

Introduction

I am a development economist with interests in public economics and political economy related issues in India. I aim to use large-scale government data sets (that have only recently begun to be collected) to better understand government capacity and to combine such data sets with field interventions to address questions of first-order causal interest. In my dissertation, I also use machine learning methods to identify strategic considerations in economic data.

My main paper uses the universe of tax returns in an Indian state to examine the effect of improved information verification capacity on tax collections. In a second ongoing project I use machine learning methods on the same set of tax data to predict tax evasion and design an intervention that uses these predictions to better target inspections. Following this, I intend to also examine strategic response of firms to the increased use of this regulatory tool. In the following sections I describe each project in detail as well as other future work.

Ongoing Research

Even though compliance issues are central to taxation policies in developing countries, convincing empirical work on tax compliance has been scarce. In my thesis I attempt to bridge some of this gap by using detailed value added tax (VAT) micro-data from Delhi, India.

Third Party Verification

A key stated advantage of value-added tax type systems is that they allow for corroboration of transactions using tax returns of interacting firms. In [Mittal and Mahajan \[2017\]](#), we 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.

Bogus Firms

In low compliance environments, a common strategy to manipulate the third-party verification system is to establish fraudulent firms ("bogus" dealers). These firms help genuine firms in reducing their tax burden by issuing fake receipts. A tax authority determines the existence of

“bogus” dealers 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. While the true revenue implication of these “bogus” firms is unknown, our conversations with the Delhi tax authority suggest that it might be considerable (₹20 billion is commonly mentioned). A key challenge in improving tax compliance then is to regularly, cheaply and reliably identify such fraudulent firms. Machine learning algorithms may have a role in meeting this challenge.

The success and increasing popularity of machine learning has largely been due to its ability to discover unspecified patterns which work well out-of-sample. These patterns are useful in characterizing heterogeneity as well as generating predictions. In my second chapter, along with Aprajit Mahajan and Ofir Reich, I use the extremely large and rich data set in our possession to construct a network of all registered firms in Delhi. We then develop a machine learning tool which uses this network data to predict the likelihood of a firm being “bogus” (based on a training data set with identified fraudulent firms).

In the second stage of this project we plan to work with the tax authority to implement a randomized control trial that will compare our machine learning tool with the “business as usual” approach followed by the authority. The “business as usual” approach is that of manual checks undertaken on an ad-hoc basis with limited analysis of the returns themselves. This comparison will allow us to make progress on a few key issues. First, from a policy point of view we can directly measure the resources saved by the tax authority if it uses our targeted approach. Second, we can credibly identify the revenue implications of removing fraudulent firms from the tax system. Third, we can analyze the behavior of firms that transact with the suspicious firms identified by our tool. Given that such data exists in many tax jurisdictions and that anecdotal evidence suggests that such false paper trails are a common problem, my second chapter should have high policy relevance both within India and elsewhere.

Bunching at Filing Thresholds

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. Therefore, estimating the magnitude of these compliance costs is an important empirical question. In chapter 3 of my dissertation, along with Jan Luksic, I will use techniques from the bunching literature to estimate the cost of complying with a filing frequency policy.

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. At each of these thresholds we find significant bunching mass just below the threshold. This implies that firms associate significant costs with filing returns and try to avoid these costs by not crossing the threshold. We also find that firms just below the threshold remit disproportionately less tax than those just above. In the years 3, 4 and 5 of our dataset, this turnover based filing policy was first weakened and then completely disbanded. Interestingly, we observe that the bunching vanishes immediately after the end of the policy.

Future Directions

For my thesis research I have been collaborating with senior bureaucrats, both at the federal and state level in India. Through this experience I have begun to understand the incentives of governments to collaborate with academic researchers. Due to advancement in technology, government departments, even in emerging economies, are now producing large complex data sets. However, the bureaucrats themselves do not have a comparative advantage in analyzing this data to inform policy design. This creates an opportunity for a symbiotic relationship between bureaucrats and researchers towards data driven policy design and large scale economic research. With additional support from partners in India, such as JPAL South Asia and Indian School of Business Hyderabad, I am pursuing potential opportunities in the areas of public procurement, public distribution system, and goods and services tax.

Public Procurement

E-procurement claims to address three common concerns with procurement: lack of access to bid information, collusion between bidders, and corruption based on opaque information processing. By ensuring public access to all procurement data, e-procurement increases transparency and public oversight. Delhi government officials have expressed interest in working to analyze data from their electronic procurement system, and to design interventions to improve performance.

Delhi government officials have highlighted some additional concerns with the current procurement system: a high rate of re-tendering and cancellation of tenders, and a few instances of inefficiency driven wastage. Machine learning applied to procurement is also a promising avenue for future research. This collaboration will allow us to explore whether information interventions can increase the competitiveness of bids. Subsequently, we intend to evaluate whether this increase in competitive bids leads to a reduction in government costs. This work directly examines the magnitude of leakages in public procurement, identifies stages that are more susceptible to leakages, and reduces leakages via information and technology tools.

Public Distribution System

Another area that has seen rapid deployment of information technology is Public Distribution Systems (PDS). PDS is the lynch-pin of the Indian government's food security strategy, yet faces many challenges. First, fraudulent or bogus transactions that divert PDS supplies away from intended beneficiaries. Second, frequent shortage of food grains at the end-point "fair-price" shops (FPS) despite availability at the warehouses. State governments are trying to address these challenges by introducing a computerized system. The system has been designed to automate the supply chain and provide a comprehensive information tracking mechanism to minimize leakages. All FPS transactions are recorded electronically (through point of sale devices) and stored on a central server. In addition, quantities released from the warehouses are also recorded and this information is shared with staff throughout the supply chain.

The states of Karnataka and Tamil Nadu have expressed an interest in collaboration with our partners at Indian School of Business and JPAL South Asia. This effort will analyze data from their PDS related technology systems, and then suggest interventions to detect fraud and to reduce leakages. We will also explore the feasibility of building a training data set based on field

inspections that would then be combined with the PDS data to create machine learning models that identify fraudulent transactions. This work directly examines a range of questions around the role of information and technology interventions in reducing leakages and in improving quality of service.

Tax Reform and Firm Productivity

In July 2017, India implemented the Good and Service Tax (GST), a major tax reform where all the federal and state level indirect taxes were replaced by a single unified tax. Along with Pierre Bachas, I am using this policy change to evaluate the incidence of consumption tax changes on consumers, on firm owners, and on workers. The tax reform also results in two separate shocks. First, a “liquidity” shock on firms due to the tax rate changes between outputs and inputs that has rich variation at the state level (as tax rates before the policy were different across states) and at the industry level. Second, a “productivity” shock as firms can now also claim input credits on services needed for production which was not possible earlier. We will use this natural experiment to evaluate the impact of “liquidity” shocks and “productivity” shocks on firm investment, productivity and growth.

Fragmentation in Agricultural Markets

Numerous sources have documented the dispersion of commodity prices, in Indian agriculture, across time and across geographies. Common wisdom suggests that market fragmentation is the primary reason for this dispersion. To tackle this problem the federal government has initiated a technology based platform which links the agricultural markets both within and across states. With Prabhat Barnwal, I am looking to analyze the impact of staggered roll-out of this linking process on reduction in dispersion of agricultural prices. I further intend to design and execute interventions to improve the take up of the platform and evaluate the subsequent welfare consequences on the farmers.

References

S. Mittal and A. Mahajan. Vat in emerging economies: Does third party verification matter? 2017.