

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Essays on Taxation in Emerging Economies

**Permalink**

<https://escholarship.org/uc/item/1c48k9wm>

**Author**

Mittal, Shekhar

**Publication Date**

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Taxation in Emerging Economies

A dissertation submitted in partial satisfaction  
of the requirements for the degree  
Doctor of Philosophy in Management

by

Shekhar Mittal

2018

© Copyright by  
Shekhar Mittal  
2018

## ABSTRACT OF THE DISSERTATION

Essays on Taxation in Emerging Economies

by

Shekhar Mittal

Doctor of Philosophy in Management

University of California, Los Angeles, 2018

Professor Romain T. Wacziarg, Co-Chair

Professor Aprajit Mahajan, Co-Chair

Even though compliance issues are central to taxation policies in emerging economies, convincing empirical research on tax compliance has been scarce. Through the three chapters of my dissertation, I bridge this gap by using detailed value added tax (VAT) micro-data from Delhi, India.

A key advantage of VAT type systems is that they allow for corroboration of transactions using returns of interacting firms. In chapter 1, co-author Aprajit Mahajan and I evaluate the effect of a technology reform that improved the Delhi tax authority's ability to cross-check buyer reports against seller reports within the VAT system. Before the technology change, such cross-checks could only be accomplished by auditing both parties, a relatively rare and time-consuming activity. After the policy change, the tax authority could (and did) relatively easily cross-check information, declared by buyers with the corresponding information from sellers, directly on its own servers without initiating an audit. We use a difference-in-difference approach to show that the policy had a large and significant effect on wholesalers relative to retailers. A wholesaler is more likely to sell to registered firms whereas a retailer is more likely to sell to final customers where the paper trail breaks down. Therefore,

on the output side, the self-enforcing mechanism of the VAT is more likely to break down for retailers compared to wholesalers. We also find significant heterogeneity with almost the entire increase being driven by changes in the behavior of the largest tax-paying firms. This result sheds light on limits of third-party verification in a context with limited audit resources and where the majority of firms do not remit any tax.

In low compliance environments, a common strategy to manipulate the third-party verification system is to establish fraudulent (“bogus”) firms. Bogus firms help genuine firms in reducing their tax burden by issuing fake receipts. A tax authority determines the existence of bogus firms by first filtering down based on a few preliminary indicators, and then undertaking physical inspections. Given the authority’s limited resources, these inspections are only done sporadically. A key challenge in improving tax compliance then is to regularly, cheaply and reliably identify such bogus firms. In chapter 2, coauthors Aprajit Mahajan, Ofir Reich and I apply a machine learning classifier to the same tax dataset to identify bogus firms which can be further targeted for physical inspections. We face a nonstandard applied machine learning scenario. First, one-sided labels: firms that are not caught as bogus are of unknown class: bogus or legitimate, and we need to not only use them to train the classifier but also make predictions on them. Second, multiple time-periods: each firm files several periodic VAT returns but its class is fixed so prediction needs to be made at the firm, not firm-period, level. Third, point in time simulation: we estimate the revenue saving potential of our model by simulating the implementation of our system in the past. We do this by rolling back the data to the state of knowledge at a specific time and calculating the revenue impact of acting on our model’s recommendations and catching the bogus firms and estimate US\$40 million in recovered revenue.

Tax authorities commonly apply size based regulations to firms. If firms are

concerned about compliance costs, then such regulations create adverse incentives for firms to stay small. These regulations also increase the monitoring effort needed from tax officials. In the first two years of our dataset, the Delhi tax system had multiple turnover based filing frequency thresholds. Firms with declared turnover (in the previous year) less than ₹1 million had to file returns annually, between ₹1 and 5 million - semiannually, between ₹5 and 50 million - quarterly, and more than ₹50 million - monthly. In the years 3, 4 and 5 of our dataset, this turnover based filing policy was first weakened and then completely disbanded. In chapter 3, coauthor Jan Luksic and I first show that this policy resulted in bunching of firms below the thresholds at all levels. Using the change in these reporting policies, we provide further evidence that such sharp bunching indeed occurs due to the VAT reporting frequency thresholds. Second, we calculate the VAT revenue losses due to such bunching and document the longer-term impact of the VAT reporting frequency thresholds. Finally, the subsequent withdrawal of the policy allows us to show that in a regime with size-dependent reporting requirements, more frequent reporting does not lead to greater levels of VAT collection.

The dissertation of Shekhar Mittal is approved.

Paola Giuliano

Adriana Lleras-Muney

Ricardo Nico Perez Truglia

Nico Voigtländer

Aprajit Mahajan, Committee Co-Chair

Romain T. Wacziarg, Committee Co-Chair

University of California, Los Angeles

2018

To Nana, my biggest champion.



## TABLE OF CONTENTS

<b>List of Figures</b> . . . . .	<b>xi</b>
<b>List of Tables</b> . . . . .	<b>xiv</b>
<b>1 VAT in Emerging Economies: Does Third Party Verification Matter?</b> . . . . .	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Related Literature . . . . .	5
1.3 Conceptual Framework . . . . .	8
1.3.1 Illustrative Example . . . . .	12
1.3.2 VAT in Delhi: Policy Change . . . . .	15
1.4 Data . . . . .	16
1.4.1 VAT Returns . . . . .	16
1.4.2 Firm Characteristics . . . . .	18
1.4.3 Audit Notices . . . . .	19
1.4.4 Sample . . . . .	19
1.5 Empirical Strategy . . . . .	19
1.5.1 Identification: Wholesalers vs Retailers Over Time . . . . .	20
1.5.2 Model Specification . . . . .	21
1.6 Results . . . . .	23
1.6.1 Description: Wholesalers vs Retailers . . . . .	23
1.6.2 Results: Difference in Difference . . . . .	25

1.7	Mechanism . . . . .	34
1.7.1	Sales to Registered Firms . . . . .	35
1.8	Conclusion . . . . .	39
<b>2</b>	<b>Who is Bogus? Using One-Sided Labels to Identify Fraudulent Firms from Tax Returns . . . . .</b>	<b>41</b>
2.1	Introduction . . . . .	41
2.2	Related Work . . . . .	45
2.2.1	Economics . . . . .	45
2.2.2	Machine Learning . . . . .	46
2.3	Value Added Tax . . . . .	47
2.4	A Mechanism of Bogus Firms . . . . .	49
2.5	System Description . . . . .	51
2.5.1	Data . . . . .	51
2.5.2	Features . . . . .	52
2.5.3	Class Labels . . . . .	54
2.5.4	Classifier . . . . .	56
2.5.5	Cross-Validated Predictions . . . . .	57
2.5.6	Multi-Period Model . . . . .	59
2.6	Evaluation of Performance . . . . .	63
2.6.1	Top Recommendations . . . . .	63
2.6.2	Top Recommendations - Maximizing Revenue . . . . .	63
2.6.3	Different Feature Sets . . . . .	65
2.6.4	Point In Time Simulation . . . . .	66

2.7	Future Work and Conclusion . . . . .	70
<b>3</b>	<b>Red Tape? The Revenue Impact of the VAT Filing Thresholds . .</b>	<b>73</b>
3.1	Introduction . . . . .	73
3.2	Related Literature . . . . .	76
3.3	Methodology . . . . .	79
3.3.1	Delhi: Policy Change . . . . .	80
3.3.2	Data . . . . .	81
3.3.3	Bunching at Thresholds . . . . .	82
3.3.4	VAT Revenue Loss to the Tax Authority . . . . .	84
3.3.5	No Benefits to More Information . . . . .	84
3.3.6	Social Welfare Analysis . . . . .	86
3.4	Results . . . . .	88
3.4.1	Bunching at Thresholds . . . . .	88
3.4.2	VAT Revenue Loss to the Tax Authority . . . . .	91
3.4.3	No Benefits to More Information . . . . .	93
3.4.4	Social Welfare Analysis . . . . .	94
3.5	Conclusion . . . . .	96
<b>A</b>	<b>“VAT in Emerging Economies: Does Third Party Verification Mat- ter?” . . . . .</b>	<b>98</b>
A.1	Data Summary . . . . .	98
A.2	Quarterly Results . . . . .	102
A.2.1	Quarterly Analysis Along with Falsification Test . . . . .	102

A.3	Flexible DID Specifications . . . . .	104
A.3.1	Econometric Model of Event Study Analysis . . . . .	104
A.3.2	Annual Analysis . . . . .	106
A.3.3	Quarterly Analysis . . . . .	107
A.4	Other Results . . . . .	108
A.5	Policy Execution: Other . . . . .	110
A.5.1	DiDiD: Proportion Registered Sales . . . . .	110
A.5.2	Revisions . . . . .	111
A.5.3	Matching . . . . .	113
A.5.4	Consolidated vs Transaction Data . . . . .	117
<b>B</b>	<b>“Return Forms” . . . . .</b>	<b>118</b>
B.1	Consolidated Form . . . . .	118
B.2	Annexures . . . . .	121
	<b>Bibliography . . . . .</b>	<b>123</b>

## List of Figures

1.1	Toy Model . . . . .	11
1.2	Third Party Verification . . . . .	15
1.3	Wholesalers vs Retailers: Annual Trends . . . . .	25
1.4	Tax Remitted by Top Percentile and Bottom 99% of Firms . . . . .	28
1.5	Heterogeneity Analysis: VAT Remitted for Wholesalers vs Retailers . . . . .	32
1.6	Top Decile: VAT Remitted for Wholesalers vs Retailers . . . . .	33
1.7	Sales to Registered Firms . . . . .	35
1.8	Top Percentile: Sales to Registered Firms . . . . .	36
2.1	One Sided Labels . . . . .	43
2.2	Stylized Example Illustrating a Value Added Tax Transaction Chain . . . . .	48
2.3	An Example Showing How Bogus Firms Facilitate Tax Evasion . . . . .	50
2.4	Comparison of Different Classifiers . . . . .	56
2.5	Cross-validated Prediction Procedure . . . . .	57
2.6	Generating Firm Level Predictions from Firm-Quarter Level Data Points . . . . .	59
2.7	Model Performance on Top 1000 Recommendations . . . . .	64
2.8	Betas Curves for Different Feature Sets . . . . .	65
2.9	Illustration of Point-in-time Simulation at $T=3$ . . . . .	67
2.10	Point-in-time Simulations Performance . . . . .	70
3.1	Distribution of Firms by Size in the Dataset . . . . .	81
3.2	Firm Revenue Distribution at the Low Threshold . . . . .	87

3.3	Firm Revenue Distribution at the Middle Threshold . . . . .	89
3.4	Firm Revenue Distribution at the High Threshold . . . . .	90
3.5	Distribution of Compliance Costs by Revenue Size . . . . .	92
A.1	Total VAT Remitted (For All Firms) . . . . .	98
A.2	Total VAT Remitted (For Firms That are Present in All Years) . . . . .	99
A.3	Lorenz Curve for All Firms in Tax Year 1 . . . . .	100
A.4	Lorenz Curve for All Firms in Tax Year 5 . . . . .	101
A.5	Wholesalers vs Retailers: Quarterly Trends . . . . .	103
A.6	Event Study: Wholesalers vs Retailers (Annual) . . . . .	106
A.7	Event Study: Wholesalers vs Retailers (Quarterly) . . . . .	107
A.8	Heterogeneity Analysis: Wholesalers vs Retailers . . . . .	108
A.9	Quarterly Analysis for Top Decile: Wholesalers vs Retailers . . . . .	109
A.10	DiDiD: Wholesalers vs Retailers . . . . .	110
A.11	Mean Revisions of All Firms . . . . .	111
A.12	Mean Revisions of Wholesalers and Retailers . . . . .	112
A.13	Matching on the Purchase Side for All Firms . . . . .	113
A.14	Matching on the Sales Side for All Firms . . . . .	114
A.15	Matching Analysis: Retailers Vs Wholesalers (Purchases, Top Percentile) . . . . .	115
A.16	Matching Analysis: Retailers Vs Wholesalers (Sales, Top Percentile) . . . . .	116
A.17	Consolidated vs Transactional Data . . . . .	117
B.1	Page 1 - Form 16 . . . . .	118
B.2	Page 2 - Form 16 . . . . .	119

B.3 Page 3 - Form 16 . . . . . 120  
B.4 Annexure 2a - Form 16 . . . . . 121  
B.5 Annexure 2b - Form 16 . . . . . 122

## List of Tables

1.1	Evasion in Toy Model . . . . .	14
1.2	Summary Statistics: Retailer and Wholesaler Sample . . . . .	24
1.3	Diff-in-Diff: Wholesalers and Retailers (Annual) . . . . .	26
1.4	Diff-in-Diff: Wholesalers and Retailers (Annual, Top Percentile) . . . . .	29
1.5	Diff-in-Diff: Wholesalers and Retailers (Annual, Bottom 99 Percentile) . . . . .	31
1.6	Diff-in-Diff for Top Decile: Wholesalers and Retailers (Annual) . . . . .	34
1.7	DiDiD: Wholesalers/Retailers; Before/After;Sales to Registered Firms . . . . .	38
2.1	Model Performance on Top Recommendations . . . . .	62
2.2	Point-in-time Simulation Performance for the 1-400 Inspection Group . . . . .	69
3.1	VAT Revenue Lost . . . . .	93
3.2	VAT Revenue vs. Annual Number of VAT Returns Regressions . . . . .	94
3.3	VAT Revenue vs. VAT Filing Categories Regressions . . . . .	95
3.4	Implicit Welfare Equalizing Per Firm Subsidies from the Thresholds . . . . .	96
A.1	Summary Stats: All Firms . . . . .	100
A.2	Summary stats: Always Present Firms . . . . .	102
A.3	Falsification Test: Wholesalers and Retailers (Quarterly) . . . . .	104
A.4	Pre-trend Analysis: Wholesalers vs Retailers . . . . .	105



## ACKNOWLEDGMENTS

I thank my advisors Romain Wacziarg and Aprajit Mahajan who are not only inspiring researchers but also approachable and caring mentors. Romain Wacziarg has given me the opportunity to explore my own agenda, even when I was clearly struggling. Many times, especially during my coursework stumbles, I have found solace in his kind and encouraging words. He has gone out of his way to provide me with financial support, and has always been ready to reach out to others on my behalf. Aprajit Mahajan has provided the much needed calm to my excitable personality. He has provided me with a safety net while I have tried to push for ambitious collaborations. He has taught me the value of integrity, rigor and reworking the draft. Most importantly, both of them have given me the freedom to fail outstandingly.

I would also like to thank my committee members - Paola Giuliano, Adriana Lleras-Muney, Ricardo Perez-Truglia, and Nico Voigtländer. Without their advice and support I could not have reached here. Paola Giuliano has more faith in my ability than me, and has always been ready to provide inputs from the “other” side. In no other class have I been as motivated to do well as in the population economics course taught by Adriana Lleras-Muney. Ricardo Perez Truglia’s research and guidance has helped me keep my work grounded in public finance challenges. Nico Voigtländer has always been insightful and constructive, not to mention kind.

This dissertation is a culmination of (or a milestone in) a journey in which many people have contributed in unexpected ways and it is impossible to name and thank them all. I am grateful to both Marianne Bertrand and Pablo Montagnes, who started me on my research journey and who were unexpectedly patient with an enthusiastic master’s student at University of Chicago. In my initial PhD years, I have greatly benefited from long conversations with Christian Dippel. I have also greatly benefitted from discussions with and advice from Pierre Bachas, Prabhat Barnwal,

Youssef Benzarti, Michael Best, Josh Blumenstock, Andrea DiMiceli, Stefano Fiorin, Vasily Korovkin, Aparna Krishnan, Temina Madon, Atishi Marlena, Dina Pomeranz, Jasmine Shah and Divya Singh. I thank them for their generosity. I also thank participants at RIDGE Conference-Public Economics (2017), the GEM seminar (Anderson School, UCLA), GEM-BPP workshop (2017), Center for Business Taxation Doctoral Meeting (Oxford, 2017), APPAM Conference (Chicago, 2017), NEUDC Conference (Tufts, 2017), and 2nd Zurich Conference on Public Finance in Developing Countries (2017). All remaining errors are mine.

Chapter 1 is a version of a working paper Mittal and Mahajan (2017), and is currently being prepared for publication. I thank coauthor Aprajit Mahajan for his contribution to the study design and the writing. Chapter 2 is the final version of Mittal et al. (2018), which is set to appear in the proceedings of the 1st Annual ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018. I thank coauthors Aprajit Mahajan and Ofir Reich for their contribution to the study design, data analysis and writing. I have and continue to benefit from Ofir Reich's friendship as well as his wisdom on challenges in applied machine learning. Chapter 3 is a work in progress with the same title coauthored with Jan Luksic. I thank Jan Luksic for his contribution to the literature review, to the analytical strategy, and to the writing. I feel blessed to have worked with such coauthors.

This dissertation would not have been possible without the generous time, access to data, and answers to questions provided by officials in the Department of Trade and Taxes of the Government of National Capital Territory of Delhi. I thank Sanjiv Sahai, Principal Secretary, Department of Finance for his undiminishing support. The work in this dissertation was funded with UK Aid from the UK Government and grants from JPAL Governance Initiative. I am also thankful for the support that has been provided by JPAL South Asia, International Growth Center (India), and University of California Center for Effective Global Action.

I thank my friends Neha Dar, Shipra Gupta, Rahul Jha, Pranesh Nagarajan, Megha Punjani, Deepak Rajanna, Parag Taneja, and Manjari Vishnoi for letting me be stupid around them. I am guilty of taking them for granted and from each of them I have gotten more than I have deserved. I am grateful for the continuing love and support of my family. Without the support of my chacha-chachi, Mahesh Mittal and Jyotsna Mittal, this journey would have never begun. My brother Anant Mittal is my 1 man entourage. My parents, Kailash Mittal and Sarika Mittal, have always provided me with more freedom and trust than I have deserved and have tried their best to not bias me with their own priors.

Finally, all this will not be worth it without my most ardent critic, my fiercest supporter, and my partner in crime, Shiwali Mohan by my side. She is my soul food. I lucked out that she settled for me.

## VITA

### Education

- 2013 MPP, Harris School of Public Policy, Chicago
- 2007 Bachelor of Engineering, Computers, Delhi University

### Awards

- 2017-2018 UCLA Dissertation Year Fellowship
- 2013-2017 UCLA Anderson Fellowship

### Presentations

- 2018 PacDev\*; 2<sup>nd</sup> International Conf. on Globalization and Development\*; PMRC (NUS)\*; \* indicates declined
- 2017 Office of Chief Economic Advisor (Govt of India); GEM-BPP Workshop; RIDGE Public Economics (Uruguay); Center for Business Taxation Doctoral Meeting (Oxford); APPAM Fall Research Conference (Chicago); NEUDC Conference 2017 (Tufts); 2<sup>nd</sup> Zurich Conference on Public Finance in Developing Countries;
- 2016 Dept of Finance (Govt of NCT of Delhi); Graduate Student Brownbag (UC-Riverside);

### Grants

- 2018        **J-PAL Governance Initiative (\$50,000)**. Improving State Response to Public Grievances.
- 2017        **J-PAL Governance Initiative (\$7,500)**. Improving the Efficacy of Public Procurement and Public Grievance Monitoring.
- 2017        **EDI consortium (£22,000)**. Who is Bogus? Catching fraudulent firms in Delhi.
- 2016        **IGC Research Grant (£50,830)**. Where's Value? Using VAT data to Improve Compliance.
- 2015        **J-PAL Governance Initiative (\$49,050)**. Information Provision and Participatory Budgeting: Mohalla Sabhas in Delhi.
- 2015        **J-PAL Governance Initiative (\$5,000)**. Improving Public Service via the Ballot Box: Evidence from Delhi.

## PUBLICATIONS

**Shekhar Mittal**, Ofir Reich and Aprajit Mahajan. Who is Bogus? Using One-Sided Labels to Identify Fraudulent Firms from Tax Returns. In Proceedings of the 1st Annual ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018. doi: <https://doi.org/10.1145/3209811.3209824>

# CHAPTER 1

## VAT in Emerging Economies: Does Third Party Verification Matter?

### 1.1 Introduction

Improving the state’s ability to tax effectively is increasingly seen as central to the development process, and the value added tax (VAT) has been proposed as a key tool towards accomplishing this goal.<sup>1</sup> However, micro-empirical evidence on its effectiveness is relatively limited.<sup>2</sup>

VAT is a broad-based tax remitted at multiple stages of production (and distribution) with taxes paid on purchases (inputs) credited against taxes withheld on sales (output). In its most common form (known as the “invoice credit” method) firms withhold taxes on sales (output tax) from which they deduct the taxes they have already paid on purchases (input credits), and finally remit the difference to the tax authority. Thus, tax revenue is collected, by the tax authority, throughout the production chain (unlike a retail sales tax) but without distorting production decisions (unlike a turnover tax).<sup>3</sup>

---

<sup>1</sup>Besley and Persson (2013).

<sup>2</sup>Almunia and Lopez-Rodriguez (2018), Naritomi (2013), Carrillo et al. (2017) and Pomeranz (2015) discussed below are notable exceptions. See Ebrill et al. (2001) for an overview of the more aggregate cross-country evidence on the effectiveness of the VAT.

<sup>3</sup>See International Tax Dialogue (2005) for more details.

The VAT system requires both parties of a transaction to report on it separately. The parties, however, face opposed incentives: the buyer has an incentive to report the transaction – in order to receive input credit and reduce her tax liability – while the seller’s incentive is to lower her tax liability by not reporting the transaction. Such opposed incentives are believed to limit the likelihood of collusion between buyer and seller. These multiple reports also enable “third-party verification” in that the tax authority can compare buyer reports against seller reports while inspecting returns. This is often cited as a key advantage of the VAT. In practice, however, there may be significant differences in the ease with which the tax authority can verify third-party reports. For instance, as is true in the initial period of our study, the tax authority may only be able to verify third-party reports after instituting (costly and rare) audits in a lengthy multi-stage process.

In this paper we evaluate the effect of a policy that significantly eased the Delhi tax authority’s ability to implement third-party verification. Delhi adopted the VAT in 2005 but until 2012 (year 3 of our data) firms were only required to file a single aggregated return (known as a consolidated return). The consolidated return contained no information on transactions at the firm level so that the tax authority could not match buyer and seller reports directly. Any matching across buyers and sellers could only be done by initiating an audit and requesting this information from the audited firm and all firms it had interacted with over the relevant tax-period.<sup>4</sup>

Starting in the first quarter of 2012-13 all firms were mandated to file additional, detailed information about transactions with other firms. Specifically, firms were

---

<sup>4</sup>The lack of automatic cross-checking of reports appears to be a common feature of VAT systems. For instance, this was true of the VAT system in Chile discussed in Pomeranz (2015). To the best of our knowledge, this seems to be the case in other countries as well – e.g. tax authorities in Tanzania and Senegal do not systematically collect transaction level information. Carrillo et al. (2017) note that the tax authorities in Ecuador only began to collect sales and purchase data from tax-payers in 2007 and only began cross-checking reports for revenue discrepancies in 2011. Almunia et al. (2017) note that while tax authorities in Uganda collect transaction level information, the information is not used in any systematic manner for cross-checking reports.

required to provide information, at the firm level, on all purchases and sales and provide tax-identification numbers for all transacting firms. The tax authority could now (and did) relatively easily cross-check information provided by buyers with the corresponding information from sellers, directly on its own servers, without initiating an audit. In case of a mismatch, notices are automatically generated and sent out to both firms who are then required to amend their respective returns.<sup>5</sup> These notifications mark the first time that third-party information was used in a systematic way by the Delhi tax authority and is likely to be adopted in other VAT regimes as they move towards electronic return filing.

The intended goal of the policy was to reduce evasion by reducing input credit claims and by increasing output tax withholdings. However, whether this will be the case is ex-ante unclear since firms can collude (e.g by co-ordinating on off-the-book transactions) to ensure matching reports, thereby subverting third-party verification.<sup>6</sup> Furthermore, since many firms are unregistered, i.e. do not file returns, registered firms can under- or over-report (or otherwise manipulate) transactions with such firms without the fear of being uncovered by third-party verification.<sup>7</sup>

We attempt to understand the effects of this policy reform by focusing on comparisons between wholesale and retail firms. Both wholesalers and retailers face

---

<sup>5</sup>After a mismatch notice is generated, firms are given up to a year to resolve the mismatch. If still unresolved, an assessment is manually carried out by an inspector. While the policy is a significant improvement over previous practice, reconciling mismatches is still a somewhat lengthy process. The nationwide Good and Services Tax (GST), which launched on July 1, 2017, is supposed to further streamline this process. Under the GST regime, if mismatches are not resolved within 180 days all relevant input tax credits will be canceled.

<sup>6</sup>As has been noted, see e.g. Pomeranz (2015), the VAT system of third-party verification breaks down at the last step since sales to final customers are not subject to third-party verification. More generally, the system can potentially unravel upwards from any firm which is significantly under-reporting revenues. In our setting third-party verification can break down at most nodes in the chain since registered firms regularly interact with unregistered firms.

<sup>7</sup>In effect this means that net revenue disclosed to the authorities at every node is a choice variable. While there is no systematic evidence on the extent of corporate tax-evasion in India, anecdotal evidence suggests that it is high.



comparable incentives on the purchase side – they can claim input credit only for purchases from registered firms and post-reform the tax authority automatically verifies these claims against the corresponding counter-claims. However, by virtue of being higher up in the production chain, a wholesaler is more likely to sell to registered firms whereas a retailer is more likely to sell to final customers from whom no verification is possible.<sup>8</sup> Therefore, on the output side, the VAT self-enforcement and third-party verification mechanisms are more likely to break down for retailers relative to wholesalers. As a result, we would expect the policy to have a stronger effect on wholesalers relative to retailers. Our findings are consistent with this hypothesis. Using a difference-in-difference strategy we show that the policy led to a 29% increase in average tax collections from wholesalers relative to retailers in real terms. This increase was largely driven by an increase in output taxes collected by wholesalers with no differential reduction in input credits.

However, focusing on averages masks significant heterogeneity in the effect of the policy. Treatment effects are substantially higher for larger wholesalers (relative to large retailers) ranked by tax remitted at baseline. A potential explanation lies in the structure of the tax authority monitoring mechanism and the low compliance environment. 96% of the top one percent of wholesalers are monitored by a special tax assessment unit which focuses solely on high taxpayers. We do not find a comparable increase in tax collections for the top one percent of retailers – 80% of whom are also monitored by the same unit. However, unlike wholesalers the bulk of these retailers’ sales are to unregistered firms. This suggests that targeted state capacity by itself may be insufficient but when combined with increased information can improve collections even in a low compliance environment.

---

<sup>8</sup>See Naritomi (2013) for an innovative program in Brazil that attempted to address this problem with final customers. Note that in our setting final customers are observationally indistinguishable from unregistered firms. In figure 1.7 we provide evidence for the claim that retailers are more likely to sell to final customers.

The remainder of this paper is as follows: Section 1.2 reviews the relevant literature, section 1.3 summarizes a theoretical framework for understanding the empirics as well as some background on the VAT in Delhi and the policy change of interest, section 1.4 describes the data, and section 1.5 describes the empirical strategy and the assumptions underlying our causal claims. Section 1.6 presents our results. Section 1.7 explores likely mechanism for our reduced form results and section 1.8 concludes.

## 1.2 Related Literature

Researchers have argued that VAT is harder to evade than a general sales tax for at least three reasons.<sup>9</sup> First, as pointed out earlier, both input sellers and intermediate good purchasers maintain reports of transactions. In principle, this provides a reliable audit trail and serves as a form of third-party verification and should act as a preventive deterrent for firms. Second, the diametrically opposed incentives between buyer and seller should reduce the scope for collusion in these reports. Finally, taxes are remitted at all stages of production rather than only at the retail level and this is thought to render VAT less vulnerable to evasion relative to taxation at a single point (e.g with a sales tax). These arguments have proved compelling to policy makers and VAT has expanded rapidly world-wide with more than 160 countries currently deploying this system. India introduced it in 2005. There is also now a fledgling micro-economic literature that seeks to evaluate the effectiveness of the VAT and its various hypothesized mechanisms.

In an influential paper, Pomeranz (2015) reports results from a randomized experiment in Chile that increases the perceived audit probability for a group of treatment firms. She finds that the treatment had a much smaller effect on firms with paper

---

<sup>9</sup>See e.g. Agha and Haughton (1996).

trails relative to firms without such trails (who would correspond to retailers in our context). We view our work as complementary to this study in several ways. First, Pomeranz’s experiment holds the tax authority’s information set constant while increasing the audit probability (or the firms perception of the probability), while our study holds the audit probabilities constant changing instead the information set available to the tax authority and firms’ knowledge of this.<sup>10</sup> <sup>11</sup> Second, as Pomeranz notes, Chile has one of the highest tax compliance rates world-wide, while our study takes place in a low compliance environment. This difference in contexts potentially helps explain some of the differences in our results - e.g. for larger firms - as we discuss below. Third, the policy change in our study is a permanent change in the tax-regime which may result in different firm responses compared to a one-time targeted intervention and we can examine the resulting changes over a two-year horizon.

Our work is also related to the recent literature on third-party verification. Kleven et al. (2011) show that evasion rates in Denmark are significant (approximately 40%) for income that is non third-party verifiable and extremely low for income that is third-party verifiable. In an earlier paper, Slemrod et al. (2001) find that audit threat letters increase tax payments in Minnesota with the results driven by tax-payers with income that is difficult to verify from a third party (e.g. self-employment or farm income). In our context, firm interactions with other registered firms in both the pre- and post-policy periods are sensu stricto third-party verifiable. However, the act of third-party verification is automated in the post-period and our results show that this automation differentially affected firms that reported more third-party

---

<sup>10</sup>In terms of cross-checking ability the Chilean tax-regime was the same as the pre-policy regime in Delhi. The Chilean tax authority could only cross-check buyer and seller reports accurately via an audit (except for a small fraction of firms filing on-line).

<sup>11</sup>With the caveat that firms in our study realized that they would have to co-ordinate on reports in order to avoid a mis-match report.

verifiable transactions. Carrillo et al. (2017) find that Ecuadorian firms increase reported revenues when informed about revenue discrepancies (based on third-party verification exercises undertaken by the tax authority). However, they also increase (non third-party verifiable) reported costs by 96 cents for every dollar of increased reported revenue. As a result, the verification exercise results in only a minor increase in tax collections. Similarly, Slemrod et al. (2015) finds that providing credit card sales information for sole-proprietorships to the IRS increased reported revenues by 24%. However, taxpayers offset the increased revenue reports by increasing reported expenses – which are harder to third-party verify – and thereby reduced taxable income.<sup>12</sup> In our setting, non third-party verifiable purchases cannot reduce tax liabilities so such direct strategies are not possible although we do find some evidence consistent with collusion between some firms that has the same effect.

Our findings are also consistent with the theoretical model in Kleven et al. (2016) in that the effects of the improved verification are most pronounced for larger wholesale firms.<sup>13</sup> Our findings of differential effects for larger wholesale firms are also complementary to the larger bunching effects documented in Almunia and Lopez-Rodriguez (2018) for Spanish firms with more third-party verifiable transactions.

A common theme in this literature is that firms are often able to circumvent monitoring policies by changing behavior along margins not visible to the authorities. Therefore, the eventual intended effect of the monitoring policy on tax collections is relatively muted. Our work is related to this literature since the policy change we examine has differential effects on firms whose activities are less visible to the authorities relative to firms whose transactions are more visible and that in addition

---

<sup>12</sup>See also Kumler et al. (2012) for an empirical study that documents improvements in payroll tax compliance by firms as a result of a policy change that established a closer link between firm reported wages and pensions.

<sup>13</sup>Although the model in Kleven et al. (2016) is described in terms of collusion between workers and firms, the logic extends to collusion between firms; we discuss the relationship in greater detail in section 1.3.

firms can use a margin unavailable in high-compliance environments - making off the book transactions.

We also contribute to a recent literature exploring mechanisms to improve the state's tax collection ability.<sup>14</sup> Finally, our work is also relevant to the literature that attempts to understand the role of technology in improving governance in developing countries.<sup>15</sup>

### 1.3 Conceptual Framework

We organize our thinking around conceptual framework provided by Allingham and Sandmo (1972); Kleven et al. (2011) for understanding the differential effects of the policy reform on wholesale and retail firms and Kleven et al. (2016) for understanding the larger effects on bigger wholesale firms.

Kleven et al. (2011) adapt the canonical Allingham and Sandmo (1972) model of tax-compliance to incorporate third-party verifiable income. The model derives the firm's optimal choice of reported income which in turn determines the (endogenized) evasion detection probability. The key trade-off here is balancing the gains arising from tax evasion by reducing reported income against the increase in the probability of detection which is assumed to be an increasing function of the tax evaded. Further, the probability of detection is assumed to be close to one for any evasion of third-party verifiable income but significantly lower for non third-party verifiable income. Consequently, tax-payers will set reported income to be at least as high as their third-party verifiable income and will under-report the non third-party verifiable component of income.

---

<sup>14</sup>See e.g. Gordon and Li (2009), Khan et al. (2016).

<sup>15</sup>See e.g. Banerjee et al. (2016), Muralidharan et al. (2016), and the review in Finan et al. (2017).

An implication of this model is that an increase in third-party verifiable income will lead to an increase in reported income.<sup>16</sup> Within the model we interpret the automation of cross-checking buyer and seller reports as an increase in third-party verifiable income. However, the increase is larger for wholesale firms because they interact with a larger number of registered firms relative to retail firms.<sup>17</sup> The model then predicts that, *ceteris paribus*, collections should increase more for wholesale firms relative to retail firms. As we shall see, tax collections from retail firms are relatively unaffected by the reform and we discuss the implications of that finding in the context of this model below.

This previous model takes third-party verifiable income as exogenously given. A complementary framework is provided by Kleven et al. (2016) who examine the choice of how much third-party verifiable income a firm will report in the context of a cooperative game played by (potentially) colluding firms.<sup>18</sup> In the model, a given (henceforth the index) firm interacts with  $N$  other firms in a pairwise manner and each pair decides on a common report to be reported separately by each member to the tax-authority. Each member of the pair knows whether the agreed upon report is true. If the (common) reported transaction differs from the truth, then a mismatch report is filed with a certain probability (that is constant across pairs) to the tax-authority.<sup>19</sup> Such denouncement events are assumed to be independent across pairs.

---

<sup>16</sup>The setting here differs from the version of the Allingham-Sandmo model explicated in e.g. Carrillo et al. (2017) in that non-third-party verifiable purchases are not eligible for input-tax credit. This difference implies that in our setting firms will not have an incentive to respond to an increase in third-party reported sales by increasing non-third party verifiable purchases. It is possible for firms to reduce their tax liability by colluding with suppliers to increase their input tax credits. We examine this possibility in the results section below.

<sup>17</sup>Shown in figure 1.7.

<sup>18</sup>The model is intended to describe the strategic considerations involved in third-party reporting of employee wages by (a) the employer and (b) employees to the tax authority. However, with a suitable caveat, discussed below, it can be adapted to examining third-party reporting between firms.

<sup>19</sup>The reasons for the whistle-blowing report are left unmodeled but could represent for instance

It then follows that the likelihood of being denounced is increasing in  $N$ . Kleven et al. (2016) then show that evasion is decreasing in  $N$ . The key insight here is that for a index firm, collusion is harder to maintain when interacting with more firms.

As pointed out earlier, collusion, and hence evasion, under a VAT is harder since the seller and the buyer face opposing incentives. However, if sales can be under-reported at any node in the VAT chain, then collusion becomes feasible at all nodes above the under-reporting node. The basic logic is that under-reporting sales reduces the strength of the opposing incentives between the under-reporting firm and its suppliers, allowing for collusion.<sup>20</sup> In particular, the declines in tax liability via reducing reported sales can now be traded-off against the declines in input tax credit from reducing reported purchases. The firm under-reporting sales and its buyer have some scope for reaching an agreement on reducing reported purchases to the tax authority. This same logic can then be extended up the VAT chain so that all nodes above can collude and evade taxes even in the presence of third-party reporting.

In high compliance environments, the under-reporting of sales is assumed to occur only at the point of sale to the final consumer since no third-party verification is possible at that final node. However, in Delhi, the majority of our study firms make some sales that are not third-party verifiable. This is because sales to unregistered firms – i.e. firms that are not registered with the tax authority and hence do not have to file returns– occur at all nodes of the VAT chain.<sup>21</sup> Therefore, firms have the potential to under-report sales at all nodes on the VAT chain. This in turn implies that in our context the pairwise analysis of Kleven et al. (2016) can be applied to any two adjacent nodes in the VAT chain. In particular, the pairwise analysis does not depend upon collusion further down the chain (upto the final sale to consumer)

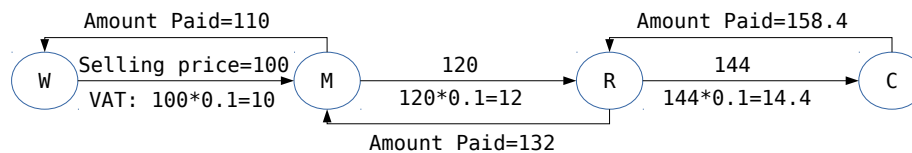
---

a breakdown in collusion, disgruntlement, error or moral concerns.

<sup>20</sup>See e.g. section 1.2 in Pomeranz (2015).

<sup>21</sup>In our study approximately 75% of firms make some positive sales to unregistered firms.

Figure 1.1: Toy Model



	W	M	R	Sale Tax(R)
Selling Price	100	120	144	144
Input Credit	0	10	12	0
Output Tax	10	12	14.4	14.4
Net Tax	10	2	2.4	14.4
Total Tax			14.4	14.4

An illustrative example describing how a value added tax system is different from a sales tax regime. Both the systems are revenue equivalent. In a sales tax system, the tax revenue is withheld and remitted only from the point at which sales are made to the final customer. However, in a VAT system, the tax revenue is withheld and remitted across all stages of production.

as is required in accounts of VAT break-down in high compliance environments.

This interpretation of the model implies that collusion, and hence tax-evasion under third-party reporting will be lower for firms that interact with a larger number of registered firms. Wholesale firms on average interact with more registered firms measured both in terms of number of firms interacted with as well as the fraction of sales that are made to registered firms when compared to retail firms. Based on the model, we hypothesize then that tax collections should increase more for wholesalers following the introduction of effective automated third-party verification.



### 1.3.1 Illustrative Example

We next outline a simple example to explicate the working of the VAT to highlight features relevant for a low compliance environment. Consider a production chain as outlined in figure 1.1 consisting of three firms and a final consumer - starting with W at the “top” of the chain on the left through to the final customer C at the “end” of the chain. Under a standard sales tax regime with full compliance with a tax rate of 10%, W and M do not withhold any tax. C pays \$14.4 (10% of 144) to R as tax and R is presumed to remit the entire amount to the tax authority. Now under a VAT regime, W withholds \$10 in tax from M (10% of 100), M withholds \$12 in tax from R, and R will withhold \$14.4 in tax from C. Finally, W will remit \$10 to the tax authority. M, however, will declare that it has already paid \$10 as tax to M and will deduct that amount from the \$12 it withheld and will remit only \$2, and similarly, R will remit only \$2.4. The amount that should finally be deposited to the the tax authority is still \$14.4. Therefore, in a system with full compliance, the VAT system collects the same tax revenue as a standard sales tax.

There are two key points worth emphasizing here. First, M (R) gets a “tax credit” (also called “input credit”) only if W (M) is registered with the tax authority. This, theoretically, should push firms which sell to registered firms to register themselves and thereby reduce informality in the system. However, in practice the effectiveness of this incentive is far from clear given the difficulties faced by developing countries in persuading firms to become formal and in monitoring the VAT system.<sup>22</sup> In our study approximately 75% of firms make some sales to unregistered firms.<sup>23</sup>

Second, as is standard in VAT systems, each firm has incentives to under-report

---

<sup>22</sup>Bird et al. (2005). See also De Paula and Scheinkman (2010) for a model showing that such incentives can also create “chains” of informality.

<sup>23</sup>Note we cannot distinguish between final consumers and unregistered firms in our data. However, sales to unregistered entities occur at all nodes on a VAT chain, not just at the end point.

sales and to over-report inputs so that a buyer and the corresponding seller have opposed incentives. For example, in the transaction between W and M, W has an incentive to not report the transaction to avoid remitting to the state any tax it has withheld while M wants to report the entire amount to maximize its tax credit. Therefore, M's incentives should act as a check on behavior of W particularly if the tax authority can credibly cross-verify M's reports with W's reports.

If, however, M can sell to unregistered firms, then the tax authority can no longer third-party verify all of its sales. This failure of third-party verification in turn opens up the possibility of collusion between W and M. For instance, if M sells \$ 60 worth of goods to unregistered firms (U), it can conceal this transaction completely since it cannot be verified by third-party reporting. This in turn provides M an incentive to collude with W to reduce reported purchases (so as to ensure that reported sales are not lower than reported purchases). For instance, they can agree upon a purchase amount of \$50. This reduces W's tax liability to \$5 relative to the no collusion scenario. This agreement does reduce M's input tax credit by \$5 as well. However, since M only reports half his sales, the decline in her output tax more than off-sets the decline in her input tax credit so that her overall tax liability falls (in this case by 50%) from \$2 to \$1.<sup>24</sup> (Refer to table 1.1)

Continuing with the example as above we can consider three different scenarios depending upon whether M and R collude. First, we assume that there is no collusion between R and M and R reports sales truth-fully. In this case, R sells output to final consumers for \$72 collecting \$7.2 in taxes of which \$6 are off-set by his input tax credit so that \$1.2 is remitted to the authority. Second, consider the case where R and M do not collude but R decides unilaterally to under-report sales to C since such sales are not subject to third-party verification. In this scenario, R can in principle

---

<sup>24</sup>Since M is not reporting the sale to U, it can give the tax-break to U by not collecting the tax or it can withhold the tax but not remit it. Anecdotally, we are aware of both the scenarios occurring in Delhi and we are indifferent between them.

Table 1.1: Evasion in Toy Model

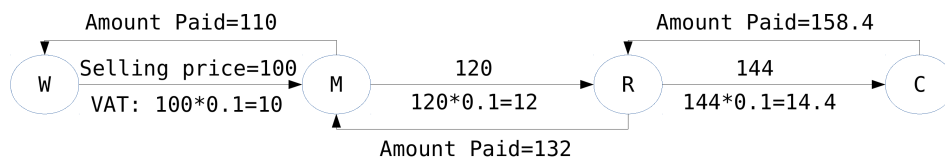
	Sales				Purchases				Tax	
	Actual		Reported		Actual		Reported		Actual	Reported
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	R	U	R	U	R	U	R	U		
W	100	0	50	0	0	0	0	0	10	5
M	60	60	60	0	100	0	50	0	$12 - 10 = 2$	$6 - 5 = 1$
R	0	72	0	60	60	0	60	0	$7.2 - 6 = 1.2$	$6 - 6 = 0$

Actual columns indicate the true values of sales or purchases. Reported columns indicate the declared values of sales or purchases. Columns (1), (3), (5), (7) show transactions with registered firms. Columns (2), (4), (6), (8) show transactions with registered firms. Tax evasion is feasible because M makes half of its sales to unregistered firms.

report any amount of sales to C but perhaps recognizing that zero (or negative) value-added in the final step may invite further scrutiny, will report the smallest amount about \$60 he is comfortable with to the tax authority. Third, M and R have incentives to collude so that the entire transaction can be carried out without involving another party U. In particular, M and R can agree to make \$60 worth of transactions on-the-books and remaining \$60 off-the-books.

The simple example above makes two points relevant for our setting. First, holding constant the probability of detection, sales to unregistered firms by a seller at any point in the chain enables pairwise collusion between the seller and his immediate supplier regardless of the presence of collusion further below (or above) the pair. This in turn implies that we can use the Kleven et al. (2016) model to analyze the implication of third-party reporting since we can limit attention to pairwise comparisons of buyers and sellers and do not need to rely on “unraveling from the bottom” arguments typically invoked to examine evasion in VAT systems. Second, one can reasonably interpret the pre-reform status quo within the model as the situation where no transactions had effective third-party verification and tax reports were not matched across buyers and sellers. The imposition of automated matching

Figure 1.2: Third Party Verification



	<b>SOLD TO</b>	<b>PURCHASED FROM</b>
W	M	NOT VERIFIED
M	R	W
R	NOT VERIFIED	M

Description of information declared by firms. For example, W will have information about M in its SOLD TO annexure and will have no information in its PURCHASED FROM annexure. Correspondingly, M will declare information about W in its PURCHASED FROM annexure, which can be used to verify sales made by W.

between buyers and sellers means that the transactions between W and M and M and R now need to be evaluated as above while the transactions between R and C are still not third-party verifiable (see figure 1.2). This difference across W and R firms in the effect of the policy along form the basis of our identification strategy.

### 1.3.2 VAT in Delhi: Policy Change

From Q1 of 2012-13 (year 3 of our data), firms were required to file two additional forms (known as Annexure 2A and Annexure 2B) in addition to their usual consolidated returns (which is referred to as Form 16, appendix B.1). The main change for our purposes is that the additional forms required firms to provide transaction details (i.e. sales and purchase information) disaggregated at the firm and tax-rate level for all registered firms and that the authority began to cross-check buyer and

seller reports automatically on its own server.<sup>25</sup>

Annexure 2B recorded all firm sales in the past tax period disaggregated at the registered buyer level for each tax rate. Annexure 2A recorded purchases disaggregated at the seller level for each tax rate (refer to appendix B.2). All firm level entries in Forms 2A and 2B had to include the tax IDs of all registered firms involved in the transaction thus enabling the tax authority to easily cross-check reports. The only across firm aggregation that was permissible was for unregistered firms (i.e. firms with no tax identification numbers which includes final consumers). The new forms meant that for the first time the tax authority could cross-check buyer and seller reports (for dis-aggregated transactions) from the submitted returns alone (i.e. without having to resort to an audit).<sup>26</sup>

## 1.4 Data

We have detailed tax data from the government of New Delhi for 5 years (from 2010 to 2015) which we describe in detail below.

### 1.4.1 VAT Returns

We have anonymized VAT returns for the entire universe of registered firms for 5 years - 2010-11 (Y1), 2011-12 (Y2), 2012-13 (Y3), 2013-14 (Y4), 2014-15 (Y5). Each firm is assigned a unique identifying number so we can follow a firm over time

---

<sup>25</sup>Different commodities are taxed at different rates. Firm A reporting transactions with Firm B would group together all transactions for commodities taxed at the same rate into a single transaction report.

<sup>26</sup>Recall that before the policy change (from 2005-2012) firms did not have to provide firm-level reports of purchases or sales but instead were only required to report total sales aggregated across all firms and correspondingly total purchases aggregated across all registered firms (and the corresponding purchase amount from unregistered firms). They were required to maintain firm-level information for their own records in case of an audit - though based on the audit notice data that we have, probability of getting audited is extremely low (less than 1%).

as well as track its presence in other firms' returns (from Y3 onwards). However, anonymization implies that we cannot link the firms to any other publicly available information on the firms.

We have detailed information on the line items in the consolidated returns (form 16) throughout the study period, and after quarter 1 of 2012-13, we have line items of form 2A and form 2B (refer to appendix B.1 for details). For the purposes of this paper, we use the following information from form 16:

1. Total turnover (sales) disaggregated by destination - (i) local (within state) sales and (ii) inter-state or international sales. Local sales are taxable and can include sales to registered firms (which are third-party verifiable) and sales to unregistered firms (which are not third-party verifiable). However, the tax forms do not require firms to differentiate between the two. After quarter 1 of 2012-13, we can use information in form 2A and form 2B to construct our own measures to categorize the sales, but form 16 does not explicitly ask for this information.
2. Total tax withheld by the firm from local sales - this is referred to as the output tax liability. This is a tax liability and needs to be deposited with the tax authority after the requisite adjustments – viz. the deduction of input credits, see below.
3. Total purchases disaggregated by destination - (i) local (within state) purchases and (ii) inter-state or international purchases. Local purchases are decomposed into purchases to registered firms (which are third-party verifiable) and purchases from unregistered firms (which are not third-party verifiable). Only local purchases from registered firms are eligible for input tax credits.
4. Total tax paid by the firm on local purchases from registered firms - this is

referred to as input credit. Input credit is subtracted from the firms' output tax liability when computing the tax the firm needs to remit to the tax authority.

5. For the three post-reform years (Y3, Y4, Y5) we also have quarterly information on sales and purchases from forms 2A and 2B as described earlier (appendix B.2). For each quarter and each tax-rate, sales made by a firm are disaggregated at the (registered) firm and tax-rate level, and likewise purchases are disaggregated at the (registered) firm and tax-rate level. Therefore, for each firm, items (1)-(4) are available at the firm quarter and tax rate level.
6. Finally, the total tax remitted by the firm to the tax authority.

Items (1)-(4) above are further broken down by tax-rates (since different goods are taxed at different rates) and we also observe some additional information such as penalties, past tax credits and liabilities.

#### **1.4.2 Firm Characteristics**

In addition to tax return information we also have basic information provided by the firm at the time of registration. We observe the date of registration, the revenue ward (i.e. the broad, largely, geographic categorization of the firm for revenue purposes), the nature of business (e.g. manufacturer, wholesaler, retail trader, exporter, importer), the legal status of the business (e.g. proprietorship, private limited company, public sector undertaking, government corporation) as well as the other tax schemes and acts the firm is subject to (e.g. the central excise act, service tax) and whether a firms is registered for international trade (import or export).

### 1.4.3 Audit Notices

We have information (for Y4 and Y5) audit notices sent out by the tax authority. These dated notices identify the targeted firm and are usually the first step in a sometimes lengthy audit procedure. We use this information to quantify the extent to which the tax authority checks on problematic returns.

### 1.4.4 Sample

We limit our analysis to firms present throughout the period of study. This sub-sample comprises 85% of all firms present in year 1 and 95% of all tax revenues in year 1. Thus, the sample consists of the near universe of revenue generating firms (see figure A.1 and figure A.2). We do not, therefore, address the effect of the policy on firm registration and exit decisions. We restrict our primary sub-sample further to firms that are exclusively wholesalers or retailers as reported on their VAT registration forms.<sup>27</sup> These comprise 27% of all firms present in year 1 and 44% of all tax revenues in year 1.<sup>28</sup> However, we use larger sub-samples for some specifications which we describe in greater detail in section 1.6. In addition, we also provide additional data description in appendix A.1.

## 1.5 Empirical Strategy

We adopt a quasi-experimental approach to examine the effectiveness of the increased information available to the tax authority.

---

<sup>27</sup>Given the previous selection rule this means we are restricting ourselves to firms registered in or before 2010-11.

<sup>28</sup>In terms of year 5, these firms comprise 19% of the sample but still contribute 43% of the tax revenues.



### 1.5.1 Identification: Wholesalers vs Retailers Over Time

Relative to wholesale firms, retail firms are more likely to sell to final consumers and conversely wholesalers are more likely to sell to registered firms.<sup>29</sup> As pointed out earlier, the change in filing requirements should therefore affect the two differentially. Following the reform, the tax authority can easily cross-check wholesaler sales to registered firms whereas previously this would only occur via an extensive and time-consuming audit process. On the other hand, sales made by retailers (any firm in general) to final consumers remain unaffected by the policy change. Finally, purchases by both types of firms from registered firms should be affected equally. This argument suggests then that if wholesalers find it harder to understate sales, then we should expect the policy to lead to an increase in taxes remitted by wholesalers driven by an increase in output tax declarations.<sup>30</sup>

Our identification strategy is thus a difference-in-difference approach comparing the difference in trends between wholesalers and retailers before and after the policy reform. The key identifying assumptions are that the timing of the third-party verification policy introduction is exogenous for our outcomes of interest and that in the absence of the reform, the (counterfactual) evolution of outcomes among wholesalers would evolve just as the (observed) evolution of outcomes among retailers (the "parallel trends" assumption). The multiple years of data before and after the reform allows us to look at the evolution of outcomes over relatively long time periods, see appendix A.3.

---

<sup>29</sup>Verified in figure 1.7.

<sup>30</sup>See also the discussion in section 1.3 which links to the relevant theoretical literature.

### 1.5.2 Model Specification

The main outcome variable is the amount of VAT remitted. However, as is typical in these settings the dispersion in tax remitted is quite large (refer to figure A.3).<sup>31</sup> Further, a large number of firms (roughly 49%) remit zero VAT. This implies that the mean may not be a representative measure of central tendency and mean regression estimates may be sensitive to outliers. We address this concern by using alternative outcome variables and estimation methods (in addition to using standard mean regressions). Specifically, we look at linear probability model using as an outcome an indicator variable equal to one if the VAT deposited is larger than zero (for the extensive margin results). In addition, we also use quantile regressions and tobit type models (although incorporating fixed-effects for a large set of firms is a computational challenge for both methods) and also estimate linear regression models over sub-samples defined by membership in deciles of relevant firm-characteristics.

Our basic specification is

$$y_{it} = \alpha_i + \nu_t + \beta * \text{Post}_{it} + \gamma * \text{Post}_{it} * \mathbb{I}\{\text{Wholesaler}_i\} + \epsilon_{it} \quad (1.1)$$

$\text{Post}_{it}$  is equal to 1 if the observation for firm  $i$  comes from the post-policy period – years 3–5 for the annual analysis and quarters 9–20 for the quarterly analysis since the policy was introduced between years 2 and 3 and quarters 8 and 9 – and zero otherwise.  $\text{Wholesaler}_i$  is a dummy variable equal to 1 if firm  $i$  self-reports as being a wholesaler and 0 if the firm self-reports as a retailer. The  $\nu_t$  are a full set of time dummies and  $\alpha_i$  are firm fixed-effects. The main outcome variables of interest are (a) an indicator for whether the firm remitted any positive amount of VAT, and (b) the amount of VAT remitted. To dig deeper into whether the effect of the policy

---

<sup>31</sup>Others e.g. Pomeranz (2015) finds similar dispersion for VAT returns.

is coming from the input side (i.e. by reducing input credits) or from the output side (i.e. by increasing the output tax liability) we also estimate regressions using (c) input credit claimed and (d) total output tax liability as outcomes. We also construct a new outcome variable which is the difference between output tax liability and input tax credits. This allows our outcome variable to be negative and takes care of the concerns arising out of there being a mass point at zero. All the outcome variables have been inflation adjusted to the corresponding baseline price levels i.e. outcome variables for the annual analysis have been indexed at the 2010 level, and outcome variables for the quarterly analysis have been indexed at the Q1-2010 level.

The parameter of interest is  $\gamma$  which captures the differential effect of the policy on wholesalers relative to retailers. Under the maintained "parallel-trends" assumption, for which we provide supportive evidence below,  $\gamma$  is consistently estimable using standard fixed-effect methods. We also present more flexible specifications that allow for time-varying differences in group means both before and after the policy reform. In particular, we include time-dummies for each period and also interaction between each time-dummy and a wholesaler dummy. Concretely,

$$y_{it} = \alpha_i + \nu_t + \gamma_t * \nu_t * \mathbb{I}\{\text{Wholesaler}_i\} + \epsilon_{it} \quad (1.2)$$

Where, in addition to the firm-fixed effects  $\alpha_i$  and the set of time-dummies (represented here in the interest of brevity and clarity by  $\nu_t$  here), we now include a full set of time-dummies interacted with the wholesaler dummy (represented here as  $\nu_t * \mathbb{I}\{\text{Wholesaler}_i\}$ ). The coefficients  $\{\gamma_s\}_{s \in S}$ <sup>32</sup> are the parameters of interest and are intended to capture the differential effect of the policy on wholesalers (relative to retailers) in period  $s$  relative to the period immediately prior to the policy's introduction. Finally, in all analysis standard errors are clustered at the firm level. In

---

<sup>32</sup>S denotes the set of post-policy time periods.

appendix A.2.1 we present additional robustness in quarterly results by adding  $Pre_{it}$  dummy which is equal to 1 for tax quarters 7 and 8 i.e. just before the introduction of the policy.

## 1.6 Results

### 1.6.1 Description: Wholesalers vs Retailers

Our main sample comprises of 32,979 retailers and 19,515 wholesalers who file a return during the entire period of our study.<sup>33</sup> Pre-policy means in year 1 along with changes in means over the pre-policy period for the two groups are shown in table 1.2. In general, wholesalers are considerably larger than retailers; we provide additional data description in appendix A.1 and discuss evidence of parallel trends in section 1.6.1.1.

As mentioned earlier, these two groups account for a substantial part of VAT collections. In year 1, both groups remitted 44% (₹46.7 billion) of total VAT collections and the corresponding number is 43% for the last year of the study. In total, they account for 55.5% of the increase in the VAT collections from the sample of firms present in each of the 5 years. Figure 1.3 and figure A.5 show the trends of VAT remitted at the annual as well as the quarterly level for the two groups (along with point-wise 95% confidence intervals).

#### 1.6.1.1 Supporting Evidence for the Parallel Trends Assumption

We begin by providing some evidence that changes in the outcome variables of interest were evolving similarly among wholesalers and retailers in the periods prior

---

<sup>33</sup>This is when returns are aggregated to the annual level. We discuss the sample size when we consider quarterly frequency regressions in appendix A.2.

Table 1.2: Summary Statistics: Retailer and Wholesaler Sample

Variables	(1) Retailers	(2) Diff	(3) Wholesalers	(4) Diff
% Positive VAT Deposited Firms	0.59 (0.00)	-0.00 (0.00)	0.53 (0.00)	-0.00 (0.00)
VAT Deposited	0.64 (0.06)	0.02 (0.02)	1.31 (0.20)	-0.01 (0.05)
Total Turnover	24.27 (1.31)	2.23 (0.74)	80.80 (6.47)	4.70 (2.59)
Turnover (Local)	18.43 (1.18)	1.89 (0.68)	49.72 (4.45)	4.66 (1.23)
Credit Claimed	0.95 (0.04)	0.10 (0.01)	1.41 (0.24)	0.23 (0.08)
Output Tax	1.53 (0.08)	0.12 (0.02)	2.63 (0.41)	0.22 (0.07)
Tax Remitted/TotalTurnover	0.01 (0.00)	-0.00 (0.00)	0.01 (0.00)	-0.00 (0.00)
Credit/TotalTurnover	0.11 (0.04)	-0.07 (0.04)	0.07 (0.03)	-0.04 (0.03)
Output Tax/TotalTurnover	0.05 (0.00)	0.00 (0.00)	0.03 (0.00)	0.00 (0.00)
Nonlocal Turnover/TotalTurnover	0.25 (0.00)	0.00 (0.00)	0.37 (0.00)	-0.00 (0.00)

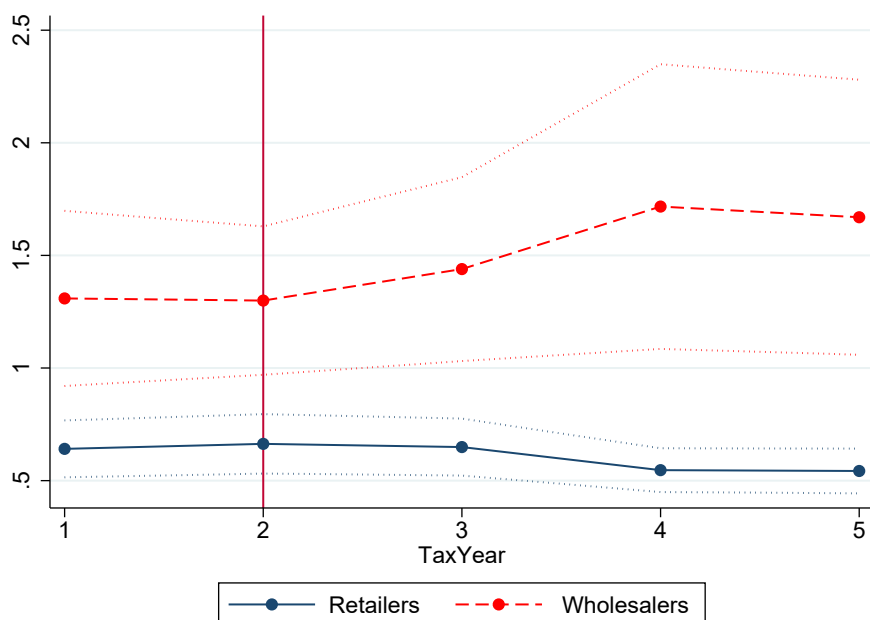
Summary statistics for selected variables in Year 1. Amounts are in million rupees, with ₹65 approximately equal to \$1.  $N_W = 32979$ ,  $N_R = 19515$ . 951 wholesalers and 521 retailers report zero turnover. Column (2) and Column (4) report mean differences between real values of year 2 and year 1. Values have been price adjusted in year 1 terms. Standard error in parenthesis.

to the reform. We examine the key outcome variable, tax remitted by firms. Figure 1.3 presents group-wise means for tax remitted in each period and figure A.6(a) graphs the coefficients for the difference between the two groups in each period from estimating equation (1.2). We cannot reject the null hypothesis that the changes in tax remitted during the first two periods were the same across wholesalers and retailers (p-value=0.61). We also carry out the same analysis at the quarterly level (Figures A.5 and A.7(a)) which yields the same conclusions.<sup>34</sup>

This gives us some confidence that the key (untestable) assumption of parallel

<sup>34</sup>In appendix A.3.1 we describe the formal statistical tests for testing the absence of differential pre-trends in more detail.

Figure 1.3: Wholesalers vs Retailers: Annual Trends



$N_W = 19515, N_R = 32979$ . Average VAT remitted is in million rupees, inflation adjusted to 2010-11 price levels, with ₹65 approximately equal to \$1. Pointwise 95% CI included. Third party verification started in year 3.

trends may be reasonable when examining tax remits. We also carry out the same analysis for the other outcomes of interest: (a) Output Tax, (b) Input Credit, (c) the Difference between the two and (d) whether any VAT was remitted. In all cases we cannot reject the null that the changes in outcomes in the pre-policy period were similar between the two groups.

### 1.6.2 Results: Difference in Difference

Figures 1.3 and A.6(a) also suggest that wholesaler tax remits increased post-policy while retailer remits remained more or less unchanged. The regressions below formally confirm these conclusions. We begin by examining changes on the “extensive” margin – changes in whether a firm remits any tax with the tax authority – before

Table 1.3: Diff-in-Diff: Wholesalers and Retailers (Annual)

VARIABLES	(1) Positive VAT Remitted Firms	(2) VAT Remitted	(3) Tax Credit	(4) Output Tax	(5) Output Tax - Input Credit
Post*Wholesaler	-0.02*** (0.00)	0.38*** (0.14)	-0.12 (0.15)	0.25** (0.11)	0.37*** (0.14)
Post	0.04*** (0.00)	-0.09* (0.05)	0.18*** (0.05)	0.09** (0.04)	-0.09** (0.05)
Mean Dep.Var.	.53 (.00)	1.31 (.20)	1.41 (.24)	2.63 (.41)	1.22 (0.20)
Observations	262,470	262,470	262,470	262,470	262,470
R-squared	0.63	0.89	0.83	0.97	0.89
Number of Firms	52,494	52,494	52,494	52,494	52,494

Robust standard errors in parentheses, clustered at firm level.  $N_W = 19515$ ,  $N_R = 32979$ . Monetary amounts are in million rupees, with ₹65 approximately equal to \$1. All monetary amounts have been inflation indexed to 2010-11 price levels. Column (1) shows linear probability regressions of the probability of depositing a positive amount. Column (2)-(4) respectively show regression of the mean VAT remitted by firms, input credit claimed by firms, and output tax collected by firms. To address the concern that VAT deposited has a significant mass at zero, Column(5) shows regression of the difference between output tax and input credit declared by firms. Mean Dep. Var. shows the mean and standard errors for wholesaler firms in year 1. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

examining changes in the amount remitted. We then examine the effect of the policy on the two key constituents of a firm’s tax obligation — input credit and output tax liability.

Table 1.3 shows the results of the difference-in-difference regressions at the firm-annual level. In column (1) the outcome is a dummy variable for whether the firm remits any tax. The proportion of wholesaler firms depositing any tax goes down by a statistically significant 2%, though this is a relatively modest decrease of 4% from a baseline of 53%. Next, VAT remitted (column 2) increases by a substantively (and statistically) significant ₹0.38m for wholesale firms. This is an increase of 29% over a year 1 mean of ₹1.31m. The wholesalers response is particularly notable since it is in start contrast to the retailer response – for retailers total tax remitted actually decreases post-policy. These result are consistent with the conceptual framework outlined in section 1.3. Wholesalers who have higher third-party verifiable income and interact with more registered firms respond much more to the improvement in third-party verification relative to retailers.

We next examine the proximate causes of the changes in tax remitted by examining changes in output tax liability and input tax credit respectively. Consistent with the arguments in section 1.3 we see that there is a substantive increase in the output tax liability for wholesalers relative to retailers – a 9.5% increase from year one. Further, we see that input credits decline for wholesalers relative to retailers though the estimate is not significant at conventional levels. Overall, the signs of the wholesaler response along each component dimension are then consistent with the predictions of the framework outlined in section 1.3. Retailer behavior is somewhat harder to rationalize. A rise in input credits is consistent with increased collusion (to minimize tax liability) or an accurate measure of purchases if under-reporting purchases is harder post-reform (recall retailers who under-report sales will have an incentive to also under-report purchases to avoid having value added estimates that are suspiciously low). We also see an increase in output tax liability for retailers post-policy and this is somewhat unexpected, particularly since retailers make most of their sales to unregistered firms. We explore this further in the next section when we examine heterogeneity in returns. Column (5) examines the difference between output tax liability and input credits and is consistent with the argument that retailers are able to offset increases in output tax liability by matching increases in input credits (as in e.g. Carrillo et al. (2017)) while wholesale firms are unable to do so.<sup>35</sup> Finally, we note that the asymmetry between effects on input credits and output tax liability suggests that that effects on wholesalers cannot be easily ascribed to a differential growth account.

As a robustness check, table A.3 repeats the regressions whose results are presented in table 1.3 but at the quarterly frequency.<sup>36</sup> The results are consistent with

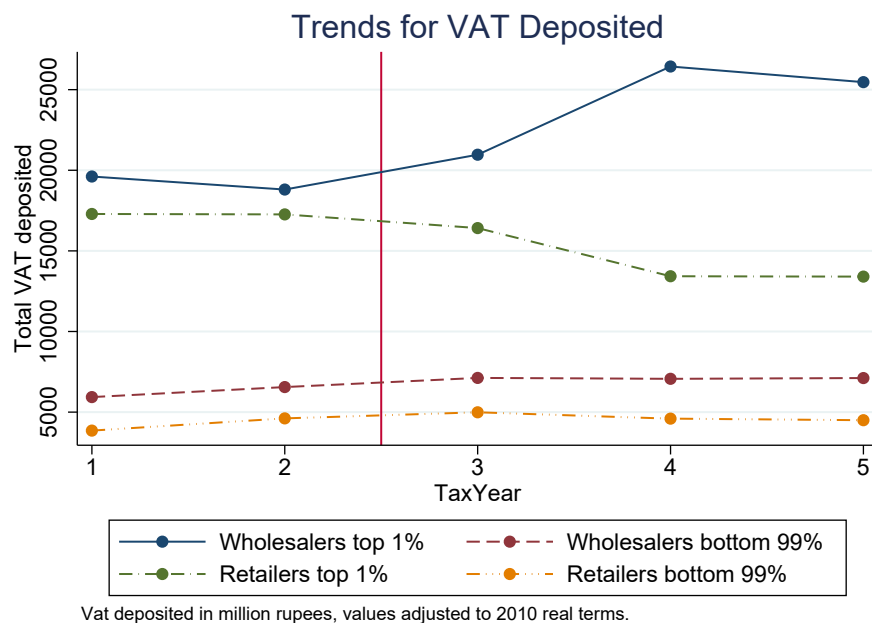
---

<sup>35</sup>The reason Col (5) is not the same as Col (2) is because some firms have negative tax liability and hence remit no tax.

<sup>36</sup>The sample is now reduced to 11,482 wholesalers and 15,337 retailers, as firms with less than ₹5 million in turnover only had to submit returns annually or semi-annually in the first two years



Figure 1.4: Tax Remitted by Top Percentile and Bottom 99% of Firms



Notes: This figure plots the total VAT remitted by the panel of top 1% and the bottom 99%, in terms of VAT remitted in year 1, of wholesalers and retailers.  $N_W = 19515$ ,  $N_R = 32979$ . Therefore, top 1% of retailers are 329 firms and top 1% of wholesalers are 195 firms. Rest of the firms form bottom 99%. Monetary amounts are in million rupees, inflation adjusted to 2010-11 price levels, with ₹65 approximately equal to \$1.

the results described at the annual frequency and so we do not discuss them here. Moreover, as described earlier, we carry out event falsification test by looking at coefficients on  $\text{Pre}_{it} * \mathbb{I}\{\text{Wholesaler}_i\}$  where  $\text{Pre}_{it}$  is 1 if tax-quarter is 7 or 8 i.e. just before the introduction of the policy. We find that pre-policy effect on wholesalers is close to zero, precisely estimated, and statistically significant.<sup>37</sup>

of our data.

<sup>37</sup>More details in appendix A.2.1.

Table 1.4: Diff-in-Diff: Wholesalers and Retailers (Annual, Top Percentile)

VARIABLES	(1) Positive VAT Remitted Firms	(2) VAT Remitted	(3) Input Credit	(4) Output Tax	(5) Output Tax - Input Credit
Post*Wholesaler	0.02* (0.01)	34.75** (13.62)	-15.94 (13.38)	19.96** (8.89)	35.90** (14.06)
Post	-0.02** (0.01)	-12.02** (4.68)	9.31** (4.07)	-3.47 (3.65)	-12.78*** (4.46)
Mean Dep.Var.	1 (0.00)	100.59 (18.55)	36.27 (23.42)	138.29 (39.17)	102.02 (18.39)
Observations	2,620	2,620	2,620	2,620	2,620
R-squared	0.42	0.88	0.84	0.98	0.88
Number of Firms	524	524	524	524	524

Robust standard errors in parentheses, clustered at firm level.  $N_W = 195, N_R = 329$ . Monetary amounts are in million rupees, inflation adjusted to annual 2010-11 price levels, with ₹65 approximately equal to \$1. Column (1) shows linear probability regressions of the probability of remitting a positive amount. Column (2)-(4) respectively show regression of the mean VAT remitted by firms, input tax credit claimed by firms, and output tax collected by firms. To address the concern that VAT remitted has a significant mass at zero, Column(5) shows regression of the difference between output tax and input credit declared by firms. Mean Dep. Var. shows the mean and standard errors for top 1% wholesaler firms in year 1. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

### 1.6.2.1 Heterogeneity

We next turn to exploring heterogeneity in our estimates. There are two main reasons for this. First, as we saw in the previous table, there is a significant point mass at zero for tax remitted. This suggests that exploring alternatives to mean regressions may be a useful exercise. Further, table 1.3 indicates that while the policy had limited extensive margin effects (i.e. on whether firms remit any tax) there were substantive intensive margin effects, suggesting the the policy may have affected different wholesalers differently.

We begin by examining the evolution of tax remitted for different sub-groups of wholesalers and retailers. In particular, we partition each group into two sub-groups ranked according to tax remitted in the first year of the study – the top 1% and the remaining 99%.<sup>38</sup> The results are graphed in figure 1.4; two points

<sup>38</sup>We use tax remitted as the stratifying variable because it is also the variable used by the authority to determine firms that will be examined by the special ward in charge of large tax-payers.

are worth noting. First, tax remits remain relatively stable for the bottom 99% of wholesale firms as compared to the top 1%. In contrast, tax remits for retail firms as a whole (both in the 1% as well as in the 99% sub-groups) are relatively stable throughout the study period. This analysis suggests that it is worthwhile to examine heterogeneity in treatment effects across different percentiles of the baseline outcome distribution.<sup>39</sup> In tables 1.4 and 1.5 we present the regression adjusted comparisons of figure 1.4. There are at least three points of interest. First, the policy has no differential effects on wholesalers relative to retailers in the bottom 99% sub-sample. Second, the response of the top 1% of wholesalers is sharply different from that of the top retailers. For the top retailers, input credits increase and output tax declines modestly<sup>40</sup> so that overall tax remits fall by 22.9% (relative to a top 1% retailer baseline of ₹52.55m) after the policy. In stark contrast, input credits decline and output tax liability increases for wholesalers (relative to top retailers) so that post-policy overall tax deposited rises by 34% (relative to the baseline mean).

These results confirm the differential effects of the policy and also help us better understand firm responses. First, the opposing signs and differences in magnitudes and significance of the output tax response between top wholesalers and retailers is consistent with the hypothesis that top wholesalers are more constrained in their responses.<sup>41</sup> Further, top wholesalers are both more likely to be monitored by a special tax assessment unit and have more third-party verifiable income. In contrast, top retailers are less likely to be monitored by the special unit and also have much less third-party verifiable income. The lack of any differential response among the bottom

---

<sup>39</sup>A natural estimation method here would be to use quantile regressions. We are currently working on implementing a stable quantile regression model with fixed effects on our data.

<sup>40</sup>The output tax response is not statistically significant at conventional levels.

<sup>41</sup>As pointed out earlier a differential growth story for top wholesalers is less plausible since input credit and output tax respond in opposite directions while increased growth via sales would be more likely to be reflected in increased purchases as well.

Table 1.5: Diff-in-Diff: Wholesalers and Retailers (Annual, Bottom 99 Percentile)

VARIABLES	(1) Positive VAT Remitted Firms	(2) VAT Remitted	(3) Input Credit	(4) Output Tax	(5) Output Tax - Input Credit
Post*Wholesaler	-0.02*** (0.00)	0.03 (0.03)	0.04 (0.06)	0.05 (0.07)	0.02 (0.02)
Post	0.04*** (0.00)	0.03*** (0.01)	0.09*** (0.02)	0.13*** (0.02)	0.03*** (0.01)
Mean Dep.Var.	0.53 (0.00)	0.31 (0.01)	1.06 (0.06)	1.26 (0.06)	0.20 (0.02)
Observations	259,850	259,850	259,850	259,850	259,850
R-squared	0.63	0.58	0.78	0.78	0.82
Number of Firms	51,970	51,970	51,970	51,970	51,970

Robust standard errors in parentheses, clustered at firm level.  $N_W = 19,320$ ,  $N_R = 32,650$ . Monetary amounts are in million rupees, inflation adjusted to 2010-11 price levels, with ₹65 approximately equal to \$1. Column (1) shows linear probability regressions of the probability of remitting a positive amount. Column (2)-(4) respectively show regression of the mean VAT remitted by firms, input tax credit claimed by firms, and output tax collected by firms. To address the concern that VAT remitted has a significant mass at zero, Column(5) shows regression of the difference between output tax and input credit declared by firms. Mean Dep. Var. shows the mean and standard errors for wholesaler firms in year 1. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

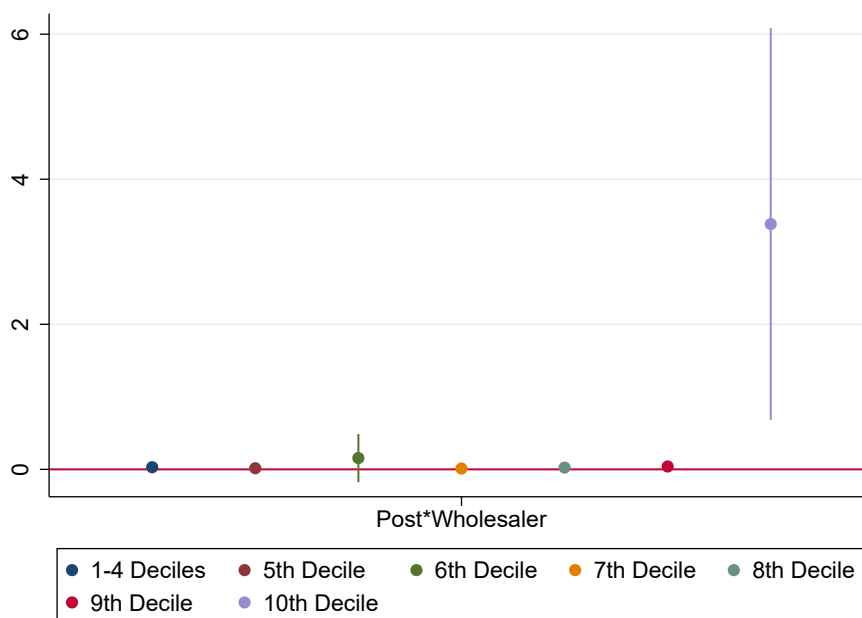
99% of wholesalers (relative to retailers) in turn suggests that differences in third-party verifiable income alone (without commensurate monitoring) are not sufficient for generating greater collections. These differential effects therefore provide some sobering evidence on the effectiveness of the VAT at increasing collections in low compliance environments.<sup>42</sup> Second, retailers (both large and small) increase input credits claimed after the policy while the wholesaler response is either much more muted or even negative. These results are consistent with the retailers relatively higher ability to collude with a smaller number of input supplies.

As an alternative approach we estimate equation (1.1) separately for different deciles based on the baseline (year one) distribution of tax remitted.<sup>43</sup> Figure 1.5 plots the coefficient of interest ( $\gamma$  from eq. 1.1) from each of the seven separate regressions though we present regression results only from the regressions using the

<sup>42</sup>Almunia et al. (2017) find substantial discrepancies in third-party reports in Uganda and also emphasize the limited enforcement capacity of the state as an important constraint. In our context, discrepancies are less of a concern post-policy, because of automatic matching, relative to collusion.

<sup>43</sup>We constructed deciles using group-specific distributions of tax remitted in year one.

Figure 1.5: Heterogeneity Analysis: VAT Remitted for Wholesalers vs Retailers



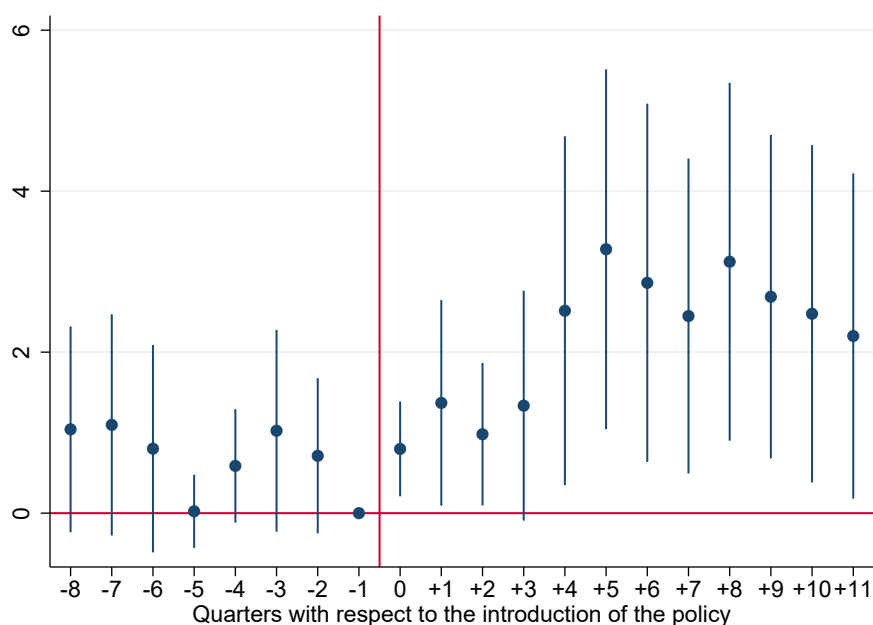
Notes: This figure plots the difference between wholesalers and retailers for different deciles, based on VAT remitted in the first year. The x axis indicates the decile. Confidence intervals at the 95% level.  $N_W = 32979, N_R = 19515$ . Coefficients are in million rupees, inflation adjusted to 2010-11 price levels, with ₹65 approximately equal to \$1. The first group consists of 4 deciles because all the firms in that group do not remit any VAT.

top decile subsample in table 1.6.<sup>44</sup> Figure 1.6 shows the event-study plots for the top decile of wholesalers and retailers on VAT remitted as an outcome variable. Other outcome variables are shown in figure A.9.

The figures and table are consistent with the earlier graph. The treatment effect is only substantively significant for the top decile, and is relatively negligible for all other deciles. Tax remitted by wholesalers in the top decile goes up by ₹3.38m, a 26.7% increase over the ₹12.6m remitted in year one. Given the relative stability of retailer outcomes across deciles, the results imply that it is only the biggest wholesalers that are driving the large increase in tax remitted in the mean regressions in

<sup>44</sup>The remaining regression results are omitted for brevity and are available upon request.

Figure 1.6: Top Decile: VAT Remitted for Wholesalers vs Retailers



Notes: Graphical event-study analysis of the effect of third party verification policy on wholesalers compared to retailers. Limiting the set of wholesalers and retailers to the top 10% of each group in terms of VAT remitted in quarter 1. Confidence intervals were constructed with heteroskedasticity-robust standard errors, clustered at the firm level. The coefficient for the wholesale group in the quarter (-1) prior the policy is normalized to zero. The regressions include firm fixed effects and time effects. The x axis indicates time, with quarterly observations and zero indicates the first year of the third party verification policy. We have 20 quarters of data from 2010-11 to 2014-15. Confidence intervals at the 95% level.  $N_W = 1148, N_R = 1533$ . Coefficients are in million rupees, with inflation adjusted to Q1 2010-11 price levels, with ₹65 approximately equal to \$1. Pretrends are not statistically significant.

table 1.3.

There are at least two possible reasons for the differential effects on the largest wholesalers. First, 96% of the top 1% of wholesalers in our sub-sample are assessed by a special tax unit (known as the Key Customers Services Unit) that is tasked with collections from large firms (the corresponding figure for the top 1% of retailers is 80%) and which has greater resources for inspection and other activities.<sup>45</sup> Second,

---

<sup>45</sup>See Das-Gupta et al. (2004) for a discussion of similar units in the context of personal income tax in India.

Table 1.6: Diff-in-Diff for Top Decile: Wholesalers and Retailers (Annual)

VARIABLES	(1) Positive VAT Remitted Firms	(2) VAT Deposited	(3) Tax Credit	(4) Output Tax	(5) Output Tax - Input Credit
Post*Wholesaler	0.02*** (0.01)	3.38** (1.38)	-1.77 (1.43)	1.68 (1.04)	3.46** (1.42)
Post	-0.06*** (0.00)	-1.11** (0.47)	1.00** (0.42)	-0.21 (0.39)	-1.20*** (0.45)
Mean Dep.Var.	1 (.00)	12.65 (1.97)	6.98 (2.40)	19.73 (4.04)	12.75 (1.95)
Observations	26,240	26,240	26,240	26,240	26,240
R-squared	0.41	0.89	0.84	0.97	0.89
Number of Firms	5,248	5,248	5,248	5,248	5,248

Equation (1.1) results of the effect of third party verification policy on wholesalers compared to retailers. Limiting the set of wholesalers and retailers to the top 10% of each group in terms of VAT remitted in year 1. Robust standard errors in parentheses, clustered at firm level. Number of wholesalers is 1951 and number of retailers is 3297. Monetary amounts are in million rupees, inflation adjusted to 2010-11 price levels, with ₹65 approximately equal to \$1. Column (1) shows linear probability regressions of the probability of remitting a positive amount of VAT. Column (2)-(4) respectively show regression of the mean VAT remitted by firms, of the input tax credit claimed by firms, and the output tax collected by firms. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

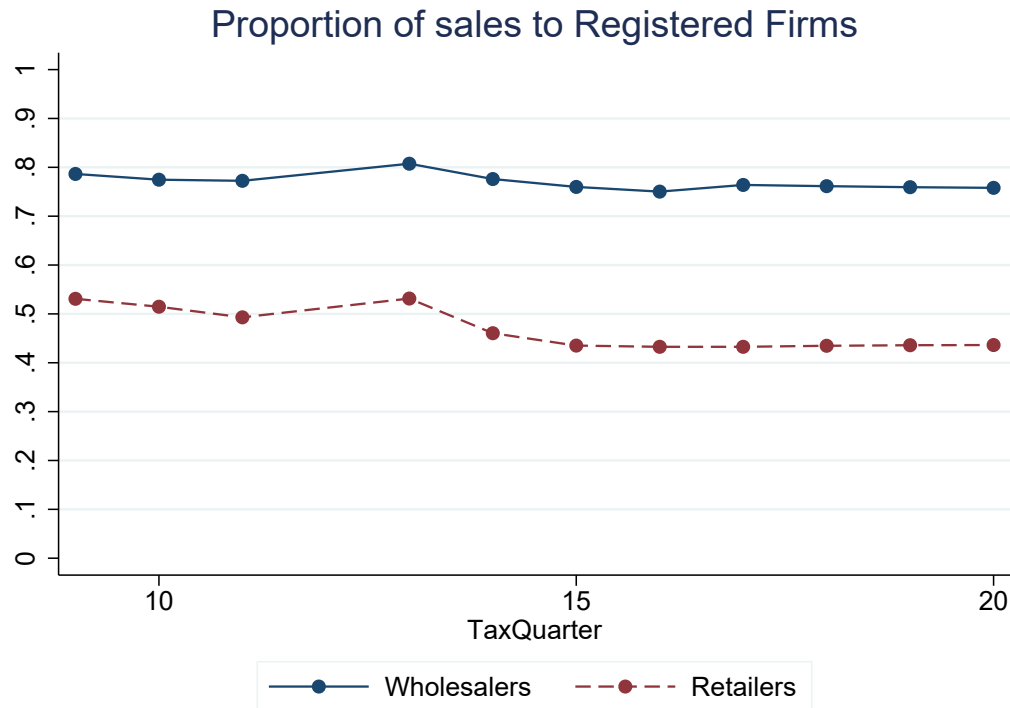
as we show below, the ease of third-party verification had a much larger effect on the top 1% of wholesalers (relative to other wholesalers and retailers) since the bulk of their sales were to other registered firms. The combination of these two factors is consistent with the results in the table above. We are currently seeking more information on the special ward assigned for larger firms.

## 1.7 Mechanism

In this section we explore potential mechanisms for the reduced form results above. Specifically, we examine the pattern of sales made to registered firms and how it evolved over time. This is important because the conceptual framework outlined in section 1.3 highlighted the role of third-party verifiable transactions as a moving force and in our context only interactions with registered firms are third-party verifiable.<sup>46</sup>

<sup>46</sup>We explore other a few other implementation details that shed light on how the policy was actually executed on the ground in appendix A.5.

Figure 1.7: Sales to Registered Firms



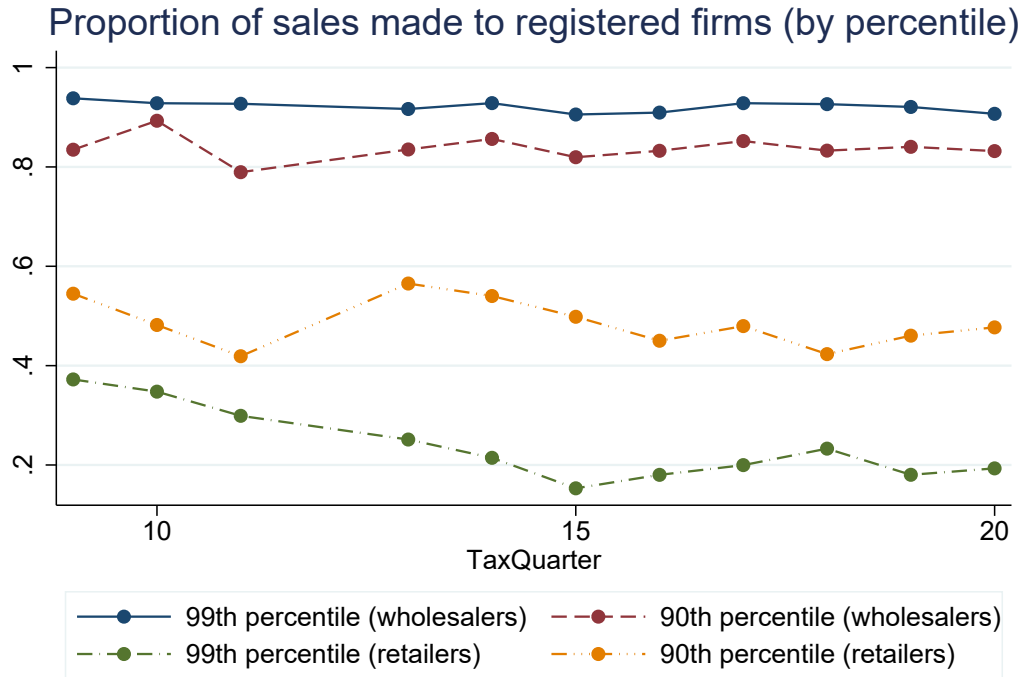
Proportion of total sales made to registered firms for wholesale and retail firms.

### 1.7.1 Sales to Registered Firms

Figure 1.7 displays quarterly sales to registered firms as a proportion of total sales separately for wholesalers and retailers. The graph is only from quarter 9 (year 3) onwards because in the pre-policy period firms were only required to report a total sales amount without further disaggregation. The lack of a complete series on this variable also limits its use as a direct measure of formality which could be used to examine heterogeneity (although we do present some suggestive evidence below). In addition, changes along the extensive margin for interacting firms also complicates the interpretation (e.g. if the policy led to more firms becoming registered). Over the entire post-policy period wholesale firms made approximately 78% of their sales to registered firms while the corresponding figure for retail firms is 43%.



Figure 1.8: Top Percentile: Sales to Registered Firms



Disregarding the 12th quarter to show a cleaner picture

Describing the proportion of sales made to registered firms. Comparing the declarations by top percentile of wholesalers and retailers with the 90th percentile.

We next examine this proportion for the 99<sup>th</sup> and 90<sup>th</sup> percentile for each group in figure 1.8. There are three points worth noting: First, over 90% of total sales are made to registered firms by the top percentile of wholesale firms (and the figure is similar, though somewhat lower for the 90<sup>th</sup> percentile). Second, a significant proportion of sales by exclusively retail firms are made to registered firms. This is prima facie surprising and we discuss its interpretation below. Finally, the proportion of total sales made to registered firms is lower ( $\approx 20\%$ ) among the top percentile of retail firms compared to the 90<sup>th</sup> percentile ( $\approx 50\%$ ), suggesting that smaller retailers are more likely to sell to registered firms.

There are at least four potential reasons for the retailer related findings. First, a larger retailer may in turn sell to smaller (registered) retailers. This could explain

the results for the top percentile of retailers but would not suffice for explaining why smaller retailers make more sales to registered firms than larger retailers.

A second possibility is that at least some of the reported sales to registered firms by retailers are in fact fraudulent – that they are created to provide fraudulent input credits and hence reduce tax collections.<sup>47</sup> We are exploring this possibility in ongoing work and in particular using network analysis to examine further the characteristics of registered firms who make purchases from small retailers.

A third explanation is that greater under-reporting of sales to unregistered firms by smaller retailers (relative to larger retailers) leads mechanically to smaller retailers reporting a higher fraction of sales to registered firms relative to larger retailers. Finally, the data may be an accurate record of retailer behavior rather than a reflection of any evasion. At the moment we do not have enough information to distinguish between these competing explanations though we are pursuing further work here as well.

We directly examine the relationship between the proportion of sales to registered firms and the change in tax collections as a result of the policy change. We conduct a triple-difference comparison by comparing the difference between wholesalers and retailers with a higher fraction of sales to registered firms to the corresponding difference between firms with a lower fraction of sales and finally doing these comparisons

---

<sup>47</sup>A simple example is the following: retailer A declares a proportion of her sales to final customers as in fact having been made to firm B. Firm A's tax obligation does not change relative to the counter-factual where all sales are reported as sales to final customers. On the other hand, Firm B reduces his tax liability so that, after accounting for the probability of detection, there is an incentive for both parties to conclude the transaction and for B to make a side payment to A.

Table 1.7: DiDiD: Wholesalers/Retailers; Before/After;Sales to Registered Firms

VARIABLES	(1) Positive VAT Remitted	(2) VAT Remitted	(3) Input Credit	(4) Output Tax	(5) Output Tax - Input Credit
Post*Wholesaler*Registered Sales	0.02** (0.01)	0.17** (0.07)	-0.04 (0.07)	0.12* (0.06)	0.16** (0.08)
Post*Wholesaler	-0.02*** (0.00)	0.06** (0.02)	-0.01 (0.02)	0.05** (0.03)	0.07*** (0.02)
Post*Registered Sales	0.01 (0.01)	0.02 (0.02)	0.02 (0.02)	0.04** (0.02)	0.03 (0.02)
Post	0.01** (0.00)	-0.07** (0.04)	0.03 (0.03)	-0.04 (0.04)	-0.07** (0.03)
Observations	536,380	536,380	536,380	536,380	536,380
R-squared	0.55	0.86	0.78	0.96	0.86
Number of Firms	26,819	26,819	26,819	26,819	26,819

Regressions run at the quarterly level on the set of firms that file returns in all quarters.  $N_W = 11482, N_R = 15337$ . All regressions include time dummies and firm fixed-effects. Column (1) displays results from a linear probability model where the outcome is a dummy for any VAT remitted. The outcomes in Column (2)-(4) are VAT remitted, input credit claimed and output tax collected. The outcome in column (5) is the difference between output tax and input credit (as one method to deal with the point mass at zero in the VAT remitted outcome). Monetary amounts are in million rupees, inflation adjusted to 2010-11 price levels, with ₹65 approximately equal to \$1. Robust standard errors in parentheses, clustered at the firm level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

before and after the policy reform for the the third difference. Specifically,

$$\begin{aligned}
 y_{it} = & \alpha_i + \nu_t + \beta * Post_{it} + \delta * Post_{it} * PropRegistered_i \\
 & + \gamma * Post_{it} * \mathbb{I}\{Wholesaler_i\} + \lambda * Post_{it} * \mathbb{I}\{Wholesaler_i\} * PropRegistered_i + \epsilon_{it}
 \end{aligned}
 \tag{1.3}$$

In equation (1.3), in addition to what we have already described for equation (1.1), we now also add a variable called  $PropRegistered_i$  which is the proportion of sales declared to be made in quarter 9 (the first quarter immediately after the introduction of the policy). Now the coefficient of interest is  $\lambda$  which tells us the effect of the policy on wholesaler firms which make greater proportion of their sales to registered firms. Another coefficient of interest is  $\delta$  which shows the effect on retailers.

The central weakness with this strategy is that, as pointed out earlier, we can construct sales to registered firms separately from total sales only in the post-policy period. In the pre-policy phase firms were only required to report total sales (which was the basis of their output tax liability). We use the fraction of sales made to

registered firms in the first post-policy quarter (quarter 9) as our measure of sales to registered firms – and hence our static measure of third-party verifiable income at the start of the policy.

The results in table 1.7 show that the treatment effect of the automation policy is stronger for wholesale firms who sell more to registered firms. Somewhat surprisingly, retailers with a larger fraction of sales to registered firms do not see any differential increases in tax remitted (or tax credits or output tax liability). This finding is again consistent with the first three explanations and high sales to registered firms is definitely not a feature of the actual transactions that retailers carry out.

## 1.8 Conclusion

In this paper we evaluate the effect of a policy reform that implemented a key pillar of the VAT system – increasing the tax authority’s ability to easily cross-check buyer and seller reports against each other. Under the previous regime such cross-checks could only be carried by instituting a rare, lengthy and expensive audit process. We interpret this change in policy as one that increased the number of transactions that were effectively third-party verifiable by the tax authority. We evaluate the effect of this policy change in the Indian state of Delhi – a low compliance environment with a large number of transactions being made to unregistered firms (i.e firms that do not file returns with the tax authority) – to shed light on the effectiveness of third-party verification in an emerging economy.

We evaluate the effect of the policy by comparing two groups of firms likely to be differentially affected by it - wholesalers and retailers. In particular, wholesalers are more likely to sell to registered firms relative to retailers. We hypothesize that requiring transacting firm tax identifiers in returns should affect wholesalers more than retailers. Further, given the invoice-credit structure of the VAT systems in

India, we expected that the primary wholesaler response would be through increases in output tax liability.

Our results confirm these hypotheses with wholesalers increasing tax remits by ₹0.38 million (an increase of 29% relative to the pre-policy period) driven by increases in output tax liability. We further find that increases in tax remits are concentrated among the very largest wholesale firms, with the top 1% of firms accounting for 89.3% of the increase in tax remitted. Next, we note that 96% of wholesale firms in the top 1% are under the jurisdiction of a special unit of the tax authority which focuses exclusively on high tax revenue firms (the corresponding figure for retail firms is 80%). Our results suggest that information and monitoring are complements in that we see the strongest effect of improved third-party verification from firms that are more likely to interact with other registered firms and are more closely monitored by the tax authority. These results suggest a more nuanced picture of the state's capacity to tax and the effectiveness of third-party verification in low compliance environments.

## CHAPTER 2

# Who is Bogus? Using One-Sided Labels to Identify Fraudulent Firms from Tax Returns

### 2.1 Introduction

Machine learning (ML) models have been effective at automating various tasks traditionally performed by humans such as character recognition, image classification, sentiment analysis, translation, and fraud detection. For many of these tasks the dataset is not constructed explicitly to provide a training set but has been re-purposed for the ML application. Such datasets may suffer from several challenges: human error in labeling (noisy labels); the labeled set could be not representative of the population (selective labels); and importantly for our application, it could record the labels of only one class (one-sided labels). For example, re-purposing a dataset that documents only criminal convictions (not all trials): the cases tried could be biased (selective labels), and it could also be unknown which cases were tried but did not end in convictions (one-sided labels). In this paper we provide solutions to some of these problems, and others, such as multiple time-period data and in-sample prediction. We do this by developing an important application on a real-world dataset: finding fraudulent firms in the Delhi value added tax (VAT) system.

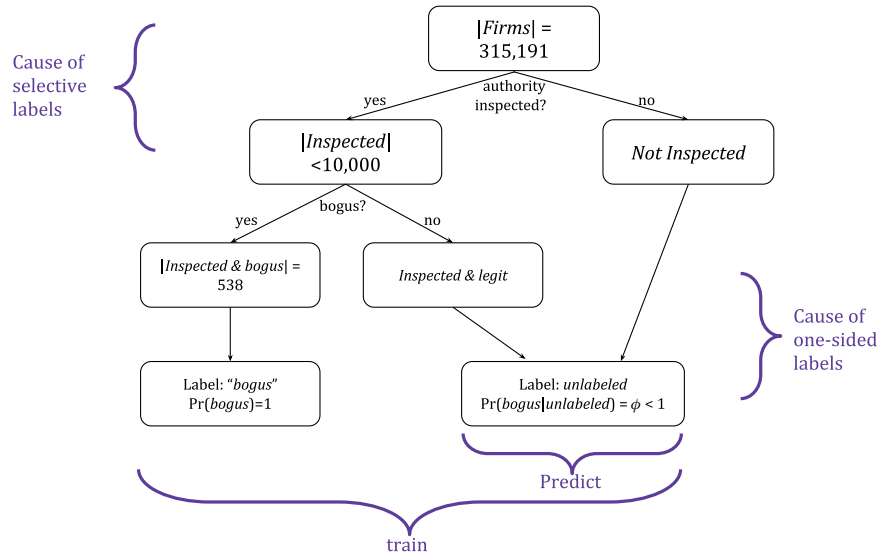
VAT implementation in many low compliance environments is plagued by firms lowering their tax liability by generating false paper trails. This demand for false paper trails has led to the creation of fraudulent firms that sell fake receipts to

genuine firms. These firms are referred to as “bogus” firms or “bill traders” by the tax authorities. A tax authority usually determines the existence of bogus firms by first filtering down based on a few preliminary indicators, and then undertaking physical inspections that verify if a firm is bogus or legitimate (“legit” for short). Given the authority’s limited resources these inspections are only done sporadically. While the true revenue implication of bogus firms is unknown, our conversations with tax officials in Delhi (India) suggest that it might be considerable (\$300 million in Delhi alone is a commonly mentioned figure). A key challenge in improving tax compliance then is to regularly, cheaply and reliably identify such fraudulent firms.

In this paper, we implement and evaluate an ML algorithm on the universe of tax returns data from the Delhi tax authority to identify such bogus firms. The classifier predicts the likelihood of a firm being bogus by using a training dataset comprising the universe of tax returns (over 3 years) and of results from past physical inspections of firms conducted by the tax officials. To our knowledge, this is the first paper to systematically study and identify these fraudulent firms in an economy with weak compliance.

We overcome several non-standard facets of the data and of our problem. First, our dataset was not constructed for the purpose of labeling for an algorithm and is, therefore, not a representative sample of the population. In our case, a small fraction of all firms are inspected, those suspected by the tax authority, and so there is a problem of selective labels Lakkaraju et al. (2017). On top of that, the tax authority only records bogus firms that were inspected, caught and canceled, but not firms that were inspected and found to be legitimate. Therefore, we are unable to distinguish between firms that were inspected and found to be legitimate and firms that were never inspected. We refer to this problem as that of one-sided labels, since no firm can be definitively labeled as legitimate (see figure 2.1). The unlabeled firms are (at least weakly) less likely to be bogus than those labeled bogus, and we leverage

Figure 2.1: One Sided Labels



Inspection is based on the tax authority’s discretion and so are biased (selective labels). Class labels are known only for firms both inspected and found to be bogus, not for the rest (one-sided labels). We use all data for training, but want to predict for those firms still unlabeled.

this difference in class distribution to train a classifier. Moreover, the authority has never identified all the bogus firms in the stock of existing VAT firms, so we want to predict the labels of all unlabeled firms. Therefore, a standard classification approach with a training set and out-of-sample prediction is not viable. We employ a form of cross-validation to use the entire dataset as a training set and make predictions on all firms in the dataset not known to be bogus.

Second, standard evaluation methods rely on a labeled test set that does not exist in our one-sided labels context. In such a scenario, the task in which the problem is embedded can provide relevant metrics to evaluate the effectiveness of classification methods. In our case, the tax authority has limited resources for inspection and so it will be able to act on only the top model recommendations. Therefore, it is reasonable to focus on the success of our top recommendations. While considering the performance evaluation results it is important to keep in mind that, according to tax



officials, bogus firms are relatively rare (i.e. the class bogus is rare in the population).<sup>1</sup> Our performance estimates remain valid regardless of the actual prevalence of bogus firms.

Third, in order to compare our algorithm’s performance to the status quo of manual targeting for inspection, we carry out what we call point-in-time simulation. We estimate not only whether our model successfully predicts bogus firms, but also whether it would have done better than the tax inspectors themselves. To do this, we test our model’s success at a point in time within the span of our dataset. We roll back the data to the state of knowledge at that time and generate predictions. We then measure the performance by using more recent data. In this manner, we can estimate the number of bogus firms our algorithm would have detected before the tax inspectors. We calculate the revenue lost to tax evasion over the period between when our algorithm could have targeted these firms and when they were actually targeted and inspected - potential revenue which we estimate at several billions of rupees (tens of millions of USD).

Finally, each firm files its returns quarterly and so supplies many data points, but its class (bogus or legitimate) is timeless. Therefore, there are several identically-formatted data points for each object of classification (the firm) and classification must be made at the object level. We train a single-period model and aggregate its predictions to the firm level. We call this approach multiple time-period prediction.

The key contributions of our work are:

- A novel solution to the classification problem with one-sided labels, where the labels of many data points are unknown and there are definite examples only of one class.
- Point-in-time simulation approach for evaluating real-world prediction systems.

---

<sup>1</sup>In future work, by closely documenting the inspection results, we intend to verify this claim.

We evaluate the impact of our predictions by calculating how much our algorithm would have increased collections by identifying bogus firms sooner than the tax authority.

- An approach for multiple time period (or multiple data-points) prediction, where several data points of similar format exist for each object of classification, but prediction is made at the object level.
- An application of machine learning on a large dataset for an emerging economy government to address an important policy problem. Our results indicate that by using our tool the tax administration can prevent fraud up to ₹1-3 billion (\$15-45 million).
- A proposed mechanism through which bogus firms may operate in a low compliance, emerging economy.

The remainder of this paper is structured as follows. In section 2.2 we review the relevant literature both in economics and machine learning. In section 2.3 we describe the basics of VAT functioning. In section 2.4 we propose the mechanism that we think is behind the existence of bogus firms. Section 2.5 describes the classification system we construct. In section 2.6 we evaluate the performance of our system and present the results of our model. Finally, in section 2.7 we describe our future goals.

## 2.2 Related Work

### 2.2.1 Economics

There is a recent and growing empirical literature on taxation and development. This project fits into a strand within this literature that explores mechanisms that improve the state's tax collection capacity (Khan et al., 2016). In contrast to this previous work, we examine the role of improved information utilization in increasing tax collec-

tions. Our paper also fits into a broader literature that seeks to improve government performance (Muralidharan and Sundararaman, 2011; Glewwe et al., 2010). Whereas this literature emphasizes the role of incentives in improving performance, we hold incentives fixed but instead improve the state’s ability to target evasion activity. Our project is also related to a nascent literature on corruption (Olken and Pande, 2012; Duflo et al., 2013). However, rather than relaxing the government’s resource constraints (e.g. in terms of inspections), we seek to reduce corruption by improving the state’s ability to detect corrupt firms by using technology to better analyze data that is already available to it. Our work also links to the “forensic economics” literature which emphasizes hidden behavior in various domains (Zitzewitz, 2012; Jacob and Levitt, 2003; Mironov and Zhuravskaya, 2014). We hope to contribute to this literature by identifying features that are predictive of fraudulent firms. Improving the state’s ability to tax effectively is increasingly seen as central to the development process and VAT has been proposed as a key tool towards accomplishing this goal.<sup>2</sup> Our work also ties in to the emerging literature on micro-empirical investigation of value added tax systems (Almunia and Lopez-Rodriguez, 2018; Mittal and Mahajan, 2017; Naritomi, 2013; Pomeranz, 2015). Although previous work has discussed VAT evasion through fraudulent firms (Keen and Smith, 2006; Pashev, 2007), this paper appears to be the first to systematically study and identify fraudulent firms in an economy with weak legal and enforcement institutions.

### **2.2.2 Machine Learning**

In recent years there has been a growing interest in using ML methods in economic development and in developing countries. These studies cover topics such as measuring poverty (Blumenstock et al., 2015; Xie et al., 2015; Chen and Nordhaus, 2011),

---

<sup>2</sup>See e.g. Besley and Persson (2013). See Ebrill et al. (2001) for an overview of the aggregate cross-country evidence on the effectiveness of the VAT.

population mapping (Deville et al., 2014), migration and mobility (Lu et al., 2016), health and epidemiology (Wesolowski et al., 2014), financial inclusion (Björkegren and Grissen, 2017), and program monitoring and evaluation (Wilson et al., 2015). These studies usually harness data sources such as satellite images of luminosity at night (Chen and Nordhaus, 2011), high resolution remote sensing (Xie et al., 2015), mobile phone usage (Blumenstock et al., 2015; Deville et al., 2014; Lu et al., 2016; Björkegren and Grissen, 2017), internet and social media (Llorente et al., 2015), and small digital sensors (Wilson et al., 2015). Very few of those studies use proprietary government data, and those that do usually do not have large scale data to benefit from ML and big-data approaches. Furthermore, very few of those studies create a system that can be used by developing country governments directly. To our knowledge, ours is the first study to use ML on large scale tax records from an emerging economy.

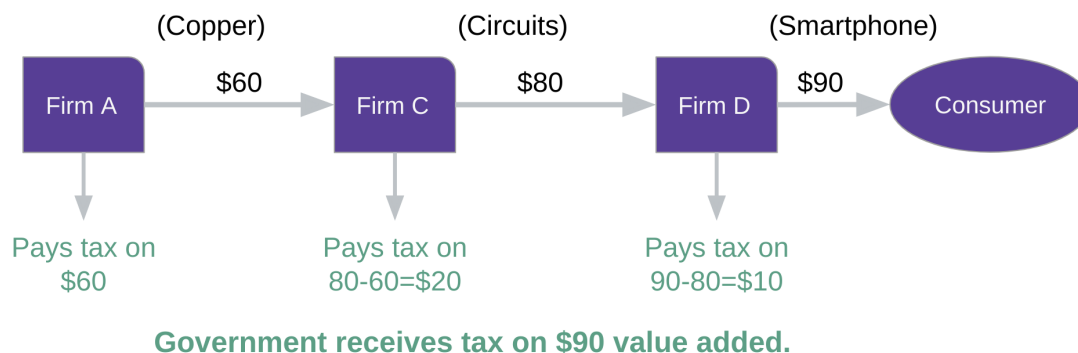
## 2.3 Value Added Tax

VAT is an indirect tax charged at multiple stages of production (and distribution) with taxes paid on purchases (inputs) credited against taxes withheld on sales (output). Firms withhold taxes on sales (output tax) from which they deduct the taxes they have already paid on purchases (input credits), and finally remit the difference to the tax authority. Thus, unlike under a retail sales tax, the tax authority collects tax revenue throughout the production chain. The VAT system also requires both firms involved in a transaction to report it independently allowing the tax authority to verify sale declarations of the seller against buyer reports while inspecting returns, known as third party verification.<sup>3</sup>

---

<sup>3</sup>See International Tax Dialogue (2005) for more details. Mittal and Mahajan (2017) further describes VAT compliance incentives, and evaluates a technology implementation intended to improve compliance in Delhi.

Figure 2.2: Stylized Example Illustrating a Value Added Tax Transaction Chain



In figure 2.2, we illustrate a simple VAT chain assuming a uniform tax rate of 10%. Firm A sells goods worth \$60 to firm C. Firm C sells it ahead to Firm D at a price of \$80. Firm D finally sells to an end customer at the price of \$90. In this example, the tax authority collects tax on \$90 of value added (\$9 of tax at tax rate 10%). \$60 of the value add (\$6 tax) comes from firm A, \$20 of the value add (\$2 tax) comes from firm C, and \$10 of the value add (\$1 tax) comes from firm D. Third-party verification works in the following manner. All firms have to report transaction level information. Firm C wants to report that it made a purchase of \$60 from firm A as that report reduces the tax that it would have to remit to the tax authority. As a result, firm A would also have to declare the sale of \$60 and subsequently pay the tax on it. Similarly, firm D will make firm C report the sale of \$80.

However, genuine firms may want to bypass the constraints imposed by third-party verification by reporting fraudulent transactions with bogus firms. Given the attention that governments in emerging economies have begun to pay to “ease of doing business” norms, registering a firm is increasingly straightforward. These factors have led to the emergence of bogus firms. In section 2.4, we explain the possible mechanism through which these bogus firms operate and lead to tax evasion.

## 2.4 A Mechanism of Bogus Firms

We now propose a mechanism that allows bogus firms to operate over multiple periods. In figure 2.3, we build on the stylized example by introducing the bogus firm B. The new transactions take place only on paper (i.e. are fraudulent). The actual transfer of goods stays the same as described in figure 2.2. Firm D diverts \$40 out of the \$90 business-to-customer sales that it was actually making to the final consumer to firm B and reports it as a business-to-business sale.<sup>4</sup> Firm D can do this because it makes sales to final consumers who are not incentivized to provide reports of such transactions to the tax authority, and because this does not increase financial liability of firm D to the tax authority.

Firm B then can sell the input credit it now has to firm A by showing a sale of an amount weakly greater than \$40 to firm A. The value add of firm A is now \$19 instead of the true value which should be \$60.<sup>5</sup> Value added by bogus firm B is \$1, and firm C and firm D have the same value added as earlier. The final value added in the system now is \$40 less relative to a fully compliant system. The surplus (tax evaded) can potentially be divided between the offending firms, i.e. A, B (bogus), and D.

A few key observations need to be highlighted. First, firm A and firm D do not need to be in the same transaction chain. It is not necessary for the eventual chain of transactions to be circular. Firm C does not benefit from the bogus transactions and need not even know about them.<sup>6</sup> Second, bogus firm B can make sales to any firm which is in need of input credits. It does not necessarily need to make such

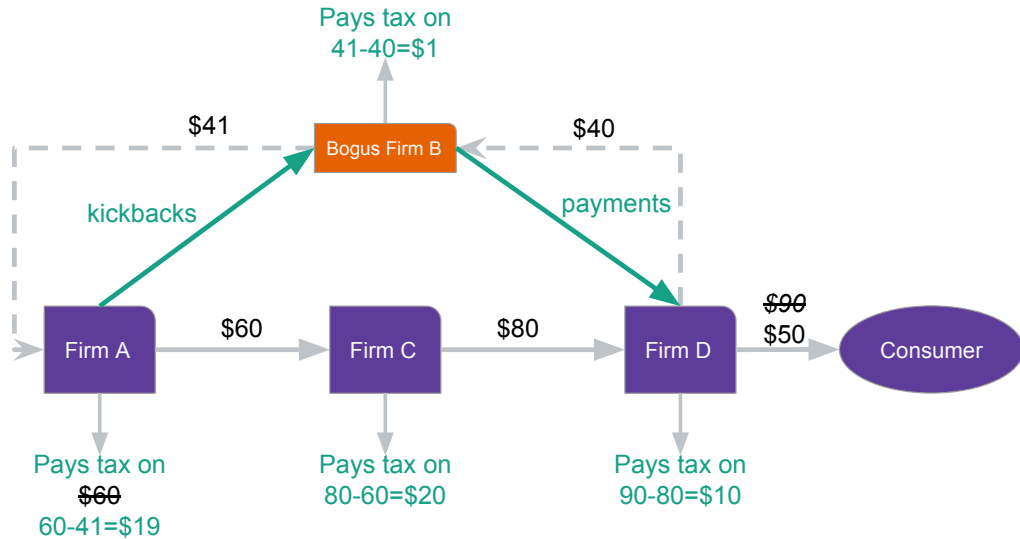
---

<sup>4</sup>Firm B may make a side payment to firm D as a small kickback for violating the law by misreporting.

<sup>5</sup>Firm A will be willing to make significant payments to firm B as this reduces firm A's tax liability.

<sup>6</sup>We include firm C to highlight that such transactions can co-exist around genuine transactions.

Figure 2.3: An Example Showing How Bogus Firms Facilitate Tax Evasion



**Government receives tax on \$50 value added. Surplus is divided between offenders.**

Firm A and Firm D need not necessarily be in the same chain. Bogus firms can make sales to any firm which needs input credits.

sales to firm A which is at the beginning of the transaction chain. Third, we believe that once detected it is not complicated to verify that firm B is bogus. It is a fake firm which should not exist physically at the address submitted to the tax authority.<sup>7</sup> Therefore, there is no need to rigorously analyze its business related paperwork (high effort) and a visit to the location (relatively low effort) should be sufficient. Finally, after identifying bogus firm B, the revenue can only be recovered if the authority pursues firm A and reverses the input credit that firm A has claimed. Any kind of a penalty only on firm B does not recoup the revenue loss. We now describe the design of our system.

<sup>7</sup>In subsequent field work we can test this assumption.

## 2.5 System Description

### 2.5.1 Data

Our dataset has almost 3 years of quarterly VAT returns from the National Capital Territory of Delhi, India from 2012-13 to 2014-15 (with personal identifying information removed). This includes consolidated returns as well as transaction level information filed by an unbalanced panel of firms,  $|Firms| = 315,191$ , and a list of firms that were previously found to be bogus by the tax authority,  $|inspected\&bogus| = 538$ . Returns are filed quarterly so we should ideally have data from 12 quarters. However, for the fourth quarter of 2012-13 the transaction level records are unusable and therefore we drop that quarter from our analysis, which leaves us with data from 11 quarters. We denote these as follows.  $Returns_{f,t}$  for firm  $f \in Firms$  in quarter  $t \in Q(f) \subseteq Q \equiv \{1, 2, 3, 5, \dots, 12\}$ , where firm  $f$  operated in quarters  $Q(f)$  ( $t=4$  was dropped). Every firm is either bogus or legitimate and this doesn't change over time:  $class_f \in \{bogus, legit\}$ .<sup>8</sup> Some firms were inspected:  $inspected_f \in \{True, False\}$  and of those inspected some were bogus and some legitimate, but this is recorded in our data only for the ones that were found to be bogus.

$$label_f = y_f = \begin{cases} \text{bogus,} & \text{if } inspected_f \text{ and } class_f = \text{bogus} \\ \text{unlabeled,} & \text{if } inspected_f \text{ and } class_f = \text{legit} \\ \text{unlabeled,} & \text{if not } inspected_f \end{cases}$$

When a bogus firm is inspected at time  $T$ , it is caught and stops operating,  $max(Q(f)) = T$ . Due to confidentiality concerns, the personally identifiable information has been removed and each firm is assigned a unique identifying number so

---

<sup>8</sup>This is an assumption, but we have good reason to believe it from discussions with the tax authority and since there are sanctions against owners of bogus firms that are found, so owners would not use their legitimate companies for this purpose.



that we can follow a firm over time as well as track its presence in other firms' returns. However, we cannot link the returns to any other publicly available information on the firms. We have detailed information on the line items in the consolidated returns,  $Returns_{f,t}$ , as well as line items from annexures,  $Transactions_{f,t}$ , where each firm reports the total purchases it made from (and sales to) each other firm at each tax rate level for that quarter.

In addition to tax return information we also have access to basic information provided by the firm at the time of registration,  $Profile_f$ . Firms are mandated to keep this information updated. We observe the date of registration, the revenue ward (broad geographic location of the firm), the nature of business (e.g. manufacturer, wholesaler, retailer), its legal status (e.g. proprietorship, private limited company), the other tax schemes and acts it is subject to (e.g. central excise act) and whether it is registered for international trade (import or export).

### 2.5.2 Features

Our unit of observation is the firm-quarter level. We have an observation for each firm in each quarter in which it filed a return: firm A in quarter 1, firm A in quarter 2, firm B in quarter 1, firm B in quarter 2, etc. We ensure that none of the features use data from different time periods - they are all within the quarterly observation. We detail them in this section. In section 2.5.6 we describe how we use the firm-quarter level observations to make predictions at the firm level.

Some of our features rely on domain expertise or anecdotes from the tax authority. For example, "VAT/Turnover ratio" is the ratio of tax paid to total turnover. To illustrate why this feature might be predictive, in figure 2.3, if firm B reported value add of \$2 instead of \$1, it would have to pay VAT for that added dollar and will also have to reduce firm A's fake input credits by \$1. Firm B is not actually carrying out

any trade, and is relying on the arbitrage between firm D and firm A to make money. This implies that it wants to minimize its declared value add or, equivalently, have a low ratio between tax paid (proportional to profit) and total turnover.

Another example of a feature we hypothesize to be predictive is the proportion of sales a firm makes to unregistered firms - these could be small firms that are not subject to reporting or they could be final customers. An unregistered firm does not claim input credits and so cannot benefit from the services of bogus firms. We would accordingly expect bogus firms to report selling a very small share of their sales to unregistered firms, and this is borne out in our (admittedly selected) data. Our features can be divided into 3 broad sets:

$$X_{f,t} = \text{featureExtraction}(\text{Profile}_f, \text{Returns}_{f,t}, \text{Transactions}_{f,t})$$

First is the set of profile features which come from the registration information provided by the firms. The second set of features come from the quarterly consolidated returns that each firm has to file. These include variables such as total turnover, within-state turnover, the amount of VAT that was actually paid, etc. Finally, the third set of features come from the annexures that firms have to file along with the consolidated returns.<sup>9</sup> These annexures cross-reference the firm's declarations of their transactions, who it sold to, and how much that other firm reported buying back. Using these annexures, we create variables such as discrepancies in reporting between clients and suppliers, the weighted VAT/Turnover ratio of client firms, the weighted VAT/Turnover ratio of supplying firms, share of sales made to biggest client, and network features such as PageRank etc.

We refrain from using the class labels of neighbors in the network in our features because it would compromise the estimates of performance by creating leakage

---

<sup>9</sup>Based on our interactions with various tax officials, tax administrations in general do not have the capacity to rigorously use the transaction level information that is now available to them. They mostly use it only for third-party verification purposes.

between training and test set, and also because we would need a point-in-time simulation to train this, since only the class labels of firms caught before a certain time period can be used as a feature when training for that time period, otherwise we create leakage from the future to the past.

### **2.5.3 Class Labels**

In this section we describe how class labels are constructed from our data, and the problems this poses.

#### **2.5.3.1 Selective Labels - Biased Training Set**

The labels in our training set come from a list of firms that were targeted for inspection by the tax department, inspected, found to be bogus, and subsequently had their registration canceled. This procedure creates a biased training set as the firms targeted for inspection are not random - in fact they were chosen explicitly by the tax department as suspicious of being bogus, so are not representative of the general population (see figure 2.1). This is the selective labels problem Lakkaraju et al. (2017). We do not have a way of completely eliminating the effect of this bias on performance. However, we *evaluate* our performance in a way that produces bounds which are not affected by this bias. In the future we plan to carry out inspections based on our model's predictions, which would provide an unbiased performance estimate.

#### **2.5.3.2 One-sided Labels**

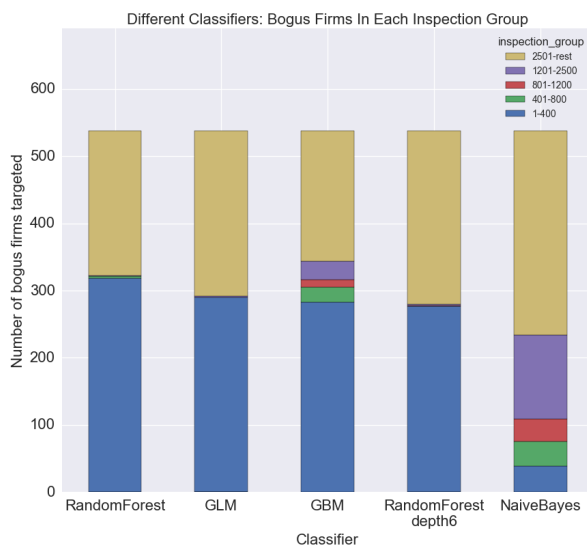
A more serious problem with our labels is that among the inspected firms we do not have records of which firms were found to be *legitimate*, and so not canceled. This is because the tax authorities do not keep records of which firms were inspected,

only which ones were found to be bogus and canceled. In our data, we are unable to distinguish the legitimate but inspected firms from those never inspected. Our labels are thus one-sided: we have firms that we know to be bogus - those canceled, and firms that we do not know for sure but are likely to be legitimate - all the rest. To further clarify the difference between selective and one-sided labels, if the tax authority randomly sampled firms for inspection, this would solve the selective labels problem but the labels might still be one-sided if they only recorded the firms found bogus.

The unlabeled firms are likely legitimate since the base rate of bogus firms in the population is not very high, though we do not know it precisely. However, they are not all legitimate as evidenced by many firms in the data being caught as bogus only after operating and facilitating tax evasion for many quarters. Therefore, we do not have a labeled training set (see figure 2.1). This is related, though not the same, as the problem of classification with noisy labels (Liu and Tao, 2016; Natarajan et al., 2013).

We define classes for all our observations in the following way. In our outcome variable, we classify firms that were found to be bogus as 1 (“bogus”) and the rest of the firms, whether never inspected or those that were inspected, found to be legitimate and not recorded, as 0 (“probably legit”). This is an acceptable starting point as bogus firms are rare in the unlabeled set. These are two classes so it seems we can solve a standard classification problem but that is not the case. There is no out-of-sample prediction to be made on some other firms that are not already labeled by this process. We are interested in predictions on the “probably legit” firms, since some of them may actually be bogus and we would like to target them for inspections. We will detail our solution in section 2.5.5 and section 2.6. While our classifiers would fit to predict inspected and bogus firms (selective labels), as in our training data, these are nonetheless bogus firms that they are finding. Additionally,

Figure 2.4: Comparison of Different Classifiers



even training on the selective set, the classifiers might learn from features not used by the tax officials and improve performance for all bogus firms.

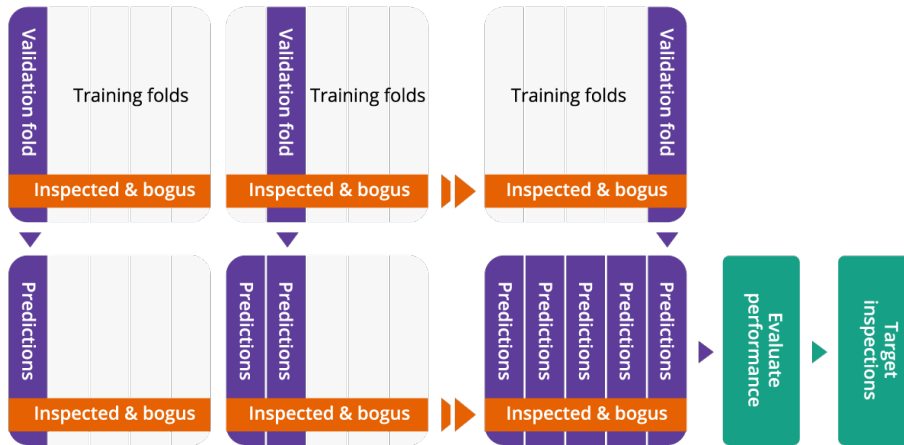
### 2.5.3.3 Multiple Time Periods

For each firm, the class does not change over time - a firm does not start legitimate and become bogus, or the other way around. However, a firm is supposed to file a return every quarter. We use our classification of firms to classify all quarterly observations of that firm with the class of the firm: “bogus” or “probably legit”.  $y_{f,t} = y_f \in \{bogus, legit\}$ .

### 2.5.4 Classifier

We use a Random Forest Liaw et al. (2002) classifier with  $n\_trees = 200$ ,  $stopping\_rounds = 2$ , and  $max\_tree\_depth = 20$  (the maximum depth is rarely reached in practice since the tree bifurcation stops when too few classification examples are left in a leaf) in H2O python implementation team (2015). We selected

Figure 2.5: Cross-validated Prediction Procedure



Random Forest for its ability to handle complex dependencies between features (since not all our features are independently predictive) and combine categorical variables together with continuous variables seamlessly.

We perform a comparison of the performance of different classifiers to see if our results are sensitive to the choice of classification algorithm. Figure 2.4 shows our results: a few standard classification algorithms obtain similar performance. For the top 400, 800 or 1200 recommendations Random Forest performs best. The one algorithm which performs worse than others is Naive Bayes, which is unsurprising since many of the features we use are not independent and since the algorithm implementation we used also assumes Gaussian distribution for numerical features conditional on class team (2018), which does not hold with our numerical features.

### 2.5.5 Cross-Validated Predictions

We have preliminarily classified all our dataset to “bogus” and “probably legit”, as explained in the section 2.5.3.2, and so it would seem that there is no more need for in-sample predictions to target inspections. However, for our real-world application, we want the model to help target inspections of existing unlabeled firms. Of the

firms we classify as “probably legit”, some are bogus and we want to find them. If we trained a classifier on all the observations (classes “bogus” vs. “probably legit”) and then made predictions on the “probably legit” firms, we would be overfitting as a result of leakage since our model would train on certain examples and then be used to make a prediction on one of those same examples. This problem comes up in cases of prediction using noisy labels, where it is required to make predictions on the noisily-labeled data that is used for training.

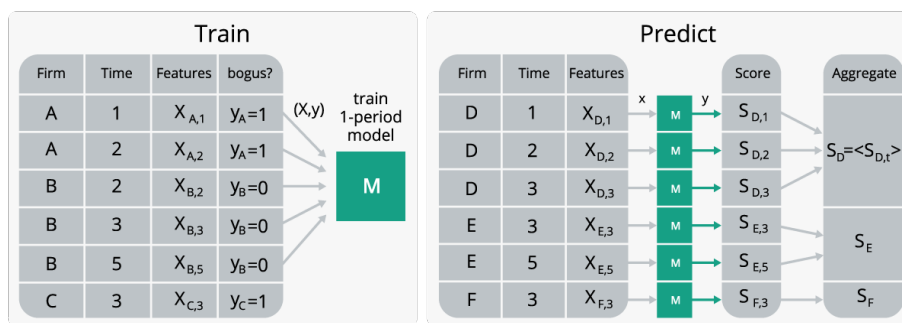
To avoid this, we carry out cross-validated holdout predictions. We randomly divide our data into 8 folds. We also ensure that all quarterly observations of a firm are within the same fold, which is crucial to avoid a different type of leakage across time-periods. Formally, for each firm we draw a fold:  $fold_f \in \{0, 1, 2, \dots, 7\}$ . We divide our data into these folds:  $Fold_j = \{(X_{f,t}, y_f) | f \in Firms, t \in Q(f), fold_f = j\}$ . We generate predictions for all of our sample by performing 8-fold cross-validation: train on 7 folds with classes “bogus” vs. “probably legit” and make predictions on the remaining fold. We then save those holdout predictions for all of the folds - for each fold from when it was the validation fold. In this way we generate predictions for our entire dataset (see figure 2.5). To obtain predictions for firm  $f$ :

1. Train on all other folds:  $Model.train(\bigcup_{j \neq fold_f} Fold_j)$
2. Make prediction:  $\hat{y}_{f,t} = Model.predict(X_{f,t})$

To evaluate performance, we will compare those predictions  $\hat{y}_{f,t}$  to the known classes  $y_f$  (see section 2.6). To target inspections, we plan to discard predictions for firms that are known to be bogus (since they were already canceled), and focus on the ones remaining with the highest model-predicted likelihood of being bogus.

More than 8 folds would make for better predictions, but increase computational expense. Increasing the number of folds from 8 to 16 would increase the size of the training set from 7/8 to 15/16 - an increase of 7% - but more than double the

Figure 2.6: Generating Firm Level Predictions from Firm-Quarter Level Data Points



computation time. The extreme ideal for prediction would be leave-one-out, which would be computationally unfeasible on our dataset. Our method is not specific to the context of bogus firms, but can be used in any scenario with noisy labels and in need of in-sample prediction.

### 2.5.6 Multi-Period Model

Each firm operates and files returns every quarter, and so we have multiple observations for each firm  $f$  - one for each quarter - with different feature values  $\{(X_{f,t}, y_f) | t \in Q(f)\}$ . However, a firm is either bogus or not and this does not change with time, so predictions should be made for a firm,  $\hat{y}_f$ , not a firm in a specific quarter,  $\hat{y}_{f,t}$ . So we need to consider these different feature values but produce one prediction per firm. We do this in three steps. First, we train a single-period model on firm-quarter observations. Second, we use the single-period model to make predictions for all firm-quarter observations,  $\{\hat{y}_{f,t}\}$ . Third, we aggregate the firm-quarter predictions to the firm level to produce a single prediction per firm,  $\hat{y}_f = Aggregate(\{\hat{y}_{f,t} | t \in Q(f)\})$  (see figure 2.6). For generality, we describe this situation as if we have full labels and separate *trainingSet* and *testSet*, but we combine this approach with our approach for one-sided labels.



### 2.5.6.1 Single-Period Model

Each firm  $f$  has multiple firm-quarter data points, all classified according to the class of that firm. We use all these points to train a single-period model:

$$Model.train(\{(X_{f,t}, y_f) | f \in trainingSet\})$$

In our case, *trainingSet* is simply all folds other than the one to which firm  $f$  belongs. Rather than taking only one observation per firm or re-weighting them so that each firm has equal total weight, this model is unaware of the fact that different observations belong to the same firm (or to different quarters) and treats them all independently (see left panel in figure 2.6). This might affect performance depending on the behavior of the same firm in different quarters and how much marginal information each observation adds to the classification problem. If features in different quarters by the same firm are identical, then our procedure only introduces unbalanced duplication. If they are different, then different time-periods add valuable training examples that are all separate pieces of evidence for what bogus firms look like and our procedure is justified. Our procedure may also result in biased data based on whether firms were caught early or late, since those caught later operate in more quarters and so have more data points on which to train on and would influence the classifier more.

### 2.5.6.2 Single-Period Predictions

We use our single-period model to make predictions on firm-quarter observations:

$$\hat{y}_{f,t} = Model.predict(X_{f,t})$$

The model is unaware of the fact that different observations may belong to the same firm. These predictions are model-predicted probabilities of being bogus - numbers between 0 and 1, where 1 is certainty of being bogus, and 0 is certainty of being legitimate. In this way we generate predictions for all the observations, using for each firm all folds that do not contain that firm as described in section 2.5.5 (see right panel in figure 2.6).

### 2.5.6.3 Aggregating Predictions to The Firm-Level

Eventually, we want to make a single prediction for a firm by combining the information available across quarters. We do this by aggregating the single-period model-predicted probabilities at the firm level to create final predictions (see figure 2.6 right panel):

$$\hat{y}_f = \text{Aggregate}(\{\hat{y}_{f,t} | t \in Q(f)\})$$

We experimented with different functions for aggregating single-period predictions. For example, if a firm’s single-period predictions were 0.1, 0.2, 0.6 then the arithmetic mean aggregating function would produce a score of 0.3, whereas the maximum aggregating function (maximum model-predicted probability of being bogus across periods) would produce 0.6. In practice, the arithmetic mean works best, only slightly ahead of the maximum score function. We therefore end up using the arithmetic mean as our aggregating function. Other functions that we tried were “take the score of the first (last) period for that firm”, and “maximal (minimal) score for that firm across all periods”.

There are other approaches that can potentially address the multiple time period problem. For example, aggregating the features before generating predictions instead of aggregating predictions, or treating each firm as a single observation and the list of its quarterly VAT returns as the corresponding feature vector. All approaches

Table 2.1: Model Performance on Top Recommendations

<i>Inspection Group</i>	<i>Firms Inspected</i>	<i>Total Bogus Firms Caught</i>	<i>Bogus Firms Caught/Inspection</i>
1 - 400	400	305	0.76
401 - 800	400	48	0.12
801 - 1200	400	24	0.06
1201 - 2500	1300	29	0.02
2501 - rest	313229	132	0.00

Using data from all time periods and cross-validated predictions.

have their own challenges and we picked the one that we seemed least problematic. If we aggregate feature values before generating the predictions, we might end up losing the variation in multiple observations of the same feature. If we combined all quarters to ensure that each firm had a single observation, then the same feature across time will be treated independently, so in effect we're reducing the number of points in the training set. Moreover, entry and exit of firms in different time periods will result in the dataset having a lot of NULL values.

Our chosen approach is not specific to multiple time periods. It is applicable whenever there are several identical-format data points for each object of classification (the firm, in our case) but classification must be made at the object level. For example, different doctor visits per patient, or fraud detection based on multiple online purchases by the same customer, or many monthly behaviors of the same object.

## 2.6 Evaluation of Performance

### 2.6.1 Top Recommendations

We are constrained by the number of firms the tax authorities can physically inspect, and so we aim to provide a ranked list of suspicious firms. Performance on other parts of the distribution has no real world implications. We rank our firms in descending order of model-predicted likelihood:  $rank(f1) \leq rank(f2)$  iff  $\hat{y}_{f1} \geq \hat{y}_{f2}$ .  $rank(f) \in \mathbb{N}$ . We take a few realistic numbers of top recommendations and check our model's success on those - of these top  $N$  firms predicted to be bogus by our model, how many are known to be bogus? Formally, bogus in top  $N = |\{f | rank(f) \leq N, y_f = bogus\}|$ .

Of our top 400 most suspicious firms, we find 75% to be actually bogus. Of the next 400 - 12% are actually bogus. Of the next 400 - 6%, of the next 1300 - 3%, and of the general population - a negligible fraction (see table 2.1). Our model finds more than 75% of firms labeled “bogus” in the top 2,500 recommendations - less than 1% of firms, which is very good performance. Moreover, recall that the “probably legit” firms in this dataset are not necessarily legitimate - they could be bogus unknown to us because they were not inspected - so these numbers for the true-bogus rate in  $N$  inspections are underestimates. See figure 2.7 for more fine-grained results on the top 1000 recommendations.

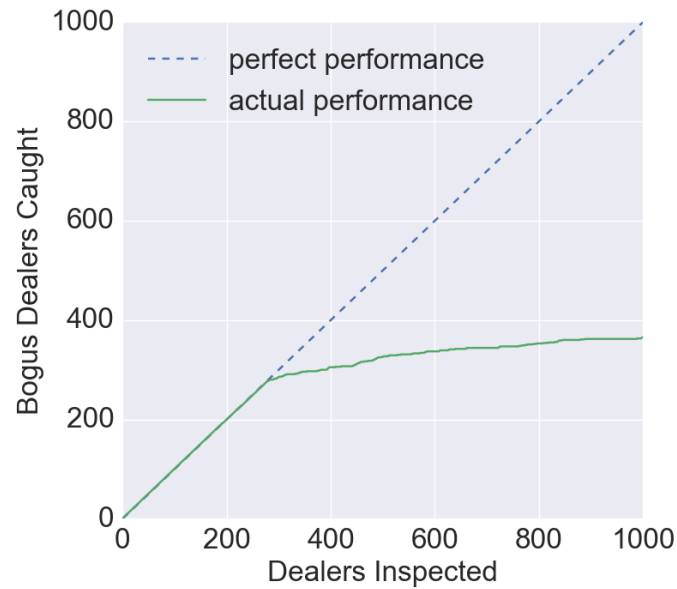
### 2.6.2 Top Recommendations - Maximizing Revenue

From tax authority's perspective, we want to maximize expected revenue captured, and not merely the likelihood of finding a bogus firm. Small firms will only account for a small amount of tax evaded, and larger bogus firms will account for a much larger amount.<sup>10</sup> We would therefore want a different ranking, in descending order

---

<sup>10</sup>More precisely, this is determined by the amount of tax input credits claimed, the \$40 in the stylized example above.

Figure 2.7: Model Performance on Top 1000 Recommendations



Using cross-validated predictions on all time periods. The x value is the rank N of suspicious firms and the y value is the number of firms out of those N that are bogus. The dashed  $y=x$  line indicates perfect performance, i.e. every firm inspected is bogus. Model performance is near-perfect for the first 250 predictions, then starts to level off.

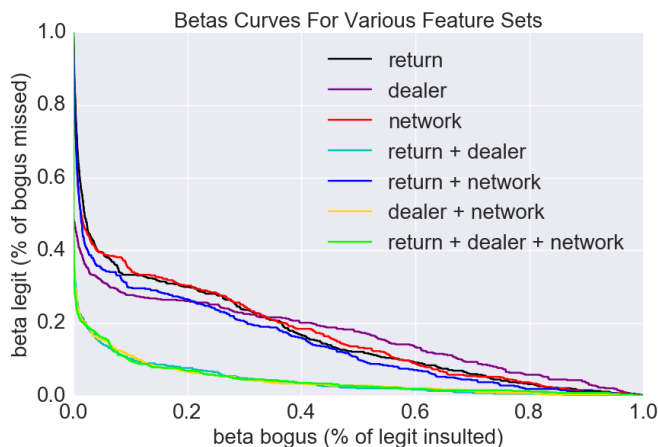
of expected recovered revenue and not of model-predicted likelihood of being bogus.

We use the following procedure to make such recommendations.

First, calibrate the model, so its score is an unbiased estimate of the actual probability of being bogus:  $E[y_f | \hat{y}_f \approx a] \approx a$ . For example, of firms with model score of 0.1, about 1 in 10 should be bogus, for those with model score 0.01 about 1 in 100 should be bogus, etc. This is not automatically the case for predictive models (e.g. Naive Bayes, which tends to extremes with more and more dependent features). Our model turns out to be very well calibrated (results not reported).

Second, calculate expected revenue recaptured by taking the amount of input tax credits the suspicious firms claim (as depicted in figure 2.3) as the lost revenue to be recaptured, and multiplying it by the calibrated model score as the probability of a

Figure 2.8: Betas Curves for Different Feature Sets



The beta curve is 1 minus the ROC curve. As we vary the classification threshold, the x value tracks the fraction of “probably legit” firms that will be insulted by getting inspected ( $\beta_{bogus}$ ); the y value tracks the fraction of bogus firms that will be missed by being classified as legit ( $\beta_{legit}$ ). A lower curve indicates better performance.

firm being bogus:  $ExpectedRevenue_f = \hat{y}_f \cdot InputCredits_f$ . We then rank all firms in descending order of the expected revenue captured, and recommend the top ones. On our data the performance of this method in terms of revenue is comparable to the earlier described method and therefore we do not report the results.

### 2.6.3 Different Feature Sets

We have constructed 3 distinct feature sets which we use together in our model: features constructed from the individual firm’s returns ( $Returns_{f,t}$ ), features constructed from the firm’s dealer profile ( $Profile_f$ ), and features constructed from the network of firms and their trading partners ( $Transactions_{f,t}$ ). We now disaggregate those feature sets and evaluate the performance of the model with each possible combination of feature sets. See figure 2.8 for the results. There is a three-way tie between feature sets that contain  $Profile_f$  and one other feature set, so that the addition of the third one makes no discernible difference.

## 2.6.4 Point In Time Simulation

Our performance estimates so far may seem unrealistic. If we use our model in a real-world scenario, we will not have access to all returns by all firms and will not be required to predict retroactively which firms were caught as bogus and which were not. In a realistic scenario, we would have the returns of all firms up to a certain point in time, and would have to predict which of the firms still operating are likely to be bogus and need to be inspected. Some of those inspected would actually turn out to be bogus. We therefore propose another metric to gauge our performance: point-in-time simulation.

We build a model based on the state of knowledge and data at a certain point in time  $T$ , when the returns for quarter  $T$  have been filed but inspections between quarters  $T$  and  $T+1$  were not yet performed. We blind our model to all information in the dataset obtained after time  $T$  - we do not consider “future” tax returns of firms from times greater than  $T$ . For our training purposes we only use firms that had already been classified as bogus and canceled by time  $T$ . This means that bogus firms caught after time  $T$  are classified in the training set as legit. We argue that this strategy simulates the state of knowledge and data at time  $T$  (figure 2.9, a panel describing  $T=3$ ).

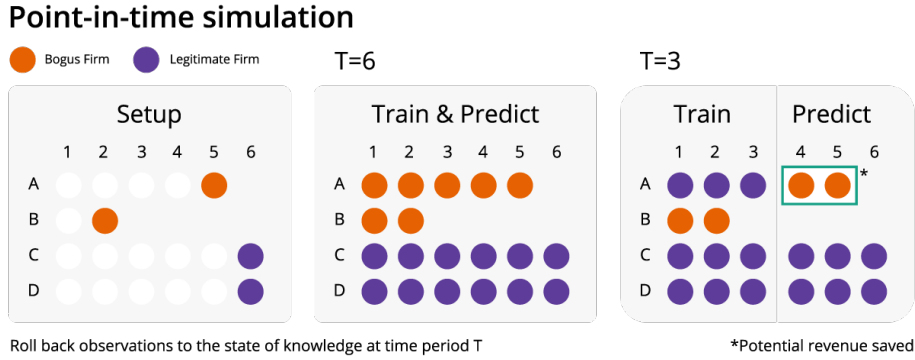
$$SimulationTrainingSet_T = \{(X_{f,t}, \tilde{y}_{f,T}) | t \leq T\}$$

$$\tilde{y}_{f,T} = \begin{cases} \text{bogus,} & \text{if } y_f = \text{bogus and } \max(Q(f)) \leq T - 1 \\ \text{legit,} & \text{if } y_f = \text{bogus and } \max(Q(f)) \geq T \\ \text{legit,} & \text{if } y_f = \text{legit} \end{cases}$$

$$SimulationTestSet_T = \{(X_{f,t}, y_f) | t \leq T, \max(Q(f)) \geq T\}$$

We then run our prediction algorithm on this state of the data to generate top

Figure 2.9: Illustration of Point-in-time Simulation at T=3



recommendations for firms still operating at time T. We evaluate performance by using, for the validation set, the real class of firms which were later caught as bogus ( $y_f = bogus$  and  $max(Q(f)) \geq T$ ). Allotting  $N$  inspections based on our simulated model’s top recommendations, we determine the number of bogus firms, which were later caught by the tax authorities, that our model would have ranked in the top  $N$  in time  $T$ . By summing the input tax credits that these recommended bogus firms claim in time periods greater than  $T$ , we can estimate how much lost revenue could have potentially been saved by using our model.

For example, suppose there are 4 firms: A, B, C, D and 6 time-periods as in figure 2.9. Firm A was inspected and identified as bogus in  $t=5$ , and firm B was inspected and identified as bogus in  $t=2$ . At  $T=6$  after all returns have been filed and inspections have taken place, we know the class of all firms and so we know that firms A & B are bogus. Our previous performance estimates use this state of knowledge to train, predict and evaluate performance ( $T=6$  panel).

Now when we simulate the state of knowledge at time  $T=3$ , we would have our features for  $t=1,2,3$ . We would also know that firm B is bogus, since it was caught and canceled after quarter  $t=2$  and so did not file a return in quarter  $t=3$ . However, we would not know that firm A, which would be caught and canceled after quarter  $t=5$ , is bogus, and so it would be labeled as “probably legit”. Firms C & D would also



be labeled as “probably legit”. We now train our model on all observations from  $t=1, 2, 3$ . We use these observations to make predictions on firms A, C, D but not on firm B - since it was already canceled and is no longer operating so there is no point in targeting it for inspections. Suppose that firm A is targeted for inspection based on its features in  $t=1, 2, 3$ . We can now discern ourselves to its real class, “bogus” - in reality only discovered at  $t=5$ , and see that we correctly identify it as bogus in  $T=3$ . If we were to conduct inspections based on these recommendations, firm A would not have been able to operate in periods 4 and 5, and the tax evasion it facilitated in those periods would have been avoided and revenue loss averted (figure 2.9,  $T=3$  panel, marked by asterisk).

This assumes no substitution, i.e. that if a firm was using the services of a bogus firm to evade taxes and that bogus firm is caught and canceled, then the client firm does not find another bogus firm to facilitate its tax evasion but instead pays the required tax. We intend to rigorously test to what extent this is the case in future work using a randomized controlled trial. To adjust our revenue implications to substitution, we can multiply the revenue captured by a factor between 0 and 1 indicating what fraction of the evasion was not substituted.

Results of point-in-time simulations for various times  $T$  are detailed in table 2.2 and plotted in figure 2.10. Even based on only a few time periods, our simulated model succeeds in targeting a sizable fraction of bogus firms in the first 400 inspections (0.12% of firms). Almost all bogus firms targeted by the model fall in the top 400 recommendations, with the other recommendations contributing very little. The performance improves when time advances from  $T=2$  to  $T=4$  since the model has more labels for training (labels from  $t$  in  $\{1,2,3\}$  vs.  $\{1\}$ ), but then declines towards the end of the dataset when there are not many bogus firms left to be detected. Note that these simulations are not disjoint - many of the firms we would have caught in  $T=2$  are the ones we would have caught in  $T=4, 6$  and so on. So we cannot simply

Table 2.2: Point-in-time Simulation Performance for the 1-400 Inspection Group

<i>T</i>	<i>Total Bogus Firms Caught</i>	<i>Bogus Firms Caught/Inspection</i>	<i>Revenue Gained by Inspecting Entire Group (USD Millions)</i>	<i>Revenue Gained per Inspection (USD 000s)</i>	<i>Total Bogus Firms in the Sample</i>	<i>Revenue Lost from All Bogus Firms (USD Millions)</i>
2	94	0.24	19.44	48.60	416	49.40
4	155	0.39	43.19	107.97	412	108.38
6	156	0.39	25.48	63.70	437	63.84
8	157	0.39	9.38	23.46	395	26.43
10	46	0.11	1.70	4.24	114	4.52
12	10	0.02	0	0	22	0

Each row shows the impact of inspecting the top 400 firms by model score: the number of known bogus firms that would have been found, and the revenue saved. The last two columns show the total number of bogus firms left to be caught at that time T, and the future revenue lost due to their activity.

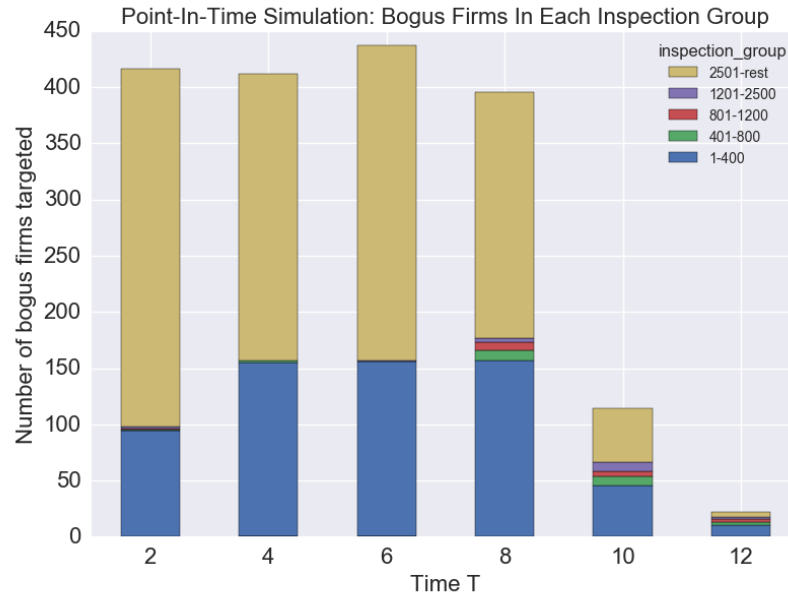
add up the numbers of firms caught and revenue captured, but have to select one. Of course, these are underestimates since some of the most suspicious “probably legit” firms are in fact bogus firms that were never inspected.

The number of bogus firms caught and the revenue saved are also calculated per inspection, showing an excellent hit rate of about 1 in 3 and extremely high returns for the first 400 inspections. In all, the top 400 recommendations catch between 20% and 40% of all known bogus firms operating at that time.

### 2.6.4.1 Revenue Implications

Point-in-time simulation also enables us to see how much additional revenue could have been gained by correctly targeting a bogus firm. We now estimate this effect by taking the top 400 firms that our point-in-time model would have recommended for inspection, and collecting the potentially gained revenue due to earlier-than-actual inspections of those of them that were later discovered to be bogus. For example if in reality a bogus firm was caught by the tax authorities after quarter 4, but our model would have recommended it for inspection in quarter 2, then the bogus firm would have been caught and its evasion in quarters 3 & 4 prevented - which is the revenue gained. We estimate revenue savings in the order of several tens of thousands US\$ recovered per inspection of the 1-400 inspection group in the earlier and middle time periods. In later periods, bogus firms were not yet identified by the tax authorities

Figure 2.10: Point-in-time Simulations Performance



For a point in time simulation in time  $T$  (x axis), out of all the bogus firms still operating, the chart shows the number of bogus firms that fall in each inspection group.

in our data and so mechanically the numbers are lower (see the results in table 2.2).

## 2.7 Future Work and Conclusion

The purpose of this paper is to assess if a machine learning tool can be effective in catching bogus firms, in reducing the effort required by tax inspectors, and in eventually increasing tax collections. By creating a model from VAT returns from the state of Delhi, India and by analyzing the performance of the model, we provide evidence that targeting inspections based on a ML model could be highly beneficial even in a low compliance high evasion setting. Our results indicate that by using our tool the tax administration can prevent fraud up to \$15-45 million. Given that such data exists in many tax jurisdictions and that anecdotal evidence suggests that such

false paper trails are a common problem, our work should have high policy relevance both within India and elsewhere.

The field level efficacy of the results that we have described in this paper needs to be investigated. We need to evaluate whether our work reduces the effort required by tax inspectors, and more importantly, does it improve revenue collections. A rigorous way to assess the reduction in administrative effort due to our model would be to target future inspections by comparing the recommendations of our model with a list of recommendations prepared by a team of tax officials and then inspecting the firms recommended by each and comparing performance. We are working with the Delhi tax authority to implement this.

Simultaneously, we aim to improve the results of our model by working closely with the tax authority in Delhi to conduct tax inspections on firms recommended by our model. We intend to stratify these inspections by our model score, inspecting firms throughout the model score distribution. Stratified inspection would provide a representative training set. On the other hand, inspecting the most likely bogus firms would provide more representatives from our rare class. We are not sure how to trade off those one vs. the other, but a similar point-in-time simulation exercise with our data where we only obtain information from our targeted inspections could shed light on this question.

Finally, recovering the revenue loss which can be directly attributed to bogus firms is not trivial. Often, the owners of these bogus firms can not be found at their declared addresses. Even if caught, they themselves have not benefited from the entire tax evasion. In fact, they have been key contributors to their trading partners who have managed to reduce their declared tax liability by interacting with these bogus firms. To do the actual revenue recovery, it is important to pursue the firms which interact with the bogus firms. Moreover, it is possible that even if the transactions that the trading partners declare with bogus firms have been canceled,

the trading partners substitute these transactions to other yet to be identified bogus firms. If our system is deployed and bogus firms try to adapt, we will face a scenario of adversarial machine learning. All these are important questions which we hope to study as part of our future work.

## CHAPTER 3

# Red Tape? The Revenue Impact of the VAT Filing Thresholds

### 3.1 Introduction

Value-added tax (VAT) is considered to be an effective tool to raise revenues by governments across the globe (Keen and Smith, 2006). However, VAT systems across the world are afflicted with size-dependent regulations. Such policies form an elementary part of the VAT administration across countries. Not only is there evidence that monitoring effort is being directed on larger firms (Almunia and Lopez-Rodriguez, 2018; Bachas and Jensen, 2017), but also both the VAT registration thresholds as well as the VAT reporting frequency thresholds depend on the reported firm revenue.

However, the benefit of size-dependent regulations, related to filing frequency, to the tax authority is unclear. On the one hand, the tax authority may prefer to receive as much information as fast as possible. The government might also be liquidity constrained and prefer to receive VAT payments from firms at as high a frequency as possible. On the other hand, the literature has provided theoretical arguments that the tax authority should economize on administrative and compliance costs by exempting small firms from taxation (Dharmapala et al., 2011). Such optimizing concern might be even more important for low and middle income countries, where compliance costs are of first order concern. Compliance costs, relative to firm size, might be greater for smaller firms - with little benefit to the tax authority

(International Tax Dialogue, 2007, 2013).

It has been shown in many countries that firms close to the VAT registration threshold, at the extensive margin, actively manipulate their reported revenues in order to avoid registering for VAT (Onji, 2009; Gebresilasse and Sow, 2016; Liu et al., 2017; Harju et al., 2016; Boonzaaier et al., 2017). Furthermore, recent work has indicated that firms may also be responding to VAT filing thresholds, at the intensive margin (Asatryan and Peichl, 2017). This work, however, only features evidence of bunching at a threshold where the results are confounded by other changes in the reporting regulation at that same threshold.

In this paper, we use an administrative dataset from the state of Delhi in India to show that a policy which mandated different frequencies of filing based on reported turnover resulted in bunching of firms below each of the thresholds. Using the change in these reporting policies in the following years, we provide further evidence that such sharp bunching indeed occurs due to the VAT reporting frequency thresholds. Second, we calculate the VAT revenue losses due to such bunching. Third, the subsequent withdrawal of the policy allows us to show that in a regime with size-dependent reporting requirements, more frequent reporting does not lead to greater levels of VAT collection. Finally, our back of the envelope social welfare calculations indicate that there are substantial benefits of the implicit subsidy, in the form of a lower frequency of return filing, offered to the smaller firms by the tax authority.

Our paper is the one of the first to carefully investigate how firms respond to multiple VAT filing thresholds on the intensive margin. We do this by taking advantage of the unique setting in the state of Delhi in India. A careful analysis of the VAT filing thresholds fills an important gap in the literature as the VAT filing thresholds are quite ubiquitous, but their impact is not well studied.

Many countries mandate a uniform rate of VAT reporting at either a monthly (e.g., Argentina, India [GST]), or a bi-monthly (e.g., Barbados), or a quarterly (e.g.,

Cyprus) frequency. At the same time, there are many countries with policies that mandate different rate of VAT return filing based on firm size. Out of a 2015 sample of 103 countries for which we could determine with certainty, 39 countries had size-dependent frequencies of VAT reporting and payments.<sup>1</sup> Among these there are both high-income (e.g., Austria, Germany, Denmark, Finland, France, Ireland, Spain, UK, etc.) as well as low and middle income (e.g., Botswana, Colombia, Mauritius, Philippines, South Africa, Swaziland) countries. If firms misreport their revenue in order to avoid filing (and remitting) VAT returns more frequently, it is important to understand the associated costs to the VAT collections.

Before answering this question empirically, it is important to understand what the optimal “first-best” policy with regards to VAT filing frequencies might be. A liquidity constrained government, which might be the case in low and middle income countries, may prefer to receive VAT reports and the associated payments at as high a frequency as possible. On the other hand, a high frequency of VAT reporting and payments may impose a significant compliance burden on firms, especially on small ones. The trade-off is relatively clear: while the government prefers, for several reasons, to obtain information and the VAT payments as frequently as possible, it is amenable to subsidizing smaller firms - those, eligible for VAT - by imposing less stringent filing requirements on them. The problem occurs when such policies lead to firms actively avoiding filing more frequently, by either underreporting their true revenues, or by under-producing and thus intentionally reducing their growth. Both of those channels are problematic: The first one creates an environment which via the choice of policies fosters a culture of evasion and avoidance, which is especially troublesome for the low- and middle-income countries. The second, underproduction channel, results in production inefficiencies, keeping the economy below its potential. In the paper, we also perform a back of the envelope welfare analysis, by calculating

---

<sup>1</sup>Based on survey of the data from Young (2015).



the minimum compliance costs needed at the level of each of the reporting categories (annual, bi-annual, quarterly, and monthly) for the policy to be optimal, despite the revenue costs incurred by the tax authority. In order for the welfare change due to the filing threshold policy to be positive, the actual compliance costs of the more frequent reporting need to be higher than the calculated minimum implicit subsidies.

This paper's unique contribution is in the simultaneous analysis of the multiple VAT filing thresholds. We observe substantial bunching of the firms below each of these VAT filing thresholds. We are the first to estimate the resulting revenue losses because of the bunching behavior due to the size-dependent VAT filing policy. Using a simple analysis we further show that more frequent VAT filing does not lead to higher revenues to the tax authority. Finally, we conduct a simple welfare analysis to show that if the compliance costs due to the VAT filing are non-negligible, the size-dependent filing policy can be justified due to the implicit subsidies to smaller firms, via the reduced filing requirements.

The paper is organized as follows. The following section describes some of the related literature. Section 3.3 describes the reforms in the VAT filing system in Delhi and the administrative level data from Delhi that we use for our analysis, and explains the methodology we use for the analysis. Section 3.4 presents the results of our analysis of the firm response, of the revenue implications, and of welfare. Section 3.5 concludes.

## **3.2 Related Literature**

We contribute to several strands of literature. We add to the literature on the effects of size-dependent policies on firm behavior. The literature has shown that size-based regulations can lead to distortions (Gollin, 1995; Guner et al., 2008). Garicano et al. (2016) look at the context of labor laws in France, which are applicable on firms with

50 or more employees. Using administrative-level firm data, they find that the cost of such regulations is equivalent to a 2.3% tax on labor, and that such regulations negatively impact welfare to the tune of 3.4% of GDP. They also show that such size-dependent regulations are effectively subsidizing small firms, at a cost to workers and to some extent large firms. In the context of taxation, Almunia and Lopez-Rodriguez (2018) show that the threshold relevant for the large tax payers unit, which monitors large firms in Spain, results in strategic bunching of the firms in order to avoid stricter monitoring by the tax authority. Finally, they show that extending stricter monitoring to small businesses would generally generate substantial welfare gains.

Such size dependent policies are ubiquitous across countries. In a cross-country study, Bachas and Jensen (2017) find that greater firm size results in increased tax enforcement and tax compliance, with negative effects on giving bribes. Important for our work, they show that such a size gradient in tax enforcement is the strongest in developing countries. In the VAT context, Mittal and Mahajan (2017) show that strengthened paper trail leads to an increase in tax collection, primarily driven by the behavior of the largest firms.

We also add on to the wide literature discussing VAT policy design, both theoretically and empirically. For instance, Keen and Smith (2006) discuss the problematic aspects of VAT policy in the EU countries, particularly those that result in a greater propensity for tax evasion and fraud. They also look at the possibility of a federal VAT system in the US. Bird and Gendron (2007) discuss the application of VAT in lower income countries, and detail many implementation challenges encountered in such settings: inappropriate thresholds, delayed refunds, and insufficient audits.

Additionally, we contribute to the literature on calculating the hassle costs affiliated with filing taxes. In this realm, Benzarti (2017) estimates the cost of filing income taxes using US income tax return data. He finds that such hassle costs are increasing with income of households. We also add on to the literature analyzing the

VAT registration thresholds: in those cases, firms below a certain size are exempt from registering for the VAT system and thus from paying VAT. In their theoretical work, Keen and Mintz (2004) look at the optimal registration threshold in the case of VAT, and show that such a threshold will necessarily lead to production inefficiencies via the bunching below the thresholds.

A growing literature has empirically analyzed the effect of the VAT registration threshold. The first paper to recognize the reaction of firms to the VAT registration threshold was Onji (2009), which documents that large firms masquerade as many small firms in response to the VAT registration threshold in Japan. Liu et al. (2017) consider the VAT registration notches and voluntary registration below the VAT registration threshold in the context of the UK. Their bunching estimates, based on separate annual observations of British firms between 2004 and 2010, are in the range between 0.815 and 1.286. They furthermore look at the determinants of voluntary registrations and of the bunching behavior, listing the low cost of inputs relative to sales, high proportion of business-to-customers sales, and high product market competition as determinants of higher bunching. Harju et al. (2016) look at the impact of the VAT registration threshold on the behavior of small Finnish firms. They find that the firms actively avoid VAT liability, and that such avoidance is directly caused by the compliance costs of VAT. Furthermore, they find that the bunching behavior persists over the long-term, implying that the VAT registration threshold permanently hinders the growth of small firms. They find a strong bunching response to the VAT notch, of 3.63.

In the context of a developing country, Boonzaaier et al. (2017) use tax register data from South Africa to look at the effects of several discontinuities in the tax schedule on the behavior of small firms. They find a moderate level of the bunching response, of 1, at the VAT registration notch. Along similar lines as the Finnish study, the work on South Africa finds that the bunching firms are less likely to show strong

growth dynamics and are more like to be 'stuck' in terms of their profits and sales. Similarly, Gebresilasse and Sow (2016) look at the firm response to VAT registration threshold in Ethiopia, and estimate substantial bunching, with the bunching response equaling 4.8.

The closest paper to ours, by Asatryan and Peichl (2017), looks at the effects of both the VAT registration threshold and the so-called administrative thresholds, on the reported revenues by Armenian firms. Their analysis shows a moderate response to the VAT filing frequency threshold, which - however - corresponds to some other regulatory thresholds. For their full sample, they find the bunching response of 1.569. Their heterogeneity analysis shows that the response is driven up by growing firms, and by firms in the primary sector and services. Furthermore, they show, using audited tax returns, that the response is mainly driven by evasion, rather than by underproduction. Finally, Asatryan and Peichl (2017) find no response to the VAT registration threshold notch.

### **3.3 Methodology**

This section describes the methodology we use to derive the distributional and other results discussed in the following section. We describe the methodology used to derive the bunching estimates, and the methodology to identify the VAT revenue losses to the tax authority due to the imposed reporting thresholds and the resulting bunching. We then describe the methodology we use to show that there are no apparent revenue benefits to the tax authority by the firms providing more frequent information in the form of the VAT reports. Lastly, we outline our methodology for the general social welfare analysis of the VAT reporting thresholds. We begin by describing the policy variation and the data that allows us to do all of the above.

### 3.3.1 Delhi: Policy Change

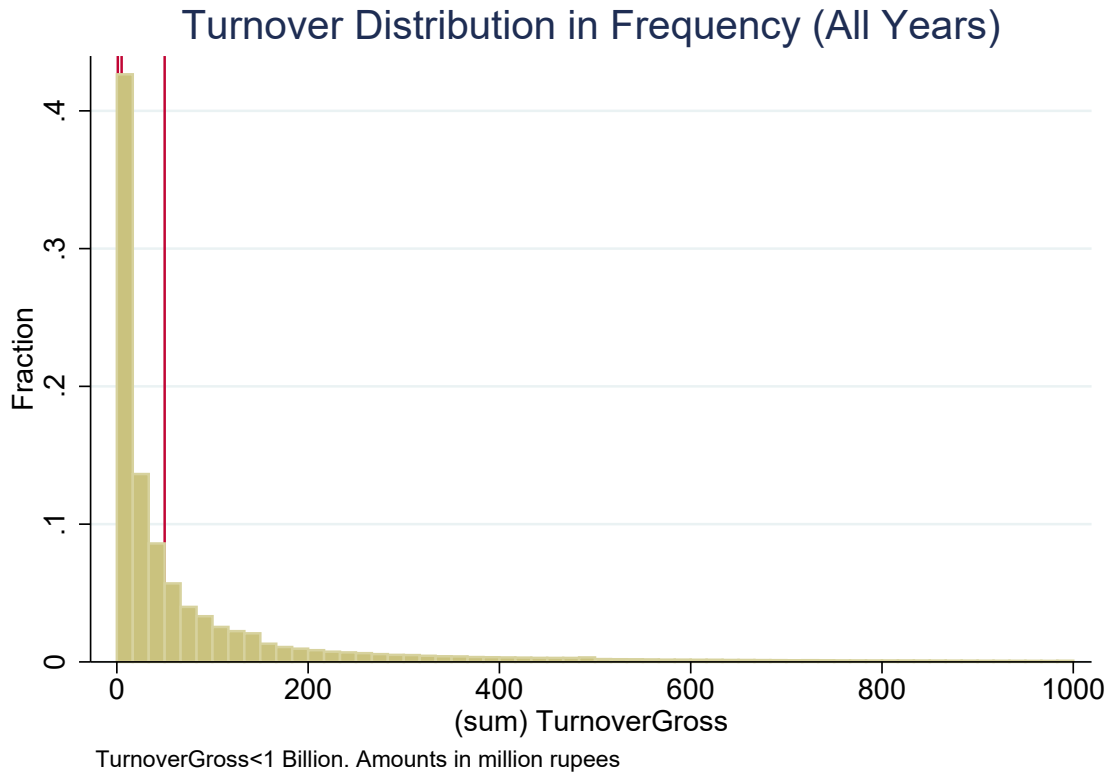
For the five years in Delhi, India for which we have the data, the return filing policy had rich variation. Firms had to file returns at a specific frequency depending their on declared turnover in the previous financial year. For the first two years, Y1 and Y2, there were 3 thresholds of interest: Threshold 1 at ₹1 million (Indian rupees; ~\$15,000), threshold 2 at ₹5 million (~\$77,000), and threshold 3 at ₹50 million (~\$770,000). For the third year, Y3, threshold 1 and threshold 2 were disbanded but the threshold 3 still existed. Y4 onwards, all firms had to file returns at a uniform quarterly frequency and there were no thresholds.

If a firm declared its turnover to be below threshold 1 (the low threshold) in the previous year, then it had to file returns at the annual frequency in year 1 and 2. If a firm declared its turnover to be between threshold 1 and threshold 2 (the middle threshold), then it had to file returns once in every 6 months, twice a year. If a firm declared its turnover to be between threshold 2 and threshold 3 (the high threshold), then it had to file returns at a quarterly frequency. Firms with declared turnover above threshold 3 had to file returns at a monthly frequency. Therefore, in years 1 and 2, the tax authority was receiving returns at annual, biannual, quarterly and monthly frequency. In year 3, the tax authority was receiving returns at quarterly and monthly frequency only. In year 4 and 5, the tax authority was receiving returns only at the quarterly frequency. Firms who were initially filing 12 returns every year, by virtue of declaring their turnover to be greater than ₹50 million the previous year, now had to file only 4 returns every year.

This change in policy allows us to pin-down that the bunching behavior that we observe around these thresholds is indeed happening due to the policy of interest and is not a confounded effect of some other policy.

### 3.3.2 Data

Figure 3.1: Distribution of Firms by Size in the Dataset



Notes: The figure shows the (relatively smooth) distribution of the firms according to their revenue in our sample, combining the samples from each of the years in our analysis (fiscal years 2010/11-2014/15). The x-axis shows the reported firm revenue, while the y-axis shows the corresponding frequency of the firms with the reported revenue shown on the x-axis.

We use the form 16 data described in section 1.4 and section 1.4.1 for our analysis. Since all thresholds of interest are defined on annual levels, if a firm is filing multiple returns in a given financial year, we aggregate the values at the financial year level. We have 5 years of tax returns from the state of Delhi, India from the fiscal year 2010-11 until the fiscal year 2014-15.

### 3.3.3 Bunching at Thresholds

Throughout our bunching analysis, we focus on the effects in the vicinity of each of the thresholds. We divide firms into bins of ₹30,000 around the low (annual firm revenue of ₹1 million  $\sim$  \$15,000) and the middle thresholds (annual revenue of ₹5 million  $\sim$  \$77,000). Similarly, we divide the firms into bins of ₹300,000 around the high threshold (annual firm revenue of ₹50 million  $\sim$  \$770,000).

The lower excluded area, starting at  $R_1$  to the left of a given threshold and ending at the threshold  $T$ , features a discontinuous increase in the distributions of firms right before the threshold. The upper excluded area, starting at the threshold  $T$  and ending at  $R_2$  to the right of a given threshold, features a missing mass in the number of firms immediately after the threshold. For each threshold, we visually observe the discontinuous increase in the number of firms before the threshold and use this visual observation to find the value of  $R_1$ . We use the convergence method, as described by Kleven and Waseem (2013), to find the values of  $R_2$  for each of the thresholds. Specifically, we choose the value of  $R_2$  such that the area above the counterfactual distribution between  $R_1$  and  $T$ , and the area below the counterfactual distribution between  $T$  and  $R_2$  are approximately equal.

As a proxy estimate of the counterfactual distribution, we draw a fitted fourth-degree polynomial across all observations, excluding the lower and upper excluded area, to the left and to the right of a given threshold. We run the following regression to estimate the smooth polynomial:

$$C_j = \sum_{i=1}^4 \beta_i (B_j)^i + \varepsilon_j, \forall B_j \leq R_1 \& B_j \geq R_2, \quad (3.1)$$

where  $C_j$  denotes the count of firms in a given bin  $B_j$ ,  $T$  denotes the threshold,  $R_1$  denotes the beginning of the lower excluded area range (before the threshold), and  $R_2$  denotes the end of the upper excluded area range (after the threshold). Once we

obtain the estimates from equation (3.1), we then use the predicted counterfactual in the excluded range as well, and predict the counterfactual number of firms as follows:

$$\hat{C}_j = \sum_{i=1}^4 \beta_i (B_j)^i \quad (3.2)$$

We then use the estimated counterfactual to estimate the bunching in the lower excluded area (to the left of the threshold), as follows:

$$b = \frac{\sum_{i \in S} (C_i - \hat{C}_i)}{\hat{C}_{lowerexcluded}}, \quad (3.3)$$

where  $S$  denotes the set of values  $i$ , for which the bins  $B_i$  are in the lower excluded area, namely:  $S = \{i \in \mathbb{N} | B_i \in [T - R_1, T]\}$ . The bunching estimate is thus the estimate of the excess mass in the distribution of firms before the threshold relative to the counterfactual distribution of firms, as a share of the counterfactual distribution of firms in the lower excluded area. In particular the latter value,  $\hat{C}_{lowerexcluded}$ , is calculated as a weighted average of the counterfactual in the lower excluded area, weighted by the distance of the actual count of firms from the counterfactual distribution, as follows:

$$\hat{C}_{lowerexcluded} = \sum_{i \in S} \mu_i \hat{C}_i, \quad (3.4)$$

where  $\mu_i$  is the weight of each bin  $i$ , constructed as follows:

$$\mu_i = \frac{C_i - \hat{C}_i}{\sum_{i \in S} (C_i - \hat{C}_i)} \quad (3.5)$$

We represent the counterfactual distribution of firms with a solid red line in all of the figures showing bunching, while we show the values of  $R_1$  and  $R_2$  using the vertical red dashed lines (refer to figure 3.2, figure 3.3, and figure 3.4).



### 3.3.4 VAT Revenue Loss to the Tax Authority

To calculate the tax revenue implications of firm level bunching to avoid increased compliance costs, we need to assume that the firms in the lower excluded area (the excess mass of firms below the threshold), and the firms in the upper excluded area (the missing mass above the threshold) are directly comparable in terms of their unobservable characteristics - apart from their reported revenue and the VAT they remit. We then determine the revenue implication of the threshold regulation as follows. First, we calculate the difference in the average VAT remitted by firms in the upper excluded area and the VAT remitted by firms in the lower excluded area. We then multiply the difference in the average VAT remitted per firm with the extra bunching density, where  $B = \sum_{i \in S} (C_i - \hat{C}_i)$ . The revenue implication,  $R$ , is thus calculated as follows:

$$R = (VAT_{above}^{mean} - VAT_{below}^{mean}) * B \quad (3.6)$$

### 3.3.5 No Benefits to More Information

One question related to the VAT reporting thresholds is why the tax authority might be interested in getting more frequent VAT reports from firms. We test whether the tax authority benefits from the firms providing more frequent information by receiving higher VAT collections when the firms provide more frequent information. We proceed as follows: First, we group all data together, so that we have a pool of observations covering 2010-2015. We then normalize the VAT collected relative to the size of the firm in terms of its revenues, and run the following regression for firm  $i$  in year  $j$ :

$$(VAT/Revenue)_{i,j} = \alpha + \beta \cdot NumberReports_{i,j} + \phi_j + \phi_i + \epsilon_{i,j}, \quad (3.7)$$

where  $VAT/Revenue$  is the ratio of a firm- and year-specific VAT remitted and the reported firm revenue,  $\alpha$  is a constant,  $NumberReports$  is the annual number of reports filed by the firm given its reported firm size,  $\phi_i$  and  $\phi_j$  are the firm and the year fixed effects, respectively (if included in the regression specification, as shown in table 3.2), and  $\epsilon$  are the heterogeneity robust standard errors.

We additionally test the responsiveness of the VAT collected as a share of revenue to the annual number of submitted reports by regressing the  $VAT/Revenue$  ratio on each of the reporting categories, while having the annual reporting as the omitted category. The regression specification for a firm  $i$  in year  $j$  is then:

$$(VAT/Revenue)_{i,j} = \alpha + \sum_{c=semiannual}^{monthly} \beta_c \cdot c + \phi_j + \phi_i + \epsilon_{i,j}, \quad (3.8)$$

where  $VAT/Revenue$  is the ratio of a firm- and year-specific VAT remitted as a share of the reported firm revenue,  $\alpha$  is a constant,  $c$  can take three values: *semiannual* is a dummy variable equal to 1 if the firm filed the VAT reports at a semiannual level in a given year, *quarterly* is a dummy variable equal to 1 if the firm filed the VAT reports at a quarterly level in a given year, *monthly* is a dummy variable equal to 1 if the firm filed the VAT reports at a monthly level in a given year,  $\phi_i$  and  $\phi_j$  are the firm and the year fixed effects, respectively (if included in the regression specification, as shown in table 3.3), and  $\epsilon$  are the heterogeneity robust standard errors.

Using these regression specifications, we can then estimate the relationship between the number of annually filed reports (or the given reporting categories) and the amount of VAT remitted (relative to the firms' reported revenue) using simple OLS regressions.

### 3.3.6 Social Welfare Analysis

We conduct a very simple back of the envelope social welfare analysis. In particular, we recognize that the overall welfare change in our context depends on the sum of the tax authority's revenue losses from the threshold policy, and the easing of compliance costs incurred by the firms in our sample due to the lower filing frequencies that these firms are now required to abide by. Therefore, we recognize that we can calculate the minimum implicit compliance subsidy needed to be given by the tax authority to the firms in a given reporting category in order to at least equalize the revenue losses stemming from the bunching behavior by the firms. If the actual compliance costs for the filing for those firms are higher than the calculated minimum implied subsidies, the overall welfare change stemming from the policy is likely to be positive *net of implicit compliance subsidies*, despite the revenue losses to the tax authority due to the bunching behavior.

We thus calculate the implicit subsidies at each threshold  $i$  in the following manner, using the values of revenue losses,  $R$ , calculated as explained in the subsection 3.3.4:

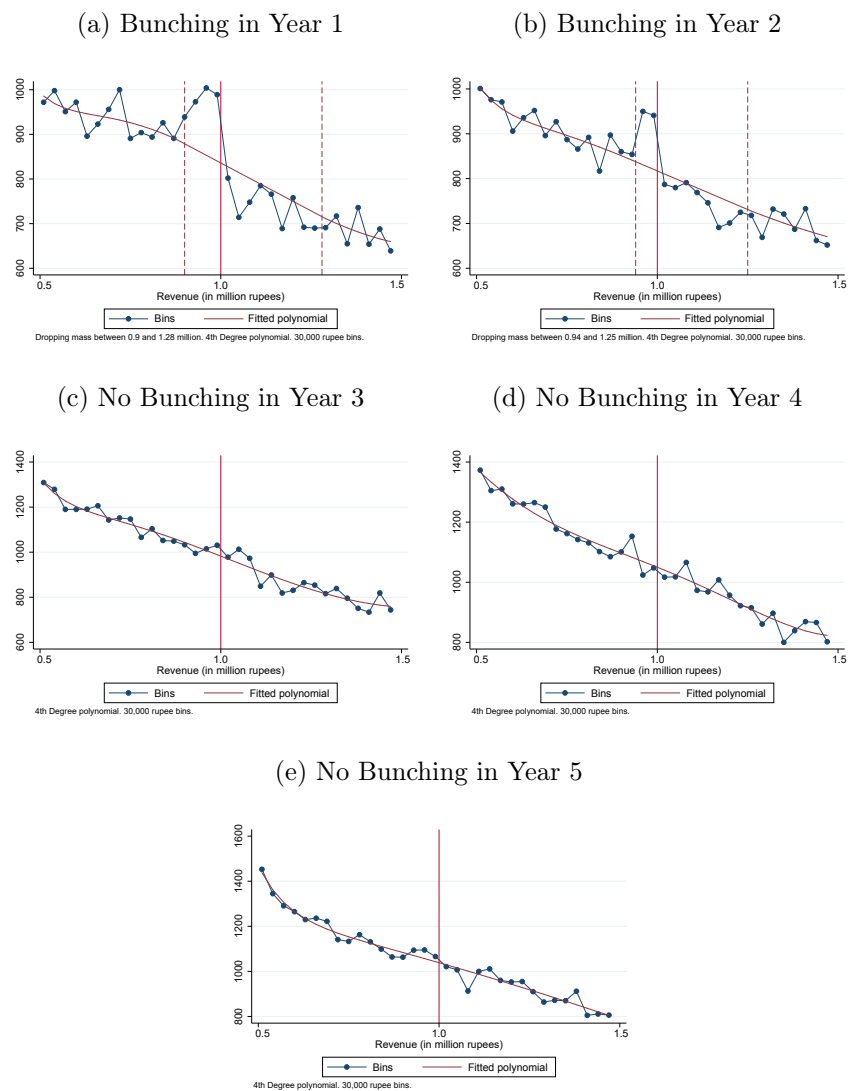
$$ImplicitSubs_i = R_i/n_j, \tag{3.9}$$

where  $n_j$  is the number of firms in the reporting category  $j$  below a given threshold  $i$ . This implies that for the low threshold, we divide the revenue losses stemming from the threshold by the number of firms reporting and paying VAT at an annual level. For the middle threshold, we divide the revenue losses stemming from the threshold by the number of firms reporting and paying VAT at a bi-annual level. Likewise, for the high threshold, we divide the revenue losses stemming from the threshold by the number of firms reporting and paying VAT at a quarterly level.

We then evaluate whether the implicit subsidies needed to equalize the revenue

losses are lower than the likely compliance costs incurred by firms reporting and paying VAT at a certain level of frequency.

Figure 3.2: Firm Revenue Distribution at the Low Threshold



Notes: The figure shows the distribution of the firms around the low threshold (₹1 million), for each of the years in our data (year 1: fiscal year 2010/11 to year 5: fiscal year 2014/15). Panels A and B show the bunching behavior by the firms for the years (year 1 and 2) with differential, size-dependent requirements of VAT filing, while panels C, D and E document that there is no bunching behavior, with the distribution of firms being smooth around the threshold once the differential reporting requirement is done away with (years 3-5).

## 3.4 Results

In this section we describe the results obtained using the methodologies detailed in the previous section. We first discuss the results of estimating the excess bunching at the relevant VAT filing thresholds. The second subsection discusses the calculation of the VAT revenue loss to the tax authority due to the existence of the arbitrary VAT filing thresholds. The third subsection shows that there are no apparent revenue benefits to the tax authority of more frequent information provision by the firms. The fourth subsection concludes by providing the relevant welfare analysis of the size-dependent VAT filing policy.

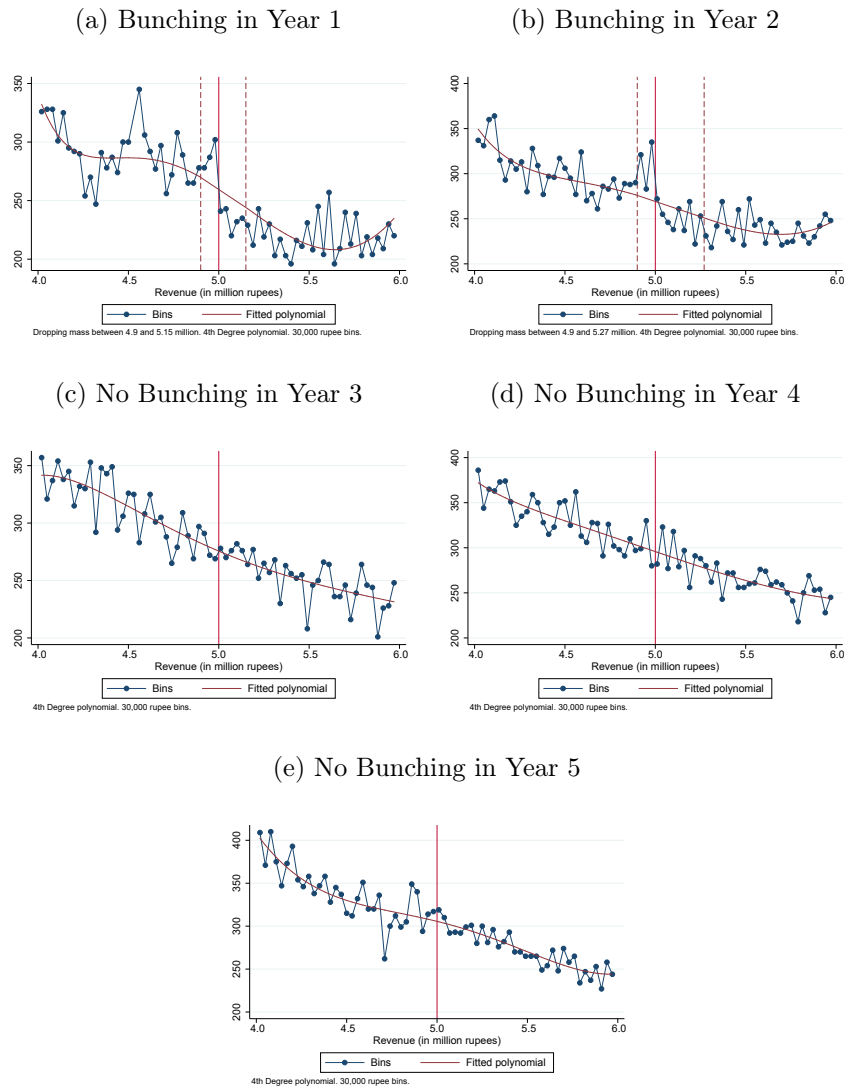
### 3.4.1 Bunching at Thresholds

This section discusses the results of estimating the excess bunching at each of the VAT filing thresholds according to the methodology described in the section 3.3.3.

The low threshold, set at the revenue of ₹1 million, was relevant for the size-dependent VAT filing policy in years 1 and 2 (2010-11 and 2011-12) and mandated the change of return filing from an annual to a bi-annual frequency. Figures 3.1a and 3.1b show that such policy resulted in excess bunching at the threshold, with the bunching estimates of 0.5 for year 1, and 0.28 for year 2, respectively. In years 3, 4, and 5, the policy was discontinued, and all the firms around the threshold were required to file their VAT reports on a quarterly basis. As figures 3.1c to 3.1e illustrate, there was - in contrast to years 1 and 2 - no excess bunching around the low threshold in the later years.

The middle threshold, set at the revenue of ₹5 million, mandated the change of return filing from a bi-annual to a quarterly frequency in years 1 and 2 (2010-11 and 2011-12). Instead of filing the VAT returns twice a year, firms with the reported revenue above ₹5 million in the previous year were required to file their reports four

Figure 3.3: Firm Revenue Distribution at the Middle Threshold

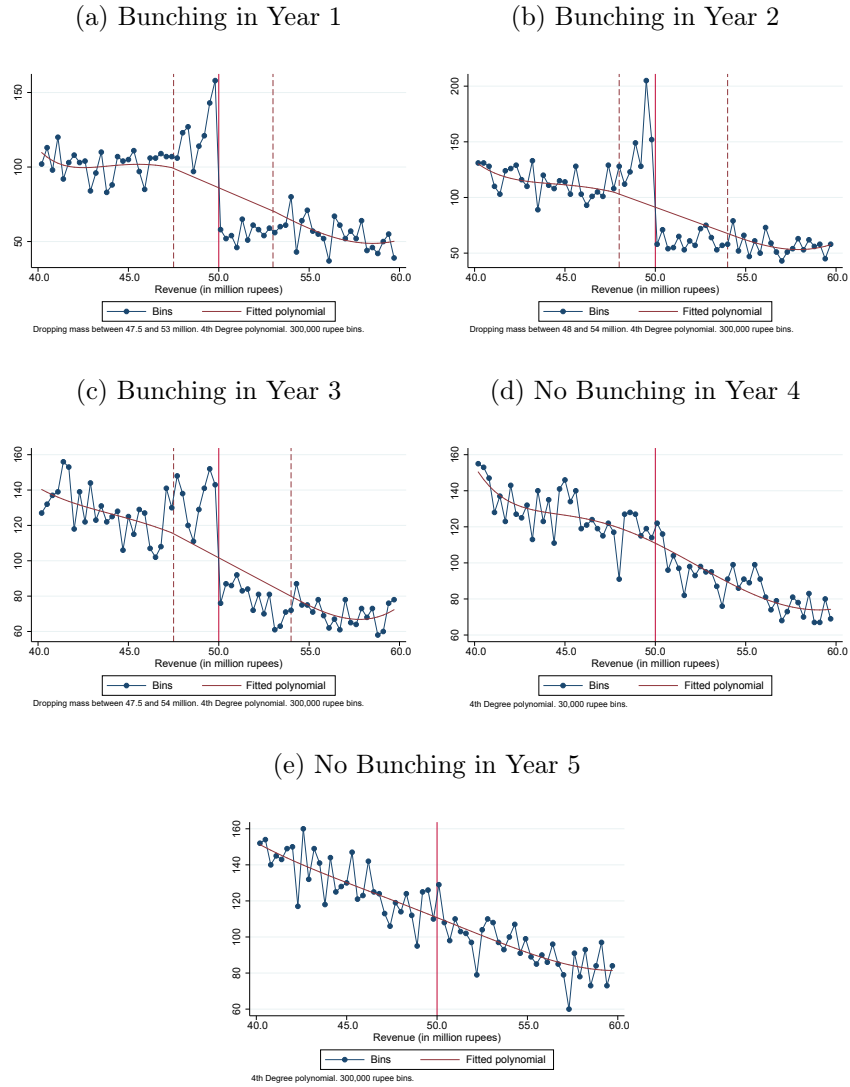


Notes: The figure shows the distribution of the firms around the middle threshold (₹5 million), for each of the years in our data (year 1: fiscal year 2010/11 to year 5: fiscal year 2014/15). Panels A and B show the bunching behavior by the firms for the years (year 1 and 2) with differential, size-dependent requirements of VAT filing, while panels C, D and E document that there is no bunching behavior, with the distribution of firms being smooth around the threshold once the differential reporting requirement is done away with (years 3-5).

times a year. Figures 3.2a and 3.2b show that such a size-dependent policy resulted in a significant bunching, with bunching estimates equaling 0.34 and 0.43, for years 1 and 2, respectively. Figures 3.2c to 3.2e further show that such bunching disappears,

with a much smoother distribution of reported firm revenues, once the threshold policy was done away with in years 3, 4, and 5.

Figure 3.4: Firm Revenue Distribution at the High Threshold



Notes: The figure shows the distribution of the firms around the high threshold (₹50 million), for each of the years in our data (year 1: fiscal year 2010/11 to year 5: fiscal year 2014/15). Panels A-C show the bunching behavior by the firms for the years (years 1, 2, and 3) with differential, size-dependent requirements of VAT filing, while panels D and E document that there is no bunching behavior, with the distribution of firms being smooth around the threshold once the differential reporting requirement is done away with (years 4 and 5).

The highest threshold, set at the revenue of ₹50 million, mandated the change

of return filing from a quarterly frequency to a monthly frequency in years 1, 2, and 3 (2010-11 to 2012-13). In these years, the firms with the reported revenue greater than ₹50 million, in the previous year, had to file VAT returns twelve times a year compared to firms just below the threshold, which had to file VAT returns four times a year. Figures 3.3a to 3.3c again indicate substantial bunching due to such filing threshold, with bunching estimates equaling 2.6, 2.98, and 1.87, in years 1, 2, and 3, respectively. Figures 3.3d and 3.3e show that after the threshold policy was done away with in years 4 and 5 - with all firms now filing the VAT reports on a quarterly basis - the distribution of reported revenues becomes much smoother, with no bunching at the relevant threshold of ₹50 million. This indicates that the observed bunching indeed occurs due to the filing policy.

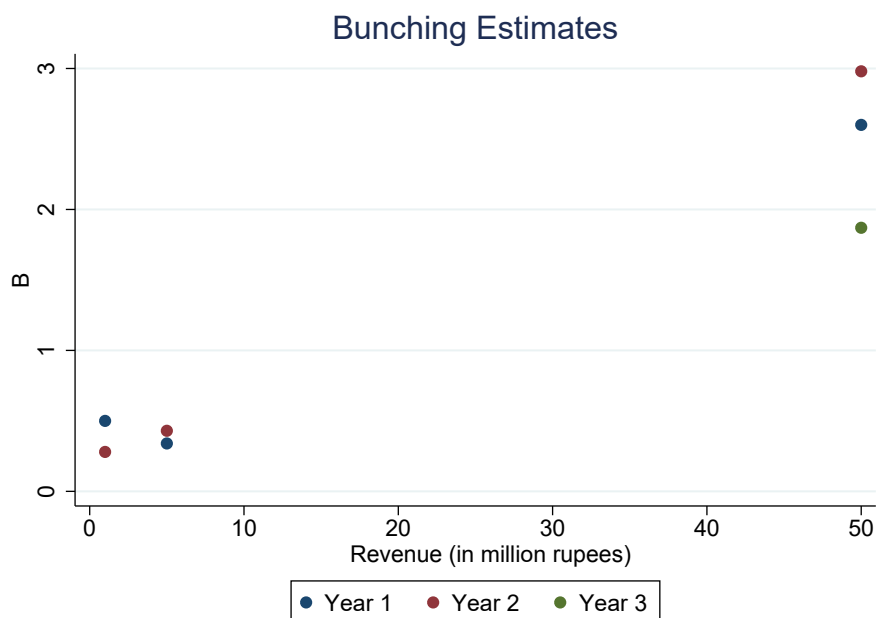
Within each of the thresholds, we see that the bunching occurs at approximately the same magnitude. For the low threshold, we see a decrease of bunching in the second year; for the middle threshold, we see a slight increase in the second year bunching; for the high threshold, we see an increase in bunching from year 1 to 2, followed by a decrease in bunching in year 3. Assuming that the level of bunching is a proxy for actual compliance costs incurred by firms, figure 3.5 uses these bunching estimates to track the compliance costs across different revenue sizes and plots the bunching estimates on the y-axis versus the revenue sizes for the relevant thresholds on the x-axis. While compliance costs seem to remain more or less at the same level across the low and middle threshold, we see a sharp increase in the apparent compliance costs at the high threshold. This implies that compliance costs of a differential VAT reporting policy are generally increasing with the reported revenue.

### **3.4.2 VAT Revenue Loss to the Tax Authority**

This section discusses the calculation of the VAT revenue loss to the tax authority due to the differential, size-dependent filing policy according to the methodology



Figure 3.5: Distribution of Compliance Costs by Revenue Size



Notes: The figure shows the bunching estimates (on the y-axis) against the reported revenue at each of the thresholds, for which the bunching is estimated. The figure provides an indication of increasing compliance costs with the size of the reported revenue thresholds.

described in the section 3.3.4.

We compare the VAT contributions of firms in the bunching region below and the region above the threshold. Table 3.1 shows that the revenue losses are substantial. For the low threshold, the losses amount to nearly ₹900,000 in year 1, and to nearly ₹500,000 in year 2. These amounts are, respectively, 2.3% and 1.8% of the bunchers' (all firms just below the threshold) total VAT contributions, therefore presenting a substantial loss of VAT revenues to the tax authority. For the middle threshold, the losses are generally less substantial, amounting to above ₹700,000 in year 1, and above ₹100,000 in year 2. These amounts respectively equal 1.8% and 0.03% of the bunchers' total VAT contributions in the respective years. The losses are most substantial for the high threshold, amounting to nearly ₹40,000,000 in year 1, and above ₹20,000,000 in years 2 and 3. These amounts are also the highest as a share of

Table 3.1: VAT Revenue Lost

Year	Threshold level		
	Low (₹1 million)	Middle (₹5 million)	High (₹50 million)
2010-11	886,017 (2.3%)	730,557 (1.8%)	38,373,960 (8.3%)
2011-12	441,600 (1.8%)	128,150 (0.03%)	21,862,749 (5.2%)
2013-14			22,871,560 (4.2%)

Notes: Values in Indian rupees. The amounts expressed as a percentage of the bunchers' (i.e. that of firms in the lower excluded area) VAT contributions are in the brackets. The values are calculated as described in the section 3.3.4.

the bunchers' total VAT contributions, amounting to 8.3%, 5.2%, and 4.2% in years 1, 2, and 3, respectively. These estimates confirm the findings in section 3.4.1 that the bunching is not only the highest, but also the costliest for the high threshold.

### 3.4.3 No Benefits to More Information

This section discusses the regressions looking at the relationship between the VAT/revenue ratio and the frequency of VAT reporting, as discussed in the methodology section 3.3.5. Table 3.2 shows the regression results looking at the relationship between the VAT/revenue ratios and the yearly number of VAT reports.

Columns 1 and 2 show that the relationship is strikingly negative without including firm fixed effects in the regression specification, implying that the greater the number of VAT reports per year, the smaller the VAT/revenue ratio. When including firm fixed effects in the regression specification, as shown by columns 3 and 4, the relationship turns positive but strongly insignificant.

Table 3.3 similarly shows the regression results looking at the relationship between

Table 3.2: VAT Revenue vs. Annual Number of VAT Returns Regressions

VARIABLES	(1) VAT/Revenue	(2) VAT/Revenue	(3) VAT/Revenue	(4) VAT/Revenue
NumberReports	-0.000148*** (3.18e-05)	-0.000314*** (0.000100)	0.000474 (0.000468)	0.000247 (0.000278)
Firm FE	NO	NO	YES	YES
Time FE	NO	YES	NO	YES
Observations	1,038,331	1,038,331	1,038,331	1,038,331
No. of firms	301,147	301,147	301,147	301,147
$R^2$	0.000	0.000	0.703	0.703

Notes: Robust standard errors, clustered at the firm level, in parentheses. The regression table presents the results of regressions performed according to the methodology described in the section 3.3.5. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

the VAT/revenue ratios and different reporting categories, with annual reporting being the omitted category. As columns 1 and 2 show, when not including firm fixed effects, all reporting categories are negatively related to the VAT/revenue ratios compared to those firm-year observations with annual reporting. Once including firm fixed effects in the regression, the estimates turn positive, but again remain strongly insignificant.

Our interpretation of the results of both of the sets of regressions is that greater than annual frequency of VAT reporting does not lead to more VAT being collected, as measured by the VAT/revenue ratios.

### 3.4.4 Social Welfare Analysis

In this section we discuss the results of the back of the envelope social welfare analysis that we conducted according to the methodology outlined in section 3.3.6. In table 3.4 we present the calculated implicit subsidies. The implicit subsidies needed to equalize welfare are very low. At the low threshold for year 1 - an implicit welfare

Table 3.3: VAT Revenue vs. VAT Filing Categories Regressions

VARIABLES	(1) VAT/Revenue	(2) VAT/Revenue	(3) VAT/Revenue	(4) VAT/Revenue
SemiAnnualCategory	-0.00685*** (0.000206)	-0.00685*** (0.000206)	0.00521 (0.00782)	0.00516 (0.00778)
QuarterlyCategory	-0.00312* (0.00162)	-0.00677*** (0.000313)	0.00696 (0.00823)	0.00606 (0.00819)
MonthlyCategory	-0.00541*** (0.000286)	-0.00713*** (0.000999)	0.00670 (0.00819)	0.00546 (0.00717)
Firm FE	NO	NO	YES	YES
Time FE	NO	YES	NO	YES
Observations	1,038,331	1,038,331	1,038,331	1,038,331
No. of firms	301,147	301,147	301,147	301,147
$R^2$	0.000	0.000	0.703	0.703

Notes: Robust standard errors, clustered at the firm level, in parentheses. Annual reporting is the omitted category. The regression table presents the results of regressions performed according to the methodology described in the section 3.3.5. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

equalizing per firm subsidy of ₹40.77 (less than \$1) implies that the welfare change of reducing the frequency would be negative only in the case that compliance costs borne by a given firm reporting and paying VAT at an annual level would be less than \$1, which is extremely unlikely. The implicit welfare equalizing per firm subsidy for year 2, at the low threshold is even smaller. Similarly, the welfare equalizing subsidies for the middle threshold are comparably low.

The implicit welfare equalizing subsidies for the high threshold are an order of magnitude higher, ranging from ₹1327.97 (about \$25) in year 3 to ₹2990.02 (about \$60) in year 1. The implication of these implicit welfare equalizing per firm subsidies is, however, similar as for the low and the middle thresholds. The actual compliance - that is, filing and paying - costs difference (quarterly level of reporting relative to monthly reporting) associated with VAT of these firms, reporting at a quarterly level is certainly higher than \$60. That means that the tax authority subsidizes these

Table 3.4: Implicit Welfare Equalizing Per Firm Subsidies from the Thresholds

Year	Threshold level		
	Low threshold ( ₹1 million)	Middle threshold ( ₹5 million)	High threshold ( ₹50 million)
2010-11	40.77	14.54	2990.02
2011-12	17.31	2.55	1505.49
2013-14			1327.97

Notes: Values in Indian rupees. The implicit minimum welfare equalizing per firm subsidies are computed according to the methodology in the section 3.3.6

firms by reducing the frequency at which they have to file, and yet increases the social welfare. Put another way, this analysis leads us to conclude that even though the tax authority incurs significant losses due to thresholds, its implicit subsidies to small and medium-sized firms are large enough to overcome the revenue losses.

### 3.5 Conclusion

In this paper, we identify bunching behavior by firms at the VAT filing thresholds, using an administrative-level dataset from the state of Delhi in India. Our unique dataset along with rich policy variation also enables us to show that the bunching completely disappears when the thresholds are done away with. The bunching is the greatest at the highest threshold indicating that compliance is most costly for those firms. We further show that the VAT revenue losses due to the bunching response by firms are substantial, up to 8.3% of the bunchers' total VAT contributions in certain years. We also suggest that there may be no benefits to the tax authority from receiving more frequent information through the VAT filing reports. Finally, our social welfare analysis shows that given that the costs of compliance by the firms are likely relatively large, the net-of-implicit-subsidies welfare impact of the thresholds is positive despite the substantial revenue losses to the tax authority.

We note two potential channels through which the firms may bunch: underpro-

duction and underreporting. The first channel, underproduction, would imply that the bunching firms halt their production or sales as soon as they start getting close to the threshold. This channel potentially brings about large welfare losses, not only because of the current reduction in the firm-level profits, but also due to the shift in the long-run growth path of the bunching firms. In the case of such channel being the dominant one, the welfare losses due to the policy (while not considering the implicit subsidies due to the reduced filing frequencies for smaller firms) would be the highest, with real and longer-term consequences in the form of stalled firm growth. The second potential bunching channel encompasses the intentional revenue underreporting by the firms. If such underreporting occurs and is substantial, the welfare losses would occur via the lost tax revenues. One potential sub-channel in the firm revenue underreporting is revenue shifting: in such a case, a portion of the bunching firms illegitimately register a part of their sales for the following filing period in order to avoid passing the relevant reporting threshold. In such a case, the welfare consequences of the policy would be the smallest of the three bunching channel options; the firms would only misreport the current revenues to avoid more frequent filing, but would not affect their actual long term real growth, and would eventually report all of their revenues. In future work, we intend to make progress on identifying which of the potential channels are in play.

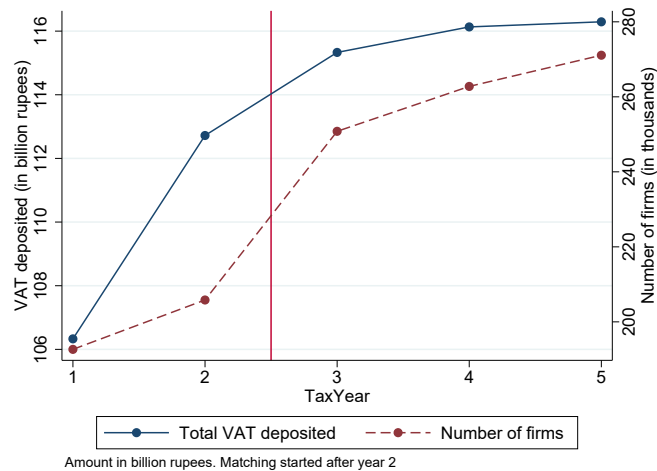
All in all, our analysis shows that firms' compliance costs to the VAT policy are quite large, with unclear benefits of more frequent information and payments being made by the VAT registered firms. Tax authorities should thus aim to reduce the compliance costs, which apparently play a substantial role in the firms' behavior.

# APPENDIX A

## “VAT in Emerging Economies: Does Third Party Verification Matter?”

### A.1 Data Summary

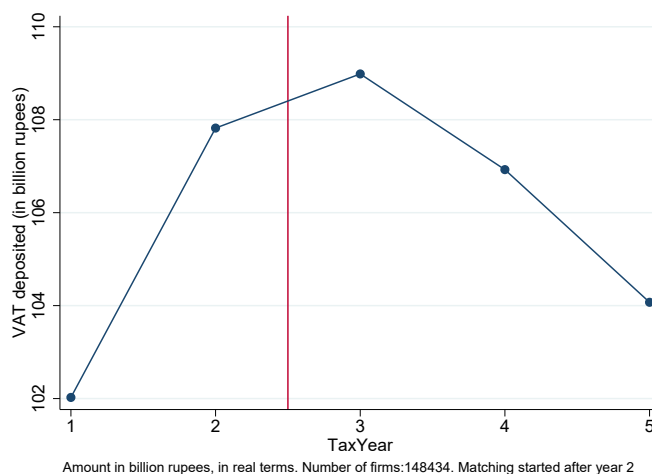
Figure A.1: Total VAT Remitted (For All Firms)



The third party verification policy began at the beginning of year 3. VAT remitted increases from ₹106.33 billion in year 1 to ₹116.29 billion in year 5. This is an average annual growth rate of 1.8% in real terms as compared to a real state level GDP growth rate of about 5.7% (Source:(Directorate of Economics and Statistics, GNCTD, 2015)). The y axis is in billion rupees, where 1\$ roughly equals ₹65. Values have been adjusted to year 1 price terms.

In this section, we describe the distribution of the firms registered in the Delhi VAT system. In figure A.1, we plot the total VAT collections and the total number of firms registered for VAT across the 5 years. The third-party verification policy was implemented at the beginning of year 3. VAT deposits increase from ₹106.33

Figure A.2: Total VAT Remitted (For Firms That are Present in All Years)



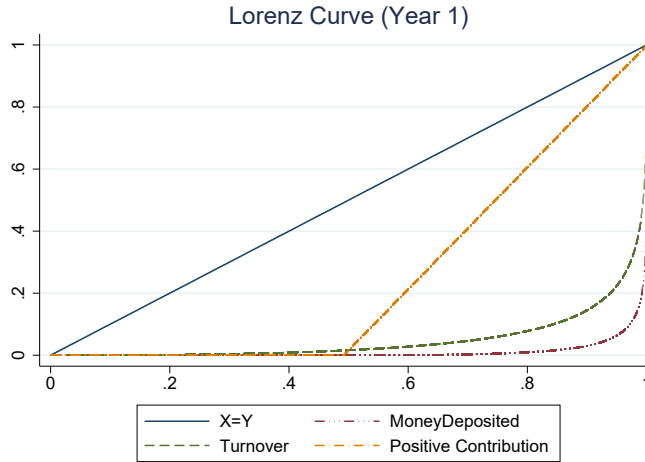
Total collection trends for firms that are present in all the years of our sample period. There are 148434 such firms. The y axis is in billion rupees, where 1\$ roughly equals ₹65. Values have been adjusted to year 1 price terms. VAT remitted increases from ₹102.02 billion in year 1 to ₹104.07 billion in year 5.

billion in year 1 to ₹116.29 billion in year 5. This is an average annual growth rate of 1.8% in real terms as compared to a real state level GDP growth rate of about 5.7%. We note that the number of registered firms go up sharply after the policy change which implies that average collections per firm actually decrease. We believe that this increase in the number of firms is driven by unrelated policy changes. Specifically, before quarter 2 of year three, firms had to deposit a surety amount between ₹50,000 to ₹100,000 as part of registration. The tax authority relaxed this restriction with the goal of improving “ease of doing business” from the second quarter of year three onwards.

The number of firms filing a return increases from 192,664 in year 1 to 271,090 in year 5. There is a wide variation in the amount deposited. To begin with, only about 50% of registered firms remit a positive VAT in any given filing period. Further, between 7 and 15% of the firms (depending upon the tax period) that file a return report a zero turnover (sales). Furthermore, between 5 and 9% of the firms declare their entire turnover to consist of interstate (or non-local) sales and about 32% firms



Figure A.3: Lorenz Curve for All Firms in Tax Year 1



Note: Only 50% of firms deposit any taxes, and 5% of firms provide 95% of total collections. Lorenz curve for all the firms in year 1 of our dataset (returns collated at annual level).

Table A.1: Summary Stats: All Firms

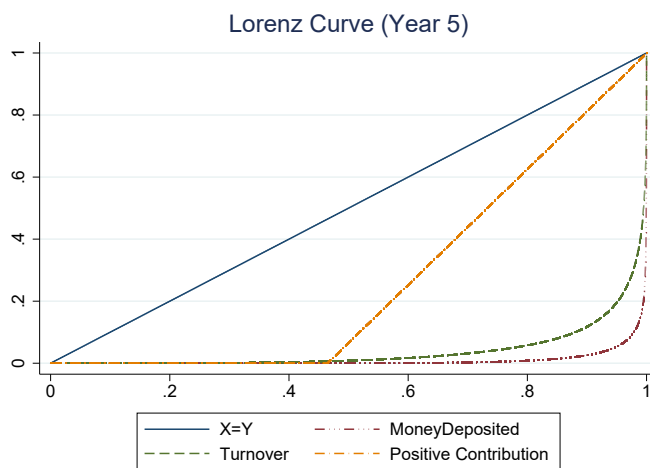
(1) Year	(2) No. of Firms	(3) VAT Remitted	(4) % Positive VAT Deposited Firms	(5) % Zero-Turnover Firms	(6) % Interstate Firms	(7) % Local Firms
1	192664	106330.3	50.88	7.10	9.03	31.26
2	205832	112720.1	48.72	9.51	7.72	31.40
3	250805	115330.6	47.57	15.05	5.94	31.68
4	262775	116132.1	49.70	13.68	5.70	32.70
5	271090	116292.4	53.60	13.98	6.00	32.64

Summary of all the firms that filed a return in the given year. Column (3) shows total VAT collected by the tax authority from all firms in that year in million rupees, with ₹65 approximately equal to \$1. Column (4) shows percentage of firms that deposited a positive amount of VAT. Column (5) show percentage of firms which filed a return but declared a turnover of zero. Column (6) shows percentage of firms that had a non-zero turnover and entire sales were interstate. Column (7) shows percentage of firms who had a non-zero turnover and all sales were local. For example, in year 1, 31.26% firms had only local sales, 9.03% had only interstate sales, and 7.1% had a turnover of 0. Therefore, roughly 53% of the firms had a non-zero turnover and had declared both local as well as inter-state sales.

declare their entire turnover to be purely local (refer to Table A.1). Note that the third party verification mechanism breaks down for inter-state sales since the transacting firm’s returns are submitted to a different tax authority and to date there has been little coordination between different tax jurisdictions on such cross-checking.<sup>1</sup>

<sup>1</sup>The GST bill legislated by the central and state governments will unify the tax administration and make it much easier to cross-check inter-state transactions.

Figure A.4: Lorenz Curve for All Firms in Tax Year 5



In figures A.3 and A.4 we plot Lorenz curves for (a) total turnover, (b) VAT remitted, and (c) a dummy for any positive VAT remitted separately for year 1 and 5 of the study. Inequality in contributions is stark with the top 5% of firms remitting roughly 95% of the total VAT collected by the tax authority. The number of firms that remit any VAT is surprisingly low, with the number hovering around 50% across the 5 years.

In figure A.2 we focus our attention on firms which are present in all 5 years of our dataset i.e. we drop firms that enter or exit during our time-frame of interest. There are 148,434 such firms which remit roughly 95% of our total tax collections in year 1 and 89% of the tax collections in year 5. VAT remits from these firms go from ₹102.02 billion in year 1 to ₹104.07 billion in year 5 (for a real growth rate of 0.39%). In this set of firms, the percentage of firms remitting a positive amount goes up marginally, compared to all firms, to about 57%. The percentage of firms that declare a turnover of zero is between 2.5 to 8.5% across the 5 years. The percentage of firms doing only interstate sales and only local sales is also comparable to the entire sample (refer to table A.2). To conclude, our estimation sample comprises the bulk of the tax collections for the state throughout the study period.

Table A.2: Summary stats: Always Present Firms

(1)	(2)	(3)	(4)	(5)	(6)
Year	VATDeposited	% Positive VAT Deposited Firms	% Zero-Turnover Firms	% Interstate Firms	% Local Firms
1	102024.5	54.60	2.50	6.97	30.76
2	107820.3	54.09	3.09	5.95	31.14
3	108985.1	57.20	3.88	5.34	30.61
4	106926.4	57.50	5.35	5.18	30.45
5	104071.8	60.49	8.50	5.22	29.74

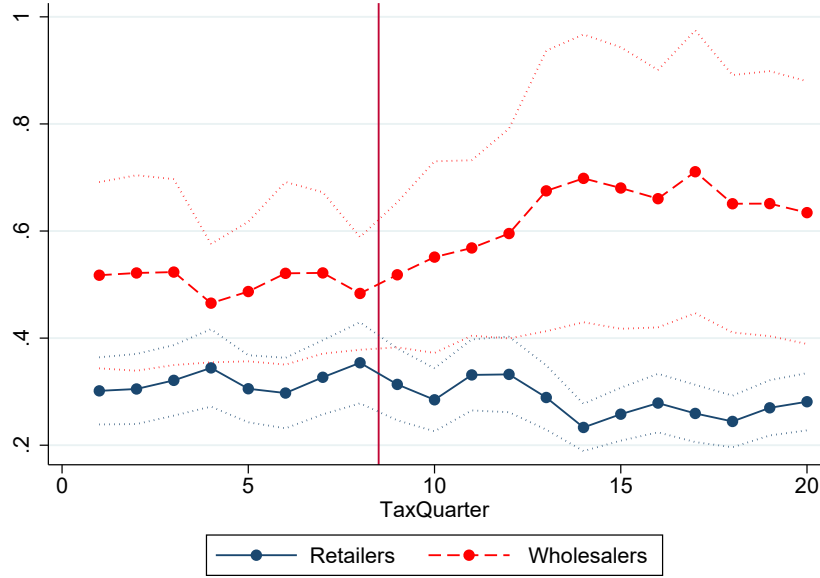
Summary of firms that filed a return in all the 5 years for which we have the data (2010-11 to 2014-15). Number of such firms in our sample is 148434. Column (2) shows total VAT remitted by the tax authority from all firms in that year in million rupees, with ₹65 approximately equal to \$1. Column (3) shows percentage of firms that remitted a positive amount of VAT. Column (4) show percentage of firms which filed a return but declared a turnover of zero. Column (5) shows percentage of firms that had a non-zero turnover and entire sales were interstate. Column (6) shows percentage of firms who had a non-zero turnover and all sales were local. For example, in year 1, 30.76% of the 148434 firms that are present in all the years of our sample, had only local sales, 6.97% had only interstate sales, and 2.5% had a turnover of 0. Therefore, roughly 60% of the firms had a non-zero turnover and had declared both local as well as inter-state sales.

## A.2 Quarterly Results

### A.2.1 Quarterly Analysis Along with Falsification Test

In addition to the basic model explained in section 1.5.2, we carry out robustness check by the falsification test described below. In equation (1.1), we now add a  $Pre_{it}$  dummy which is equal to 1 if the quarter is 7 or 8 i.e. just before the introduction of the third party verification policy. We also interact the  $Pre_{it}$  dummy with  $\mathbb{I}\{\text{Wholesaler}_i\}$  which is a dummy for the firm  $i$  being a wholesaler. The falsification test is provided by coefficient  $\mu$  corresponding to the interaction between  $Pre_{it}$  dummy and  $\mathbb{I}\{\text{Wholesaler}_i\}$  dummy. The coefficient indicates whether, before the introduction of the policy, the outcome variable is evolving similarly between wholesalers and retailers during the 2 quarters before the introduction of the policy. Consistent with the quarterly event-study analysis (figure A.7), the pre-policy effect

Figure A.5: Wholesalers vs Retailers: Quarterly Trends



$N_W = 11482$ ,  $N_R = 15337$ . VAT remitted is in million rupees, with ₹65 approximately equal to \$1. VAT remitted has been inflation adjusted to Q1-2010-11 price levels. Sample smaller than the annual frequency sample because in year 1 and year 2 firms with turnover less than 5 million had to file at annual or semi-annual frequency.

on wholesalers is close to zero, precisely estimated,<sup>2</sup> and statistically insignificant.

Results shown in table A.3.

$$y_{it} = \alpha_i + \nu_t + \beta * Post_{it} + \delta * Pre_{it} + \gamma * Post_{it} * \mathbb{I}\{\text{Wholesaler}_i\} + \mu * Pre_{it} * \mathbb{I}\{\text{Wholesaler}_i\} + \epsilon_{it} \quad (\text{A.1})$$

---

<sup>2</sup>The standard errors are smaller than those for  $\gamma$  coefficient.

Table A.3: Falsification Test: Wholesalers and Retailers (Quarterly)

VARIABLES	(1) Positive VAT Remitted	(2) VAT Remitted	(3) Tax Credit	(4) Output Tax	(5) Output Tax - Tax Credit
Post*Wholesaler	-0.0147*** (0.00355)	0.158*** (0.0491)	-0.0240 (0.0496)	0.132*** (0.0392)	0.156*** (0.0501)
PrePolicy*Wholesaler	-0.00278 (0.00372)	-0.0312 (0.0304)	0.0385 (0.0308)	0.00530 (0.0269)	-0.0332 (0.0320)
Post	0.0139*** (0.00349)	-0.0604* (0.0322)	0.0382* (0.0220)	-0.0194 (0.0392)	-0.0576** (0.0289)
PrePolicy	0.0189*** (0.00352)	0.0288 (0.0210)	0.0637*** (0.0240)	0.0928*** (0.0359)	0.0291 (0.0201)
Mean Dep.Var.	0.44 (0.00)	0.52 (0.09)	0.54 (0.15)	1.02 (0.22)	.48 (0.09)
Observations	536,380	536,380	536,380	536,380	536,380
R-squared	0.549	0.86	0.78	0.96	0.86
Number of Firms	26,819	26,819	26,819	26,819	26,819

Robust standard errors in parentheses, clustered at firm level.  $N_W = 11482$ ,  $N_R = 15337$ . Monetary amounts are in million rupees, inflation adjusted to price levels of Q1 of 2010-11, with ₹65 approximately equal to \$1. Column (1) shows linear probability regression of the probability of remitting a positive amount. Column (2)-(4) respectively show regression of the mean VAT remitted by firms, of the input tax credit claimed by firms, and the output tax collected by firms. To address the concern that VAT remitted has a significant mass at zero, column(5) shows regression of the difference between output tax and input credit declared by firms. Dependent variables have been price adjusted in Q1 of 2010-11 terms. Row “Mean Dep.Var.” shows mean and standard errors for wholesalers in quarter 1. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## A.3 Flexible DID Specifications

### A.3.1 Econometric Model of Event Study Analysis

We can take advantage of the large N and moderate T dimensions in our data set to estimate richer treatment effect models. In particular, we can examine flexibly the differential evolution of outcomes between wholesalers and retailers over the entire time-period under study. In particular, we estimate equation (1.2) as outlined in section 1.5.2. We normalize  $\gamma_2$  in that equation ( $\gamma_8$  for quarterly analysis) to zero so that the coefficients  $\{\gamma_s\}_{s \in S}$  measure differential changes in the outcome and relative to the year (quarter) prior to the policy introduction.

We report the coefficients in graphical form, with their corresponding 95% confidence intervals (see for example figure A.6). If the coefficients  $\{\gamma_s\}_{s>2}$  (i.e., the coefficients after the introduction of the third party verification) are positive, that implies that outcomes increase on average for wholesalers relative to retailers after the policy introduction. If  $\gamma_s$  is positive only for some  $2 < s \leq \bar{s} \leq 5$ , then the

Table A.4: Pre-trend Analysis: Wholesalers vs Retailers

VARIABLES	(1) Annual	(2) Quarter	(3) Top Decile (Annual)	(4) Top Decile (Quarter)
Positive VAT Deposited	0.68	0.27	0.02	0.00
VAT Deposited	0.61	0.20	0.46	0.27
Tax Credit	0.12	0.36	0.69	0.43
Output Tax	0.23	0.65	0.98	0.28
Output Tax - Tax Credit	0.56	0.54	0.56	0.69

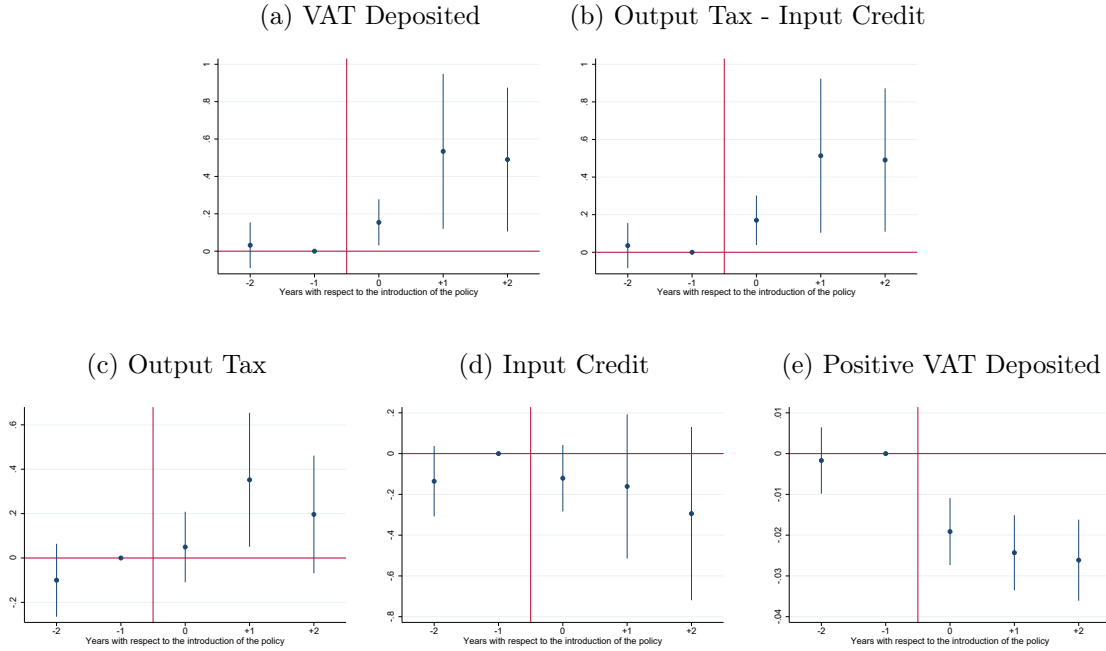
To formally test the hypothesis that pre-policy trends between wholesalers and retailers are not different, we test the null hypothesis  $\gamma_1 = \gamma_2 = \dots = \gamma_{l-1} = 0$  where  $l$  denotes the time period in which the automatic third party verification policy was introduced.  $l$  is 3 in column (1) and (3), and 9 in column (2) and (4). Column (1) does the test for returns data at annual frequency, column (2) does the test for returns at quarterly frequency, and column (3) and (4) do the test for returns data at annual and quarterly frequency but only for firms in the top decile (of both retailers and wholesalers) of VAT remitted in year/quarter 1.

increase in outcomes is transitory, lasting up to  $\bar{s}$  years after the policy introduction. If the coefficient  $\gamma_1$  (i.e., the coefficient in the figure to the left of the policy introduction) is positive, then the outcome variable was declining before the introduction of the policy.

To formally test the hypothesis that pre-policy trends between wholesalers and retailers are not different, we test the null hypothesis  $\gamma_1 = \gamma_2 = \dots = \gamma_{l-1} = 0$  where  $l$  denotes the time period in which the automatic third party verification policy was introduced. For the analyses below (carried out for the quarterly and annual frequencies) we cannot reject that pre-treatment trends are equal between wholesalers and retailers.

### A.3.2 Annual Analysis

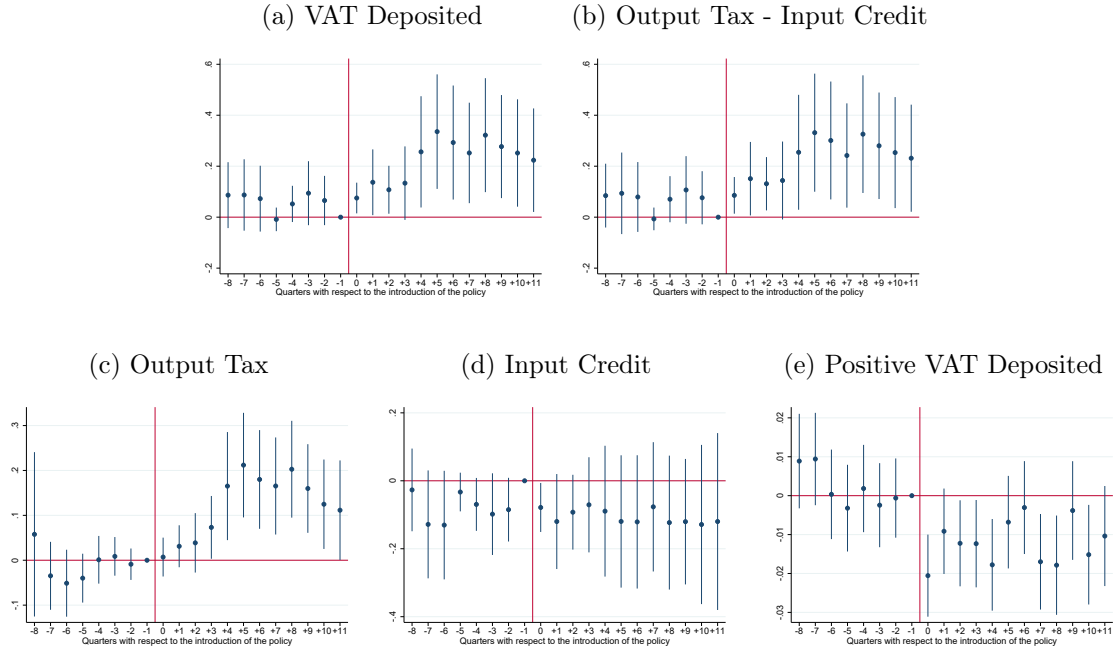
Figure A.6: Event Study: Wholesalers vs Retailers (Annual)



Notes: Graphical event-study analysis of the effect of third party verification policy on wholesalers compared to retailers. Confidence intervals were constructed with heteroskedasticity-robust standard errors, clustered at the firm level. The coefficient for the year “-1” (i.e., the year prior the policy) was normalized to zero. The regressions include firm fixed effects and time effects. The x axis indicates time, with annual observations and zero indicates the first year of the third party verification policy. Confidence intervals at the 95% level.  $N_W = 32979$ ,  $N_R = 19515$ . Coefficients in panel (a), (b), (c) and (d) are in million rupees with dependent variables price adjusted to the first year levels. ₹65 approximately equal to \$1. Pretrends are not statistically significant.

### A.3.3 Quarterly Analysis

Figure A.7: Event Study: Wholesalers vs Retailers (Quarterly)

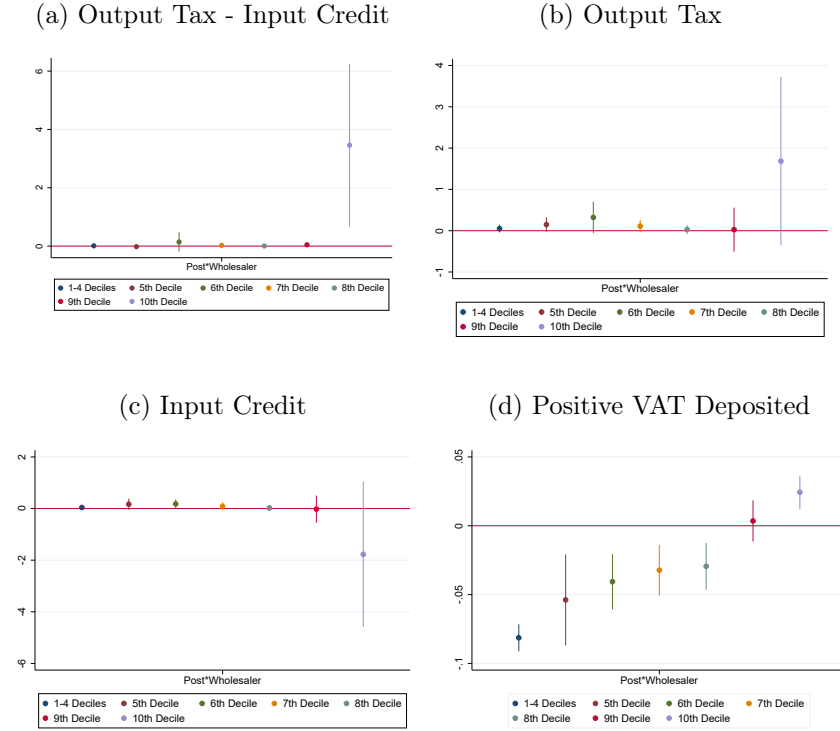


Notes: Graphical event-study analysis of the effect of third party verification policy on wholesalers compared to retailers. Confidence intervals were constructed with heteroskedasticity-robust standard errors, clustered at the firm level. The coefficient for the quarter “-1” (i.e., the quarter prior to the policy) was normalized to zero. The regressions include firm fixed effects and time effects. The x axis indicates time, with quarterly observations and zero indicates the first quarter of the third party verification policy. Confidence intervals at the 95% level.  $N_W = 11482$ ,  $N_R = 15337$ . Coefficients in panel (a), (b), (c) and (d) are in million rupees with dependent variables price adjusted to the first quarter of 2010 levels. ₹65 approximately equal to \$1. Pretrends are not statistically significant.



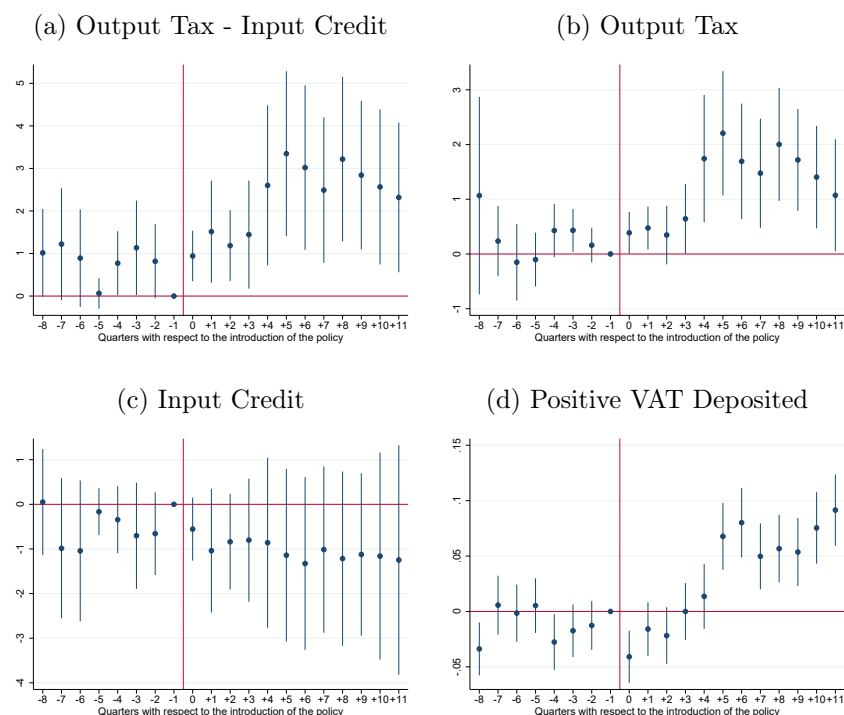
## A.4 Other Results

Figure A.8: Heterogeneity Analysis: Wholesalers vs Retailers



Notes: This figure plots the difference between wholesalers and retailers for different deciles, based on VAT remitted in the first year. The x axis indicates the decile. Confidence intervals at the 95% level. Number of retailers is 32979 and number of wholesalers is 19515. Coefficients in panel (a), (b), and (c) are in million rupees, price adjusted to 2010-11 levels, with ₹65 approximately equal to \$1. Pretrends are not statistically significant.

Figure A.9: Quarterly Analysis for Top Decile: Wholesalers vs Retailers

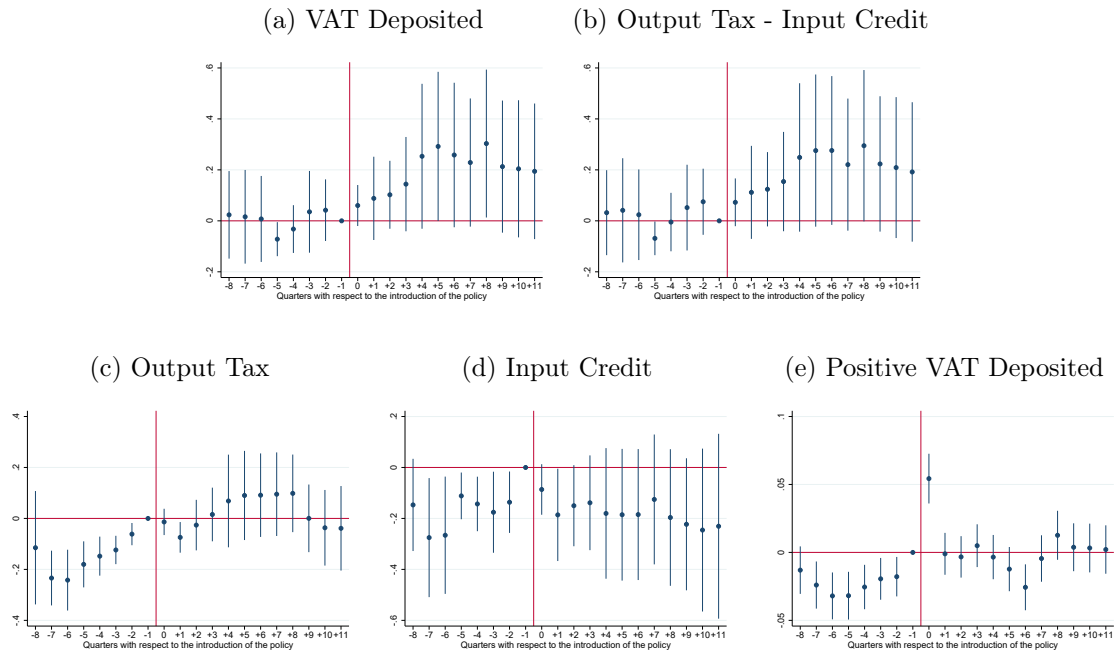


Notes: Graphical event-study analysis of the effect of third party verification policy on wholesalers compared to retailers (for the top decile). Confidence intervals were constructed with heteroskedasticity-robust standard errors, clustered at the firm level. The coefficient for the group “-1” (i.e., the year prior the policy) was normalized to zero. The regressions include firm fixed effects and time effects. The x axis indicates time, with quarterly observations and zero indicates the first year of the third party verification policy. We have 20 quarters of data from 2010-11 to 2014-15. Confidence intervals at the 95% level. Number of retailers is 1533 and number of wholesalers is 1148. Coefficients in panel (a), (b), (c) and (d) are in million rupees, price adjusted to Q1 of 2010-11 levels, with ₹65 approximately equal to \$1. Pretrends are not statistically significant.

## A.5 Policy Execution: Other

### A.5.1 DiDiD: Proportion Registered Sales

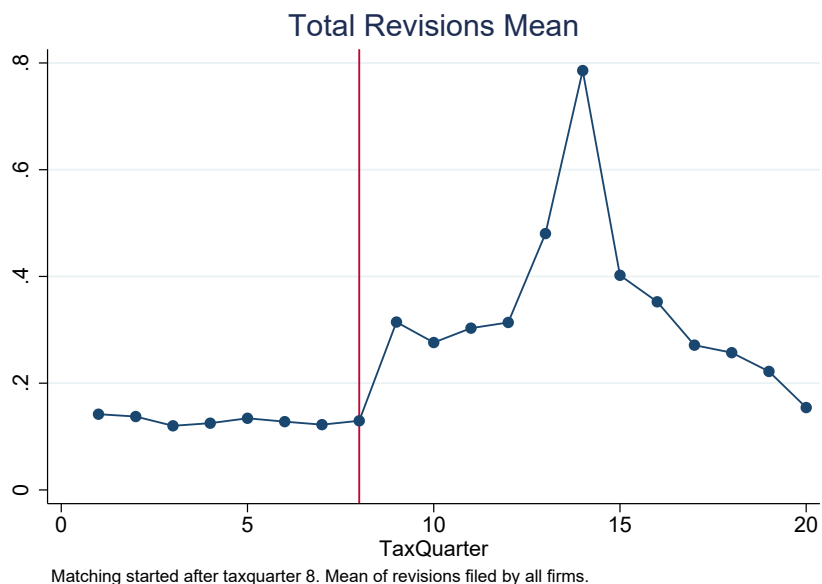
Figure A.10: DiDiD: Wholesalers vs Retailers



Notes: This figure plots the difference between wholesalers and retailers explained in equation (1.3). The x axis indicates time, with quarterly observations and zero indicates the first quarter of the third party verification policy. Confidence intervals at the 95% level. Number of retailers is 15337 and number of wholesalers is 11482. Coefficients in panel (a), (b), (c) and (d) are in million rupees, price adjusted to Q1 of 2010-11 levels. Pretrends are not statistically significant.

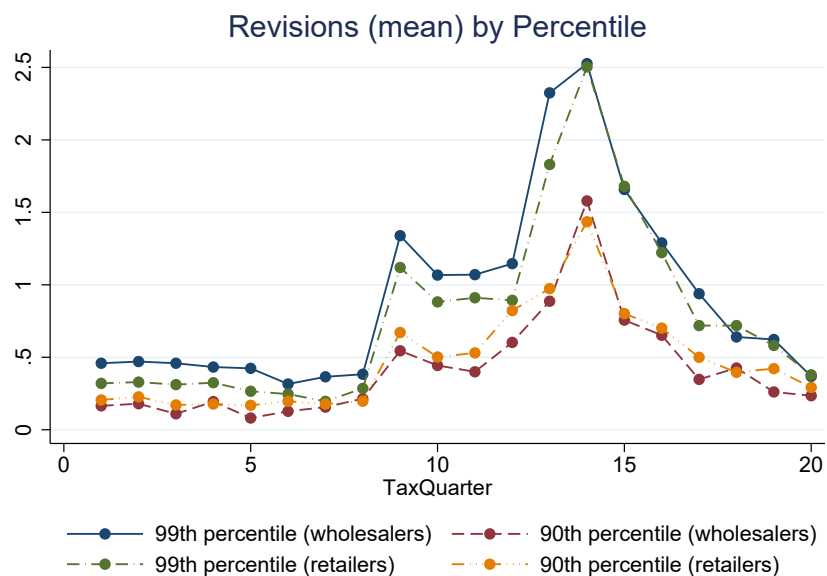
## A.5.2 Revisions

Figure A.11: Mean Revisions of All Firms



Firms are allowed to revise filed returns until the end of next financial year. Before the start of the monitoring policy the average revision rates were around 13% i.e. the mean of the total number of times a firm filed its returns was 1.13 (in Y1 and Y2). This revision rate was constant in the pre-period for both retailers and wholesalers. However, immediately after the introduction of the policy, the revision rates shot up to 30% i.e. the mean of the total number of returns filed by a firm was now 1.3 (in Y3). As the issue mentioned between the consolidated and transaction returns was fixed in Y4, this number further shot up temporarily in Y4 (Q13) up to 78% in Q14 and subsequently started coming down but remained higher than the average amount in Y1 and Y2. This happened as now the firms had to file transaction level as well as the consolidated information (refer to figure A.11). This behavior points towards two scenarios. Either the cost of complying with the tax policy is going up, or the firms are colluding and the increase in revisions is due to

Figure A.12: Mean Revisions of Wholesalers and Retailers



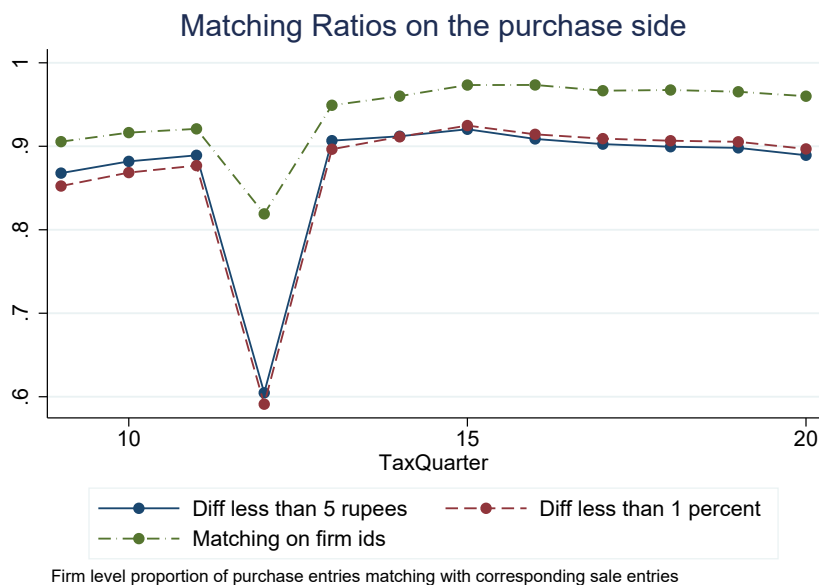
Describing revision trends by percentile. Comparing the firms in the top percentile with the firms in the 90th percentile

coordination costs. Either ways, it is important to think through the efficacy of the third party verification policy. Specifically, if most of the gain in revenue is from the top percentile of firms, then increasing the cost of compliance for firms across the board may not be cost efficient, both for the firms as well as the tax authority.

There seems to be size based variation in the revision trends as well. When we compare wholesalers and retailers, we see that the 99th percentile (in terms of VAT deposited) of both the wholesalers and retailers revise their returns at a greater frequency than the 90th percentile firms (refer to figure A.12). This again hints towards the increase in revisions being driven by the increased cost of compliance.

### A.5.3 Matching

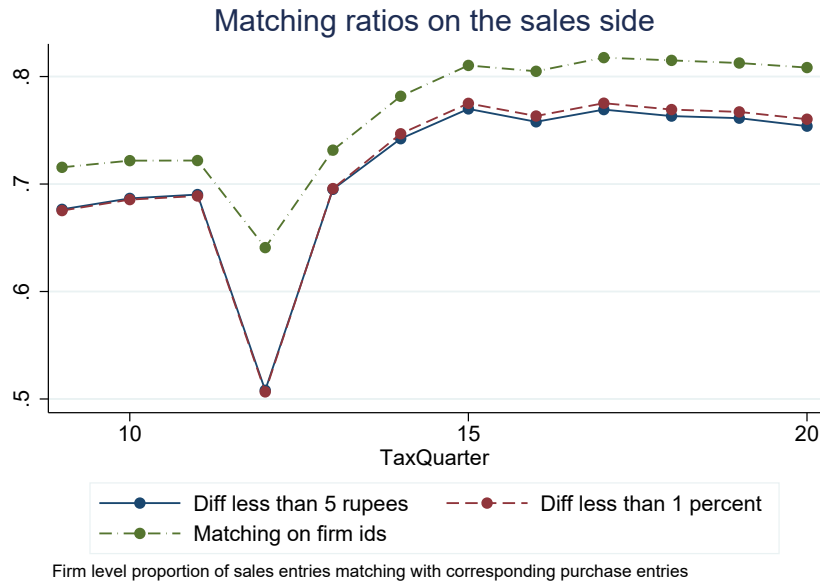
Figure A.13: Matching on the Purchase Side for All Firms



Firm level mean of purchases transactions matching with transactions declared by selling firms

In figure A.13, figure A.14, figure A.15, and figure A.16 we show the accuracy with which the sale and purchase transactions of firms match. Our ex-ante expectation was that after the policy was mandated, the transaction records will perfectly align. However, this does not seem to be the case. Figure A.13 shows the average matches of the purchase declarations of a firm with the corresponding sale declarations of the selling firm. If we match only at the firm-id level, without considering the amount and tax rate declared, the most generous specification possible, the matching started around 90% in the first year (Y3 of our dataset) and is around 96% in year 5 of our analysis. 4% of transactions are still unaccounted for ex-post.

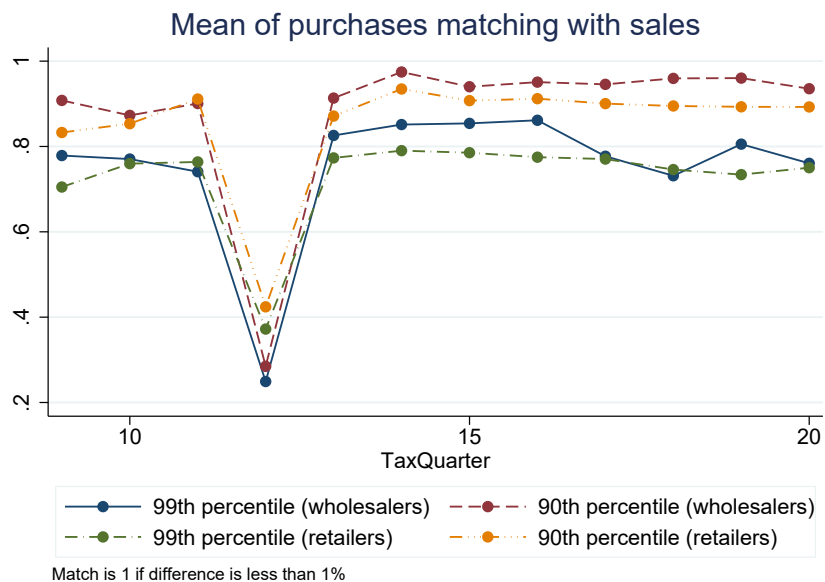
Figure A.14: Matching on the Sales Side for All Firms



We further narrow our analysis and consider the differences in amount. We try two specifications and the results are similar in both the specifications. We classify a transaction as a match if the difference in the total purchases declared by a firm (A) and the total sales declared by the corresponding firm (B) is less than 5 rupees or 1% of the total purchases made by the firm A from firm B. One can assume that the mismatches that happen within this classification, are mostly driven by human error as the revenue implication is minimal. With this specification, the matching rate goes down to roughly 90% across all the quarters. Therefore, roughly 10% of the purchase declarations do not match in our sample in a serious manner. Figure A.14 repeats the analysis but now we are comparing the sales transactions declared by a firm with the corresponding purchase transactions of the buying firms. The results are similar except that now the firm-id level matching has gone down to 80% across quarters.

In figure A.15 and figure A.16 , we repeat the purchase and sale matching analysis

Figure A.15: Matching Analysis: Retailers Vs Wholesalers (Purchases, Top Percentile)

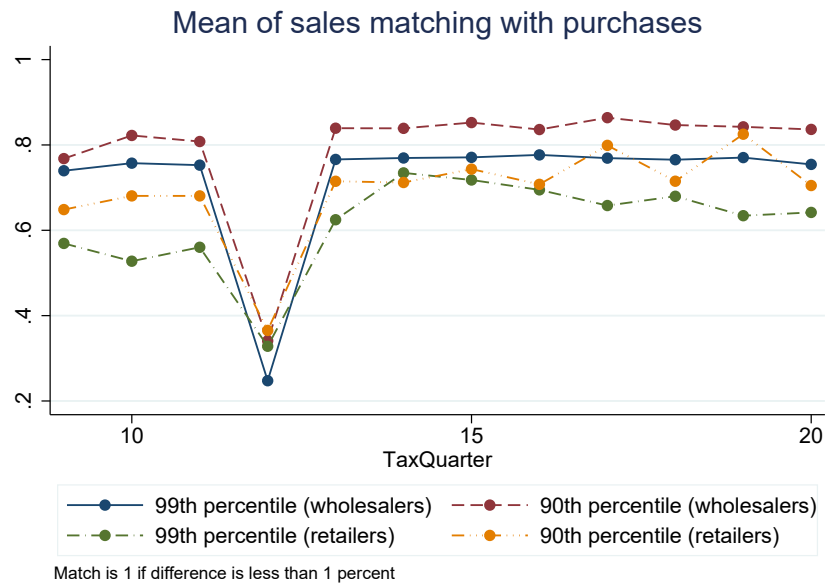


Describing revision trends by percentile. Comparing the firms in the top percentile with the firms in the 90th percentile

but limiting ourselves only to our difference-in-difference sample of wholesalers and retailers. An interesting insight that is clearly visible is that matching for 90th percentile firms for both wholesalers as well as retailers is higher than the matching for the 99th percentile for the corresponding group. This is unexpected and further highlights that just the third party verification information may not be sufficient to reduce evasion and increase tax collections. Some sort of human monitoring effort on top of it is also needed, as despite this lower matching, most of the tax deposit growth is coming from the top percentile of wholesalers.



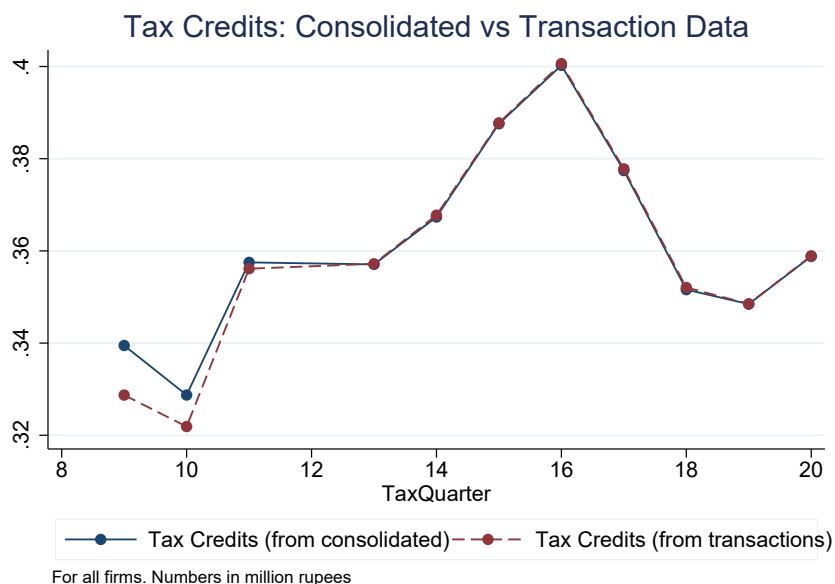
Figure A.16: Matching Analysis: Retailers Vs Wholesalers (Sales, Top Percentile)



Describing revision trends by percentile. Comparing the firms in the top percentile with the firms in the 90th percentile

#### A.5.4 Consolidated vs Transaction Data

Figure A.17: Consolidated vs Transactional Data



In the first year, transaction data was not matched with the consolidated returns. Firms were clearly fudging, which was fixed in the subsequent years. We drop Q12 as unexplained behavior (possibly unrelated) is skewing the image.

In the first year of the third party verification policy, transaction records filed were independently from the consolidated returns and these were not required to be consistent to each other. Therefore, it was possible that the aggregated transaction records did not match the consolidated returns – which is what we observe. Specifically, total input credit claimed in the consolidated forms is on average higher than that implied by the annexures. The tax authority began requiring mechanical reconciliation of the two forms in the subsequent year at which point such divergence ceases by definition – see figure A.17).

# APPENDIX B

## “Return Forms”

### B.1 Consolidated Form

Figure B.1: Page 1 - Form 16

Department of Trade & Taxes  
Government of NCT of Delhi

**Form DVAT 16**  
[See Rule 28 and 29]  
**Delhi Value Added Tax Return**

Refund Claimed?

 Yes  
 No

If revised - **Original/Revised**

(i) Date of filing original return \_\_\_\_\_

(ii) Acknowledgment Receipt No. \_\_\_\_\_

(iii) Date of discovery of mistake or error \_\_\_\_\_

Specify the reasons for revision \_\_\_\_\_

**R1 Tax Period** From: 

	/		/		/		/		/		/			
Dd		mm		yy		T		o		dd		mm		yy

**R2.1 TIN**

**R2.2 Full Name of Dealer**

**R2.3 Address of Principal Place of Business**

**R2.4 Mobile No.**

R3 Description of top items you deal in <i>(In order of volume of sales for the tax period or till the aggregate of sale volume reaches at least 80% - 1- highest volume to 5-lowest volume)</i>	Sl. No.	Commodity Code	Description of Goods	Tax Rate	Tax contribution
	1				
	2				
	3				
	4				
	5				

**R4 Turnover details**

R4.1 Gross Turnover

R4.2 Central Turnover

R4.3 Local Turnover

R5 Computation of output tax	Turnover (Rs.)	Output tax (Rs.)
R5.1 Goods taxable at 1%		
R5.2 Goods taxable at 5%		
R5.3 Goods taxable at 12.5%		
R5.4 Goods taxable at 20%		
R5.5 Works contract taxable at 5%		
R5.6 Works contract taxable at 12.5%		
R5.7 Exempted Sales (Tax Free)		
R5.8 Charges towards labour, services and other fee charges		
R5.9 Charges towards cost of land, if any, in civil works contracts		
R5.10 Sale of Diesel & Petrol as have suffered tax in the hands of various Oil Marketing Companies in Delhi		
R5.11 Sales within Delhi against Form 16		
R5.12 Output Tax before adjustments	Sub Total	
R5.13 Adjustments to output tax (Complete Annexure and enter Total A2 here)		
R5.14 Total Output Tax (R5.12 + R5.13)		

Total A2 from Annexure

R6 Turnover of Purchases in Delhi (excluding tax) & tax credits	Purchases (Rs.)	Tax Credits (Rs.)
R6.1 Capital goods		
R6.2 Other goods		
R6.2(1) Goods taxable at 1%		
R6.2(2) Goods taxable at 5%		
R6.2(3) Goods taxable at 12.5%		
R6.2(4) Goods taxable at 20%		
R6.2(5) Works contract taxable at 5%		
R6.2(6) Works contract taxable at 12.5%		

Figure B.2: Page 2 - Form 16

<b>R6.3 Local purchases not eligible for credit of Input Tax</b>													
R6.3(1) Purchase from Unregistered dealers													
R6.3(2) Purchases from Composition dealers													
R6.3(3) Purchase of Non creditable goods (Schedule-VII)													
R6.3(4) Purchase of Tax Free Goods (Exempted)													
R6.3(5) Purchases of labour and services related to works contract													
R6.3(6) Purchases against tax invoices not eligible for ITC													
R6.3(7) Purchase of goods against retail invoices													
R6.3(8) Purchase of Diesel & Petrol taxable in the hands of various Oil Marketing Companies in Delhi													
R6.3(9) Purchases from Delhi dealers against Form 'H'													
R6.3(10) Purchase of Capital Goods (Used for manufacturing of non-creditable goods)													
											R6.4 Tax credit before adjustments		Sub Total
R6.5 Adjustments to tax credits (Complete Annexure and enter Total A4 here)													
											R6.6 Total Tax Credits (R6.4 + R6.5)		

Total A4 from Annexure

R7.1 Net Tax													(R5.14) – (R6.6)	
R7.2 Interest @ 15% if payable													(B)	
R7.3 Penalty, if payable													(C)	
R7.4 Tax deducted at source (attach TDS certificates (downloaded from website) with Form DVAT 56)														
Sl. No.	Form DVAT-43 ID No.		Date		Amount									
R7.5 Tax credit carried forward from previous tax period														
R7.6 Adjustment of excess balance under CST towards DVAT liability														
R7.7 Balance payable [(R7.1+R7.2+R7.3) – (R7.4+R7.5 +R7.6)]														
R7.8 Amount deposited by the dealer (attach proof of payment with Form DVAT-56)														
S.No.	Date of deposit		Challan No.		Name of Bank and Branch						Amount (Rs.)			
R8 Net Balance*													(R7.7-R7.8)	

\* The net balance should not be positive as the amount due has to be deposited before filing the return.

<b>IF THE NET BALANCE ON LINE R8 IS NEGATIVE, PROVIDE DETAILS IN THIS BOX</b>												
R9 Balance brought forward from line R8 (Positive value of R 8)												
R9.1 Adjusted against liability under Central Sales Tax												
R9.2 Refund Claimed												
R9.3 Balance carried forward to next tax period												

<b>IF REFUND IS CLAIMED, PROVIDE DETAILS IN THIS BOX (Also fill Annexure-2E)</b>												
R10 Details of Bank Account												
R10.1 Account No.												
R10.2 Account type (Saving/Current etc.)												
R10.3 MICR No.												
R10.4												
(a) Name of Bank												
(b) Branch Name												

<b>R11 Inter-state trade and exports/ imports</b>													<b>Inter-state Sales/Exports</b>		<b>Inter-state Purchases / Imports</b>	
R11.1 Against C Forms (Other than Capital Goods)																
R11.2 Against C+E1/E2 Forms																

Figure B.3: Page 3 - Form 16

R11.3 Inward/outward Stock Transfer ( Branch) against F Forms																														
R11.4 Inward/outward Stock Transfer (Consignment) against F Forms																														
R11.5 Own goods received/transferred after job work against F Forms																														
R11.6 Other dealers goods received/returned after job work against F Forms																														
R11.7 Against H Forms (other than Delhi dealers)																														
R11.8 Against I Forms																														
R11.9 Against J Forms																														
R11.10 Exports to / Imports from outside India																														
R11.11 Sale of Exempted Goods (Schedule I)																														
R11.12 High Sea Sales/Purchases																														
R11.13 Sale/Purchases without Forms																														
R11.14 Capital goods purchased against C Form																														
R11.15 Total																														

**R12 Verification**  
 I/We \_\_\_\_\_ hereby solemnly affirm and declare that the information given hereinabove is true and correct to the best of my/our knowledge and belief and nothing has been concealed there from.  
  
 Signature of Authorised Signatory \_\_\_\_\_  
  
 Full Name (first name, middle, surname) \_\_\_\_\_  
  
 Designation/Status \_\_\_\_\_

Place

Date  Day  Month  Year

**Instructions for filling Return Form:**

1. Please complete all the applicable fields in the Form.
2. The fields, which are not applicable, may be left blank.
3. Return should be filed electronically, on the departmental website, within the stipulated period as prescribed under rule 28 of the DVAT Rules.
4. Transmit (i) quarter wise and invoice wise Purchase and Sales data maintained in Form DVAT-30 & 31 OR (ii) quarter wise and dealer wise summary of purchase and sales in Annexure-2A & 2B appended to this Form. Purchase/Sale made from un-registered dealers may be entered in one row for a quarter. However, sale detail of goods sold to Embassies/Organizations specified in Sixth Schedule should be reported invoice wise in case opted for Form DVAT-30 & 31 or Embassies/Organizations wise, if opted for Annexure 2A & 2B, as the case may be.
5. In case of refund, the information in Annexure -2E appended to this Form should be furnished electronically, on departmental website, at the time of filing online return.
6. All dealers to file tax rate wise details of closing stock in hand as on 31<sup>st</sup> March, with the second quarter return of the following year, in Annexure 1D
7. Transmit the information relating to issue of debit/credit note in Annexure 2C & 2D.

## B.2 Annexures

Figure B.4: Annexure 2a - Form 16

**Annexure – 2A**  
(See instruction 6)  
**SUMMARY OF PURCHASE / INWARD BRANCH TRANSFER REGISTER**  
(Quarter wise)

(To be filed along with return)

TIN: \_\_\_\_\_ Name of the Dealer: \_\_\_\_\_

Purchase for the Tax Period: From \_\_\_\_\_ to \_\_\_\_\_

**Summary of Purchase (As per DVAT-30)**

(All amounts in Rupees)

Sr. No.	Quarter & Year	Seller's TIN	Seller's Name	Rate of Tax under DVAT Act (for all columns)
1	2	3	4	5

Inter-State Purchase/Stock Transfer/Import not eligible for credit of input tax									
Import from Outside India	High Sea Purchase	Capital Goods purchased against C-Forms	Goods (Other than capital goods) purchased against C-Forms	Purchase against H-Form (other than Delhi dealers)	Purchases without Forms	Inward Stock Transfer (Branch) against F-Form	Inward Stock Transfer (Consignment) against F-Form	Own goods received back after job work against F-Form	Other dealers goods received for job work against F-Form
6	7	8	9	10	11	12	13	14	15

Local Purchases not eligible for credit of input tax									
Purchase From Unregistered dealer	Purchases from Composition Dealer	Purchase of Non-creditable goods (Schedule-VII)	Purchase of Tax free goods	Purchase of labour & services related to Works Contract	Purchase against tax invoices not eligible for ITC *	Purchase of Goods against retail invoices	Purchase of Petrol & Diesel from Oil Marketing Companies in Delhi	Purchase from Delhi dealers against Form-H	Purchase of Capital Goods (Used for manufacturing of non-creditable goods)
16	17	18	19	20	21	22	23	24	25

Local Purchases eligible to credit of input tax					
Capital Goods		Others (Goods)		Others (Works Contract)	
Purchase Amount (excluding VAT)	Input Tax Paid	Purchase Amount (excluding VAT)	Input Tax Paid	Purchase Amount (excluding VAT)	Input Tax Paid
26	27	28	29	30	31

Note: - Data in respect of unregistered dealers may be consolidated tax rate wise for each Quarter.

\* will include purchase of DEPB (for self-consumption), consumables goods & raw material used for manufacturing of tax free goods in Column No.21.

Signature of Dealer /  
Authorized Signatory

Figure B.5: Annexure 2b - Form 16

**Annexure – 2B**  
 (See instruction 6)  
**SUMMARY OF SALE / OUTWARD BRANCH TRANSFER REGISTER**  
 (Quarter wise)  
 (To be filed along with return)

TIN: \_\_\_\_\_ Name of the Dealer: \_\_\_\_\_  
 Address: \_\_\_\_\_ Sale for the Tax Period: From \_\_\_ to \_\_\_\_\_

**Summary of Sales (As per DVAT-31)**

(All amounts in Rupees)

Sr No.	Quarter & Year	Buyer's TIN / Embassy/Organisation Regn. No.	Buyer/Embassy/Organisation Name	Tax Rate (DVAT) (for all columns)
1	2	3	4	5

Turnover of Inter-State Sale/Stock Transfer / Export (Deductions)												
Export	High Sea Sale	Own goods transferred for Job Work against F-Form	Other dealers' goods returned after Job work against F-Form	Stock transfer (Branch ) against F- Form	Stock transfer (Consignment) against F- Form	Sale against H-Form	Sale against I-Form	Sale against J-Form	Sale against C+E-I/E-II	Sale of Exempted Goods [Sch. I]	Sales covered under provision to [Sec. 9(1)] Read with Sec. 8(4)]	Sales of Goods Outside Delhi (Sec. 4)
6	7	8	9	10	11	12	13	14	15	16	17	18

Turnover of Inter-State Sale (Taxable)					Turnover of Local Sale						
Rate of Tax (CST)	Sale against C-Form excluding sale of capital assets	Capital Goods sold against C-Forms	Sale without forms	Tax (CST)	Turnover (Goods) (excluding VAT)	Turnover (WC) (excluding VAT)	Output Tax	Charges towards labour, services and other like charges, in civil works	Charges towards cost of land, if any, in civil works contracts	Sale against H-Form to Delhi dealers	Sale of Petrol/Diesel suffered tax on full sale price at OMC level
								contracts			
19	20	21	22	23	24	25	26	27	28	29	30

Note:- Data in respect of unregistered dealers may be consolidated tax rate wise for each Quarter. Data of Embassies/Organisations listed in Sixth Schedule shall be provided entity wise.

Signature of Dealer /  
Authorized Signatory

## Bibliography

- Ali Agha and Jonathan Haughton. Designing vat systems: Some efficiency considerations. The Review of Economics and Statistics, pages 303–308, 1996.
- Michael G. Allingham and Agnar Sandmo. Income Tax Evasion: A Theoretical Analysis. Journal of Public Economics, 1(3-4):323–338, 1972. ISSN 00472727. doi: 10.1016/0047-2727(72)90010-2.
- 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.
- Miguel Almunia, François Gerard, Jonas Hjort, Justine Knebelmann, Dorothy Nakymbadde, Claude Raisaro, and Lin Tian. An analysis of discrepancies in tax declarations submitted under value-added tax in uganda. Technical report, 2017.
- Zareh Asatryan and Andreas Peichl. Responses of firms to tax, administrative and accounting rules: Evidence from armenia. Working paper, 2017.
- Pierre Bachas and Anders Jensen. Size-dependent tax enforcement and compliance: Global evidence. Working paper, 2017.
- Abhijit Banerjee, Esther Duflo, Clement Imbert, Santhosh Mathew, and Rohini Pande. E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in india. Technical report, National Bureau of Economic Research, 2016.
- Youssef Benzarti. How taxing is tax filing? using revealed preferences to estimate compliance costs. Technical report, National Bureau of Economic Research, 2017.



- Timothy Besley and Torsten Persson. Taxation and development. In Handbook of public economics, volume 5, pages 51–110. Elsevier, 2013.
- Richard Bird and Pierre-Pascal Gendron. The VAT in Developing and Transitional Countries. Cambridge University Press, 2007.
- Richard Bird, J Martínez-Vázquez, and B Torgler. Value-added taxes in developing and transitional countries: lessons and questions. ITP paper, 505, 2005.
- Daniel Björkegren and Darrell Grissen. Behavior revealed in mobile phone usage predicts loan repayment. arXiv preprint arXiv:1712.05840, 2017.
- Joshua Blumenstock, Gabriel Cadamuro, and Robert On. Predicting poverty and wealth from mobile phone metadata. Science, 350(6264):1073–1076, 2015.
- Wian Boonzaaier, Jarkko Harju, Tuomas Matikka, and Jukka Pirttila. How do small firms respond to tax schedule discontinuities? evidence from south african tax registers. Working paper, 2017.
- Paul Carrillo, Dina Pomeranz, and Monica Singhal. Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. American Economic Journal: Applied Economics, 9(2):144–164, 2017. doi: 10.3386/w20624.
- Xi Chen and William D Nordhaus. Using luminosity data as a proxy for economic statistics. Proceedings of the National Academy of Sciences, 108(21):8589–8594, 2011.
- Arindam Das-Gupta, Shanto Ghosh, and Dilip Mookherjee. Tax Administration Reform and Taxpayer Compliance In India. International Tax and Public Finance, 11(5):575–600, 2004. ISSN 09275940. doi: 10.1023/B:ITAX.0000036692.95048.00.

- Áureo De Paula and Jose A Scheinkman. Value-Added Taxes, Chain Effects, and Informality. American Economic Journal: Macroeconomics, 2(October):195–221, 2010. ISSN 1556-5068. doi: 10.1257/mac.2.4.195.
- Pierre Deville, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R Stevens, Andrea E Gaughan, Vincent D Blondel, and Andrew J Tatem. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111(45):15888–15893, 2014.
- Dhammika Dharmapala, Joel Slemrod, and John Douglas Wilson. Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs. Journal of Public Economics, 95(9-10):1036–1047, 2011.
- Directorate of Economics and Statistics, GNCTD. Delhi Statistical Handbook. 2015.
- Esther Duflo, Michael Greenstone, Rohini Pande, and Nicholas Ryan. Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from india. The Quarterly Journal of Economics, page qjt024, 2013.
- Liam Ebrill, Michael Keen, Jean-paul Bodin, and Victoria Summers. The Modern VAT. International Monetary Fund Publication Services, 2001. ISBN 1-589-06026-1.
- F Finan, BA Olken, and R Pande. The personnel economics of the developing state. Handbook of Economic Field Experiments, 2:467–514, 2017.
- Luis Garicano, Claire Lelarge, and John Van Reenen. Firm size distortions and the productivity distribution: Evidence from france. American Economic Review, 106(11):3439–79, 2016.
- Mesay M Gebresilasie and Soule Sow. Firm response to vat registration threshold in ethiopia. Technical report, 2016.

- Paul Glewwe, Nauman Ilias, and Michael Kremer. Teacher incentives. American Economic Journal: Applied Economics, 2(3):205–227, 2010.
- Douglas Gollin. Do Taxes on Large Firms Impede Growth? Evidence from Ghana. Economic Development Center, Department of Economic, Minneapolis Department of Agricultural and Applied Economics, St. Paul, University of Minnesota, 1995.
- Roger Gordon and Wei Li. Tax structures in developing countries: Many puzzles and a possible explanation. Journal of Public Economics, 93(7):855–866, 2009.
- Nezih Guner, Gustavo Ventura, and Yi Xu. Macroeconomic implications of size-dependent policies. Review of Economic Dynamics, 11(4):721–744, 2008.
- Jarkko Harju, Tuomas Matikka, and Timo Rauhanen. The effects of size-based regulation on small firms: Evidence from vat threshold. Working paper, 2016.
- International Tax Dialogue. The value added tax: Experiences and issues. In Conference Paper, Rome, March 2005.
- International Tax Dialogue. Taxation of small and medium enterprises. Background paper for the International Tax Dialogue Conference, 2007.
- International Tax Dialogue. Key issues and debates in vat, sme taxation and the tax treatment of the financial sector. Background paper for the International Tax Dialogue Conference, 2013.
- Brian A Jacob and Steven D Levitt. Rotten apples: An investigation of the prevalence and predictors of teacher cheating. The Quarterly Journal of Economics, 118(3): 843–877, 2003.
- Michael Keen and Jack Mintz. The optimal threshold for a value-added tax. Journal of Public Economics, 88(3-4):559–576, 2004.

Michael Keen and Stephen Smith. Vat fraud and evasion: What do we know and what can be done? National Tax Journal, pages 861–887, 2006.

Adnan Q Khan, Asim I Khwaja, and Benjamin A Olken. Tax farming redux: Experimental evidence on performance pay for tax collectors. The Quarterly Journal of Economics, 131(1):219–271, 2016.

Henrik J Kleven and Mazhar Waseem. Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. The Quarterly Journal of Economics, 128(2):669–723, 2013.

Henrik Jacobsen Kleven, Martin B Knudsen, Claus Thustrup Kreiner, Søren 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.

Todd Kumler, Eric Verhoogen, and Judith Frias. Enlisting workers in monitoring firms: Payroll tax compliance in mexico. Columbia University Department of Economics Discussion Papers, 96, 2012.

Himabindu Lakkaraju, Jon Kleinberg, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 275–284. ACM, 2017. doi: <https://doi.org/10.1145/3097983.3098066>.

Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. R news, 2(3):18–22, 2002.

- Li Liu, Ben Lockwood, and Miguel Almunia. Vat notches, voluntary registration and bunching: Theory and uk evidence. Working paper, 2017.
- Tongliang Liu and Dacheng Tao. Classification with noisy labels by importance reweighting. IEEE Transactions on pattern analysis and machine intelligence, 38(3):447–461, 2016.
- Alejandro Llorente, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. Social media fingerprints of unemployment. PloS one, 10(5):e0128692, 2015.
- Xin Lu, David J Wrathall, Pål Roe Sundsøy, Md Nadiruzzaman, Erik Wetter, Asif Iqbal, Taimur Qureshi, Andrew Tatem, Geoffrey Canright, Kenth Engø-Monsen, et al. Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in bangladesh. Global Environmental Change, 38:1–7, 2016.
- Maxim Mironov and Ekaterina Zhuravskaya. Corruption in procurement: Evidence from financial transactions data. Unpublished working paper. Social Science Research Network, 2014.
- Shekhar Mittal and Aprajit Mahajan. Vat in emerging economies: Does third party verification matter? 2017.
- Shekhar Mittal, Ofir Reich, and Aprajit Mahajan. Who is bogus? using one-sided labels to identify fraudulent firms from tax returns. In Proceedings of the 1st Annual ACM SIGCAS Conference on Computing and Sustainable Societies. ACM, 2018. doi: <https://doi.org/10.1145/3209811.3209824>.
- Karthik Muralidharan and Venkatesh Sundararaman. Teacher performance pay: Experimental evidence from india. Journal of political Economy, 119(1):39–77, 2011.

- Karthik Muralidharan, Paul Niehaus, and Sandip Sukhtankar. Building State Capacity: Evidence from Biometric Smartcards in India. American Economic Review, 106(10):2895–2929, 2016. ISSN 00028282.
- Joana Naritomi. Consumers as tax auditors. Job market paper, Harvard University, 2013.
- Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. In Advances in neural information processing systems, pages 1196–1204, 2013.
- Benjamin A Olken and Rohini Pande. Corruption in developing countries. Annu. Rev. Econ., 4(1):479–509, 2012.
- Kazuki Onji. The response of firms to eligibility thresholds: Evidence from the japanese value-added tax. Journal of Public Economics, 93(5-6):766–775, 2009.
- Konstantin V Pashev. Countering cross-border vat fraud: the bulgarian experience. Journal of Financial Crime, 14(4):490–501, 2007.
- Dina Pomeranz. No taxation without information: Deterrence and self-enforcement in the value added tax. The American Economic Review, 105(8):2539–2569, 2015.
- 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, Brett Collins, Jeffrey Hoopes, Daniel Reck, and Michael Sebastiani. Does credit-card information reporting improve small-business tax compliance? Technical report, National Bureau of Economic Research, 2015.
- The H2O.ai team. h2o: Python Interface for H2O, 2015. URL <http://www.h2o.ai>. Python package version 3.1.0.99999.

The H2O.ai team. h2o: Naive Bayes Classifier, 2018. URL <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/naive-bayes.html>.

Amy Wesolowski, Gillian Stresman, Nathan Eagle, Jennifer Stevenson, Chrispin Owaga, Elizabeth Marube, Teun Bousema, Christopher Drakeley, Jonathan Cox, and Caroline O Buckee. Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. Scientific reports, 4, 2014.

Daniel Lawrence Wilson, Mohammed Idris Adam, Omnia Abbas, Jeremy Coyle, Angeli Kirk, Javier Rosa, and Ashok J Gadgil. Comparing cookstove usage measured with sensors versus cell phone-based surveys in darfur, sudan. In Technologies for Development, pages 211–221. Springer, 2015.

Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. Transfer learning from deep features for remote sensing and poverty mapping. arXiv preprint arXiv:1510.00098, 2015.

Ernst & Young. Worldwide vat, gst and sales tax guide. 2015. URL [http://www.ey.com/Publication/vwLUAssets/Worldwide-VAT-GST-and-sales-tax-guide-2015/\\$FILE/Worldwide%20VAT,%20GST%20and%20Sales%20Tax%20Guide%202015.pdf](http://www.ey.com/Publication/vwLUAssets/Worldwide-VAT-GST-and-sales-tax-guide-2015/$FILE/Worldwide%20VAT,%20GST%20and%20Sales%20Tax%20Guide%202015.pdf).

Eric Zitzewitz. Forensic economics. Journal of Economic Literature, pages 731–769, 2012.