level.

The rest of the paper is organised as follows. Section 3.2 provides empirical analysis, where I explain the Indonesian data, estimate firm-level productivity and document LBD mechanism in the estimated productivity measures. Section 3.3 describes the theoretical framework, where I discuss the implication of introducing LBD mechanism into a standard partial equilibrium model. Section 3.4 concludes.

3.2 Empirical Analysis

In this section I study firms' productivity process whether it exhibits signs of learning-by-doing (LBD). First, I estimate firms' productivity level. Then I study its evolution over time. To do so, I use Indonesian firm-level annual manufacturing census between 1990-2010, which covers firms with 20 or more employees and has information on their sales, capital and labour inputs to calculate the revenue productivity measures.

Measuring firm-level productivity is not straightforward. Numerous methods of estimating productivity are proposed in the literature, and a large fraction of existing work focuses on estimating revenue-based productivity of firms using the value of sales because physical quantities are seldom observed in typical firm-level datasets.

In this paper, I estimate not only revenue-based productivity but also physical productivity measures for each firm using product-level price data from Indonesia, where I observe firms' sales and physical output of each product. In particular, I estimate two different measures of revenue-based productivity: one is simple revenue productivity based on Cobb-Douglas production function assumption, and the other one is estimated following a method proposed in Wooldridge (2009)[2] Moreover, I estimate three versions of physical productivity using product-level price data. The first one is the simple physical productivity based on the Cobb-Douglas production function, where I divide the estimated simple revenue productivity by prices. Similarly, the second measure is the ratio between Wooldridge revenue productivity and prices. The last measure is borrowed from Hsieh and Klenow (2015), which uses functional form assumptions to calculate physical productivity of firms. I adopt Hsieh and Klenow (2009) approach, partly because I compare model implication with their result in Section 3.3. Based on these five productivity measures, I examine firms' productivity process differ across firms with different characteristics such as age and size.

3.2.1 Data

The main data I use in this paper is an Indonesian firm-level panel data from the Manufacturing Survey of Large and Medium-Sized Firms (Statistik Industri).[3] Statistik Industri

[2]This estimator is more efficient compared to semi-parametric approaches introduced by Olley and Pakes (1996), Levinsohn, Petrin (2003) and Ackerberg et al. (2006). Also this method doesn't use bootstrapping techniques to obtain standard errors for the estimates. See Van Beveren (2010) for a detailed comparison of TFP estimation methods.

[3]The Statistik Industri has also been used in several studies, please see Amiti and Konings (2007), Blalock et al. (2008), Yang (2012) and Peters(2013) for detailed description of the dataset.

encompasses all manufacturing firms with twenty or more employees on an annual basis between 1990 and 2010, and contains information on firms' sales, value-added, capital stock, number of workers, total wages, material costs and other firm characteristics. To exclude the outliers, I winsorise the main variables such as firms' sales, value-added, capital stock, number of workers, total wages, material costs, at 1% in each tail for each year and 5 digit industry. I deflate nominal variables using industry level price deflators. Specifically, I use U.S. CPI (base year 1990) to deflate firms' sales, value-added, total wages, and material costs; and U.S. price level of the capital stock (base year 1990) to deflate value capital stock.[4] The final unbalanced full sample has 296,300 observations and 44,000 unique firms.[5]

More importantly, I complement the firm-level data with product-level data where I observe product prices. In particular, I have information on product name, 9-digit ISIC.Rev.3 industry code, quantity sold, unit of measurement, monetary value of sales for each product that a firm sells in a year. Product name is very detailed and it allows me to distinguish similar products such as palm kernel, crude palm oil, crude palm kernel oil, which have different 9-digit industry codes. However, this price data covers the period between 1994 and 2010. I winsorise the sales value and quantity of each type of product at 1% and 99% each year and deflate sales value using U.S. CPI (base year 1990). The price sample size is 452,000 and it contains 37,300 unique firms.

To calculate physical productivity, I focus on single-product firms. Single-product firms are defined as firms that produce only one good, which is measured by product name and 9-digit industry code, each year. In principle, single-product firms can switch its product over time.[6] On average 35% of the firms are single-product firms each year. They are younger and smaller compared to multi-product firms. Table 3.1 presents summary statistics. These datasets allow me to estimate not only revenue-based productivity but also physical productivity of firms each year.

[4] U.S. price deflators are downloaded from Penn World Table 9.0.

[5] Unfortunately, the data does not contain information on capital stock for the year 2006. Even though I calculate the value of capital stock in 2006 as the average of the capital stock in 2005 and 2007, the final sample contains 40% fewer observations for 2006: 8,300 observations compared to 14,100, which is the average number of observations.

[6] On average, firms produce 2 different types of products over the sample period. For reference, the largest number of products produced by single firm in the sample period is 14.

Table 3.1: Summary statistics: Multi vs. single-product firms

	Multi	Single	All
Age	17	16	17
Sales	38,125	13,382	29,352
Observations	150,189	82,514	232,703

Note: Table 3.1 summarises age and size of firms. The sample period covers 1994 and 2010. Column 1 and 2 represents multi- and single-product firms, respectively. Column 3 shows average age and size of all firms. Size is annual sales value of firms, measured in 1000 Indonesian rupiah.

3.2.2 Productivity estimation

I estimate five different measures of firm-level productivity: two revenue-based firm-level productivity estimates, $TFPR$s, and three physical productivity measures, $TFPQ$s, for each firm each year.[7]

A few remarks about distinction between $TFPR$ and $TFPQ$ are useful here. First, $TFPR$ is the most commonly used firm-level productivity measure in the literature, which is based on firm nominal revenue divided by industry-level price deflator. This implies firm-level prices are captured in firm-level productivity $TFPR$. In other words, this revenue-based productivity reflects both physical productivity $TFPQ$ and demand structure of the firm. On the other hand, firm-level prices make it possible to calculate $TFPQ$, which more closely corresponds to notion of productivity that reflect firms' ability to turn production inputs to output. Under constant returns to scale $TFPQ$ equals the ratio between $TFPR$ and firm-level prices. One of the drawback of this physical-productivity does not reflect quality of products. Quality of the products are manifested in firm-level prices, in this sense $TFPR$ can be more suitable for representing firms' productivity. Therefore, I use both types of firm-level productivity measures to test if they exhibit LBD process. When estimating firm-level productivity, I assume that firms in the same industry face the same production function when I estimate firm-level productivity below.

I introduce the five productivity measures below. The first revenue-based productivity is based on Cobb-Douglas production function assumption, where the labour share is assumed to be equal to U.S. labour share in each industry.[8] In particular, I estimate the

[7] Notations $TFPR$ and $TFPQ$ are borrowed from Foster et al. (2008).

[8] In other words, labour share is allowed to vary across industries, but not within an industry. Moreover, the reason I use U.S. labour shares instead of calculating labour shares Indonesian data is because it is considered that distortions are potentially substantial in Indonesia (the same as in Hsieh and Klenow (2009)) and it is not possible to separately identify industry-level labour distortion and labour elasticity in each industry.

productivity as the ratio between firms' value-added and their composite inputs:

$$TFPR_{simple,it} \equiv \frac{P_{it}Y_{it}}{K_{it}^{\alpha}L_{it}^{1-\alpha}} \tag{3.1}$$

Here, $P_{it}Y_{it}$ is value-added for firm i in time t. I use book value of total fixed assets for K_{it}. As the labour inputs, L_{it}, I use total wage bill which includes both wages and benefits to the workers. $1-\alpha$ is the labour share, which is calculated based on NBER-CES Manufacturing Industry Database for each 4-digit industry.[9]

For the second revenue productivity measure, one of the widely used revenue-based productivity measure in the literature. I use Wooldridge (2009) methodology to control for endogeneity and selection bias problems when calculating firms' productivity level, $TFPR_{W,i}$. Wooldridge (2009) shows that a one-step generalised method of moments (GMM) can generate a productivity estimate that is not only consistently but also efficiently.[10]

Next, I estimate three versions of physical productivity. The first two measures of physical productivity takes advantage of the price and physical quantity data, which I observe at the product level. Namely, the first one is the simple physical productivity based on Cobb-Douglas production function, where I divide the estimated simple revenue productivity $TFPR_{simple}$ by product price for each single-product firms and year:

$$TFPQ_{simple,it} \equiv \frac{Y_{it}}{K_{it}^{\alpha}L_{it}^{1-\alpha}} = \frac{TFPR_{simple,it}}{P_{it}} \tag{3.2}$$

The second measure is calculated by dividing the Wooldridge revenue productivity by the single-product firms' prices:

$$TFPQ_{W,it} \equiv \frac{TFPR_{W,it}}{P_{it}} \tag{3.3}$$

The last measure is based on Hsieh and Klenow (2009), which uses some strong assumptions about firms' demand structure on top of supply side assumptions. Specifically, firms face an isoelastic residual demand curve and their marginal cost curves are both flat (invariant to quantity) and are negative unit elastic with respect to TFPQ. I adopt Hsieh

[9] I use the NBER-CES Manufacturing Industry Database from 1990 to 2010, to calculate the labour shares for each industry each year. The reason for I use U.S. labour share from NBER-CES is that U.S. is considered to be relatively undistorted economy compared to developing countries such as Indonesia. Similar argument is made in Hsieh and Klenow (2009), where they use the U.S. labour share to calculate productivity of firms in China and India.

[10] The estimator is more efficient compared to semi-parametric approaches introduced by Olley and Pakes (1996), Levinsohn, Petrin (2003) and Ackerberg et al. (2006). Also this method doesn't use bootstrapping methods to obtain standard errors for the estimates. See Van Beveren (2010) for detailed comparison of TFP estimation methods.

and Klenow (2009) approach, mainly because I compare model implication with their result, where it is assumed that firms with high real output have a lower price to explain why buyers would demand the higher output. Therefore, output in physical unit can be inferred from value-added of a firm, by using demand elasticity σ, which equals 3 as in Hsieh and Klenow (2009).

$$TFPQ_{HK,it} \equiv \frac{Y_{it}}{K_{it}^{\alpha} L_{it}^{1-\alpha}} = \frac{(P_{it}Y_{it})^{\frac{\sigma}{\sigma-1}}}{K_{it}^{\alpha} L_{it}^{1-\alpha}} \tag{3.4}$$

3.2.3 Learning-by-doing in firm-level productivity

This subsection tests if firms' productivity level, measured by the previous five measures, exhibit LBD process. In particular, I test if young and small firms have lower level of productivity but have higher productivity growth.

I start by examining the relation how they relate to one another. For single-product firms I use five productivity measures: $TFPR_{simple}$, $TFPR_W$, $TFPQ_{simple}$, $TFPQ_W$ and $TFPQ_{HK}$. However, for multi-product firms (as well as single-product firms), I calculate only the following three productivity measures: $TFPR_{simple}$, $TFPR_W$ and $TFPQ_{HK}$. This is because each firm produces more than one product and it is not clear which price should be used to deflate the revenue productivity measures.

Table 3.2: Correlation coefficients between productivity measures

	$TFPR_{simple}$	$TFPR_W$	$TFPQ_{HK}$
$TFPR_{simple}$	1		
$TFPR_W$	0.00984***	1	
$TFPQ_{HK}$	0.793***	0.0563***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(a) Full sample

	$TFPR_{simple}$	$TFPQ_{simple}$	$TFPR_W$	$TFPQ_W$	$TFPQ_{HK}$
$TFPR_{simple}$	1				
$TFPQ_{simple}$	0.0000832	1			
$TFPR_W$	0.00128	0.000530	1		
$TFPQ_W$	0.0000383	0.993***	0.0280***	1	
$TFPQ_{HK}$	0.994***	0.0000929	0.0237***	0.000510	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) Single-product firms

The correlation coefficients of the productivity measures are presented in Table 3.2. From Table 3.2a we can see that correlation between $TFPR_{simple}$ and $TFPQ_{HK}$ are strong, and it is due to their functional form. Relation of $TFPR_W$ to other productivity measures are weak. For single-product firms, again $TFPR_{simple}$ and $TFPQ_{HK}$ are strongly correlated. Similarly, $TFPQ_{simple}$ and $TFPQ_W$ have strong positive relations. Other productivity measures have weak positive correlations.

To see whether the small and young firms have lower productivity level I run the following regressions:

$$log(TFP_{sit}) = \beta_0 + \beta_1 log(Char_{sit}) + \delta'\mathbf{Z} + \mu_s + \mu_p + \mu_l + \mu_t + \varepsilon_{sit} \tag{3.5}$$

Here $log(TFP_{sit})$ is the log of productivity level of firm i in sector s at time t. $log(Char_{sit})$ represents firms' size and age. I use volume of sales to measure firms' size because I use value-added to calculate the $TFPR$ and $TFPQ$. It is worth noting that $log(Sales_{sit})$ is not cumulative past production, which is often used in the LBD literature. Instead, it is the current revenue which is closely linked to the theoretical part in Section 3.3. For alternative measures of size, I use average number of workers per working day as well as value-added, and the results are qualitatively similar. $\mathbf{Z}$ denotes a vector of firm characteristics, which includes export status, capacity utilization, capital intensity, whether the firm has any investment financed through equity issue, by foreign loans, by FDI, and any investment financed on capital markets. μ_s is a vector of industry fixed effects which are included in order to control for all time-invariant industry characteristics, μ_p is a vector of province fixed effects, which control for time-invariant location specific characteristics, μ_l is a vector of legal status fixed effects (for example whether the firm is a state enterprise or a limited liability company), and μ_t is a vector of time dummies, included in order to control for all factors affecting all firms in the same way in a given year. Moreover I cluster the error terms at industry and province level since firms in the same industry and province are likely to be correlated with each other.[11]

The main interest is the sign and statistical significance of β_1. Notice that regressions results in this paper does not imply causations. It only shows the correlation between dependent variable and the right hand side variables because there is no exogenous variation in the explanatory variables. Specifically, β_1 captures the correlation between firm characteristics, such as firm size, and its productivity level controlling for other firm characteristics.

Table 3.3 presents the regression using the full, unbalanced sample. The results imply that

[11] If I cluster the disturbances at industry and regency, industry and sub-regency, and industry and village level, significance levels are roughly the same and the results survive.

Table 3.3: Productivity level - full, unbalanced sample

	(1) $log(TFPR_{simple})$	(2) $log(TFPR_W)$	(3) $log(TFPQ_{HK})$
log(Sales)	0.214***	0.416***	0.673***
	(0.00907)	(0.00568)	(0.0123)
log(Age)	0.0505***	0.0258***	0.0396***
	(0.0114)	(0.00406)	(0.0121)
Observations	296324	296324	296324
Adjusted R^2	0.218	0.760	0.631

Note: This table presents regression results specified in equation 3.5 using full, unbalanced sample. The dependent variables in column 1, 2 and 3 are the simple revenue productivity defined in equation 3.1, revenue-based productivity measure based on Wooldridge (2009) estimation approach, and the log of physical productivity based on Hsieh and Klenow (2009) approach, defined in equation 3.4, respectively. I include year, industry, province and legal status fixed effects in all regressions. Control variables include firm characteristics such as export status, capacity utilization, capital intensity, whether the firm has any investment financed through equity issue, by foreign loans, by FDI, and any investment financed on capital markets. The standard errors are clustered at both year and industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

larger and older firms have higher productive level. This is true for all three measures of productivity.[12] It is worth mentioning the survival bias caused by firms' entry and exit. It is well documented that there is higher frequency of entrance and exits among smaller/younger firms and most of the firms that exit the market have lower productivity level.[13] Given these facts, the estimations above are likely to be underestimating the difference in productivity level between small/young and large/old firms, and that will strengthen the results further.

Next I turn to relation between productivity growth rate and firm characteristics. The Indonesian dataset has a panel dimension, which gives me an opportunity to analyse the dynamic characteristics of the firm productivity level. Productivity growth is calculated by taking the log differences between productivity level in the current and the last period.

The regression equation are the following:

$$gr_{sit+1} = \beta_0 + \beta_1 log(Char_{sit}) + \delta'\mathbf{Z} + \mu_s + \mu_p + \mu_l + \mu_t + \varepsilon_{sit} \tag{3.6}$$

This equation is exactly the same as regression (3.5), except that the dependent variable, gr_{sit+1}, is now growth rate of productivity level for firm i in sector s and at time t. The results are presented in Table 3.4. It shows that firm size and age are negatively correlated

[12] Table C1 shows regression results where I associate age and size of firms to their productivity separately. The results are qualitatively similar to Table 3.3.

[13] For example see Syverson (2011).

with productivity growth rate, which means productivity growth decreases as firm grow older and larger. In other words, small and young firms have higher productivity growth.[14]

Table 3.4: Productivity growth - full, unbalanced sample

	(1) $\Delta log(TFPR_{simple})$	(2) $\Delta log(TFPR_W)$	(3) $\Delta log(TFPQ_{HK})$
log(Sales)	-0.0496***	-0.0483***	-0.0835***
	(0.00219)	(0.00228)	(0.00350)
log(Age)	-0.0101***	-0.0146***	-0.0305***
	(0.00282)	(0.00257)	(0.00453)
Observations	252298	252298	252298
Adjusted R^2	0.025	0.028	0.035

Note: This table shows estimated coefficients from regression equation 3.6. The dependent variables in column 1, 2 and 3 are growth rates in the simple revenue productivity defined in equation 3.1, revenue-based productivity measure based on Wooldridge (2009) estimation approach, and the log of physical productivity based on Hsieh and Klenow (2009) approach, defined in equation 3.4, respectively. I include year, industry, province and legal status fixed effects in all regressions. Control variables include firm characteristics such as export status, capacity utilization, capital intensity, whether the firm has any investment financed through equity issue, by foreign loans, by FDI, and any investment financed on capital markets. The standard errors are clustered at both year and industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

It is important to consider the higher exit rate of small and young firms. As previously discussed, it is known in the literature that small and young firms exit more often that large and old firms. It is also documented that firms that exit the market have low productive level. Hence, it is possible that I overestimate the productivity growth rate of small and young firms because those with low productivity growth leave the market. Therefore, I study the extent of the potential selection issue of small and young firms. In particular, I examine if there is differential survival rates for young vs. old and small vs. large firms. Survival rate is defined as the share of firms appearing in the next year. Using this definition, I calculate survival rate for each quantiles of size and age distribution each year. Then I run survival rates on quantiles with and without year fixed effects and results are reported in Table 3.5. It shows that the higher quantiles in size and age distribution, the higher the survival rate. In Figure C1 I plot the survival rates of firms in different size and age quantiles and it also shows that firms in higher quantiles of both size and age distribution have (slightly) higher probability of appearing in the next year. For the exit rates of different size and age quantiles tells the same story. Table C3 show that rank of firms in age and size distributions is negatively correlated with exit rates. Similarly, Figure C2 shows that firms in lower age and size quantiles have higher exit rates compared to firms in higher quantiles.

[14] Table C2 shows regression results where I associate age and size of firms to their productivity separately. The results are qualitatively similar to Table 3.3 and imply that larger and older firms have lower

Table 3.5: Survival rates by size and age quantiles

	Size		Age	
	(1)	(2)	(3)	(4)
	Survival rate	Survival rate	Survival rate	Survival rate
Quartiles	0.0189**	0.0189***	0.00875	0.00875***
	(0.00854)	(0.00330)	(0.00778)	(0.00146)
Year FE	No	Yes	No	Yes
Observations	80	80	80	80
Adjusted R^2	0.053	0.899	0.003	0.972

Note: The dependent variables in the first 2 columns are survival rates of firms by size quantiles. The last 2 columns are survival rates of firms by age quantiles. In even columns I include year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Lower survival rate for young and small firms could result in over-estimation of productivity because lower productivity growth firms exit the market. To correct for the survival bias, I multiply the mean growth rates of each quartiles with its survival rates and study if young and small firms still have higher adjusted productivity growth compared to old and large firms. Table 3.6 reports the regression results and it confirms that smaller and younger firms have higher productivity growth rate compared to larger and older firms even after correcting for the survival bias.[15]

To stock, I documented that smaller and younger firms have lower productivity level but higher productivity growth rate compared to larger and older firms using several measures of firm-level productivity. This suggests that firms' productivity process exhibit signs of LBD mechanism. Therefore I incorporate LBD mechanism in the theoretical model in the next section and examine its implications on conditions that govern the efficient allocation of resources in the economy.

productivity growth.

[15]There could still be some concerns about my estimation of firms' productivity growth. For example, some large firms grow by buying small and productive firms. In this case, I over-estimate productivity level of large firms but under-estimate their productivity growth. However, I cannot test this because the available data does not have information about firms' merger and acquisitions.

Table 3.6: Productivity growth corrected by survival bias - full, unbalanced sample

	$\Delta log(TFPR_{simple})$		$\Delta log(TFPR_W)$		$\Delta log(TFPQ_{HK})$	
	(1) Size	(2) Age	(3) Size	(4) Age	(5) Size	(6) Age
log of sales	-0.0432***	-0.0429***	-0.0414***	-0.0419***	-0.0715***	-0.0722***
	(0.00185)	(0.00186)	(0.00196)	(0.00196)	(0.00298)	(0.00300)
lage	-0.00828***	-0.00874***	-0.0123***	-0.0117***	-0.0259***	-0.0252***
	(0.00244)	(0.00241)	(0.00225)	(0.00220)	(0.00393)	(0.00387)
Observations	252298	252298	252298	252298	252298	252298
Adjusted R^2	0.024	0.025	0.026	0.027	0.034	0.034

Note: This table shows estimated coefficients from regression equation 3.6. The dependent variables in odd (even) columns are productivity growth rates multiplied by survival rates of corresponding size (age) quantiles. The dependent variables in column 1and 2 are based on growth rates in the simple revenue productivity defined in equation 3.1. Columns 3 and 4 use revenue-based productivity measure based on Wooldridge (2009) estimation approach, and the last two columns use the log of physical productivity based on Hsieh and Klenow (2009) approach, defined in equation 3.4, respectively. I include year, industry, province and legal status fixed effects in all regressions. Control variables include firm characteristics such as export status, capacity utilization, capital intensity, whether the firm has any investment financed through equity issue, by foreign loans, by FDI, and any investment financed on capital markets. The standard errors are clustered at both year and industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.3 Theoretical Framework

This section presents a dynamic model where forward-looking firms maximise their present discounted future profits.

3.3.1 Model Set Up

The general set-up of the model is based on a partial equilibrium model in Hsieh and Klenow (2009), to which I add the LBD mechanism. The LBD process links firms' future productivity level to its current productivity level and production value and thus it creates an inter-temporal link in firms' maximisation problem. In other words, firms solve a dynamic profit maximization problem. For simplicity, I focus on a one-sector economy but it can easily be changed back to the multi-sector economy by using the Cobb-Douglas aggregation over industries as in Hsieh and Klenow (2009) and the analytical results would not change. Moreover, to illustrate the fact that dispersion in marginal (revenue) products is present even in an undistorted economy, I abstract from any distortions.

I assume there is a single final good Y_t produced by the representative firm in a perfectly competitive final output market. This firm combines M differentiated goods by CES function:

$$Y_t = \left(\sum_{i=1}^{M} Y_{it}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

The differentiated goods are produced by monopolistically competitive firms with Cobb-Douglas technology. The only source of heterogeneity across differentiated good producers is their level of productivity, A_{it}. That is, capital and labour, K_{it} and L_{it} are choice variables that depends on A_{it} of the monopolistically competitive firms:

$$Y_{it} = A_{it} K_{it}^{\alpha_t} L_{it}^{1-\alpha_t}$$

As seen in the previous section 3.2.3, that larger/older firms have higher productivity level but lower productivity growth rate compared to the smaller/younger firms. This can be modeled using a general Cobb-Douglas LBD function, where future productivity level is a function of current productivity level and current production.[16] In other words, accumulation of knowledge occurs as a by-product of production:

$$A_{it+1} = A_{it}^{\theta} Y_{it}^{\phi} \text{ where } 0 < \theta < 1, \;\; 0 < \phi < 1. \tag{3.7}$$

This model captures the fact that smaller and younger firms have low productivity level but higher productivity growth compared to larger and older firms. This is because I implicitly represent the larger and older firms as firms with high initial productivity level A_{si0} and smaller and younger firms as firms with low A_{si0}. Note that I assume LBD effect works only through size, but not through age as can be seen from the LBD equation 3.7. Moreover, I abstract from entry and exit of firms. Since this is a partial equilibrium model, wages and interest rates are exogenously given every period and they are constant across periods.[17] Hence the producer of differentiated good i solve the following infinite horizon

[16]If we iterate backwards then this LBD function can we written as below:

$$A_{it+1} = A_{i0}^{\theta^{t+1}} (Y_{i0}^{\theta^t} Y_{i1}^{\theta^{t-1}} \dots Y_{it}^{\theta^0})^{\phi}$$

Therefore this functional form can be interpreted as the future productivity level is a function of current and all past productions.

[17]The predictions of the model is that firms become more productive over time until the economy reaches the steady state. And firms demand more capital and labour as they become more productive. Therefore, constant wages and interest rates over time might cause some problems. I plan to extend the model into general equilibrium model to explore this further.

dynamic profit maximization problem:

$$
\begin{aligned}
V(A_{it}) = \max_{\{L_{it},K_{it}\}} & P_{it}Y_{it} - wL_{it} - RK_{it} + \beta V(A_{it+1}) \qquad (3.8)\\
\text{subject to} \quad & P_{it} = \left[\frac{Y}{Y_{it}}\right]^{\frac{1}{\sigma}} P \\
& Y_{it} = A_{it}K_{it}^{\alpha}L_{it}^{1-\alpha} \\
& A_{it+1} = A_{it}^{\theta}Y_{it}^{\phi} \\
& A_0 \ \text{ given}
\end{aligned}
$$

Here the monopolistically competitive firm producing good i internalise the demand function and LBD function into the optimization problem. And the FOCs are:

$$[K_{it}]: \quad \left(1-\frac{1}{\sigma}\right)\alpha\frac{P_{it}Y_{it}}{K_{it}} + \alpha\beta\phi V'(A_{it+1})\frac{A_{it+1}}{K_{it}} = R \qquad (3.9)$$

$$[L_{it}]: \quad \left(1-\frac{1}{\sigma}\right)(1-\alpha)\frac{P_{it}Y_{it}}{L_{it}} + (1-\alpha)\beta\phi V'(A_{it+1})\frac{A_{it+1}}{K_{it}} = w \qquad (3.10)$$

From these optimality conditions we can see differences between this dynamic optimization model with LBD effect and the static model. The first terms on the left hand side of equations (3.9) and (3.10) are the marginal (revenue) products of capital (MRPK) and marginal (revenue) products of labour (MRPL), respectively. The second terms reflect the inter-temporal link in the model through the LBD mechanism. The intuition is that by increasing current capital K_{it} we increase the future value of the firm $V(A_{it+1})$ through higher future productivity A_{it+1}, not only increasing current production by MRPK. On the other hand, in the static profit maximization problem, because there is no inter-temporal link, the second terms would be equal to zero. We can see it by setting $\phi = 0$ in the LBD function and by making productivity process completely exogenous to firms in equation (3.7). Then the second terms in FOCs will disappear. Both MRPK and MRPL should be equalised across firms every period because they face the same interest rate and wage. Depending on the characteristics of these second terms, we can determine whether we should see any dispersion in MRPK and MRPL across firms.

One thing that is clear at this stage is that if these monopolistically competitive firms have exactly the same productivity level as well as the same productivity growth rate then MRP should be equalised across firms despite the LBD effect. The reason is that if firms have the same A_{it} then they would choose same level of capital and labour and follow the same productivity process over time. Hence MRPK/MRPL as well as the second terms are equalised across firms. On the other hand, as long as there is heterogeneity in productivity level of firms then generally we would expect dispersion in MRPK/MRPL.

We can rewrite the FOCs as below to draw more intuition about the optimality conditions:[18]

$$\frac{K_{it}}{L_{it}} = \frac{\alpha w}{(1-\alpha)R} \tag{3.11}$$

$$R - (1-\frac{1}{\sigma})\alpha\frac{P_{it}Y_{it}}{K_{it}} = \beta\cdot\alpha\frac{Y_{it}}{K_{it}}\cdot\phi\frac{A_{it+1}}{Y_{it}}\left[(1-\frac{1}{\sigma'})\frac{P'_{it}Y'_{it}}{A_{it+1}} + \frac{\theta'}{\alpha'\phi'}\frac{K'_{it}}{A_{it+1}}\left(R' - \alpha'(1-\frac{1}{\sigma'})\frac{P'_{it}Y'_{it}}{K'_{it}}\right)\right] \tag{3.12}$$

As we can see from equation (3.11), it is required to have constant capital-labour ratio for all firms every period. This is also true for the static model. On the other hand, one can see the difference between the dynamic and static model from equation (3.12). The left hand side of the equation represents trade-off between interest rate and MRPK. When we increase capital by one unit we incur additional interest rate cost but at the same time increase in capital produces more output, represented by MRPK. As we know, the static optimality conditions require this difference to be equal to zero. But in dynamic setting firms can increase their future productivity by producing more and hence have lower MRPK in the current period. Therefore, the gap between interest rate and MRPK is typically positive. The right hand side of the equation shows this inter-temporal link. By increasing current capital, we would be able to increase future productivity level. The terms outside of the square brackets on the right hand side shows this channel: incremental increase in capital will cause output to increase by $\alpha\frac{Y_{it}}{K_{it}}$ and increase in output will raise future output by $\phi\frac{A_{it+1}}{Y_{it}}$ through LBD mechanism. Since this is in the next period we discount it by β. The increase in the future productivity will directly increase the output next period given the inputs. This is given by the first term inside the square brackets on the right hand side. Moreover, the increase in productivity will also increase the optimal size of capital inputs next period. Because of this we face the same trade-off between interest rate and MRPK in the next period. This is given by the second and third terms inside the square brackets.

3.3.2 Calibration and Simulation Results

I solve the model quantitatively and simulate the dynamic model over time. Moreover, since I am mainly interested in within-sector resource allocation I run model simulation within each sector. First, I calibrate/estimate the parameters. I take most of the param-

[18]Derivation is in Appendix C.1

eter values from Hsieh and Klenow (2009) partly because I compare the dynamic model to their static model. I take interest rate $R = 0.10$ as in Hsieh and Klenow (2009). Here it is assumed that real interest rate is 5% and depreciation is 5%. Moreover, elasticity of substitution between plant production $\sigma = 3$ and wage $w = 1$, also taken from Hsieh and Klenow (2009). Following Midrigan & Xu (2013), I set the intertemporal discounting rate $\beta = 0.92$. Time period I used for simulation is a year.

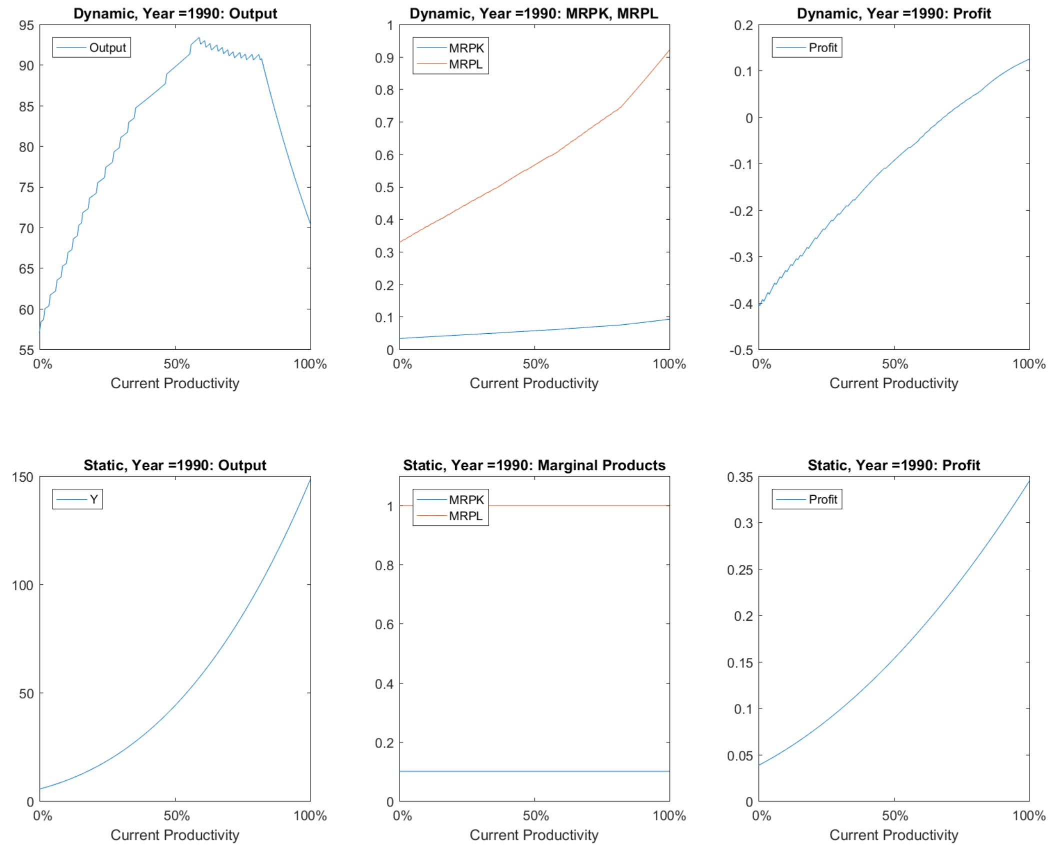
The most important parameters are the LBD parameters θ and ϕ, which governs how the plant level productivity process evolve over time in equation (3.7). Note that, estimating the LBD parameters from the Indonesian data doesn't help much. This is because it is very likely that Indonesian firms face larger distortions compared to developed economies and their productivity process is somehow deviated from its normal evolution. Hence the regression is not very informative.[19]

In order to determine values for LBD parameters, I use the following approach: I start simulations with a productivity distribution, expressed by grid points. When I choose the productivity distribution I take the Indonesian data as a reference. Specifically, I take the lowest and highest productivity of the Indonesian plants as the lower and upper limit of the grid, respectively. It is important to note that upper limit of the productivity distribution exogenously sets the steady state productivity level of the firms in the sense that firms cannot reach higher productivity level than the upper limit. This limit on productivity level together with LBD equation (3.7) determines how much firms produce at the steady state.[20] In order to make simulation of the dynamic and static models comparable, I choose the LBD parameters so that firms in dynamic model produce the same amount of production at the steady state as firms in static model. As a result, I set $\theta = 0.7$ and $\phi = 0.35$. I experiment with several other values for LBD parameters as well to see how sensitive my results are.

[19] Nevertheless, I present the regression results in Table C4 in Appendix C.2.

[20] Also competition level in the market, the number of firms in the market, also affects how much the plants produce at the steady state. Hence the simulation results are also affected by it.

Table 3.7: Simulation — Comparison between dynamic and static models — First year



Note: The top row shows the simulation results of the dynamic model and the bottom row presents the static model simulation. 1990 is the first year in the sample period as well as the first year of model simulations.

In the numerical simulations below I focus on the Indonesian textile manufacturing industry, which has the largest number of permanent plants: 443 (permanent) plants in 1990.[21] Table 3.7 shows the simulations of both dynamic model and static model in the first year. In particular, I feed the same productivity distribution of firms estimated from 1990 data into the model. The first row presents the dynamic model, and the second row shows the static model simulations.

Column 1 in Table 3.7 shows that low (high) productivity firms in the dynamic economy produce more (less) than their counterparts in the static model. The reason for low productivity firms in the dynamic economy produce more than the low-productive firms in the static economy is the LBD effect: the low-productive firms have more capacity to increase their future productivity compared to firms with high productivity. Hence, they invest in their future productivity level by producing more in the current period. This

[21]ISIC Rev2 code is 3211 and corresponding activity of the industry is spinning, weaving and finishing textiles.

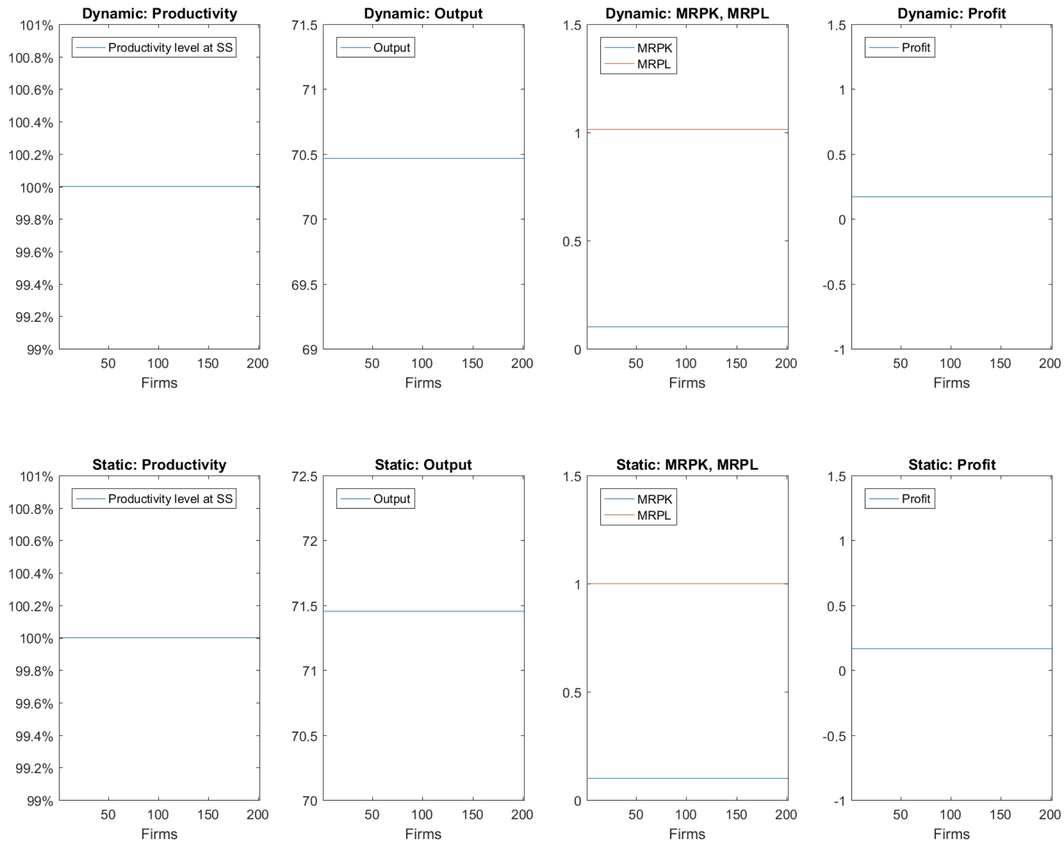
fact is reflected in their profit in Column 3 as well, where firms with low productivity earn negative profits in the dynamic model as opposed to positive profits in the static model. On the other hand, high-productive firms in the dynamic setting produce less than what static optimality conditions require because their market share is shrunk due to overproduction by the low-productive firms. The implication of the LBD effect on firms' marginal (revenue) products is shown in Column 2. The dynamic efficiency requires firms' marginal (revenue) products to be increasing function of their current productivity level. Therefore, it results in heterogeneous MRPK/MRPL across firms. In contrast, the static efficiency requires marginal (revenue) products to be equalised across firms.

Table C5 and Table C6 in Appendix C.2 present the simulations of the first four years of the dynamic and static models, respectively. Simulations are done in the following steps. For the dynamic model, first I the estimated productivity distribution of firms from 1990 data and I solve the dynamic model numerically. The resulting policy function gives the productivity distribution for the next period, for the year 1991. I feed it for the next period simulation and so on. I do it until 1994 and plot it in Table C5. It shows the convergence of the productivity distribution over time, on the x-axis of the graphs. More importantly, it shows that there is dispersion in MRPK/MRPL across firms along the transitional dynamics. On the other hand, for the static model, I use the same productivity distributions as the dynamic model each year. For example, I start with the same estimated productivity distribution in year 1990, and for 1991 I use the productivity distribution generated from the dynamic model's policy function. The results are shown in Table C6 and it shows that there is no dispersion in MRPK/MRPL across firms along the transitional dynamics.

Overtime low productivity firms in dynamic model catch up with high productivity firms through LBD effect and the firms become homogenous when the economy reaches the steady state. Table 3.8 shows the steady state of both dynamic model and static model and it is clear that there is no difference between the two models at the steady state. Row 1 represents the dynamic model, and since all firms have the same level of productivity level they produce the same amount of production, have same marginal (revenue) products and earn same amount of profit.

Therefore I conclude this section with the following statement: as long as there is heterogeneity in firms' productivity level, dynamic efficiency requires some level of dispersion in firms' marginal (revenue) products. The optimal level of dispersion in MRPK/MRPL depends on underlying productivity distribution. In contrast, static efficiency requires

Table 3.8: Simulation — Comparison between dynamic and static models — Steady state



Note: This table shows steady state of the dynamic model in the top row and the static model in the bottom row.

firms to have the same marginal (revenue) products no matter what is their productivity distribution.

3.4 Conclusion

It is well-known that misallocation of resources across firms reduce the aggregate productivity and hinders the long-run economic growth. There is considerable consensus in the literature on misallocation to regard variation in firms' MRPs as a sign of misallocation.

In this paper I show theoretically that some dispersion in MRPs could be beneficial for the economy in the long run in the presence of LBD mechanism. I take a standard model with monopolistically competitive firms, as in Hseih and Klenow (2009), and incorporate the LBD mechanism in the model. In particular, I assume that firms' future productivity level is endogenous and it depends on firms' input choices in the current period.

More importantly, the currently low-productive firms are assumed to be more likely to have high productivity growth. Therefore, the LBD mechanism turns the standard static model into a dynamic model, where the inter-temporal link is emerging from the LBD effects in firms' productivity process. The Indonesian firm-level panel data supports these assumptions. I document that smaller and younger firms have lower productivity level but their productivity growth is higher than that of larger and older firms.

To see the implications in the long run I simulate both the static and dynamic models over time. I feed both models with the same arbitrary productivity distribution and compare their transitional dynamics and the speed of reaching the steady state. Both economies start with firms, who are heterogeneous in terms of their initial productivity level, and converges to the steady state, where all firms reach the highest possible productivity level in the productivity distribution. Even though firms in the repeated static model do not internalise the LBD effects their productivity in the next period increases because of the LBD effect. In contrast, in the dynamic model, lower productive firms produce more to utilise the LBD effects and thus their productivity increases faster and reach the steady state faster than the static model. The main conclusion from this exercise is that some dispersion in MRPs could be beneficial in the long run, and it has important policy implications. For example, to promote long-run economic growth, it would be wise to help young and small firms grow and produce more since they can reach higher productivity through LBD. This could be done in many ways, such as introducing small business tax incentives or having fewer and weaker rules and regulations for small and young firms. This will allow them to use more resources and hire more workers, which results in small and young firms having lower MRPs.

References

Philippe Aghion and Peter Howitt. A Model of Growth Through Creative Destruction. *Econometrica*, 60(2):323, 1992.

Costas Arkolakis, Theodore Papageorgiou, and Olga A. Timoshenko. Firm learning and growth. *Review of Economic Dynamics*, 27:146–168, 2018.

Costas Arkolakis, Theodore Papageorgiou, Olga A Timoshenko, Emanuel Ornelas, Andres Rodriguez-Clare, Kim Ruhl, Erick Sager, and James Tybout. Firm Learning and Growth. Technical report, 2017.

Kenneth J. Arrow. The Economic Learning Implications of by Doing. *The Review of Economic Studies*, 29(3):155–173, 1962.

John Asker, Allan Collard-wexler, Jan De Loecker, and John Asker. Dynamic Inputs and Resource (Mis) Allocation. *Journal of Political Economy*, 122(5):1013–1063, 2014.

Yan Aw, Mark J Roberts, and Daniel Yi Xu. R & D Investment , Exporting , and Productivity Dynamics. *American Economic Review*, 101:1312–1344, 2011.

Salvador Barrios and Eric Strobl. Learning by Doing and Spillovers: Evidence from Firm-Level Panel Data. *Review of Industrial Organization*, 25(2):175–203, 2004.

Eric Bartelsman, John Haltiwanger, and Stefano Scarpetta. Cross-Country Differences in Productivity: The Role Of Allocation and Selection.

Eric J Bartelsman and Mark Doms. Understanding Productivity: Lessons from Longitudinal Microdata. *Journal of Economic Literature*, 38(3):569–594, 2000.

Bento Pedro and Diego Restuccia. Misallocation, Establishment Size, and Productivity. *NBER Working Paper*, 2016.

Mark Bils, Peter J. Klenow, and Cian Ruane. Misallocation or Mismeasurement? *Working Paper*, pages 1–48, 2017.

Nick Bloom, Stephen Dorgan, John Dowdy, and John Van Reenen. Management Practice & Productivity: Why they matter. *Center for Economic Performance and McKinsey policy release*, (July), 2007.

Benjamin Bridgman, Shi Qi, and James Andrew Schmitz. The Economic Performance of Cartels: Evidence from the New Deal U.S. Sugar Manufacturing Cartel, 1934–74. *Federal Reserve Bank of Minneapolis, Research Department Staff Report*, (November), 2009.

Ricardo J. Caballero, Takeo Hoshi, and Anil K. Kashyap. Zombie lending and depressed restructuring in Japan. *American Economic Review*, 98(5):1943–1977, 2008.

Ulrich Doraszelski and Jordi Jaumandreu. R and D and productivity: Estimating endogenous productivity. *Review of Economic Studies*, 80(4):1338–1383, 2013.

Jonathan Eaton and Samuel Kortum. TECHNOLOGY, GEOGRAPHY, AND TRADE By Jonathan Eaton and Samuel Kortum 1. *Econometrica*, 70(5):1741–1779, 2002.

Ana M. Fernandes and Alberto E. Isgut. Learning-by-exporting effects: Are they for real? *Emerging Markets Finance and Trade*, 51(1):65–89, 2015.

Lucia Foster, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. Macro and Micro Dynamics of Productivity: From Devilish Details to Insights. *National Bureau of Economic Research Working Paper Series*, 2017.

Lucia Foster, John Haltiwanger, and Chad Syverson. Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *Source: The American Economic Review*, 98(1):394–425, 2008.

Lucia Foster, Censusgov John Haltiwanger, and Chad Syverson. The Slow Growth of New Plants: Learning about Demand? 2013.

Gita Gopinath, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. Capital Allocation and Productivity in South Europe. *NBER Working Paper*, 2015.

John Haltiwanger. Firm Dynamics and Productivity:: TFPQ, TFPR, and Demand-Side Factors. *Economía*, 17(1):3–26, 2016.

Hugo A. Hopenhayn. *Firms, misallocation, and aggregate productivity: A review*, volume 6. 2014.

Chang-Tai Hsieh and Peter J Klenow. Misallocation and Manufacturing TFP in China and India. *THE QUARTERLY JOURNAL OF ECONOMICS*, 2009.

Chang-Tai Hsieh and Peter J Klenow. The Life Cycle of Plants in India and Mexico. 2012.

J Bradford Jensen, Robert H Mcguckin, and Kevin J Stiroh. The Impact of Vintage and Survival on Productivity : Evidence from Cohorts of U . S . Manufacturing Plants. *The MIT Press*, 83(2):323–332, 2001.

Chad John Haltiwanger, Robert Kulick and Syverson. Misallocation Measures: The Distortion That Ate The Residual. *NBER WORKING PAPER SERIES*, (1):430–439, 2018.

Wolfgang Keller and Stephen R. Yeaple. Multinational enterprises, international trade, and productivity growth: Firm level evidence from the United States. *Review of Economics and Statistics*, 91(4):821–831, 2009.

Hyeog Ug Kwon, Futoshi Narita, and Machiko Narita. Resource reallocation and zombie lending in Japan in the 1990s, 2015.

Steven D Levitt, John A List, and Chad Syverson. Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant. *Journal of Political Economy*, 121(4):643–681, 2013.

Jens Matthias Arnold and Beata S. Javorcik. Gifted kids or pushy parents? Foreign direct investment and plant productivity in Indonesia, 2009.

Marc J. Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725, 2003.

Stephen L. Parente. Technology Adoption, Learning-by-Doing, and Economic Growth, 1994.

Nina Pavcnik. Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. *Review of Economic Studies*, 69(1):245–276, 2002.

Diego Restuccia and Richard Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720, 2008.

Diego Restuccia and Richard Rogerson. Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1–10, 2013.

Diego Restuccia and Richard Rogerson. The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3):151–174, 2017.

Chad Syverson. What determines productivity. *NBER Working Paper*, 2010.

Peter Thompson. *Learning by doing*, volume 1. Elsevier BV, 1 edition, 2010.

Peter Thompson. The Relationship between Unit Cost and Cumulative Quantity and the Evidence for Organizational Learning-by-Doing. *Journal of Economic Perspectives*, 26(3):203–224, 2012.

Rebecca Achee Thornton and Peter Thompson. Learning from Experience and Learning from Others: An Exploration of Learning and Spillover in Eartime Shipbuilding. *The American Ecnomic Review*, 91(5):1350–1368, 2001.

Bart Van Ark, Mary O'Mahony, and Marcel P. Timmer. The productivity gap between Europe and the United States: Trends and causes. *Journal of Economic Perspectives*, 22(1):25–44, 2008.

Ilke van Beveren. Total factor productivity estimation: A practical review. *Journal of Economic Surveys*, 26(1):98–128, 2012.

C Appendix

C.1 Derivations

Differentiate the value function in (3.8) with respect to productivity A_{it}:

$$V'(A_{it}) = \left(1 - \frac{1}{\sigma}\right) \frac{P_{it}Y_{it}}{A_{it}} + \beta V'(A'_{it})\theta \frac{A'_{it}}{A_{it}} \tag{13}$$

Find $V'(A'_{it})$ from equation (3.9):

$$V'(A'_{it}) = \frac{R - \alpha\left(1 - \frac{1}{\sigma}\right)\frac{P_{it}Y_{it}}{K_{it}}}{\alpha\beta\phi\frac{A'_{it}}{K_{it}}} \tag{14}$$

Substitute equation (14) into equation (13)

$$\begin{aligned} V'(A_{it}) &= \left(1 - \frac{1}{\sigma}\right) \frac{P_{it}Y_{it}}{A_{it}} + \beta\theta \frac{A'_{it}}{A_{it}} \frac{R - \alpha\left(1 - \frac{1}{\sigma}\right)\frac{P_{it}Y_{it}}{K_{it}}}{\alpha\beta\phi\frac{A'_{it}}{K_{it}}} \\ &= \left(1 - \frac{1}{\sigma}\right) \frac{P_{it}Y_{it}}{A_{it}} + \frac{\theta}{\alpha\phi}\frac{K_{it}}{A_{it}}\left(R - \alpha(1 - \frac{1}{\sigma})\frac{P_{it}Y_{it}}{K_{it}}\right) \end{aligned} \tag{15}$$

Move equation (15) forward by one period:

$$V'(A'_{it}) = (1 - \frac{1}{\sigma'})\frac{P'_{it}Y'_{it}}{A'_{it}} + \frac{\theta'}{\alpha'\phi'}\frac{K'_{it}}{A'_{it}}\left(R' - \alpha'(1 - \frac{1}{\sigma'})\frac{P'_{it}Y'_{it}}{K'_{it}}\right) \tag{16}$$

Substitute $V'(A'_{it})$ in equation (3.9) by using equation (16):

$$\left(1 - \frac{1}{\sigma}\right)\alpha\frac{P_{it}Y_{it}}{K_{it}} + \beta \cdot \alpha\frac{Y_{it}}{K_{it}} \cdot \phi\frac{A'_{it}}{Y_{it}} \cdot \left[(1 - \frac{1}{\sigma'})\frac{P'_{it}Y'_{it}}{A'_{it}} + \frac{\theta'}{\alpha'\phi'}\frac{K'_{it}}{A'_{it}}\left(R' - \alpha'(1 - \frac{1}{\sigma'})\frac{P'_{it}Y'_{it}}{K'_{it}}\right)\right] = R \tag{17}$$

C.2 Other empirical results

Table C1: Productivity level - full, unbalanced sample

	(1) $log(TFPR_{simple})$	(2) $log(TFPR_{simple})$	(3) $log(TFPR_W)$	(4) $log(TFPR_W)$	(5) $log(TFPQ_{HK})$	(6) $log(TFPQ_{HK})$
log(Sales)	0.214***		0.416***		0.672***	
	(0.00890)		(0.00562)		(0.0121)	
log(Age)		-0.0292***		0.0156		0.0273
		(0.0101)		(0.0103)		(0.0177)
Observations	296324	296324	296324	296324	296324	296324
Adjusted R^2	0.217	0.095	0.760	0.458	0.631	0.210

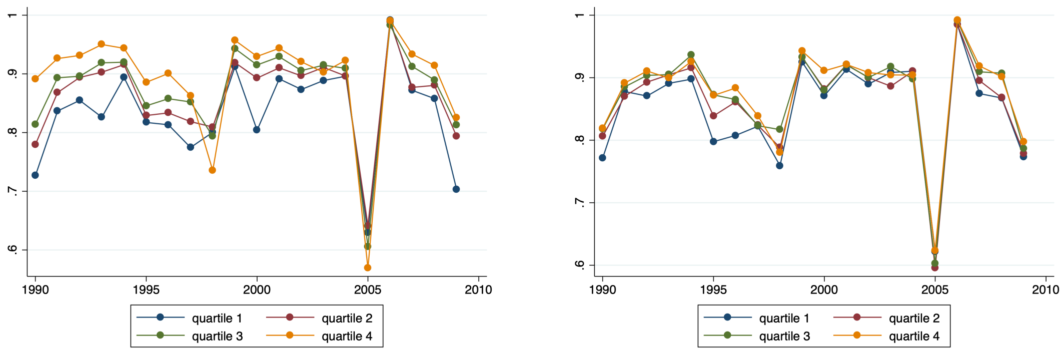
Note: This table presents regression results specified in equation 3.5 using full, unbalanced sample. The dependent variables in column 1, 2 and 3 are the simple revenue productivity defined in equation 3.1, revenue-based productivity measure based on Wooldridge (2009) estimation approach, and the log of physical productivity based on Hsieh and Klenow (2009) approach, defined in equation 3.4, respectively. I include year, industry, province and legal status fixed effects in all regressions. Control variables include firm characteristics such as export status, capacity utilization, capital intensity, whether the firm has any investment financed through equity issue, by foreign loans, by FDI, and any investment financed on capital markets. The standard errors are clustered at both year and industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C2: Productivity growth - full, unbalanced sample

	(1) $\Delta log(TFPR_{simple})$	(2) $\Delta log(TFPR_{simple})$	(3) $\Delta log(TFPR_W)$	(4) $\Delta log(TFPR_W)$	(5) $\Delta log(TFPQ_{HK})$	(6) $\Delta log(TFPQ_{HK})$
log(Sales)	-0.0498***		-0.0485***		-0.0841***	
	(0.00218)		(0.00227)		(0.00349)	
log(Age)		-0.0150***		-0.0194***		-0.0389***
		(0.00229)		(0.00215)		(0.00354)
Observations	252298	252298	252298	252298	252298	252298
Adjusted R^2	0.025	0.018	0.028	0.018	0.035	0.024

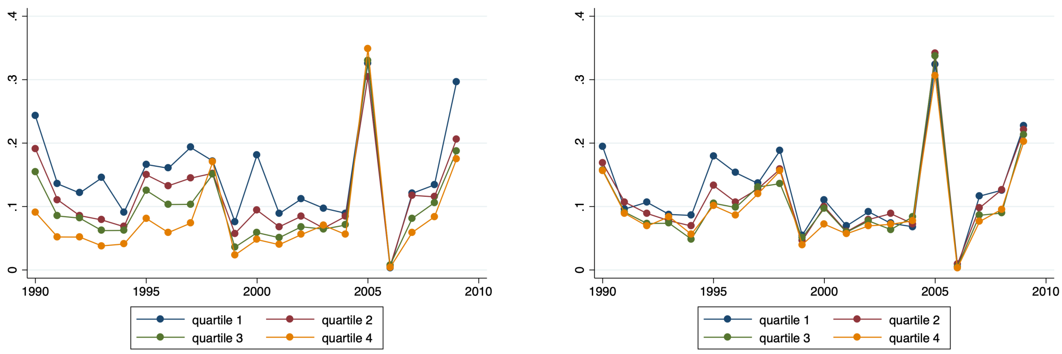
Note: This table shows estimated coefficients from regression equation 3.6. The dependent variables in column 1, 2 and 3 are growth rates in the simple revenue productivity defined in equation 3.1, revenue-based productivity measure based on Wooldridge (2009) estimation approach, and the log of physical productivity based on Hsieh and Klenow (2009) approach, defined in equation 3.4, respectively. I include year, industry, province and legal status fixed effects in all regressions. Control variables include firm characteristics such as export status, capacity utilization, capital intensity, whether the firm has any investment financed through equity issue, by foreign loans, by FDI, and any investment financed on capital markets. The standard errors are clustered at both year and industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure C1: Survival rate by firm age and size groups



(a) Size quantiles (b) Age quantiles

Figure C2: Exit rate by firm age and size groups



(a) Size quantiles (b) Age quantiles

Table C3: Exit rates by size and age quantiles

	Size		Age	
	(1) Survival rate	(2) Survival rate	(3) Survival rate	(4) Survival rate
Quartiles	-0.0216***	-0.0216***	-0.00856	-0.00856***
	(0.00739)	(0.00271)	(0.00678)	(0.00136)
Year FE	No	Yes	No	Yes
Observations	80	80	80	80
Adjusted R^2	0.096	0.911	0.007	0.970

Note: The dependent variables in the first 2 columns are exit rates of firms by size quantiles. The last 2 columns are exit rates of firms by age quantiles. In even columns I include year fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C4: LBD parameters - single-product firms, unbalanced sample

	(1) $log(TFPQ_{simple})$	(2) $log(TFPQ_{HK})$	(3) $log(TFPR_{simple})$	(4) $log(TFPR_W)$	(5) $log(TFPQ_W)$
log(Sales)	0.0365**	0.468***	0.150***	0.176***	0.0585***
	(0.0162)	(0.0186)	(0.0123)	(0.00982)	(0.0123)
Past $log(TFPQ_{simple})$	0.803***				
	(0.0129)				
Past $log(TFPQ_{HK})$		0.389***			
		(0.00957)			
Past $log(TFPR_{simple})$			0.529***		
			(0.0100)		
Past $log(TFPR_W)$				0.682***	
				(0.0182)	
Past $log(TFPQ_W)$					0.834***
					(0.0128)
Observations	66428	66428	66428	66428	66428
Adjusted R^2	0.667	0.668	0.424	0.717	0.714

Note: This table estimates LBD coefficients expressed in equation 3.7 using single-product firms. The notations of dependent variables are explained in Section 3.2. I include year, industry, province and legal status fixed effects in all regressions. Control variables include firm characteristics such as export status, capacity utilization, capital intensity, whether the firm has any investment financed through equity issue, by foreign loans, by FDI, and any investment financed on capital markets. The standard errors are clustered at both year and industry level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C5: Simulation — Dynamic model — Transitional dynamics

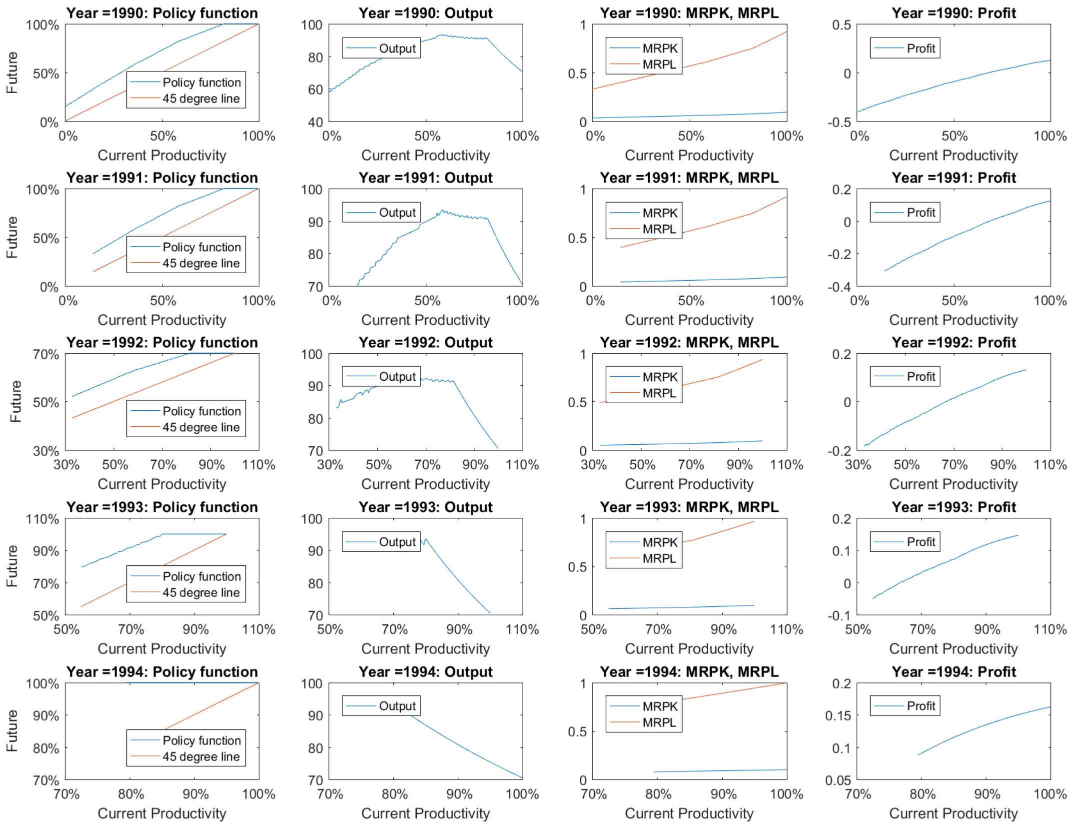


Table C6: Simulation — Static model — Repeated cross section