Three essays on policy evaluation in developing countries

Gechter, Michael David
BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

THREE ESSAYS ON POLICY EVALUATION IN DEVELOPING COUNTRIES

by

MICHAEL DAVID GECHTER
B.A., Pomona College, 2005
M.A., Boston University, 2012

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
2015
Acknowledgments

I would like to first thank the co-chairs of my dissertation committee: Dilip Mookherjee and Hiroaki Kaido. This dissertation would not have been possible without their tireless support, advice and patience. Thanks to Dilip, I learned the importance and power of economic theory combined with careful empirical work and how to fit both into a way of thinking about development policy. His feedback, built on a profound knowledge of theory and context, pushed me to continuously improve and expand the scope of my work. I learned a tremendous amount from working with Hiro and deeply appreciate the trust and faith he placed in me from the start of our working relationship. I benefited from his ability to draw on a broad and deep base of knowledge to suggest ways forward when facing difficult problems. I would also like to thank Kevin Lang for many hours of advice and support. Kevin’s insights helped me bring together the econometric and applied sides of my research through his thorough understanding of economic theory, applied empirical work and econometric theory. I could not ask for three better role models as I begin my academic career.

I owe thanks to the Department of Economics at Boston University as a whole and especially to Kehinde Ajayi, Sam Bazzi, Iván Fernández-Val, Claudia Olivetti and Pierre Perron, who contributed a great deal to this dissertation and to my development as an economist. I have benefited from having a great group of fellow students to discuss with and learn from, starting with my coauthor on two of these chapters: Amrit Amirapu. I am also grateful to have had the chance to discuss regularly with Aparna Dutta, Debbie Goldschmidt, Huailu Li, Anusha Nath, Russell Weinstein, Nate Young and Max Zhao. I owe thanks to Tahir Andrabi and Fernando Lozano as well, for helping me take my first steps as a development economist.

I would also like to thank many friends and family members from outside Boston University. Teresa Bruno-Niño provided me with more love, support, strength and crucial advice
than I could ever hope for. Anne and Jerry Gechter and Karin and Sanjeev Francis have always supported me in any way they could, with care and useful advice. Finally, Keith Irving provided a wealth of valuable insights. Without all your support and belief in me, I could never have completed this dissertation.

For the first chapter, I again owe special thanks to my advisors, Dilip Mookherjee, Hiroaki Kaido and Kevin Lang for guidance, support and encouragement. I would also like to thank Kehinde Ajayi, Manuel Arellano, Sam Bazzi, Xavier d’Haultfoeuille, Rajeev Dehejia, Isabella Dobrescu, Iván Fernández-Val, Claudio Ferraz, Andy Foster, Patrik Guggenberger, Rema Hanna, Stefan Hoderlein, Asim Khwaja, Horacio Larreguy, David Lam, Leigh Linden, Michael Manove, Shanthi Nataraj, Claudia Olivetti, Rohini Pande, Justin Sandefur, Johannes Schmieder, Jeff Smith, Duncan Thomas, Nate Young and seminar participants at Boston University, NEUDC and the Western Economic Association International Graduate Student Workshop for useful conversations and feedback. I especially thank David McKenzie for discussions and sharing the data from McKenzie and Woodruff (2008). I gratefully acknowledge funding support from the Boston University Department of Economics and able research assistance from Gloria Sarmiento-Becerra, supported by the Boston University Department of Economics MA-RA program.

For the second and third chapters, special thanks go to Arvind Subramanian, Kevin Lang, Hiroaki Kaido, Kehinde Ajayi, Sam Bazzi and Marc Rysman for input, guidance and encouragement. Laveesh Bhandari, Aditya Bhattacharjea, Sharon Buteau, Areendam Chanda, Urmila Chatterjee, Nancy Chau, Bivas Chaudhuri, Bibek Debroy, Francois Gourio, Jayinthi Harinath, Rana Hasan, Ravi Kanbur, Saibal Kar, Dan Keniston, Devashish Mitra, Sandip Mitra, Shanthi Nataraj, G. Sajeevan, Johannes Schmieder, Pronab Sen, Anushree Sinha, Sandip Sukhtankar, Shyam Sundar, Balasekhar Sudalaimani, John Van Reenen and conference and seminar participants at NEUDC, Boston University, University of Passau, Cornell, ISI Kolkata and IGIDR provided many useful comments. Representatives in the state governments of Bihar, Kerala, Maharashtra, Tamil Nadu and West Bengal provided
invaluable assistance in understanding the data collection process for the Economic Census.
Financial support from the Weiss Family Program Fund for Research in Development Eco-
nomics is gratefully acknowledged. Rahul Gupta, Liu Li, Stefan Winata and Haiqing Zhao
provided excellent research assistance, supported by the Boston University Department of
Economics MA-RA mentor program.
THREE ESSAYS ON POLICY EVALUATION IN DEVELOPING COUNTRIES

MICHAEL DAVID GECHTER

Boston University, Graduate School of Arts and Sciences, 2015

Major Professor: Dilip Mookherjee, Professor of Economics

ABSTRACT

This dissertation is a collection of three independent essays on the theory and application of methods of policy analysis in development economics. The first chapter tackles the methodological problem of external validity: the extent to which a study’s findings apply to settings outside of the study’s original context. There has recently been much debate within development economics over the practical usefulness of randomized policy evaluations outside of the places and times they were originally implemented. The chapter offers a practical solution to the problem of assessing generalizability: I derive bounds on the average causal effect in a context of interest using data from a previous experimental study.

The second chapter, coauthored with Amrit Amirapu, investigates the costs that labor regulations impose on Indian firms. Regulatory burden has been viewed as the cause of many economic problems in India, including misallocation of resources across firms and slow growth. Many Indian regulations do not apply to firms with fewer than 10 workers. We use a downward discontinuity in the size distribution of firms at the 10-worker threshold to quantify regulatory costs imposed on firms. We find that these costs are substantial on average but vary widely across states, with firms in more corrupt states facing higher implied costs.

The third chapter, also coauthored with Amrit Amirapu, investigates changes in the production of certain goods in India in response to the elimination of the Small Scale
Industry Reservation policy. This policy mandated that only firms maintaining less than a particular level of capital stock (about $1 million) were eligible to produce specific goods. The policy was gradually phased out over the period from 1997 to 2008. Different goods were de-reserved in different years, facilitating the comparison of markets for recently de-reserved goods relative to never-reserved and still-reserved goods. We investigate the effects of de-reservation on small incumbent firms owing to competition from large formal entrants.
## Contents

### 1 Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India

1.1 Introduction .................................................. 1
1.2 Intuition for the methodology: a simple example ................. 5
1.3 Econometric setup .............................................. 11
  1.3.1 Previous methods ........................................ 14
  1.3.2 Example: remedial education in India ...................... 17
1.4 Bounds on $ATE^a$ using differences in the untreated outcome distributions . 19
  1.4.1 Identification ............................................ 19
  1.4.2 Estimation ............................................... 30
  1.4.3 Inference ............................................... 36
1.5 Transfers to Mexican microenterprises .............................. 37
1.6 Remedial education in India ..................................... 40
  1.6.1 Using Mumbai to predict Vadodara ......................... 44
  1.6.2 Using Vadodara to predict Mumbai ........................ 46
1.7 Conclusions .................................................. 47
1.8 Definition of copula ........................................... 49
1.9 Proof of equivalence of bounds in proposition 2 and linear programming representation ........................................ 49

### 2 Indian Labor Regulations and the Cost of Corruption: Evidence from the Firm Size Distribution

2.1 Introduction .................................................. 55
2.2 Institutional Background: Size-Based Regulations in India ........ 60
2.3 Data and the Size Distribution in India .......................... 62
3.9 Appendix: Linear Differences in Differences - the effect of varying the estimation sample ........................................... 126

3.10 Appendix: Fractions of plants producing in each 2-digit ASICC product class by category ............................................. 132

3.11 Appendix: example of desvasion within a 3-digit ASICC product class . 135

Bibliography ................. 136

Curriculum Vitae .......... 141
List of Tables

1.1 The distribution of enrollment with and without the CCT is unknown in location $e$ .................................................. 8
1.2 Case 1: there are no wage effects ........................................ 9
1.3 Case 2: wage effects only impact those who enroll without the CCT .......................... 9
1.4 Case 3: wage effects impact the same fraction of both groups .............................. 9
1.5 Choice variables, $e =$Mumbai ........................................... 34
1.6 Contribution of choice variables to the objective, $e =$Mumbai ......................... 34
1.7 $P^e(y_{0j}, y_{1k}|$competency on entering third grade $= 0), \rho^L = 1 e =$Mumbai .... 35
1.8 Bounds on the average return to cash transfers in urban Mexico in 2012 using experimental data from McKenzie and Woodruff (2008) ...................... 40
1.9 Vadodara ........................................................................ 42
1.10 Mumbai ........................................................................ 42
1.11 City-specific average effects on maximum math grade level competency ........ 43
1.12 City-specific average effects on maximum verbal grade level competency ...... 43
1.13 Controls - $P( \text{competency on exiting grade 3} | \text{competency on entering grade 3})$ ......................................................... 45
1.14 Bounds on the change in average grade level competency in Vadodara using experimental results from Mumbai and untreated outcomes from Vadodara 46
1.15 Bounds on change in average grade level competency in Mumbai using experimen
tal results from Vadodara and untreated outcomes from Mumbai .................... 47

2.2 Tau vs Other Measures of Regulations ................................... 80
2.3 Tau vs Strikes and Lockouts .............................................. 82
2.4 Tau vs Enforcement of Regulations ...................................... 83
2.5 Tau vs Transparency International Corruption Score ......................... 85
2.6 Tau vs Transmission and Distribution Losses .............................. 86
2.7  Tau vs State Level Corruption Interacted with Industry Level “Dependence on Regulation” (with Industry FEs) .................................................. 89
2.8  Manufacturing Employment Growth (2005 - 2010) vs Tau and Other Measures 93
2.9  Manufacturing Productivity Growth (2005 - 2010) vs Tau and Other Measures 94
2.10 Estimates of Tau by State ................................................................. 99
2.11 Estimates of Tau by Industry ............................................................ 100
2.12 Estimates of Tau by Owership Type .................................................. 101

3.1 Historical definition of Small Scale Industry Sector .............................. 110
3.2 CIC results: dereserved vs. never reserved ......................................... 120
3.3 CIC results: 2006 impacts - dereserved 2001-2006 vs. dereserved before 2001 122
3.4 CIC results: 2008 impacts - dereserved 2001-2006 vs. dereserved before 2001 123
3.5 Number of Factories Over The Period: 2001 and 2006 ......................... 125
3.6 Capital vs Dereservation Status ......................................................... 127
3.7 Capital vs Dereservation Status (interacted with readymade garments indicator) ................................................................. 128
3.8 Capital vs Dereservation Status (including interaction with entrant/incumbent status) ........................................... 129
3.9 Capital vs Dereservation Status (including interaction with entrant/incumbent status - excluding readymade garments) ......................... 130
3.10 Capital vs Dereservation Status (including interaction with entrant/incumbent status - including only readymade garments) ......................... 131
3.12 Product class (subdivision) 443 - Leather Footwear & Parts Thereof .... 135
List of Figures

1.1 Permissible distributions for $P(\text{enrolled with CCT} \mid \text{enrollment without CCT})$ in location $a$ ................................................................. 7
1.2 Perfect positive dependence of $F_{Y_0|X}(y_0|x)$, $F_{Y_1|X}(y_1|x)$ ............................ 22
1.3 Perfect negative dependence of $F_{Y_0|X}(y_0|x)$, $F_{Y_1|X}(y_1|x)$ ............................ 23
1.4 Distribution of profits: McKenzie and Woodruff (2008) control group and 2012 ENAMIN ................................................................. 50
1.5 Bounds on the average return to cash transfers in urban Mexico in 2012 using experimental data from McKenzie and Woodruff (2008) ............................ 51
1.6 controls - grade level competency on exiting 3rd grade conditional on grade level competency on entering 3rd grade ........................................ 52
1.7 Bounds on the change in average grade level competency in Vadodara using experimental results from Mumbai and untreated outcomes from Vadodara 53
1.8 Bounds on change in average grade level competency in Mumbai using experimental results from Vadodara and untreated outcomes from Mumbai 54
2.1 Distribution of establishment size for establishments with 1-200 total workers, 2005 ................................................................. 64
2.2 Distribution of establishment size for establishments with 5-25 total workers, 2005 ................................................................. 65
2.3 Distribution of establishment size, 2005, log scale ........................................ 65
2.4 Downward shift at the 10-worker threshold in the distribution of establishment size, 2005, log scale (omitting establishments with more than 100 workers) 66
2.5 Theoretical Model of Misreporting, log scale (thick line = true distribution; thin line = reported distribution; dashed line = counterfactual distribution) 73
2.6 Downward shift at the 10-worker threshold in the distribution of establishment size estimated on nonparametric density estimates, 2005, log scale (including all establishments). Black points = actual data; Grey = smoothed data. ................................................................. 74
2.7 Variation in the distribution of establishment size across time: .............. 95
2.8 Decision Tree ............................................................... 105
3.1 2001 density of log(factor cost ratio) ...................................... 114
3.2 2001 density of log(capital) ............................................... 115
3.3 2006 density of log(capital) ............................................... 116
3.4 2006 density of log(capital) - incumbents and entrants ................... 117
3.5 2001, factor cost ratio (capital/labor) by log(capital) for producers of goods to be dereserved ......................................................... 118
3.6 2006, factor cost ratio (capital/labor) by log(capital) for dereserved products - incumbents ......................................................... 118
3.7 2006, factor cost ratio (capital/labor) by log(capital) for dereserved products - entrants ......................................................... 119
3.8 Distribution of value of log(real capital stock), 2001 (before) and 2006 (after) dereserved and never reserved .............................. 121
3.9 2001 density of log(capital) ............................................... 123
3.10 2006 density of log(capital) ............................................... 124
3.11 2008 density of log(capital) ............................................... 124
Chapter 1

Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India

1.1 Introduction

What do causal effects measured in one place tell us about causal effects in another place or at another time? It is clear that not every finding applies in every context. Some authors have recently protested against policy recommendations they see as based on implicit extrapolation from a small number of experiments to a wide variety of dissimilar contexts (Deaton (2010); Pritchett and Sandefur (2013)). Empirically, a growing body of work finds different effects of identical policies among individuals with the same observed characteristics living in different contexts (e.g. Allcott (2015); Attanasio, Meghir, and Szekely (2003)). Unobserved differences between populations remain, even when considering individuals with the same observed characteristics.

In this paper, causal effects from one place may be only partially informative about effects elsewhere. I derive bounds on the average causal effect in a context of interest using experimental evidence from another context. I use differences in outcome distributions for individuals with the same characteristics and treatment status in the original study and the context of interest to learn about unobserved differences across contexts\(^1\). Greater differences in outcome distributions generate wider bounds. The bounds represent a practical solution to the problem of assessing generalizability of experimental results from one context to another and are easily computed using software provided by the author for any pair of contexts. They formalize the idea that the conclusions we can draw about the average causal effect in the context of interest and the strength of assumptions required to do so

\(^1\)When we do not have experiments with context-level characteristics we believe are sufficiently similar to the context of interest, unobserved differences necessarily include differences in context-level characteristics.
depends on the similarity between the two contexts\(^2\).

I consider settings where we have run a randomized evaluation of a pilot program and wish to know what we can conclude about the effect of the program in another context. The experimental treatment group has access to the program, while the control group does not. As part of the evaluation, we collected data on characteristics and outcomes of individuals participating in the experiment. We also have data on outcomes and characteristics of individuals in the alternative context, possibly coming from a separate survey. Since the program is a pilot, individuals in the alternative context do not have access to the program\(^3\). For each distinct set of characteristics, we thus have the distributions of treated and untreated outcomes from the experiment and the distribution of untreated outcomes from the alternative context.

The bounds I derive on the average causal effect in the context of interest for each set of characteristics are based on the assumption that the distribution of treated outcomes for a given untreated outcome in the context of interest is consistent with the experimental results. This is a weak restriction on the average causal effect because the experiment does not rule out any level of dependence between treated and untreated outcomes\(^4\). Except in extreme cases, we expect positive dependence between treated and untreated outcomes, to varying degrees depending on the nature of the program. Most programs cannot cause those well-off without the program to switch places with those poorly-off in absolute terms.

I therefore develop tighter bounds, indexed by the minimum level of dependence between an individual’s treated and untreated outcomes we are willing to consider. When treated and untreated outcomes are perfectly dependent, differences in untreated outcome

\(^2\)See Heckman, Moon, Pinto, Savelyev, and Yavitz (2010) and McKenzie and Woodruff (2008) who assess the external validity of experimental results on the basis of the similarity of the experimental populations to larger populations of interest.

\(^3\)The analysis can easily be extended to the case when individuals choose their treatment status and an experiment denies treatment to a random subset of individuals who would wish to be treated (see Bitler, Domina, and Hoynes (2014) for an example of such an experiment).

\(^4\)The literature on distributions of causal effects consistent with experimental results generates similarly wide bounds on functionals of interest (Heckman, Smith, and Clements (1997); Djebbari and Smith (2008); Fan and Park (2010); Kim (2014)).
distributions are not a problem because each untreated outcome is linked to a single treated outcome. As we move away from perfect dependence, different associations between treated and untreated outcomes become possible. These different associations produce uncertainty about the average causal effect in the new context that is increasing in the difference between the distributions of untreated outcomes in the experiment and the context of interest. The width of the bounds for a given minimum dependence level provide a measure of uncertainty about the average causal effect. They also allow us to assess the assumptions on dependence between treated and untreated outcomes necessary to draw specific conclusions about the effect of the program in the context of interest, such as its ability to exceed a cost-effectiveness threshold.

I empirically evaluate the results of my bounding procedure compared to existing methods for extrapolating causal effects to new contexts. The current benchmark method (Hotz, Imbens, and Mortimer (2005), henceforth HIM) also uses outcome distributions for individuals with the same characteristics to assess generalizability, but does so within a testing framework. If we reject that the untreated outcome distributions for individuals with the same characteristics are the same, we conclude that the experiment teaches us nothing about causal effects in the context of interest. Otherwise, the HIM framework concludes the experiment is perfectly predictive for the causal effect of interest.

I first examine the generalizability of a small experiment on the returns to loosening credit constraints by providing cash transfers to very small-scale entrepreneurs in Leon, Mexico in 2006 documented in McKenzie and Woodruff (2008). We would like to know what the large estimated returns (an increase in monthly profits equal to roughly 40% of the transfer in baseline specifications) in Leon in 2006 tell us about the average return for similarly small-scale microentrepreneurs in urban Mexico in 2012, as represented by that year’s national microenterprise survey. The distributions of untreated outcomes are fairly similar in the Leon and 2012 urban Mexico samples so the estimated bounds are narrow for a wide range of assumptions on dependence between profits with and without the
transfer. Properly accounting for the unobserved differences between the populations along with sampling variation in the small experimental sample and the national microenterprise sample leads to wide confidence intervals around the bounds. Testing equality of control outcome distributions, in contrast, would lead us to be overconfident in our prediction of the average return. Perversely, using the HIM method, we would compute a narrower confidence interval on the predicted causal effect for urban Mexico in 2012 than on the causal effect in the original experiment.

Second, to check the predictions of different methods against measured causal effects, I use data from randomized evaluations of a remedial education program implemented in two Indian cities and described in Banerjee, Cole, Duflo, and Linden (2007). I find different average causal effects for individuals with the same observed characteristics in the two cities. The two cities’ student populations are sufficiently different that equality of their untreated outcome distributions is rejected, which, in the HIM framework would lead us to believe we cannot learn anything about the causal effect in one city based on experimental results from the other. However, I show that if we assume treated and untreated outcomes are sufficiently dependent, we can exclude a substantial range of average causal effects - such as a zero effect - in one city using the results from the other. The observed causal effects in both cities are consistent with predictions based on strong dependence between the treated and untreated outcomes.

This paper extends the literature on generalizing causal effects to new contexts based on invariance assumptions on average treated outcomes or causal effects for individuals with the same observed characteristics (HIM, Attanasio et al. (2003); Angrist and Fernández-Val (2013); Angrist and Rokkanen (2013); Cole and Stuart (2010); Stuart, Cole, Bradshaw, and Leaf (2011); Pearl and Bareinboim (2014); Flores and Mitnik (2013)). In interpreting differences in untreated outcome distributions as indicative of unobserved differences in populations, I follow a long line of literature interpreting outcome quantiles as representing the effect of unobserved heterogeneity in non-separable models (see, for example, Matzkin
(2007) and the references therein). Most directly, Athey and Imbens (2006) make use of this interpretation when generalizing the standard difference-in-differences estimator and derive an estimator that is equivalent to mine under perfect dependence between the treated and untreated outcomes. In moving from a testing framework to an approach based on quantifying assumptions required to draw conclusions about causal effects, my paper relates to work by Altonji, Elder, and Taber (2005) and Altonji, Conley, Elder, and Taber (2013). Altonji et al. (2005) and Altonji et al. (2013) move from testing whether observed covariates related to an outcome are also related to a candidate instrument to providing bounds on the average causal effect whose width depends on the magnitude of the relationship between the covariates and the instrument.

The rest of the paper is organized as follows. Section 1.2 describes the intuition behind the proposed methods by means of a simple example. Readers uninterested in the technical details behind the methods in their full generality may wish to read section 1.2 then skip to the empirical results in sections 1.5 and 1.6. Beginning the theoretical discussion, section 1.3 sets up the problem and notation and provides a review of existing approaches to extrapolation on the basis of experimental results. In section 1.4, I present the derivation of the bounds. Section 1.5 presents the empirical results for generalizing from the 2006 Leon microenterprise experiment to urban locations in Mexico in 2012. Section 1.6 investigates using the results from one of the two remedial education experiments to try to predict the results in the other experiment. Section 1.7 concludes.

1.2 Intuition for the methodology: a simple example

To illustrate the intuition behind the methodological contributions, I begin by laying out a simple example involving a fictional conditional cash transfer program (CCT) that incentivizes parents to enroll children in school. Suppose we have obtained experimental results that tell us the CCT program caused a large increase in the enrollment rate in location $e$, from $\frac{1}{3}$ of all children to $\frac{2}{3}$ of all children. We observe only outcomes and no characteristics.
We would like to know what the results from location e tell us about the causal effect we can expect in location a, where no CCT was implemented. Whereas \( \frac{1}{3} \) of children were enrolled without the CCT program in location e, \( \frac{1}{2} \) of children are enrolled without the CCT in location a. We would like to know what impact the difference in the no-CCT enrollment rates will have on the average causal effect in location a. The law of total probability allows us to decompose the average effect of the CCT program in a, denoted \( ATE^a \), as follows.

\[
ATE^a = P(\text{enrolled with CCT} | \text{enrolled without CCT}) \times P(\text{enrolled without CCT}) \\
+ P(\text{enrolled with CCT} | \text{out of school without CCT}) \times P(\text{out of school without CCT}) \\
- P(\text{enrolled without CCT}) \\
= P(\text{enrolled with CCT} | \text{enrolled without CCT}) \times \frac{1}{2} \\
+ P(\text{enrolled with CCT} | \text{out of school without CCT}) \times \frac{1}{2} \\
- \frac{1}{2}
\]

The average causal effect in a depends on two unknown probabilities: (1) the probability that an individual who does not enroll without the CCT would instead enroll with the CCT and (2) the probability that an individual who enrolls in school without the CCT would also enroll with the CCT.

The rationale behind (1), individuals who do not enroll without the CCT enrolling with a CCT, is clear: the program provides cash incentives for parents to enroll children in school and some parents respond to these incentives. The rationale behind (2), individuals who enroll without the CCT but would not enroll with the CCT, is less straightforward. Attanasio, Meghir, and Santiago (2012) show that CCT programs can increase wages for children by lowering the supply of child labor. An increased wage for children works against the enrollment incentives. Further, Attanasio et al. (2012) show that enrollment subsidies and child wages do not have equal opposite effects on households’ enrollment decisions, as they would if only the net child wage entered into the enrollment decision. So we can think of some fraction of households who are more sensitive to child wages than they are to
enrollment subsidies and would respond to a CCT by having children work. To maintain the simplicity of this example, I will refer to forces that cause children who would enroll without the CCT but would not enroll with a CCT in place as wage effects, although in principle there may be other ways for the CCT to cause children who would otherwise enroll to not enroll.

Figure 1.1: Permissible distributions for $P(\text{enrolled with CCT} \mid \text{enrollment without CCT})$ in location $a$

I will assume that $P(\text{enrolled with CCT} \mid \text{enrolled without CCT})$ in location $a$ is consistent with the experimental results. There are many possible pairs of conditional enrollment probabilities that are consistent with the experimental results. The possible pairs are given in Figure 1.1. To see why a continuum of pairs is possible, recall that

$$P(\text{enrolled with CCT} \mid \text{enrolled without CCT}) = \frac{P(\text{enrolled with CCT} \& \text{enrolled without CCT})}{P(\text{enrolled without CCT})}.$$
We see that $P(\text{enrolled with CCT} \mid \text{enrolled without CCT})$ relies on knowledge a child’s enrollment status with and without the CCT at the same time, knowledge that we cannot have. If a child is in one of the treated localities, we only observe her enrollment decision with the CCT. If she is in one of the control localities, we only observe her enrollment decision without the CCT. The question marks in table 1.1 indicate the unknown fractions of the population of location $e$ falling into each of the four possible combinations of enrollment decisions with and without the CCT. The sums across rows and down columns show the information we do have from the experiment. The rows of table 1.1 must sum to the control group results and the columns to the treatment group results.

Table 1.1: The distribution of enrollment with and without the CCT is unknown in location $e$

<table>
<thead>
<tr>
<th></th>
<th>CCT</th>
<th></th>
<th>All Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Out of school</td>
<td>Enrolled</td>
<td></td>
</tr>
<tr>
<td>No CCT</td>
<td>?</td>
<td>?</td>
<td>$\frac{3}{4}$</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>?</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td>All Treatment</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{3}$</td>
<td></td>
</tr>
</tbody>
</table>

Our assumptions about the way the wage effects of the CCT impact the two groups of children (those enrolling without the CCT and those who do not enroll without the CCT) will generate different predictions for the causal effect of the CCT program in location $a$. To see this, first consider the case where there are no wage effects or wage effects only impact children who do not enroll without the CCT. Then there are no children who enroll without the CCT but would not enroll when the CCT is in place. Our assumption allows us to fill in all the entries of table 1.1, as shown in table 1.2. The probability of enrolling with the CCT if a child is out of school without the CCT is $\frac{1}{2}$ and the increase in the fraction enrolled in location $a$ is $\frac{1}{4}$.

Now consider another assumption about the wage effects: they only impact those who enroll without the CCT and they are so strong that all children who would enroll without the CCT drop out. To match the distribution of control and treated group outcomes in location $e$, all children who are out of school without the CCT must enroll with the CCT.
Table 1.2: Case 1: there are no wage effects

<table>
<thead>
<tr>
<th></th>
<th>CCT</th>
<th>Out of school</th>
<th>Enrolled</th>
<th>All Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No CCT</td>
<td>Out of school</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td></td>
<td>Enrolled</td>
<td>0</td>
<td>$\frac{2}{3}$</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td></td>
<td>All Treatment</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
<td>$\frac{1}{3}$</td>
</tr>
</tbody>
</table>

Again, we can fill in the unknown entries of table 1.1, as shown in table 1.3. In this rather unbelievable case, we predict no change in the fraction enrolled in location $a$.

Table 1.3: Case 2: wage effects only impact those who enroll without the CCT

<table>
<thead>
<tr>
<th></th>
<th>CCT</th>
<th>Out of school</th>
<th>Enrolled</th>
<th>All Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No CCT</td>
<td>Out of school</td>
<td>0</td>
<td>$\frac{2}{3}$</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td></td>
<td>Enrolled</td>
<td>$\frac{1}{3}$</td>
<td>0</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td></td>
<td>All Treatment</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
<td>$\frac{1}{3}$</td>
</tr>
</tbody>
</table>

Assuming that wage effects impact the same fraction of both groups is somewhat more believable. To be consistent with the experimental results, this fraction must be $\frac{1}{3}$. The entries of table 1.1 can be filled in as shown in table 1.4. The predicted increase in the fraction employed is $\frac{1}{6}$.

Table 1.4: Case 3: wage effects impact the same fraction of both groups

<table>
<thead>
<tr>
<th></th>
<th>CCT</th>
<th>Out of school</th>
<th>Enrolled</th>
<th>All Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No CCT</td>
<td>Out of school</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td></td>
<td>Enrolled</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td></td>
<td>All Treatment</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
<td>$\frac{1}{3}$</td>
</tr>
</tbody>
</table>

While more believable than assuming that those enrolled with and without the CCT exchange places when the CCT is in place, assuming that wage effects have the same impact on both groups is still not very convincing. Intuitively, we believe that wage effects would have a stronger impact on enrollment decisions for children who do not enroll without the CCT. Formally, we expect positive dependence between enrollment with the CCT and en-
rollment without. In this paper, I measure dependence using the rank correlation\(^5\) between treated and untreated outcomes for any individual. The first assumption on wage effects we considered, that there are none or they only affect enrollment decisions for children who would enroll without the CCT, generates the maximum possible rank correlation between a child’s enrollment decision with and without the CCT. The third assumption, that wage effects have the same impact regardless of enrollment status without the CCT, generates a rank correlation of zero. As we have seen, different rank correlations generate different predictions for the change in enrollment caused by the CCT in location \(a\).

How close should the rank correlation we use to predict the effect of the CCT on enrollment in location \(a\) be to the maximum possible? I consider two options. First, we might specify a range of plausible values. In this example, we might be conservative and consider rank correlations between zero and the maximum possible. Then, the gain in enrollment in location \(a\) lies between \(\frac{1}{6}\) and \(\frac{1}{4}\). A second option is to explore the strength of assumptions on dependence required to draw specific conclusions about the effect of the program. For example, we might consider what we need to assume about dependence to conclude that the CCT will have a positive effect on enrollment. With an enrollment rate of \(\frac{1}{2}\) in location \(a\), a zero effect on enrollment in location \(a\) is only possible when the rank correlation between enrollment with and without the CCT is the minimum possible, which is what occurs in the second case we considered, when children who enroll without the CCT all drop out with the CCT. Since this case is highly implausible, we would feel confident in our conclusion that the CCT will have a positive effect on enrollment location \(a\).

Note the key role played by the enrollment rate without the CCT in location \(a\). If instead of \(\frac{1}{2}\), the enrollment rate in location \(a\) were \(\frac{2}{3}\), choosing a rank correlation between zero and the maximum possible would predict an increase in the enrollment rate due to the CCT between 0 and \(\frac{1}{6}\). We would need stronger, but still believable, assumptions on dependence to predict a positive effect on enrollment.

In the following two sections, I generalize the intuition developed here to settings where

---

\(^5\)The standard Pearson product-moment correlation measures only linear dependence.
we also have information about observed characteristics in the two populations, where outcomes are non-binary and where our data about locations \( e \) and \( a \) come from samples. Readers uninterested in the details of generalization may wish to skip to the empirical results in sections 1.5 and 1.6.

1.3 Econometric setup

Suppose we are interested in the causal effect of a binary treatment \( T \in \{0, 1\} \) on an observable outcome \( Y \in \mathcal{Y} \subseteq \mathbb{R} \). Each individual is associated with two potential outcomes: \( Y_1 \in \mathcal{Y}_1 \subseteq \mathcal{Y} \) is her outcome if she receives treatment and \( Y_0 \in \mathcal{Y}_0 \subseteq \mathcal{Y} \) is her outcome if she does not. Only one of these two outcomes is ever observed, the other is hypothetical. Mathematically, the observed outcome \( Y \) can be written as:

\[
Y = TY_1 + (1 - T)Y_0.
\]

Because both the observed and hypothetical outcome are defined for each individual we can also define an individual’s own treatment effect \( \Delta \subseteq \mathbb{R} \):

\[
\Delta = Y_1 - Y_0
\]

Our data come from two populations, indexed by \( D \in \{e, a\} \). \( e \) is the population in which the experimental evaluation of \( T \) was conducted and \( a \) is the alternative population of interest. \( d \)-superscripts will index population-specific distributions and their attributes. In population \( e \), the experimental evaluation assigns \( T \) at random independently of all other random variables with perfect compliance\(^6\). Therefore, we can identify the marginal

---

\(^6\)Putting perfect compliance with treatment assignment another way, the estimands of interest will be intention-to-treat (ITT) effects, including any participation decisions. The ITT is often thought to be the object of policy interest since compliance can rarely be mandated in policy settings.
distribution of untreated outcomes in population $e$, $F^e_{Y_0}(y_0)$, from the equality

$$F^e_{Y_0}(y_0) = F^e_{Y|T}(y|T = 0)$$

where $F^e_{Y|T}(y|T = 0)$ denotes the marginal distribution of $Y$ conditional on the treatment indicator being equal to zero. The equality follows from the independence of the treatment indicator from the potential outcomes. We can also identify the marginal distribution of treated outcomes:

$$F^e_{Y_1}(y_1) = F^e_{Y|T}(y|T = 1).$$

We can additionally identify any functionals of the outcome distributions, which allows us to identify the average individual-specific treatment effect $\Delta$ in population $e$:

$$E^e[\Delta] = E^e[Y_1 - Y_0]$$

$$= E^e[Y_1] - E^e[Y_0]$$

$$= E^e[Y_1|T = 1] - E^e[Y_0|T = 0] = E^e[Y|T = 1] - E^e[Y|T = 0].$$

$E^d$ stands for the expectation with respect to the distribution in $D = d$.

As in previous sections, I maintain the assumption that all members of the alternative population are untreated for concreteness. So $T = 0$ for all individuals in population $a$. This means that in population $a$, we identify that distribution of untreated outcomes:

$$F^a_{Y_0}(y_0) = F^a_{Y|T}(y|T = 0) = F^a_Y(y).$$

We are, however, interested in the average treatment effect in alternative population, $E^a[\Delta]$, which depends on our ability to identify $E^a[Y_1]$:

$$E^a[\Delta] = E^a[Y_1 - Y_0]$$
If the treatment effect were constant for all individuals and equal to \( \Delta \), \( E^a[\Delta] \) would simply be equal to \( E^e[\Delta] \). However, theory rarely implies a constant treatment effect and we can often reject it empirically, see e.g. Heckman et al. (1997); Djebbari and Smith (2008). In fact, theory usually predicts heterogeneity in treatment response depending on the individual and her context’s observed and unobserved attributes.

To demonstrate the role of heterogeneity in observed and unobserved characteristics on the average treatment effect in \( a \), I now introduce some additional notation. Suppose we observe a vector of covariates \( X \in \mathcal{X} \subseteq \mathbb{R}^{d_X} \) for each individual. Additionally, suppose there is a vector of unobserved covariates \( U \in \mathcal{U} \subseteq \mathbb{R}^{d_U} \) that we believe affects the outcome. Concretely, we can think of the observed covariates in the remedial education example from the introduction: the student’s grade level competency when entering third grade, class size and gender. The unobserved covariates might be her latent ability and any parental inputs. An equivalent representation for the potential outcomes is that treatment status and covariates combine to produce the outcome through a function common across populations, \( g : \{0, 1\} \times \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R} \). In this representation, the potential outcomes are:

\[
Y_0 = g(0, X, U) \\
Y_1 = g(1, X, U).
\]

The individual-specific treatment effect is

\[
\Delta = Y_1 - Y_0 = g(1, X, U) - g(0, X, U),
\]
which will in general depend on both $X$ and $U$. Our target, $E^a[\Delta]$ can be written as:

$$ATE^a = E^a[Y_1 - Y_0]$$

$$= \int_{X \times U} g(1, x, u) - g(0, x, u)dF_{X,U}(x, u)$$

where $F_{X,U}(x, u)$ denotes the joint distribution of observed and unobserved covariates in population $a$. Note that $F_{X,U}(x, u)$ in general differs from $F_{X}(x, u)$. Iterating expectations, $ATE^a$ can be written in three equivalent ways:

$$ATE^a = \int_X \left[ \int_{U} g(1, x, u) - g(0, x, u)dF_{U|X}(u|x) \right] dF_X(x) \quad (1.1)$$

$$= \int_X \left[ \int_{\mathbb{R}^2} y_1 - y_0 dF_{Y_0,Y_1|X}(y_0, y_1|x) \right] dF_X(x) \quad (1.2)$$

$$= \int_X \left[ \int_{\mathbb{R}} \delta dF_{\Delta|X}(\delta|x) \right] dF_X(x) \quad (1.3)$$

Equations (1.1) and (1.2) show that $ATE^a$ depends on the distribution of $Y_0, Y_1|X, D = a$, which itself depends on the distribution of $U|X, D = a$. Equation (1.3) makes the connection to the distribution of treatment effects for individuals with a particular value of the observed covariates. Note that the equivalence of equations (1.1) and (1.2) shows that the invariance to the population indicator of the function generating outcomes is without loss of generality, since the dimension of $U$ is unrestricted and could include a separate indicator for each population, analogous to defining the $d$-index of $F_{Y_1,Y_0|X}(y_1, y_0|x)$ as an element of $U$.

1.3.1 Previous methods

Within this general setup, I now describe previous methods for using the distributions from the experimental population to identify the average treatment effect in the alternative population.
Conditional independence of the gains

The standard approach to extrapolating the results of social experiments has been to reweight the average treatment effects conditional on each value of the observed covariates by the distribution of observed covariates in the population of interest. That is:

\[ ATE^a = \int_{\mathcal{X}} E^e[Y_1 - Y_0|x] dF^e_x(x). \]  

(1.4)

This estimator is justified on the basis of the following assumptions (Allcott (2015)):

\[ \mathcal{X}^a \subseteq \mathcal{X}^e \]  

(1.5)

\[ \Delta \perp \perp D|X \]  

(1.6)

where \( \mathcal{X}^a \) denotes the support of \( X \) in the alternative population, \( \mathcal{X}^e \) denotes the support in the experimental population and \( \perp \perp \) denotes statistical independence\(^7\). (1.5) is a standard condition required for non-parametric extrapolation. (1.6) is the key identification assumption. Note that under (1.6), \( \Delta = Y_1 - Y_0 \) is independent of any difference between the conditional distributions of untreated outcomes, \( F^e_{Y_0}(y_0|x) \) and \( F^e_{Y_0}(y_0|x) \). With a bounded outcome, the conditional distributions of control outcomes may be such that (1.6) is impossible. For one extreme example, consider the case where \( Y \in \{0, 1\} \), and the outcomes in population in population \( e \) are as in section 1.2, with \( E^e[Y_0] = \frac{1}{3} \) and \( E^e[Y_1] = \frac{2}{3} \). \( E^e[\Delta] = \frac{1}{3} \). If \( Y = 1 \) for all individuals in population \( a \), (1.6) cannot hold. Predictions will also depend on the scaling of \( Y \), for example, whether it is measured in levels or logs\(^8\).

Even more substantively, differences in the conditional distributions of control outcomes are indicative of some unobserved differences between the experimental population and the

---

\(^7\)The estimator in equation (1.4) can be justified on the basis of a weaker mean-independence assumption, but I will focus on the assumptions considered in the literature.

\(^8\)(1.4) is analogous to the counterfactual portion of a difference-in-differences estimator, where the assumption is that the mean difference in outcomes is conditionally independent of the population indicator. Hence, these standard criticisms of difference-in-differences estimators as non-invariant to scaling of the outcome and possibly delivering predictions outside the support of the outcome variable, as described in Athey and Imbens (2006) for example, apply here as well.
population of interest. To see this, note that:

\[ F_{Y_0|X}(y_0|x) = F_{g(0,x,U)}^{d}(g(0, x, U)). \]

Then

\[ F_{Y_0|X}(y_0|x) \neq F_{Y_0|X}^{e}(y_0|x) \implies F_{U|X}^{a}(u|x) \neq F_{U|X}^{e}(u|x). \]

If the elements of \( U \) whose difference in conditional distribution produce the difference in the conditional distribution of control outcomes also influence the individual-specific treatment effect, (1.6) will not hold.

**Conditional independence of the potential outcomes**

Due to some combination of these criticisms, the primary assumption used in the theoretical literature on extrapolation of experimental results combines (1.5) with the assumption that the joint distribution of potential outcomes is independent of the population conditional on the observed covariates:

\[ (Y_0, Y_1) \perp \perp D|X \quad (1.7) \]

or equivalently, that all unobserved covariates determining the outcome are independent of the population indicator:

\[ U \perp \perp D|X \]

It is straightforward to show that (1.7) implies \( E^a[Y_1|x] = E^e[Y_1|x] \) so that we can identify the average treatment effect in the population of interest by reweighting the expectation of the treated outcome from the experimental population conditional on covariates by the distribution of covariates in the population of interest and subtracting the expected control
outcome from the population of interest:

\[
ATE^a = \int_{\mathcal{X}} E^e[Y_1|x]dF^e_X(x) - E^a[Y_0].
\]

For (1.7) to hold, the conditional distributions of control outcomes must be the same in the two populations. Therefore Hotz et al. (2005) and papers following them have suggested testing equality of the distributions or their moments. Two issues come up when testing \( F^e_{Y_0|X}(y_0|x) = F^a_{Y_0|X}(y_0|x) \) and using the result to conclude whether or not we can generalize results from the experiment to the population of interest. First, considering the small sample sizes of many social experiments, we may often be underpowered to reject equality of the conditional outcome distributions, as raised in Flores and Mitnik (2013). Second, if we do reject the null hypothesis, we must conclude that the experiment tells us nothing about \( ATE^a \). Again, this may be an issue of sample size: with large samples from both the experimental population and the population of interest we will in all likelihood reject the null. Furthermore, there is an issue of degree. Suppose we have two alternative populations of interest \( a \) and \( a' \) and our samples are large enough to reject both \( F^e_{Y_0|X}(y_0|x) = F^a_{Y_0|X}(y_0|x) \) and \( F^{e'}_{Y_0|X}(y_0|x) = F^{a'}_{Y_0|X}(y_0|x) \) but \( F^a_{Y_0|X}(y_0|x) \) is quite similar to \( F^{e'}_{Y_0|X}(y_0|x) \) while \( F^{a'}_{Y_0|X}(y_0|x) \) is quite different, it seems inappropriate to conclude that the results from \( e \) are equally (and completely) uninformative in predicting the average causal effect in both \( a \) and \( a' \). In the following section, I depart from the testing framework and derive bounds on the average causal effect in the population of interest as a function of the differences in the conditional distributions of control outcomes between the population of interest and the experimental population. I conclude this section with a simple example.

1.3.2 Example: remedial education in India

To make the above discussion concrete, I now describe a simple parametric model using the example of remedial education India. Suppose students from the city of Mumbai represent the experimental population, \( e \), and students from the city of Vadodara the alternative
population, \( a \), where we would like to predict the average treatment effect. We will leave the observed covariates \( X \) as a vector, but break the vector \( U \) into the two components discussed above, latent skill \( S \) and parental input \( I \). \( g(\cdot) \) is a linear production function with different parameters depending on treatment status

\[
g(0, X, S, I) = \beta_0 + \beta_{0X}X + \beta_{0S}S + \beta_{0I}I = Y_0
\]

\[
g(1, X, S, I) = \beta_1 + \beta_{1X}X + \beta_{1S}S + \beta_{1I}I = Y_1
\]

Note that once we assume linearity, the commonality of \( g(\cdot) \) across populations is no longer without loss of generality. In this case, the individual-specific treatment effect, \( \Delta \), is

\[
\Delta = Y_1 - Y_0 \\
= (\beta_1 - \beta_0) \\
+ (\beta_{1X} - \beta_{0X})X \\
+ (\beta_{1S} - \beta_{0S})S \\
+ (\beta_{1I} - \beta_{0I})I
\]

Our objective is to identify:

\[
ATE^a = E^a[Y_1 - Y_0]
\]

\[
= (\beta_1 - \beta_0) \\
+ E^a[(\beta_{1X} - \beta_{0X})X] \\
+ E^a[(\beta_{1S} - \beta_{0S})S] \\
+ E^a[(\beta_{1I} - \beta_{0I})I]
\]

The four elements of \( ATE^a \) are, respectively, a treatment effect common to all students, the average deviation from the common treatment effect due to observables in population
$a$, the average deviation from the common effect due to latent skill in population $a$ and the average deviation from the common effect due to the parental input. When $\beta_1' X \neq \beta_0' X$, there is treatment effect heterogeneity due to observable covariates and when $\beta_1 S \neq \beta_0 S$ or $\beta_1 I \neq \beta_0 I$ there is treatment effect heterogeneity due to unobservables.

$ATE^e$ alone will in general be biased as an estimator for $ATE_a$, with the bias taking the following form:

$$ATE^e - ATE_a = (\beta_1' X - \beta_0' X)(E^e[X] - E^a[X])$$
$$+ (\beta_1 S - \beta_0 S)(E^e[S] - E^a[S])$$
$$+ (\beta_1 I - \beta_0 I)(E^e[I] - E^a[I])$$

The bias depends on the differences between sites in the marginal distributions of characteristics along which treatment effects are heterogeneous.

In this simple example, we need $E^a[S|x] = E^e[S|x]$ if $\beta_1 S \neq \beta_0 S$ and $E^a[I|x] = E^e[I|x]$ if $\beta_1 I \neq \beta_0 I$ for conditional independence of the gains, (1.6), to hold. We need $E^a[S|x] = E^e[S|x]$ if $(\beta_0 S, \beta_1 S) \neq (0, 0)$ and $E^a[I|x] = E^e[I|x]$ if $(\beta_0 I, \beta_1 I) \neq (0, 0)$ for conditional independence of the potential outcomes, (1.7), to hold. We will return to this parametric model to build intuition for key points in the next section as well.

1.4 Bounds on $ATE^a$ using differences in the untreated outcome distributions

1.4.1 Identification

In investigating the role of the conditional untreated outcome distributions in determining the average causal effect in the population of interest, recall first that since we can already identify $E^a[Y_0]$ (simply the expected outcome in the population of interest), what we need to identify $E^a[Y_1] - E^a[Y_0]$ is the counterfactual $E^a[Y_1]$. The expected value of the treated
outcome in the population of interest can be written as follows:

\[
E^a[Y_1] = \int_X \left( \int_\mathbb{R} \left[ \int_\mathbb{R} y_1 \, dF_{Y_1|Y_0,X}^a(y_1|y_0,x) \right] \frac{dF_{Y_0|X}^a(y_0|x)}{\text{identified}} \right) \frac{dF_X^a(x)}{\text{identified}}
\]

(1.8)

We are missing information on the distribution of treated outcomes that individuals with a particular untreated outcome would experience in the population of interest. Since no one is treated in the population of interest, for information on this object, we must turn to the experimental population.

For the experiment to tell us anything about \( F_{Y_1|Y_0,X}^a(y_1|y_0,x) \), we must first impose two support conditions.

**Assumption 1.** The support of \( X \) in the population of interest is a subset of the support in the experimental population: \( \mathcal{X}^a \subseteq \mathcal{X}^e \).

**Assumption 2.** The support of \( Y_0|X = x \) in the population of interest is a subset of the support in the experimental population for all values of \( X \) in the support of \( X \) in the population of interest: \( \text{Supp}^a(Y_0|X = x) \subseteq \text{Supp}^e(Y_0|X = x) \) \( \forall x \in \mathcal{X}^a \).

Assumption 1 is the same as employed in the previous literature (see equation (1.5)). Assumption 2 will be needed to nonparametrically tie differences in the conditional distributions of untreated outcomes to differences in the conditional distributions of treated outcomes. I will explore alternative assumptions when these are violated in an extension.

Turning now to the question of identification of \( F_{Y_1|Y_0,X}^a(y_1|y_0,x) \) using information from the experiment, we first observe that there are many possible covariate-and-untreated-outcome-conditional distributions \( F_{Y_1|Y_0,X}^c(y_1|y_0,x) \) associated with the covariate-conditioned marginal untreated outcome \( F_{Y_0|X}^c(y_0|x) \) and treated outcome distributions \( F_{Y_1|X}^c(y_1|x) \).

Specifically, \( F_{Y_1|Y_0,X}^c(y_1|y_0,x) \) is a valid conditional distribution for the marginal distributions \( F_{Y_0|X}^c(y_0|x) \) and \( F_{Y_1|X}^c(y_1|x) \) if

\[
F_{Y_1|Y_0,X}^c(y_1|y_0,x) = C_1(F_{Y_0|X}^c(y_0|x), F_{Y_1|X}^c(y_1|x)|x)
\]
where $C : [0, 1]^2 \to [0, 1]$ is a copula function (see appendix 1.8 for the definition), and $C_1(v, w|x) = \frac{\partial C(v, w|x)}{\partial v}$. Informally, a copula function is a bivariate CDF where both arguments are defined on the unit interval which fully determines a dependence structure between the untreated and treated outcomes in the experimental population for individuals with the same covariates. A copula function combined with the marginal distributions of untreated ($F^e_{Y_0|X}(y_0|x)$) and treated outcomes ($F^e_{Y_1|X}(y_1|x)$) defines a joint distribution ($F_{Y_0, Y_1|X}(y_0, y_1|x)$) consistent with those marginal distributions. $F_{Y_1|Y_0, X}(y_1|y_0, x)$ is the conditional distribution associated with the joint distribution $F_{Y_0, Y_1|X}(y_0, y_1|x)$. Let $C$ denote the set of valid copula functions.

I will assume that the distribution of treated outcomes conditional on an untreated outcome and observed covariates in the alternative population of interest is consistent with the experimental results.

**Assumption 3.** Consistency of the conditional distribution of treated outcomes in the population of interest with the experimental results:

$$F^{e}_a|Y_0, X(y_1|y_0, x) = C_1(F^e_{Y_0|X}(y_0|x), F^e_{Y_1|X}(y_1|x)|x)$$

for some copula function $C \in C$.

Assumption 3 states that we must be able to express the distribution of the treated outcome conditional on an untreated outcome and covariates as one of the conditional distributions consistent with the distributions of untreated and treated outcomes in the experiment.

To make Assumption 3 more concrete, I illustrate two examples of copula functions and show how they define a joint distribution of potential outcomes $F_{Y_0, Y_1|X}(y_0, y_1|x)$. Let $Q^e_{Y_0|X}(\alpha|x)$ denote the $\alpha$-quantile of $Y_0|X$ in the experimental population and $Q^e_{Y_1|X}(\alpha|x)$ the $\alpha$-quantile of $Y_1|X$ in the experimental population. Figures 1.2 and 1.3 show two possible copulas and the joint distributions they define. The arrows in the figures represent...
dependence relationships between \( F_{Y_0|X}(y_0|x) \) and \( F_{Y_1|X}(y_1|x) \) defined by the copulas. The horizontal arrows in figure 1.2 represent the joint distribution \( Y_0, Y_1|X \) in the experimental population when the treatment preserves individuals’ ranks in the outcome distributions perfectly. In the example of remedial education in India, the highest-scoring student without a remedial education teacher assigned to her school would still be the highest-scoring student with a remedial education teacher assigned. The crossing arrows in figure 1.3 represent the case when the treatment reverses ranks: the highest scoring student without the treatment would be the lowest-scoring student without the treatment.

Figure 1.2: Perfect positive dependence of \( F_{Y_0|X}(y_0|x) \), \( F_{Y_1|X}(y_1|x) \)

A joint distribution \( F_{Y_0,Y_1|X}(y_0, y_1|x) \) consistent with the experimental marginal distributions of control and treated outcomes also determines the extent of heterogeneity in treatment effects for individuals with covariates \( x \). When the treatment perfectly preserves individuals’ ranks in the outcome distributions, treatment effect heterogeneity due to unobservables is minimized. That is, conditional on \( x \), the individual-specific treatment effects \( \Delta \) have the the smallest magnitude possible. In contrast, when the treatment inverts individuals’ ranks in the outcome distributions, the treatment effects have the largest possible
Figure 1.3: Perfect negative dependence of $F^e_{Y_0|X}(y_0|x)$, $F^e_{Y_1|X}(y_1|x)$

The relationship between $Y_0|X$ and $Y_1|X$ under perfect positive dependence is known as comonotonicity, which is defined as follows.

**Definition 1.** Comonotonicity. When two random variables $V$ and $W$ are comonotonic

$$F_{V,W}(v,w) = \min \{F_V(v), F_W(w)\}.$$  

A necessary condition for Assumption 3 is that if the control outcomes conditional on a value of the covariates have the same distribution in the experimental population and the population of interest, the conditional treated outcomes have the same distribution as well. That is,

$$F^a_{Y_0|X}(y_0|x) = F^e_{Y_0|X}(y_0|x) \implies F^a_{Y_1|X}(y_1|x) = F^e_{Y_1|X}(y_1|x).$$

A sufficient condition but stronger than necessary condition is that the distribution of
the treated outcomes be the same across populations once we have conditioned on a value of the control outcome and the observed covariates, an assumption also used in Athey and Imbens (2006). Formally:

\[ Y_1 \independent D \mid Y_0, X \] (1.9)

This is the relevant condition to answer the hypothetical, what would the conditional distribution of treated outcomes have been in the experiment had the distribution of control outcomes been the same as in the population of interest (see Fortin, Lemieux, and Firpo (2011))? In terms of the underlying unobservables, a sufficient condition for (1.9), in turn, is:

\[ U \independent D \mid g(0, x, U) = y_0, X = x. \]

Finally, we require existence of the expectation of \( Y_1 \) in \( e \).

**Assumption 4.** \( Y_1 \) has finite expectation in \( e \): \( E^e[|Y_1|] < \infty. \)

Combining assumptions 1, 2, 3 and 4, we state the following result.

**Proposition 1.** Under assumptions 1, 2, 3 and 4:

\[
E^a[Y_1 - Y_0|x] \in \left[ \min_{C \in \mathcal{C}} \int \left( \int y_1 dC_1(F_{Y_0}^e(y_0|x), F_{Y_1}^e(y_1|x)|x) \right) dF_{Y_0}^a(y_0|x) \right] - E^a[Y_0|x],
\]

\[ \left\{ \max_{C \in \mathcal{C}} \int \left( \int y_1 dC_1(F_{Y_0}^e(y_0|x), F_{Y_1}^e(y_1|x)|x) \right) dF_{Y_0}^a(y_0|x) \right\} - E^a[Y_0|x] \right] \]

Bounds on the unconditional average treatment effect in the population of interest can then be recovered by weighting the minimal and maximal conditional average treatment effects by the distribution of covariates in the population of interest.

\[
ATE^a \in \left[ \int_{\mathcal{X}} \min_{x \in \mathcal{X}} E^a[Y_1 - Y_0|x] dF_X^a(x), \int_{\mathcal{X}} \max_{x \in \mathcal{X}} E^a[Y_1 - Y_0|x] dF_X^a(x) \right] \] (1.10)
All of the objects in proposition 1 are identified, with the exception of the copula $C$. We minimize and maximize over the set of possible copulas $C$ to obtain the bounds. The bounds defined in proposition 1 are sharp by construction, since each element of $C$ defines a valid possible conditional distribution $F_{Y_1|Y_0,X}(y_1|y_0, x)$.

By considering the full set of possible copulas, we consider copulas that may not be credible, however. In particular, the dependence structure shown in figure 1.3 is not realistic in most applications. In the remedial education example, it is clearly unrealistic to believe that the highest-performing students when no remedial education teacher is assigned to their school become the lowest-performing when a remedial education teacher is assigned. Unless remedial education is so effective that a poor-performing student without treatment becomes the best-performing student, the best-performing student without treatment’s rank in the outcomes distribution is likely unaffected: she is not assigned to work with the remedial education teacher and remains the highest-performing. We typically anticipate some positive dependence between outcomes with and without treatment for any one individual, with the degree of dependence (and thus of unobserved treatment effect heterogeneity) depending on the application.

We therefore index copulas by the degree of dependence in the joint distributions of control and treated outcomes they generate. We use Normalized Spearman’s $\rho$, defined below, to measure dependence.

**Definition 2.** For any two random variables $V$ and $W$, Normalized Spearman’s $\rho$ is given by:

$$\rho(V,W) = \frac{Cor_C(R(V), R(W))}{Cor_M(R(V), R(W))}$$

where $R(V) = F_V(v)$ when $V$ is continuously distributed and $R(V) = \frac{F_V(v) + F_V(v^-)}{2}$ when $V$ takes a finite number of values and equivalently for $W$. The notation $F_V(v^-)$ denotes $P(V < v)$ and equivalently for $W$. $Cor_C(R(V), R(W))$ refers to the product-moment correlation between $R(V)$ and $R(W)$ under copula $C$: $\int (R(V) - \frac{1}{2}) (R(W) - \frac{1}{2}) \ dC(F_V(v), F_W(w))$. 

\( \text{Cor}_M(R(V), R(W)) \) is the product-moment correlation between \( R(V) \) and \( R(W) \) under comonotonicity: \( \int (R(V) - \frac{1}{2}) (R(W) - \frac{1}{2}) d (\min \{ F_V(V), F_W(W) \}) \).

The definition of Normalized Spearman’s \( \rho \) coincides with the standard calculation of Spearman’s \( \rho \) in the numerator (see Nešlehová (2007)). In the denominator, when \( V \) and \( W \) are continuously distributed, \( \int (R(V) - \frac{1}{2}) (R(W) - \frac{1}{2}) d (\min \{ F_V(V), F_W(W) \}) \) equals 1 so that the calculation is completely standard. However, when \( V \) and \( W \) take a finite number of values, \( \int (R(V) - \frac{1}{2}) (R(W) - \frac{1}{2}) d (\min \{ F_V(V), F_W(W) \}) \) may be less than 1. So the only difference with the standard calculation is the normalization in the discrete case.

We can produce bounds on \( E^a[Y_1 - Y_0|x] \) subject to the restriction that we only consider copula functions generating dependence greater than a specified level. This is represented in the following assumption and proposition.

**Assumption 5.** \( C \) is an element of \( \mathcal{C}(\rho^L) \), the set of copula functions such that \( \rho(Y_0, Y_1|X = x) \geq \rho^L \) where \( \rho^L \in [0, 1] \).

**Proposition 2.** Under Assumptions 1, 2, 3, 4 and 5:

\[
E^a[Y_1 - Y_0|x] \in \left[ \min_{C \in \mathcal{C}(\rho^L)} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} y_1 d C_1(F^c_{Y_0}(y_0|x), F^c_{Y_1}(y_1|x)|x) \right) d F^a_{Y_0}(y_0|x) - E^a[Y_0|x], \right. \\
\left. \max_{C \in \mathcal{C}(\rho^L)} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} y_1 d C_1(F^c_{Y_0}(y_0|x), F^c_{Y_1}(y_1|x)|x) \right) d F^a_{Y_0}(y_0|x) - E^a[Y_0|x] \right]
\]

Bounds on the unconditional \( \text{ATE}^a \) can be computed in the same way as under proposition 1 (equation (1.10)). \( C(1) \) is a singleton and the bounds shrink to a point. We now investigate the structure underlying the potential outcomes as a means of interpreting the results and assumptions.

**1-dimensional unobservables generate comonotonicity**

Suppose an individual’s control and treated potential outcomes, \( Y_0 \) and \( Y_1 \), are both generated by a single unobserved characteristic of the individual so that \( U \) is one-dimensional
and the structural functions $g(0, x, u)$ and $g(1, x, u)$ are each weakly increasing in $u$. It is a standard result that this implies comonotonicity of the potential outcomes (see, for example, the proof of proposition 5.16 in McNeil, Frey, and Embrechts (2005)).

Athey and Imbens (2006) use this characterization of $Y_t$ (however, in their difference-in-differences setting $T$ indexes time, rather than treatment), along with assumptions 1, 2 and 3 and the condition $U \perp T$ to yield an estimator they refer to as the changes-in-changes model with conditional independence (see section 4.2 of Athey and Imbens (2006)). $U \perp T$ by design in the experiment ($T$ is randomly assigned independently of any other random variable), so the changes-in-changes model with conditional independence is a valid estimator for the point defined under proposition 2 when $\rho_L = 1$. When outcomes are continuous, Athey and Imbens (2006) point out that assumption 3 is implied by monotonicity in $u$ of the function generating outcomes and thus does not need to be separately imposed.

**Example.** To gain some intuition for the identifying power of assuming $g(0, x, u)$ and $g(1, x, u)$ are strictly increasing in 1-dimensional $u$, we return to the parametric example introduced in section 1.3.2. Assume the parental input $I$ is excluded from the production function so unobservables are one-dimensional and the potential outcomes can be written as

$$Y_0 = \beta_0 + \beta_{0X} X + \beta_{0S} S$$
$$Y_1 = \beta_1 + \beta_{1X} X + \beta_{1S} S$$

This is not the only way to generate 1-dimensional unobservables in the linear production function described in section 1.3.2. We could make use of a single index specification for the unobservables where

$$Y_0 = \beta_0 + \beta_{0X} X + \beta_{0S} S + \beta_{0I} I$$
$$Y_1 = \beta_1 + \beta_{1X} X + \kappa(\beta_{0S} S + \beta_{0I} I)$$

Alternatively, if $S$ and $I$ have a Pearson product-moment correlation of 1, we can write $I$ as a linear function of $S$ ($I = bS$) so that:

$$Y_0 = \beta_0 + \beta_{0X} X + (\beta_{0S} + \beta_{0I} b)S$$
$$Y_1 = \beta_1 + \beta_{1X} X + (\beta_{1S} + \beta_{1I} b)S$$
In this section I illustrate that with a one-dimensional unobservable, the way in which the distributions of observables \( F_{X,Y}(x,y) \) in the experimental population change with treatment status can be mapped into differences in the treatment and control structural functions. This knowledge of the changes in the structural function can be applied to differences in the distributions of observables in the control state, \( F_{X,Y_0}(x,y_0) \) and \( F_{X,Y_0}(x,y_0) \), across populations to recover \( E^a[Y_1] \).

Let \( \alpha = F_{Y_0|X}(y_0|x) \) for a given value of \( y_0 \). Consider the \( \alpha \) quantiles of \( Y_1|X \) and \( Y_0|X \) in \( e \):

\[
Q_{Y_1|X}(\alpha|x) = \beta_1 + \beta_1'X + \beta_1S Q_{S|X}(\alpha|x)
\]

\[
Q_{Y_0|X}(\alpha|x) = \beta_0 + \beta_0'X + \beta_0S Q_{S|X}(\alpha|x)
\]

Making use of the linear functional form, we can subtract the \( x \)-subgroup, \( t \)-specific mean from each quantile to remove the common and \( x \)-specific structural effects:

\[
Q_{Y_1|X}(\alpha|x) - E^e[Y_1|x] = \beta_1S \left( Q_{S|X}(\alpha|x) - E^e[S|x] \right)
\]

\[
Q_{Y_0|X}(\alpha|x) - E^e[Y_0|x] = \beta_0S \left( Q_{S|X}(\alpha|x) - E^e[S|x] \right)
\]

By dividing the \( e \) treatment group \( \alpha \)-quantile-specific deviation from the \( x \)-subgroup specific mean from the corresponding \( \alpha \)-quantile-specific deviation in the \( e \) control group, we obtain the ratio of the effects of the latent skill \( S \) in the treated and control states.

\[
\frac{Q_{Y_1|X}(\alpha|x) - E^e[Y_1|x]}{Q_{Y_0|X}(\alpha|x) - E^e[Y_0|x]} = \frac{\beta_1S \left( Q_{S|X}(\alpha|x) - E^e[S|x] \right)}{\beta_0S \left( Q_{S|X}(\alpha|x) - E^e[S|x] \right)} = \frac{\beta_1S}{\beta_0S}
\]

\[ (1.11) \]

Knowing the ratio of the effects of latent math skill across treatment and control states allows us to map differences in the distributions of latent skill and pre-test score \( F_{X,S}(x,s) \)
and $F^e_{X,S}(x,s)$ identified by differences in the joint distributions of the control outcomes $F^e_{X,Y_0}(x,y_0)$ and $F^a_{X,Y_0}(x,y_0)$ into differences in the observed treatment group distribution in $e$, $F^e_{X,Y_1}(x,y_1)$, and the unknown treated group distribution in $a$, $F^a_{X,Y_1}(x,y_1)$. Specifically, consider:

$$
E^a[Y_0|x] - E^e[Y_0|x] = \beta_{0S} (E^a[S|x] - E^e[S|x]).
$$

Then we can use the change in the effect of unobservables from equation (1.11) to identify the unknown expected value of the treated outcome conditional on covariates $x$.

$$
E^a[Y_1|x] - E^e[Y_1|x] = \frac{\beta_{1S}}{\beta_{0S}} (E^a[Y_0|x] - E^e[Y_0|x])
$$

$$
E^a[Y_1|x] = \frac{\beta_{1S}}{\beta_{0S}} (E^a[Y_0|x] - E^e[Y_0|x]) + E^e[Y_1|x]
$$

Finally, the conditional average treatment effect is obtained by subtracting the conditional expectation of the test score in the population of interest.

$$
E^a[Y_1 - Y_0|x] = \frac{\beta_{1S}}{\beta_{0S}} (E^a[Y_0|x] - E^e[Y_0|x]) + E^e[Y_1|x] - E^a[Y|x]
$$

**Multidimensional heterogeneity**

However, when we introduce multidimensional heterogeneity, we can no longer cleanly apply the knowledge we gain from the experiment about how the structural function $g(t,x,u)$ changes with treatment to the differences in $F^e_{X,Y_0}(x,y_0)$ and $F^a_{X,Y_0}(x,y_0)$.

**Example.** This is easy to see in the parametric illustration when we reintroduce independent variation in $I$. Consider the treatment-to-control ratio of $\alpha$-quantile deviations from the $x$-specific subgroup means in the experimental population:

$$
\frac{Q^e_{Y_1|x}(\alpha|x) - E^e[Y_1|x]}{Q^e_{Y_0|x}(\alpha|x) - E^e[Y_0|x]} = \frac{Q^e_{\beta_{1S}\alpha + \beta_{1I}|x}(\alpha|x) - E^e[\beta_{1S}\alpha + \beta_{1I}|x]}{Q^e_{\beta_{0S}\alpha + \beta_{0I}|x}(\alpha|x) - E^e[\beta_{0S}\alpha + \beta_{0I}|x]}
$$
Whereas previously this ratio simplified to the treatment-to-control ratio of effects of latent skill on the test score at the end of third grade, it no longer identifies any specific change in the structural function. Put more generally, the $\alpha$-quantile of $Y_t|x$ in the experimental population now provides no structural information.

We will see in the next section that for very small deviations from 1-dimensional unobserved heterogeneity, the bounds on the average treatment effect in the population of interest expand substantially, depending on the extent of difference in the conditional distributions of the control outcomes between the population of interest and the experimental population. Only when unobserved heterogeneity is exactly, and not approximately, 1-dimensional do differences in the conditional distributions of the control outcomes not lead to a loss in identification. This motivates considering the bounds from proposition 2 and investigating how they change with $\rho^L$.

1.4.2 Estimation

In estimation, I will consider the case when outcomes and covariates are discrete or discretized. I will illustrate both possibilities in the empirical work. When outcomes and covariates are discrete, we can represent the optimization over the restricted space of copulas $C(\rho^L)$ as a linear programming problem. In particular, the bounds on the average causal effect in context $a$ for individuals with covariates $x$ admit a representation as the solution to a discrete optimal transportation problem with a non-standard cost function and an additional linear constraint on dependence (see Villani (2009) for a comprehensive discussion of optimal transportation problems). Very efficient algorithms are available to solve linear programs (see e.g. Boyd and Vandenberghe (2004)), so the bounds can be computed quickly using software provided by the author.

A similar representation as a continuous optimal transportation problem exists when outcomes are continuous, but there is no analogous tractable method to compute the solution, which involves optimization over an infinite-dimensional space ($C(\rho^L)$). It may be
possible to represent $C(\rho^L)$ with a sieve space $C_n(\rho^L)$, which would be finite-dimensional and compact, becoming dense as $n \to \infty$. Exploring this possibility is left to future research. I therefore impose the following assumption on outcomes and covariates.

**Assumption 6.** Finite support of the potential outcomes and covariates. Let $J, K \in \mathbb{N}$. $Y_0$ and $Y_1$ take values in $\mathcal{Y}_0 = \{y_{0,1}, \ldots, y_{0,j}, \ldots, y_{0,J}\}$ and $\mathcal{Y}_1 = \{y_{1,1}, \ldots, y_{1,k}, \ldots, y_{1,K}\}$, respectively. Further, $X$ takes values in a finite set $\mathcal{X}$.

**Linear programming representation**

I first describe the linear programming representation of the bounds in Proposition 2. I leave conditioning on $x$ implicit to economize on notation. Given $\rho^L$, the upper bound is obtained by solving the following linear programming problem with solution $\tau^U(\rho^L)$ (the lower bound, $\tau^L(\rho^L)$ is obtained by replacing the max operator with min).
\[ z^L(\rho^L) = \max_{C(\rho^L)} E^a[Y_1 - Y_0] \]

\[ = \max_{\{P^e(y_{0j}, y_{1k})\}_{j=1,...,J}^{k=1,...,K}} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{1k} \frac{P^a(y_{0j})}{P^e(y_{0j})} \times P^e(y_{0j}, y_{1k}) \]

\[ - \sum_{j=1}^{J} y_{0j} P^a(y_{0j}) \]  

(1.12)

subject to

\[ \sum_{k=1}^{K} P^e(y_{0j}, y_{1k}) = P^e(y_{0j}) \forall j \in \{1, ..., J\} \]  

(1.14)

\[ \sum_{j=1}^{J} P^e(y_{0j}, y_{1k}) = P^e(y_{1k}) \forall k \in \{1, ..., K\} \]  

(1.15)

\[ \sum_{j=1}^{J} \sum_{k=1}^{K} \left( R(y_{0j}) - \frac{1}{2} \right) \left( R(y_{1k}) - \frac{1}{2} \right) P^e(y_{0j}, y_{1k}) \]

\[ \geq \rho^L \left[ \max_{\{P^e(y_{0j}, y_{1k})\}_{j=1,...,J}^{k=1,...,K}} \sum_{j=1}^{J} \sum_{k=1}^{K} \left( R(y_{0j}) - \frac{1}{2} \right) \left( R(y_{1k}) - \frac{1}{2} \right) P^e(y_{0j}, y_{1k}) \right] \]  

(1.16)

Maximization is with respect to the elements of the matrix defining the joint distribution of \( Y_0 \) and \( Y_1 \) in population \( e \), \( \{P^e(y_{0j}, y_{1k})\}_{j=1,...,J}^{k=1,...,K} \). Line (1.13) is simply a normalization so that the value of the objective function of the problem can be interpreted as \( E^a[Y_1 - Y_0] \).

Constraints (1.14) and (1.15) require that the minimizing/maximizing joint distribution be consistent with the marginal outcome distributions in \( e \). Constraint (1.16) enforces that Normalized Spearman’s \( \rho \) (see Definition 2) applied to the potential outcomes \( Y_0 \) and \( Y_1 \), \( \rho(Y_0, Y_1) \), may not be below \( \rho^L \). Constraints (1.14), (1.15) and (1.16) make maximizing over the elements of the joint distribution of \( Y_0 \) and \( Y_1 \) equivalent to maximizing over the restricted space of copulas, \( C(\rho^L) \) (proof in Appendix 1.9).

The coefficients on the elements of \( \{P^e(y_{0j}, y_{1k})\}_{j=1,...,J}^{k=1,...,K} \) are \( \left\{ \frac{y_{1k} P^a(y_{0j})}{P^e(y_{0j})} \right\}_{j=1,...,J}^{k=1,...,K} \). Together with constraint (1.15), this shows the role of the distributions of control outcomes \( \{P^a(y_{0j})\}_{j=1,...,J} \) and \( \{P^e(y_{0j})\}_{j=1,...,J} \) in determining the bounds. If \( P^a(y_0) = P^e(y_0) \),
$P^n(y_0) \over P^n(y_0) = 1$ and constraint (1.15) implies that the counterfactual $E^a[Y_1] = E^e[Y_1]$ \(^{10}\). All else equal, in order to maximize the objective function, we would like to assign higher probability to high values on the support of $Y_1$ (high $k$) when $P^n(y_0j) / P^n(y_0j)$ is large and to low values on the support of $Y_1$ (low $k$) when $P^n(y_0j) / P^n(y_0j)$ is small. However, constraint (1.16) limits our ability to do so.

**Example.** Table 1.5 shows the choice variables and constraints 1.14 and 1.15 in the context of the remedial education in India example where the city of Mumbai is treated as $e$ and Vadodara as $a$. As will be discussed in more detail in section 1.6, I do not use the test score directly as an outcome, but rather the discrete grade level competency of third graders when completing third grade. In table 1.5, I condition on a competency level of zero on entering third grade. The row and column labeled “All” represents the constraints on the marginal distributions $P^e(y_0|x)$ and $P^e(y_1|x)$. Without further constraints, the values of the choice variables are restricted only by the requirement that the sums across rows (for the untreated outcomes) equal the probability in the column labeled “All control” and that the sums down the columns (for the treated outcomes) equal the probability in the row labeled “All treated.”

Table 1.6 shows the coefficient on each choice variable $P^e(y_0j, y_1k)$ when Mumbai is treated as $e$ and we condition on students’ grade-level competency being zero on entering third grade. We can see that the differences in the distributions of control outcomes mean

\[ \sum_{j=1}^{J} \sum_{k=1}^{K} y_{1k} P^e(y_0j, y_{1k}) = \sum_{j=1}^{J} y_{1k} \sum_{k=1}^{K} P^e(y_0j, y_{1k}) = \sum_{j=1}^{J} y_{1k} P^e(y_{1k}) = E^e[Y_1] \]

where the second equality follows from substituting in constraint (1.15). \(\square\)
that we would maximize the objective function by ascribing the highest treatment effects to individuals with $Y_0 = 1$ and the lowest treatment effects to individuals with $Y_0 = 3$.

Constraint (1.16) on the dependence between $Y_0$ and $Y_1$ in Mumbai limits our ability to do so arbitrarily. Recall that $\rho^L$ governs the allowed deviations from 1-dimensional heterogeneity. To gain some intuition for the joint distributions implied by different values of $\rho^L$, table 1.7 shows the joint distribution implied by $\rho^L = 1$ when Mumbai is treated as $e$ and we condition on students’ grade-level competency being zero on entering third grade. When $\rho^L = 1$, the 1-dimensional heterogeneity case, the majority of the mass in the joint distribution lies on the principal diagonal. Most individuals (88%) have a treatment effect of zero, with a few individuals experiencing a positive treatment effect of at most 1 competency level.

Table 1.6: Contribution of choice variables to the objective, $e =$Mumbai

<table>
<thead>
<tr>
<th>Remedial education</th>
<th>Competency on exiting grade 3</th>
<th>All control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No remedial ed</td>
<td>$P^e(0, 0)$</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>$P^e(1, 0)$</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>$P^e(2, 0)$</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>$P^e(3, 0)$</td>
<td>0.03</td>
</tr>
<tr>
<td>All treated</td>
<td>0.66</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Table 1.7: Choice variables, $e =$Mumbai

<table>
<thead>
<tr>
<th>Remedial education</th>
<th>Competency on exiting grade 3</th>
<th>All control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No remedial ed</td>
<td>$P^e(0, 0)$</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>$2 \times 0.71$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$3 \times 0.71$</td>
<td></td>
</tr>
<tr>
<td>All treated</td>
<td>0.71</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$2 \times 2.26$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$3 \times 2.26$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$1.16$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$2 \times 1.16$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$3 \times 1.16$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.60$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$2 \times 0.60$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$3 \times 0.60$</td>
<td></td>
</tr>
</tbody>
</table>
Table 1.7: $P_{ex}(y_{0j},y_{1k}|\text{competency on entering third grade} = 0), \rho^L = 1 \iff \text{Mumbai}$

<table>
<thead>
<tr>
<th>Remedial education</th>
<th>Competency on exiting grade 3</th>
<th>All Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.66 0.07 0.0 0 0.73</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0 0.13 0.04 0 0.17</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0 0 0.06 0.01 0.07</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0 0 0 0.03 0.03</td>
<td></td>
</tr>
<tr>
<td>All Treatment</td>
<td>0.66 0.2 0.1 0.04</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Sample counterparts estimator

The solutions to the linear programming representation conditional on observed covariates $x$ and minimum rank correlation $\rho^L$, $\tau^L_x(\rho^L)$ and $\tau^U_x(\rho^L)$ when minimizing and maximizing respectively, are functions of the population objects $\{P_{ex}(y_{0j}|X = x)\}_{j=1,...,J}$, $\{P_{ex}(y_{1k}|X = x)\}_{k=1,...,K}$ and $\{P_{tr}(y_{0j}|X = x)\}_{j=1,...,J}$. We can write

$$\tau^L_x(\rho^L) = \phi^L \left( \{P_{ex}(y_{0j}|X = x)\}_{j=1,...,J}, \{P_{ex}(y_{1k}|X = x)\}_{k=1,...,K}, \{P_{tr}(y_{0j}|X = x)\}_{j=1,...,J}; \rho^L \right)$$

(1.17)

where $\phi^L : \Delta(\mathcal{Y}_0) \times \Delta(\mathcal{Y}_1) \times \Delta(\mathcal{Y}_0) \rightarrow \mathbb{R}$ and $\Delta(Z)$ denotes the unit simplex on an arbitrary finite set $Z$. We can similarly write

$$\tau^U_x(\rho^L) = \phi^U \left( \{P_{ex}(y_{0j}|X = x)\}_{j=1,...,J}, \{P_{ex}(y_{1k}|X = x)\}_{k=1,...,K}, \{P_{tr}(y_{0j}|X = x)\}_{j=1,...,J}; \rho^L \right)$$

(1.18)

where $\phi^U : \Delta(\mathcal{Y}_0) \times \Delta(\mathcal{Y}_1) \times \Delta(\mathcal{Y}_0) \rightarrow \mathbb{R}$. In terms of $\phi^L(\cdot)$ and $\phi^U(\cdot)$, the bounds on the unconditional $ATE^a (\tau(\rho^L))$ with $\rho^L$ specified are as follows.

$$\tau(\rho^L) \in [\tau^L(\rho^L), \tau^U(\rho^L)]$$

$$\left[ \sum_{x \in \mathcal{X}} \phi^L \left( \{P_{ex}(y_{0j}|X = x)\}_{j=1,...,J}, \{P_{ex}(y_{1k}|X = x)\}_{k=1,...,K}, \{P_{tr}(y_{0j}|X = x)\}_{j=1,...,J}; \rho^L \right) P^a(x), \right.$$  

$$\left. \sum_{x \in \mathcal{X}} \phi^U \left( \{P_{ex}(y_{0j}|X = x)\}_{j=1,...,J}, \{P_{ex}(y_{1k}|X = x)\}_{k=1,...,K}, \{P_{tr}(y_{0j}|X = x)\}_{j=1,...,J}; \rho^L \right) P^a(x) \right]$$
The bounds can be estimated by replacing population objects with their sample counterparts, denoted with hats.

\[
[\hat{\tau}_L(\rho^L), \hat{\tau}_U(\rho^L)] = \\
\left[ \sum_{x \in X} \phi_L\left( \left\{ \hat{P}^e(y_{0j}|X = x) \right\}_{j=1,\ldots,J}, \left\{ \hat{P}^e(y_{1k}|X = x) \right\}_{k=1,\ldots,K}, \left\{ \hat{P}^a(y_{0j}|X = x) \right\}_{j=1,\ldots,J}; \rho^L \right) \hat{P}^a(x), \\
\sum_{x \in X} \phi_U\left( \left\{ \hat{P}^e(y_{0j}|X = x) \right\}_{j=1,\ldots,J}, \left\{ \hat{P}^e(y_{1k}|X = x) \right\}_{k=1,\ldots,K}, \left\{ \hat{P}^a(y_{0j}|X = x) \right\}_{j=1,\ldots,J}; \rho^L \right) \hat{P}^a(x) \right].
\]

1.4.3 Inference

Imbens and Manski (2004) provide confidence intervals with a fixed asymptotic coverage probability of containing the true value of a partially-identified parameter under the high-level assumption that the joint distribution of the bounds on the parameter is bivariate Gaussian. These could in principle be used to compute confidence intervals covering the true value of the average causal effect in context \(a\), conditional on a specific value for \(\rho^L\), with fixed probability. However, the asymptotic distribution of the bounds is not available in closed form, so I compute them using the bootstrap. The distribution of the bounds in bootstrap samples will be asymptotically normal, satisfying the assumption in Imbens and Manski (2004), under the following assumption.

**Assumption 7.** (i) Sampling. \((Y_i, T_i)\) for \(i = 1, \ldots, N^e\) in population \(e\) are i.i.d. conditional on \(X_i = x\). \((Y_i, T_i)\) for \(i = 1, \ldots, N^a\), in population \(a\) are i.i.d, where \(T_i = 0\) \(\forall i\) conditional on \(X_i = x\). (ii) For each \(x\) in \(X\), there exists a neighborhood of \(V_x\) of \(\left\{ P^e(y_{0j}|X = x) \right\}_{j=1,\ldots,J}, \left\{ P^e(y_{1k}|X = x) \right\}_{k=1,\ldots,K}, \left\{ P^a(y_{0j}|X = x) \right\}_{j=1,\ldots,J} \) such that

\[
\phi_L\left( \left\{ P^e(y_{0j}|X = x) \right\}_{j=1,\ldots,J}, \left\{ P^e(y_{1k}|X = x) \right\}_{k=1,\ldots,K}, \left\{ P^a(y_{0j}|X = x) \right\}_{j=1,\ldots,J}; \rho^L \right)
\]

and

\[
\phi_U\left( \left\{ P^e(y_{0j}|X = x) \right\}_{j=1,\ldots,J}, \left\{ P^e(y_{1k}|X = x) \right\}_{k=1,\ldots,K}, \left\{ P^a(y_{0j}|X = x) \right\}_{j=1,\ldots,J}; \rho^L \right)
\]
(defined in equations 1.17 and 1.18) are differentiable on $V_x$ for all $\rho^L$ in $[0, 1]$.

**Proposition 3.** Suppose Assumptions 1, 2, 3, 5, 6$^{11}$ hold. Let $\mathcal{P}$ be the set of distributions for which Assumption 7 holds. Then, $\lim_{N \to \infty} \inf_{P \in \mathcal{P}, \tau(\rho^L) \in [\tau_L(\rho^L), \tau_U(\rho^L)]} P(\tau(\rho^L) \in CI_N(\rho^L)) \geq 1 - \alpha$.

1.5 Transfers to Mexican microenterprises

McKenzie and Woodruff (2008) (henceforth MW) document the results of an experiment they carried out in 2006 (baseline Oct. 2005) in Leon, Mexico. The experiment was intended to investigate the returns to measured profits of loosening credit constraints for small scale male microentrepreneurs by giving the microentrepreneurs transfers. The authors collected data over the course of five quarterly waves, including the baseline. A treated group of entrepreneurs was randomly assigned to receive a transfer and, conditional on assignment to receiving a transfer, randomly assigned a wave to receive the transfer. The transfers were valued at 1,500 pesos (about $140). Half the transfers were randomly determined to be in-kind, which meant that a member of the survey team accompanied the entrepreneur to purchase equipment or inputs of his choice.

To ensure that the transfers be significant relative to each firm’s scale of operation, the authors restricted their initial sample to entrepreneurs with a capital stock valued at less than 10,000 pesos and no paid employees. Entrepreneurs had to be working full-time on their firm (35 or more hours per week). They further restricted the sample to entrepreneurs working in retail between the ages of 22 and 55. In baseline specifications, the authors find that the transfers increase average monthly profits by about 40% of the transfer.

I explore the extent to which we can generalize this striking finding to microentrepreneurs with the same characteristics in urban Mexico in 2012. The Leon experiment is uniquely suited to this exercise because the questionnaire used in the experiment was based on the national microenterprise survey: Encuesta Nacional de Micronegocios (ENAMIN). This en-

---

$^{11}$Assumption 6 implies Assumption 4.
sures that variables are measured in approximately the same way, which has been shown to be important when using information from one dataset to learn about counterfactual potential outcomes in another - in this case treated outcomes (see e.g. Heckman, Ichimura, and Todd (1997); Diaz and Handa (2006)). I exclude entrepreneurs from the 2012 ENAMIN using the same criteria as MW, additionally requiring that the entrepreneurs be working in urban areas since ENAMIN also captures entrepreneurs in rural area. Since sample selection already chooses a restricted set of individuals, I do not condition on any covariates in the analysis.

I trim profit reports of more than 20,000 pesos in both samples. This trimming keeps slightly more observations than MW who exploit the panel structure and base their trimming procedure on percentile changes in reported profits. Since ENAMIN is a cross-section, I cannot implement a similar procedure and therefore choose a specific value for trimming. Results are robust to choosing different values for trimming. After implementing the trimming, I am left with 903 observations from the ENAMIN sample and 207 unique microentrepreneurs from the experiment.

Figure 1.4 (this and subsequent figures are collected at the end of the paper) shows the outcome distributions in ENAMIN and the control group from the experiment, which provide one key ingredient for the bounds. Since heaping is a substantial issue in reported profits, particularly in ENAMIN, I first smooth profits using a kernel density estimator with a Gaussian kernel and a bandwidth of 750 pesos before discretizing to 500 peso (about $50) bins. Figure 1.4 shows that the experimental control group and the ENAMIN sample have similar outcome distributions, although the ENAMIN sample has substantially more very low profit realizations.

We now explore implications of the differences in the distributions of untreated profits for what we can learn about the average return to cash transfers in urban Mexico in 2012 on the basis of the findings in MW. Figure 1.5 shows bounds (in black) on the average monthly return to providing cash transfers as a function of the minimum rank correlation between
untreated and treated outcomes allowed, $\rho^L$. The bounds shrink to a point when the rank correlation between profits with and without transfers is the maximum possible. Imbens and Manski (2004) 95% confidence regions (in translucent gray) are computed using 100 bootstrap replications for each $\rho^L$, clustering at the firm level for the experiment\textsuperscript{12}. The information in the plot is repeated in table 1.8.

We can draw two conclusions from the results. First, the overall similarity of the control outcome distributions yield narrow bounds on the average return to transfers for male microentrepreneurs in urban Mexico in 2012 for a wide range of possible dependence between outcomes with and without cash transfers. And, second, the experimental sample size is sufficiently small that the 95% confidence interval includes a zero effect on monthly profits at all levels of dependence. We cannot reject a zero effect because the confidence interval around the bounds takes into account three sources of uncertainty: 1) the small sample size of the experiment (207 entrepreneurs), 2) the fact that our information on the distribution of control outcomes in urban Mexico in 2012 also comes from a finite sample (903 entrepreneurs) and 3) the difference in the distribution of untreated profits, particularly for low profit reports.

Previous work (discussed in detail in section 1.3.1) suggested taking into account the differences in the distributions of untreated outcomes by testing their equality (Hotz et al. (2005)). The small size of the experimental sample renders us unable to reject equality of the distributions (the p-value from a Kolmogorov-Smirnov-based test is 0.92). Having been unable to reject the equality of the untreated outcome distributions due to the small size of the experimental sample, we would predict the average profits for male microentrepreneurs in urban Mexico in 2012 to be equal to the average profits for the treated group measured in the experiment, with the same confidence interval as in the experiment. The confidence interval for the difference in treated and untreated profits would be smaller because the sample from ENAMIN is larger so we would be able to reject a zero effect on transfers, ignoring the existence of differences in the distributions of control outcomes. I am able to

\textsuperscript{12}This requires replacing the individual-level indicator $i$ with a cluster-level indicator $g$ in Assumption 7.
separately quantify the uncertainty due to the difference in the control outcome distributions and the uncertainty due to the small sample in the Leon experiment\footnote{I do not take into account the substantial sample attrition that affected the experiment and is explored in MW. MW conclude that the possibility of differential attrition between the experimental treatment and control groups would not dramatically affect their results. Taking into account the possibility of differential attrition would lead to wider bounds on the average return to the transfers than reported in figure 1.5 and table 1.8.}. Considering that the small sample size of the experiment led MW to be cautious in drawing conclusions from their results in-sample, it seems unintuitive that we should be able to draw stronger conclusions about the returns in all urban Mexico. Of course, we do not know the returns to transfers in urban Mexico in 2012, so we now turn to a setting where we can compare predictions and measured causal effects.

Table 1.8: Bounds on the average return to cash transfers in urban Mexico in 2012 using experimental data from McKenzie and Woodruff (2008)

<table>
<thead>
<tr>
<th>Rank correlation</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ATE^a$ lower bound</td>
<td>0.008</td>
<td>0.020</td>
<td>0.034</td>
<td>0.052</td>
<td>0.077</td>
<td>0.222</td>
</tr>
<tr>
<td>$ATE^a$ upper bound</td>
<td>0.436</td>
<td>0.427</td>
<td>0.416</td>
<td>0.392</td>
<td>0.354</td>
<td>0.222</td>
</tr>
<tr>
<td>95% Imbens and Manski (2004) confidence interval lower bound</td>
<td>-0.264</td>
<td>-0.247</td>
<td>-0.245</td>
<td>-0.224</td>
<td>-0.174</td>
<td>-0.125</td>
</tr>
<tr>
<td>95% Imbens and Manski (2004) confidence interval upper bound</td>
<td>0.726</td>
<td>0.723</td>
<td>0.693</td>
<td>0.705</td>
<td>0.638</td>
<td>0.569</td>
</tr>
</tbody>
</table>

1.6 Remedial education in India

Banerjee et al. (2007) (henceforth BCDL) evaluated a remedial education program implemented by the same NGO, Pratham, in two Indian cities: Mumbai and Vadodara. Under the program, Pratham provides government schools with a teacher to work with 15-20 students in the third and fourth grade who have been identified as falling behind. The teacher works with these students for about half the school day.

BCDL carried out the experimental evaluations in Mumbai and Vadodara over the course of three years, from 2001 to 2003. The last year was primarily used to investigate the persistence of effects of the program on learning, so I focus on the first two. In Mumbai,
the experiment was carried out only among third graders in the first year of the evaluation, while in the second year there were compliance issues, with only two-thirds of Mumbai schools agreeing to participate. In Vadodara, both grade levels were represented in each of the first two years but during the first year communal riots disturbed part of the school year. Because of the compliance issues in Mumbai year 2 and the shorter duration of the program in Vadodara year 1, it is harder to interpret the programs being evaluated in the two cities as actually being the same in these periods. Therefore, I consider the Mumbai population as made up of third graders surveyed during the first year of the experiment and the Vadodara population as third graders surveyed in the second year of the experiment.

The researchers administered different achievement tests for both math and verbal skills in the two samples, which poses a challenge in applying the bounds proposed here or existing extrapolation methods in this dataset. Along with different questions, the two tests featured different numbers of questions as well, with 30 questions on the Mumbai test and 50 on the Vadodara test. As an alternative to using the raw test scores, I take advantage of the fact that the test scores were mapped to the students’ grade level competency. Grade level competency measures whether the student successfully answered questions showing mastery of the subjects taught in each grade. This measure of achievement is used in the Annual Status of Education Report, also affiliated with Pratham, to compare achievement across Indian states. One final complication is that students may not achieve all competencies below their maximum competency. For simplicity, I consider the maximum competency as the outcome of interest.

With the exception of the test score and competency at baseline, relatively little data on students are available consistently across the two samples. Tables 1.9 and 1.10 show summary statistics for the maximum competency at baseline in the two populations as well as students’ class size and gender. The populations are relatively balanced on gender, while Mumbai classes are notably larger than those in Vadodara. BCDL find no evidence of treatment effect heterogeneity on either of these characteristics, so I ignore them and focus
on the maximum competency level at baseline.

Table 1.9: Vadodara

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test: maximum math competency</td>
<td>0.276</td>
<td>0.361</td>
</tr>
<tr>
<td>Pre-test: maximum verbal competency</td>
<td>.613</td>
<td>.678</td>
</tr>
<tr>
<td>Male</td>
<td>0.497</td>
<td>0.5</td>
</tr>
<tr>
<td>Number of students in class</td>
<td>62.109</td>
<td>26.516</td>
</tr>
<tr>
<td>N</td>
<td>5819</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.10: Mumbai

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test: maximum math competency</td>
<td>0.543</td>
<td>0.641</td>
</tr>
<tr>
<td>Pre-test: maximum verbal competency</td>
<td>1.991</td>
<td>1.113</td>
</tr>
<tr>
<td>Male</td>
<td>0.473</td>
<td>0.499</td>
</tr>
<tr>
<td>Number of students in class</td>
<td>89.506</td>
<td>40.233</td>
</tr>
<tr>
<td>N</td>
<td>4429</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.11 shows the difference across cities in the unconditional effect of the remedial education program on the average maximum math grade level competency. The first line shows the average effect in Vadodara. In Vadodara, the program raised students’ maximum grade level competency in math by 0.16 grade levels. The third line shows the unconditional bias in using the average treatment effect in Mumbai as an estimator for the average treatment effect in Vadodara. The average effect in Mumbai is estimated at 0.059 grade levels, 0.103 less than the Vadodara $ATE$ and the difference is significant.

Table 1.12 shows the equivalent results for the maximum verbal competency. Here the average effect again differs across cities, but the difference is not significant. For this reason, I focus on examining the ability of extrapolation methods to account for the significant difference in the effect of the remedial education program on the maximum grade-level competency in math across cities.
Table 1.11: City-specific average effects on maximum math grade level competency

<table>
<thead>
<tr>
<th></th>
<th>Post-test: maximum math competency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mumbai</td>
<td>0.020 (0.026)</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.162*** (0.024)</td>
</tr>
<tr>
<td>Treatment*Mumbai</td>
<td>−0.103*** (0.036)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.709*** (0.017)</td>
</tr>
<tr>
<td>Observations</td>
<td>10,248</td>
</tr>
<tr>
<td>R²</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Notes: ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 1.12: City-specific average effects on maximum verbal grade level competency

<table>
<thead>
<tr>
<th></th>
<th>Post-test: maximum verbal competency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mumbai</td>
<td>0.947*** (0.028)</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.071*** (0.026)</td>
</tr>
<tr>
<td>Treatment*Mumbai</td>
<td>0.049 (0.039)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.230*** (0.018)</td>
</tr>
<tr>
<td>Observations</td>
<td>10,248</td>
</tr>
<tr>
<td>R²</td>
<td>0.199</td>
</tr>
</tbody>
</table>

Notes: ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.
1.6.1 Using Mumbai to predict Vadodara

We now move to trying to use the results from Mumbai and the Vadodara control group to predict the average outcome level in the Vadodara treatment group. We can think of this as the policy-making exercise of using the results from Mumbai year 1 to try to infer the average treatment effect on math test scores of implementing the remedial education program among Vadodara third graders in the following year. As in previous work, I find that the average treatment effect in Vadodara predicted using by reweighting Mumbai average treatment effects conditional on grade level competency on entering third grade is biased, with the bias equal to half the Vadodara average treatment effect (bias of 0.081 grade level competencies with a standard error of 0.033).

The first step in the extrapolation methodology developed in Hotz et al. (2005) is testing equality of the distributions of maximum grade level competency in math for the two control groups. Visual inspection of the conditional distributions in figure 1.6 shows that they are quite different. Table 1.13 confirms this impression statistically. The table shows the distributions of grade level competency in math on leaving third grade in the control groups in both cities in the BCDL experiments conditional on their grade level competency in math on entering third grade. The last column of the panel labeled Vadodara shows the p-value associated with a $\chi^2$ test of equality of each conditional distributions representing a grade level competency on entering third grade. The test rejects at the 5% level for all values grade level competencies on entering third grade. Following the Hotz et al. (2005) methodology, we would conclude that we cannot learn anything about the causal effect in Mumbai from the causal effect in Vadodara: the students in the two cities are too different.

Turning to the bounds developed in this paper, figure 1.7 plots bounds on the predicted values of the average effect of the remedial education program on maximum math grade level competencies in Vadodara as a function of the minimum rank correlation, $\rho^L$, between outcomes with and without the remedial education for individuals with the same grade level competency on entering third grade. The bounds are plotted in black, while the translucent
Table 1.13: Controls - P(competency on exiting grade 3 | competency on entering grade 3)

<table>
<thead>
<tr>
<th></th>
<th>Pre-competency</th>
<th>Post-competency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mumbai</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 0.73 0.17 0.07 0.03 1246</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 0.39 0.28 0.19 0.13 468</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 0.28 0.20 0.28 0.23 254</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 0.12 0.22 0.14 0.53 51</td>
<td></td>
</tr>
<tr>
<td>Vadodara</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 0.52 0.38 0.08 0.02 2094</td>
<td>&lt;2.2e-16</td>
</tr>
<tr>
<td></td>
<td>1 0.28 0.50 0.15 0.07 647</td>
<td>3.834e-12</td>
</tr>
<tr>
<td></td>
<td>2 0.18 0.39 0.22 0.22 51</td>
<td>0.03195</td>
</tr>
</tbody>
</table>

gray region represents a 95% Imbens and Manski (2004) confidence interval, based on 100 bootstrap replications. Table 1.14 replicate the key results from figure 1.7 in tabular form.

A notable feature of the bounds is that they widen quickly with only small deviations from the maximum possible rank correlation. This is due to the fact that the conditional distributions of control outcomes differ substantially between Mumbai and Vadodara, as we saw in figure 1.6 and table 1.13. A zero average treatment effect in Vadodara can only be rejected using the Mumbai results if $\rho > .925$. The light gray line plots the measured average effect of remedial education on maximum grade level competency in math from table 1.11, while the dashed lines show the 95% confidence interval. In terms of the prediction of the average increase in maximum grade level competency in math, we see that though the point estimate with maximum rank correlation under-predicts the sample mean of the maximum competency on leaving 3rd grade in Vadodara, the two estimates are fairly close and the difference between the two is not statistically different from zero. Simply allowing for 1-dimensional heterogeneity goes a long way toward accurately predicting the Vadodara results.

---

14 Additional replications, to be added, would smooth out the irregularities in the confidence intervals.
Table 1.14: Bounds on the change in average grade level competency in Vadodara using experimental results from Mumbai and untreated outcomes from Vadodara

<table>
<thead>
<tr>
<th>Rank correlation</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>0.925</th>
<th>0.95</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ATE^d$ lower bound</td>
<td>-0.145</td>
<td>-0.107</td>
<td>-0.062</td>
<td>-0.017</td>
<td>0.030</td>
<td>0.042</td>
<td>0.058</td>
<td>0.109</td>
</tr>
<tr>
<td>$ATE^d$ upper bound</td>
<td>0.366</td>
<td>0.364</td>
<td>0.345</td>
<td>0.321</td>
<td>0.287</td>
<td>0.278</td>
<td>0.268</td>
<td>0.109</td>
</tr>
<tr>
<td>95% Imbens and Manski (2004) confidence interval lower bound</td>
<td>-0.193</td>
<td>-0.155</td>
<td>-0.107</td>
<td>-0.058</td>
<td>-0.017</td>
<td>-0.011</td>
<td>0.007</td>
<td>0.039</td>
</tr>
<tr>
<td>95% Imbens and Manski (2004) confidence interval upper bound</td>
<td>0.427</td>
<td>0.431</td>
<td>0.410</td>
<td>0.395</td>
<td>0.353</td>
<td>0.342</td>
<td>0.316</td>
<td>0.179</td>
</tr>
</tbody>
</table>

1.6.2 Using Vadodara to predict Mumbai

Figure 1.8 and table 1.15 show the results of using Vadodara to predict Mumbai. The results show the difficulty that arises when there are some observed characteristics in the region for which we want to predict the average causal effect that are not present in the experimental results (Assumption 1). As shown in table 1.13, Vadodara does not include any students who enter grade three with a third grade level competency while Mumbai includes a small fraction of such students. The results in figure 1.8 assign these students the lower bound of the support of the maximum grade level competency (0) when computing the lower bound on the average causal effect in Mumbai and the upper bound of the support of the competency (3) when computing the upper bound. As a result, we can only reject zero average treatment effect in Mumbai using the Vadodara results under an even smaller range of rank correlations between outcomes with and without the remedial education program ($< .975$). Setting the mean treated outcome at zero competency for students with a competency of three on entering third grade is almost surely too severe even when computing the lower bound on the average treatment effect in Mumbai. I am currently exploring alternatives such as assuming that the distribution of treated outcomes for this group first-order stochastically dominates the distribution for students entering third grade with a grade-level competency of two.
Table 1.15: Bounds on change in average grade level competency in Mumbai using experimental results from Vadodara and untreated outcomes from Mumbai

<table>
<thead>
<tr>
<th>Rank correlation</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>0.925</th>
<th>0.975</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATE&lt;sup&gt;a&lt;/sup&gt; lower bound</td>
<td>-0.063</td>
<td>-0.051</td>
<td>-0.039</td>
<td>-0.020</td>
<td>0.006</td>
<td>0.019</td>
<td>0.049</td>
<td>0.089</td>
</tr>
<tr>
<td>ATE&lt;sup&gt;a&lt;/sup&gt; upper bound</td>
<td>0.370</td>
<td>0.338</td>
<td>0.304</td>
<td>0.265</td>
<td>0.223</td>
<td>0.209</td>
<td>0.180</td>
<td>0.165</td>
</tr>
<tr>
<td>95% Imbens and Manski (2004) confidence interval lower bound</td>
<td>-0.120</td>
<td>-0.098</td>
<td>-0.087</td>
<td>-0.067</td>
<td>-0.036</td>
<td>-0.028</td>
<td>-0.002</td>
<td>0.027</td>
</tr>
<tr>
<td>95% Imbens and Manski (2004) confidence interval upper bound</td>
<td>0.421</td>
<td>0.383</td>
<td>0.352</td>
<td>0.309</td>
<td>0.260</td>
<td>0.253</td>
<td>0.227</td>
<td>0.226</td>
</tr>
</tbody>
</table>

1.7 Conclusions

The methods derived in this paper offer researchers a formal and tractable way of assessing the extent to which experimental results generalize to contexts outside the original study. More broadly, this paper provides a first step away from seeing generalizability as an all-or-nothing proposition. I empirically demonstrated the problems with testing for unobserved differences across contexts among individuals with the same observed characteristics and taking the test results as sanctioning or prohibiting extrapolation to a particular context. In the Mexican microenterprise example, the test grants the researcher license to extrapolate broadly based on a very small experiment. In the remedial education example, testing leads us to conclude that the experimental results from one site teach us nothing about causal effects in the other.

In contrast, the bounds developed here quantify our uncertainty about effects in the context of interest due to unobserved differences across the contexts. In the Mexican microenterprise case, the narrow bounds showed us that the Leon 2006 results appear largely representative of effects for similar entrepreneurs in urban Mexico in 2012. However, the small size of the experiment should make us cautious about extrapolating, which shows up in the wide confidence intervals around the bounds. In the remedial education example, the bounds showed that under assumptions of strong dependence between a student’s grade-level competency with and without a remedial education teacher assigned to her school, we can learn quite a bit about about the effect of remedial education in one city using
results from the other. The experimental effects in the two cities are consistent with the assumption of strong dependence.

Since experimental sites must often be chosen for reasons of cost or convenience, the methods proposed in this paper have broad applicability. In addition to assessing what can be learned about causal effects in new contexts on the basis of existing experimental results, they may be used when researchers have some leeway to select experimental sites. Based on an assumed distribution for treated outcomes, a researcher could estimate prospective bounds on causal effects in contexts of interest with different possible experimental sites\textsuperscript{15}.

\textsuperscript{15}This procedure would be akin to the power calculations commonly undertaken in determining the necessary sample size for an experiment, but for identification.
1.8 Definition of copula

A copula function $C : [0, 1]^2 \rightarrow [0, 1]$ satisfies:

1. Boundary conditions:
   
   (a) $C(0, v) = C(u, 0) = 0 \forall \, u, v \in [0, 1]$
   
   (b) $C(u, 1) = u$ and $C(1, v) = v \forall \, u, v \in [0, 1]$

2. Monotonicity condition:

   3. $C(u, v) + C(u', v') - C(u, v') - C(u', v) \forall \, u, v, u', v' \text{ s.t. } u \leq u', v \leq v'$

1.9 Proof of equivalence of bounds in proposition 2 and linear programming representation

Proof. By the definition of a copula, any $C \in \mathcal{C}$ defines a joint distribution $F_{Y_0, Y_1}(y_0, y_1) = C(F_{Y_0}(y_0), F_{Y_1}(y_1))$ satisfying $F_{Y_0, Y_1}(y_0, \infty) = F_{Y_0}^{c}(y_0)$ and $F_{Y_0, Y_1}(\infty, y_1) = F_{Y_1}^{c}(y_1)$. This is exactly what is required by constraints 1.14 and 1.15. The equivalence of the bounds in Proposition 2 and the full linear programming representation follows immediately from the definition of $\rho(V, W)$ and constraint 1.16. \qed
Figure 1.4: Distribution of profits: McKenzie and Woodruff (2008) control group and 2012 ENAMIN

Note: distribution of profits in 2005 pesos for control firms in McKenzie and Woodruff (2008) and the 2012 Encuesta Nacional de Micronegocios, using the same sample selection criteria as in McKenzie and Woodruff (2008). The distribution of profits is smoothed using a kernel density estimator with a Gaussian kernel and a bandwidth of 750 pesos before discretizing to 500 peso bins.
Figure 1.5: Bounds on the average return to cash transfers in urban Mexico in 2012 using experimental data from McKenzie and Woodruff (2008).

Note: For each lower bound on the dependence between profits with and without cash transfers, $\rho_L$, the solid black region shows the bounds on the return to cash transfers in urban Mexico in 2012 for microentrepreneurs selected according to the criteria in McKenzie and Woodruff (2008), $ATE^a$, derived from the experimental results in McKenzie and Woodruff (2008). The translucent gray region is a Imbens and Manski (2004) 95% confidence interval for $ATE^a$, based on 100 bootstrap replications, clustered at the firm level.
Figure 1.6: controls - grade level competency on exiting 3rd grade conditional on grade level competency on entering 3rd grade
Figure 1.7: Bounds on the change in average grade level competency in Vadodara using experimental results from Mumbai and untreated outcomes from Vadodara

Note: For each lower bound on the dependence between a student’s maximum grade level competency with and without a remedial education teacher assigned to her school, $\rho^L$, the solid black region shows the bounds on the average gain in maximum grade level competency in Vadodara, $ATE^a$, derived from the experimental results in Mumbai. The translucent gray region is a Imbens and Manski (2004) 95% confidence interval for $ATE^a$, based on 100 bootstrap replications. The light gray line shows the point estimate of the actual average gain in Vadodara, using the experimental results. The dashed line shows a 95% confidence interval for the actual average gain.
Figure 1.8: Bounds on change in average grade level competency in Mumbai using experimental results from Vadodara and untreated outcomes from Mumbai

Note: For each lower bound on the dependence between a student’s maximum grade level competency with and without a remedial education teacher assigned to her school, \( \rho^L \), the solid black region shows the bounds on the average gain in maximum grade level competency in Mumbai, \( ATE^a \), derived from the experimental results in Vadodara. The translucent gray region is a Imbens and Manski (2004) 95% confidence interval for \( ATE^a \), based on 100 bootstrap replications. The light gray line shows the point estimate of the actual average gain in Mumbai, using the experimental results. The dashed lines show the 95% confidence interval on the actual average gain.
Chapter 2

Indian Labor Regulations and the Cost of Corruption: Evidence from the Firm Size Distribution

2.1 Introduction

India’s labor and industrial regulations have been blamed for many of the country’s ills, including low levels of aggregate productivity, slow growth of productivity, and lackluster job creation in the formal sector\(^1\) (e.g. Hsieh and Klenow (2009); Kochhar, Kumar, Rajan, Subramanian, and Tokatlidis (2006); Besley and Burgess (2004); Hasan and Jandoc (2012)). Of particular note is Hsieh and Klenow (2009)’s landmark study, in which the authors argue that aggregate total factor productivity in India could be 40%-60% higher if not for significant misallocation of resources across firms. They go on to suggest that India’s labor regulations may be to blame for the observed misallocation, although they leave the job of fully corroborating this link to others. In fact, the view that labor regulations are of primary importance is not universally held. Many argue that the laws as written are rarely enforced so that, in practice, firms are effectively unconstrained.\(^2\) Others argue that the existing evidence on the detrimental impact of labor regulations is flawed (Bhattacharjea (2006, 2009)). Still others point out that the vast majority of regulations have gone unstudied while nearly all of the attention from economists and the press has focused on a single regulation (Chapter VB of the Industrial Disputes Act) - one that is not likely to constrain any but the very largest firms (Bardhan (2014)).\(^3\)

It is the goal of this paper to address the above aspects of this debate while avoiding

\(^1\)Here and elsewhere in the paper, the formal sector refers to business enterprises that are registered with some branch of the government.

\(^2\)For instance, in a recent paper, Chaurey (2015) provides evidence that firms seem to hire contract workers as a way of avoiding certain regulations.

\(^3\)Chapter VB of the Industrial Disputes Act (IDA) stipulates that firms in the industrial sector with 100 or more workers must obtain permission from the relevant governmental authority before laying off workers. Bardhan (2014) points out that 92 percent of firms in the garment sector have less than 8 workers.
some of the criticisms that have been leveled at previous work. In particular, we use a novel methodology and a uniquely well-suited dataset to study the behavior of firms in response to regulatory thresholds in order to determine whether and to what extent firms are in fact constrained by regulations.⁴ We proceed in the following steps. First, we generate the establishment-size distribution using data from the Economic Census of India (EC), which, importantly, aims to be a complete enumeration of all non-farm business units, regardless of size or status (formal or informal). From the distribution, we observe that it closely follows a power law, except for a discontinuous and proportional decrease in the density of establishments with 10 or more workers. This is precisely the threshold at which a multitude of regulations become legally binding, so we take this observation as evidence that firms with 10 or more workers do seem to be constrained in size by certain regulations - although these are not the same regulations that most others have focused on. We then develop a model of firm size choice under regulatory thresholds which is based on Garicano, Lelarge, and Van Reenen (2013) (henceforth GLV), but augmented to explicitly allow for the possibility of misreporting.⁵ We model the regulations as causing an increase in the unit labor costs of those firms that report having exceeded the 10 worker threshold⁶, and then use the observed distortion in the size distribution to estimate these costs. Under our primary estimation method at the All-India level, we find that firms behave as if operating at or above the 10-worker threshold entailed a 35% increase in their per-worker costs.

Our next step is to document substantial heterogeneity in the size of our estimated costs along several dimensions including state, industry and ownership type. For example, we find that the state with the highest estimated regulatory costs is Bihar and that privately-

⁴Note: for expositional purposes we occasionally refer to “firms”, although it would be more correct to refer to “factories” or “establishments”, since all of the data and most of the regulations are at the factory/establishment level rather than the firm level. Regardless, the distinction is almost moot: nearly all Indian firms are single factory/establishment firms.

⁵Misreporting was a lesser concern in GLV’s original setting, as they had access to administrative data. In contrast, the data in the Economic Census are self-reported, which makes the threat of deliberate misreporting more significant in our case.

⁶This is the only way to generate a proportional decrease in the theoretical density, at least in a static model.
owned establishments have the highest costs, while government-owned establishments have the lowest. Exploring this variation further, we find that our estimated costs turn out to be correlated with some previous state-level measures of labor regulation reforms (in particular, certain measures from Dougherty (2009)), though not with others (for example, the Besley-Burgess measure from Aghion, Burgess, Redding, and Zilibotti (2008)). Moreover, we find strong and robust correlations between our estimated costs and two quite distinct measures of corruption, even after controlling for a number of factors including state GDP per capita. As further support for our state-level results, we show that industries with greater “regulatory dependence” have higher estimated costs, but only when they are located in more corrupt states. We take these correlations to be suggestive of the fact that the true cost of the regulations may have more to do with bureaucracy and corruption, rather than the content of labor and industrial regulations themselves.

Finally, we turn to a brief discussion of the possible dynamic consequences of the costs we estimate. We show that, while higher costs are associated with slower growth in employment and productivity in the registered manufacturing sector, this association is more muted - or even in the opposite direction - in the unregistered manufacturing sector, where the regulations are less salient. This suggests that the costs we estimate may play a role in the “informalization” of the Indian economy, by pushing workers from the formal to the informal sector.

This paper aims to contribute to at least three important strands of literature. The first, which we have already mentioned, is the literature on misallocation of resources and total factor productivity (TFP), as exemplified by Hsieh and Klenow (2009). Our contribution

\footnote{This may reflect the fact that the Besley-Burgess measures focus on the IDA, while the regulations we study are entirely different. On the other hand, if the Besley-Burgess measures are meant to capture the general effect of labor laws at the state level, one might expect the two measures to be correlated.}

\footnote{These corruption measures include a subjective, perceptions-based measured of corruption from Transparency International and a measure of the percentage of electricity that is lost in transmission and distribution as reported by the Reserve Bank of India (this latter measure has been used as a proxy for government corruption and ineffectiveness in, for example, Kochhar et al, 2006).}

\footnote{We measure “regulatory dependence” by taking the industry average of the number of inspector visits among Indian firms in the 2005 World Bank Enterprise Surveys.}
is to provide direct evidence that at least some of the misallocation of resources across firms in India is tied to regulations or the enforcement thereof.\textsuperscript{10} In particular, we show that size-based regulations (or at least the ways in which they are enforced) lead firms to fall short of their optimal scale, thus distorting the allocation of labor among firms in the economy and, likely, lowering TFP.

Another strand of literature to which we aim to contribute relates to corruption in the enforcement of government policies. Most previous studies (eg: Besley and McLaren (1993); Mookherjee and Png (1995)) have modeled such corruption as collusion between inspectors and firms or citizens: corrupt inspectors allow firms to avoid the de jure costs of abiding by regulations in exchange for bribes. Hence, in these frameworks, corruption lowers the costs associated with regulations. However, our results suggest that the costs associated with size-based regulations are higher in more corrupt environments, and are thus more in line with an alternative framework in which corruption takes the form of extortion between inspectors and firms (i.e.: corrupt inspectors take advantage of bureaucratic regulations in order to extract higher rents from firms in the form of harassment bribes).\textsuperscript{11} We present a theoretical model as well as anecdotal evidence from “ipaidabribe.com” to support this interpretation and view the support we provide for this alternative conception of corruption to be another contribution of the paper.

Lastly, this paper is also clearly related to the large literature that more generally investigates the impact of Indian labor regulations on economic outcomes. The literature dates back to at least Fallon and Lucas (1993), but the more recent proliferation seems to be due to the work of Besley and Burgess (2004). In that paper, the authors first interpret state-level amendments to the Industrial Disputes Act (IDA) as either “pro-worker” or “pro-employer” and then aim to show that Indian states that amended the IDA in a “pro-worker” direction experienced slower growth in output, employment, investment and productivity.

\textsuperscript{10}In future work we hope to determine what portion of the TFP loss from misallocation estimated by Hsieh and Klenow (2009) can be attributed to the regulations we study.

\textsuperscript{11}This finding echoes Novosad and Asher (2014), in which it is argued that regulations can provide a means through which politicians can impose costs on businesses.
in registered manufacturing. The paper, though extremely influential, has been criticized by Bhattacharjea (2006) and Bhattacharjea (2009) on a number of grounds. One of Bhattacharjea’s major criticisms is that Besley and Burgess’s interpretations of amendments as “pro” or “anti-worker” are subjective and debatable (ie: different people might read and code them in a different way). This criticism affects most of the subsequent academic work on this topic, since most papers use the Besley-Burgess codings, but it is a criticism we are able to sidestep with our methodology. Since our analysis is based only on firm level data and size-thresholds stated explicitly in the laws themselves, it has the advantage of objectivity.

The second contribution we make to this literature is to focus on a set of regulations that have been almost entirely ignored even though they effect a much larger proportion of firms than Chapter VB of the IDA.\(^{12}\) The only other papers of which we are aware that study regulations that kick in at the 10-worker threshold are Dougherty (2009), Dougherty, Frisancho, and Krishna (2014) and Kanbur and Chatterjee (2013). The latter investigates the Factories Act, which applies to all manufacturing firms that use power and have 10 or more workers (or don’t use power and have 20 or more workers), but their focus is to document non-compliance under the act, which we see as complementary to our approach of estimating the costs of the regulations.\(^{13}\) The papers by Dougherty and co-authors employ state-level indices of labor reforms that differ from the Besley-Burgess codes in that they include consideration of non-IDA regulations such as the Factories Act, but they are constructed from surveys of industry experts and, as such, are by and large subject to similar concerns regarding subjectivity.

Another way in which we distinguish ourselves from the previous literature on Indian

\(^{12}\)Chapter VB of the IDA only applies to manufacturing firms with 100 or more workers. In contrast, the regulations we study affect all firms with 10 or more workers and are thus relevant for a much larger share of firms. We have also tried analyzing Chapter VB of the IDA using the same methodology we employ for the regulations with the 10 worker threshold, but find no effects. I.e.: there does not seem to be a proportional decrease in the density of establishments with more than 100 workers. We also fail to observe “bunching” of firms at sizes just below 100, although the presence of rounding may make such bunching impossible to discern even if it exists.

\(^{13}\)Our estimated costs are robust to the possibility of noncompliance.
regulations is that we explore the effect of regulations in all non-farm segments of the Indian economy - not just in registered manufacturing, on which nearly all previous academic studies have focused. A final contribution of the paper is to provide suggestive evidence that improper government enforcement of regulations may play a role in shifting employment from the registered to the unregistered sector.

In the next section (Section 2.2), we provide an overview of the relevant institutional details regarding Indian labor and industrial regulations. Section 2.3 introduces the data and covers some basics about the size distribution of enterprises in India. In Section 2.4 we go over the theoretical model and our corresponding empirical strategy. Section 3.6 provides the main results. In Section 2.6, we interpret the findings, explore the multiple dimensions of variation in our results, and investigate the connection between our estimated costed and corruption. Section 2.7 concludes.

2.2 Institutional Background: Size-Based Regulations in India

In this paper we attempt to investigate the effects of certain size-based industrial and labor regulations in India. These are regulations that only apply to establishments that exceed a certain size, measured either in terms of a firm’s revenue, the amount of fixed capital invested, or the number of workers employed. One of the most significant such thresholds occurs when establishments employ 10 or more workers, after which they must register with the government and meet various workplace safety requirements (under the Factories Act\textsuperscript{14}, for example), pay social security taxes (under the Employees’ State Insurance Act), distribute gratuities (under the Payment of Gratuity Act) and bear a greater administrative burden (under, e.g., the Labor Laws Act).

Not only are the laws numerous, it has been argued that certain components of the laws are antiquated and/or arbitrary. For example we read in the “India Labour Report” that\footnote{Technically the Factories Act applies for 10-plus worker establishments only if they use power. For establishments that do not use power, the Factories Act does not apply until they employ 20 workers.}
“Rules under the Factories Act, framed in 1948\textsuperscript{15}, provide for white washing of factories. Distemper won’t do. Earthen pots filled with water are required. Water coolers won’t suffice. Red-painted buckets filled with sand are required. Fire extinguishers won’t do... And so on” TeamLease Services (2006). Firm owners who choose not to comply with such regulations may face costs if discovered and convicted.\textsuperscript{16}

In addition to - or in lieu of - the explicit costs of complying with the regulations, establishments with 10 or more workers may be subject to implicit costs associated with increased interaction with labor inspectors, et al, who may have the power to extract bribes and tighten (or ease) the administrative burden firms face. Indeed, inspectors in India have a large amount of discretion regarding the enforcement of administrative law. For example, in some cases, the definition of what constitutes a “day” is at the discretion of the inspector, and it is a commonly held view that “[w]hile grave violations are ignored, minor errors become a scope for harassment” (TeamLease Services (2006)).

This kind of behaviour has been referred to as “harassment bribery” (Basu (2011)). Anecdotal evidence of inspectors using the complexity and sheer amount of paperwork as a way to extract bribes is easy to come by. For example, we have included a selection of citizen reports from “ipaidabribe.com” in Appendix 2, which demonstrate just this kind of behaviour.\textsuperscript{17} Interestingly, some of the reports suggest that the size of the bribe paid is a direct linear function of the number of employees - which will be relevant to our estimation procedure later.

As we alluded to earlier, the 10 worker threshold is not the only one relevant; there are other cutoffs at which different regulations become binding. For example, the threshold that seems to have received the most attention, both from academics and the press, is that of 100 workers, at which enterprises in most states become subject to Chapter VB of the Industrial Disputes Act, under which they must be granted government permission to lay

\textsuperscript{15}The Factories Act itself dates to 1948, but the origins of the law go back another 100 years at least, to Britain’s first Factory Acts.

\textsuperscript{16}These costs may include fines and/or prison sentences.

\textsuperscript{17}We thank Andrew Foster for this suggestion.
off workers. There are other cutoffs still,\textsuperscript{18} but in this paper we will focus on estimating some of the costs and effects associated with the regulations that come into force at the 10 worker cutoff. One important limitation of our analysis is that we will not be able to address issues regarding the efficacy of any regulations in promoting worker welfare.

2.3 Data and the Size Distribution in India

2.3.1 Data

The data we rely on to investigate the 10-worker threshold comes from the Economic Census (EC) of India. The EC is meant to be a complete enumeration of all (formal \textit{and} informal) non-farm business establishments\textsuperscript{19} in India at a given time, \textit{regardless of their size}. It is this last clause that makes the EC different from every other data source available and precisely suited to our needs. Although the 2005 dataset contains a large number of observations (almost 42 million), there is not very detailed information collected on each observation. For each establishment in the data, there is only information on a handful of variables including the total number of workers usually working, the number of non-hired workers (such as family members working alongside the owner), the registration status, the 4-digit NIC industry code, the type of ownership (private, government, etc) and the source of funds for the establishment. There is no information on capital, output or profits, and the data is cross-sectional.

The EC has rarely been used in academic papers - possibly because it is cross-sectional, contains a significant amount of measurement error, and only contains information on the handful of variables just enumerated, so that better data sources exist for most purposes.

\textsuperscript{18}For example, firms with 20 or more workers must abide by the Provident Funds Act. Firms with 50 or more workers must comply with Chapter VA of the Industrial Disputes Act, which requires them to provide compensation and notice to employees prior to lay-offs.

\textsuperscript{19}The EC refers to these as “entrepreneurial units” and defines them as any unit “engaged in the production or distribution of goods or services other than for the sole purpose of own consumption.” As is common in the literature, we occasionally refer to them as “firms” even though the unit of observation in the data is actually a factory or an establishment, rather than a firm (i.e.: multiple establishments may belong to the same firm). We do this for expositional purposes and justify our use of this convention with the observation that the proportion of establishments that belong to multi-establishment firms is minute.
The EC is ideal for our purpose, however, since it includes information on employment size and covers the entire universe of establishments. Other more commonly used datasets, such as the CMIE’s Prowess Database, the Annual Survey of Industries (ASI) or the National Sample Survey’s (NSS) Unorganized Manufacturing Surveys cover only certain parts of the distribution and thus cannot be used for our purpose. The ASI, for example, only covers factories in the manufacturing sector that have registered with the government under the Factories Act. However, registration under this Act is only required for establishments with 10 or more workers if the unit uses power (20 or more workers if the factory uses no power). Therefore, the selection of the ASI varies discontinuously at precisely the point of interest. Similar limitations on coverage make the other datasets - other than the EC - unsuitable.

Aside from the Economic Census, we also supplement our analysis with data from a variety of other sources. From the ASI we get employment and productivity in the registered manufacturing sector. We generate those same variables for the unregistered sector with data from the Ministry of Statistics and Programme Implementation (MOSPI) and the Reserve Bank of India (RBI). We get data on state and industry level corruption from a) Transparency International’s “India Corruption Study 2005”, b) the RBI, and c) the World Bank Enterprise Survey for India (2005). Data on State-level regulatory enforcement come from the Indian Labour Year Book.\textsuperscript{20} Other measures of state-level regulations come from Aghion et al. (2008) and Dougherty (2009), while industry-level measures of exposure to trade liberalization come from Ahsan and Mitra (2014).

\subsection{The Size Distribution of Establishments in India}

Figure 2.1 below shows the distribution of establishments by the number of total workers (hired and non-hired workers) for establishments with up to 200 total workers in 2005. Perhaps the most striking feature of figure 2.1 is the extraordinary degree to which the distribution is right-skewed. Indeed, about half of all establishments are single person enterprises,

\textsuperscript{20}We would like to thank Anushree Sinha and Avantika Prabhakar for their considerable and generous help in obtaining these data.
Figure 2.1: Distribution of establishment size for establishments with 1-200 total workers, 2005

while the densities for establishments with 10 or more workers are almost imperceptible. Figure 2.2 shows the drop in density for establishments with 10 or more workers in detail and figure 2.3 shows the full distribution of establishment size frequencies according to a log scale. Each point represents one bar in the earlier histograms.

Two things are most striking about figure 2.3. First, the natural log of the density is a linear function of the natural log of the number of total workers. This implies that the unlogged distribution follows an inverse power law in the number of total workers. This pattern will be important for the analysis that follows but it is not very surprising in and of itself: power law distributions in firm sizes have been documented in many countries (e.g. Axtell (2001) and Hernández-Pérez, Angulo-Brown, and Tun (2006)). The second and more unique feature of the distribution is that there appears to be a level shift downward in the log frequency for establishment sizes greater than or equal to 10. Figure 2.4 shows this effect for establishments with fewer than 100 workers by running an OLS regression of the log density against log firm size and allowing the intercept to vary for firms with 10 or more workers.

---

21The densities for establishments with more than 200 workers are also imperceptible. We have omitted them only for clarity in the figure.
Figure 2.2: Distribution of establishment size for establishments with 5-25 total workers, 2005

Figure 2.3: Distribution of establishment size, 2005, log scale
Figure 2.4: Downward shift at the 10-worker threshold in the distribution of establishment size, 2005, log scale (omitting establishments with more than 100 workers)

workers. To the best of our knowledge, ours is the first paper to document this phenomenon in India.

Also of note from the figures above is that there appears to be a significant amount of non-classical measurement error, seemingly due to rounding of establishment sizes to multiples of 5 and 10. The existence of rounding is not surprising given that the data are self-reported and that respondents are asked to give the “number of persons usually working [over the last year]”. Partially to alleviate concerns that the non-classical measurement error due to rounding might bias our results (and partially for other reasons to be made explicit shortly), we will employ an estimation procedure which first smooths the data non-parametrically.

2.4 Model and Empirical Strategy

2.4.1 Basic Model

To interpret the downward shift from Figure 2.4 in economic terms, we turn to the model in GLV. In their framework, size-based regulations are assumed to increase the unit labor costs of firms that exceed the size threshold, which results in a downshift in part of the
theoretical firm size distribution. From the magnitude of the downshift they observe in the empirical distribution they attempt to estimate the additional labor costs imposed by the regulations.

GLV begin with a distribution of managerial ability \( \alpha \sim \phi(\alpha) \) as the primitive object, following Lucas (1978). As is common in the literature (e.g. Eaton, 2011), they assume that the distribution of managerial ability follows a power law (e.g. \( \phi(\alpha) = c_\alpha \alpha^{-\beta_\alpha} \)). It is this that will generate a power law in the theoretical firm size distribution. A firm with productivity or managerial ability \( \alpha \) faces the following profit-maximization problem:

\[
\pi(\alpha) = \max_n \alpha f(n) - w\tau n
\]

where \( n \) is the number of workers a firm employs, \( f(n) \) is a production function (with \( f'(n) > 0 \) and \( f''(n) < 0 \)), \( w \) is a constant wage paid to all workers, and \( \tau \) is a proportional tax on labor that takes the value 1 if \( n \leq N \) and \( \tau \) if \( n > N \), where \( \tau > 1 \).

From the first order condition on this maximization problem, \( \alpha = \frac{w\tau}{f'(n)} \), one can see that higher productivity establishments/managers will employ more workers, and that firms which cross the threshold \( (N) \) and must therefore pay higher labor costs will hire fewer workers than they would otherwise. This latter feature is built to match the observed “downshift” in the actual firm size distribution to the right of the regulatory threshold.

One can informally characterize the solution as follows: one set of managers with particularly low productivity (below some threshold \( \alpha_1 \)) will be effectively unconstrained. These managers would have chosen to hire fewer than 10 workers whether or not the regulation was present. Another set of managers with slightly higher productivity (between some thresholds \( \alpha_1 \) and \( \alpha_2 \)) would, in the absence of the regulation, have chosen to hire 10 or more workers - but who, in the presence of the regulation, obtain higher profits by hiring only 9 workers to avoid the discontinuous increase in costs implied by crossing the threshold. These managers should be “bunched up” at 9. The last set of managers are those with high enough productivity \( (\alpha > \alpha_2) \) that it is not worth it to avoid the regulation and so they choose to
exceed the threshold and pay the tax. However, these managers face higher marginal costs than they would in the absence of the regulation and therefore employ fewer workers by a constant proportion (resulting in a “downshift” in the logged firm size distribution).

An exact expression for the distribution of firm size, $\chi(n)$, can be recovered as a transformation of the distribution of managerial ability, $\phi(\alpha)$, since the first-order conditions on the firms’ maximization problems imply a monotonic relationship between $\alpha$ and $n$. The key result is that a function of the tax enters multiplicatively in the expression for the density of firms size $n$ (for all $n > 9$). Therefore, the function of the tax enters additively in the log density for all firms large enough to be subject to the tax.

Formally, the density of firms with $n$ total workers, $\chi(n)$ is given by:

$$
\chi(n) = \begin{cases} 
(\frac{1-\theta}{\theta})^{1-\beta} (\beta - 1)n^{-\beta} & \text{if } n \in [n_{\min}, N) \\
(\frac{1-\theta}{\theta})^{1-\beta} (N^{1-\beta} - \tau^{-\frac{\beta-1}{1-\beta}}n_{u}^{1-\beta}) & \text{if } n = N \\
0 & \text{if } n \in (N, n_{u}) \\
(\frac{1-\theta}{\theta})^{1-\beta} (\beta - 1)\tau^{-\frac{\beta-1}{1-\beta}}n^{-\beta} & \text{if } n \geq n_{u}
\end{cases}
$$

where $\theta$ measures the degree of diminishing returns to scale, capturing both features of the production function and market power, $\beta$ represents the negative slope of the power law and $\tau$ is the implicit per worker tax. Taking logs and combining the first and last cases\textsuperscript{22} leads to:

$$
\log(\chi(n)) = \log\left[\left(\frac{1-\theta}{\theta}\right)^{1-\beta} (\beta - 1)\right] - \beta \log(n) + \log(\tau^{-\frac{\beta-1}{1-\beta}})1\{n > 9\}
$$

This leads to an estimating equation:

$$
\log(\chi(n)) = \alpha - \beta \log(n) + \delta 1\{n > 9\} \tag{2.1}
$$

\textsuperscript{22}In other words, we ignore the bunching at $N$ and the valley directly after, since these are features that are not easily observable in the data. Instead we focus on the ranges $n \in [n_{\min}, N)$ and $n > n_{u}$.
We can identify $\tau$ according to:

$$\tau = \exp(\delta) \frac{1-\theta}{\beta-1}$$

$\tau$ is thus a function of $\theta, \beta$ and $\delta$. We get estimates for $\alpha, \beta$ and $\delta$ from equation 2.1. Knowing $\alpha$ and $\beta$ pins down $\theta$, which allows us to identify $\tau$.

### 2.4.2 Concerns Regarding Misreporting

Before proceeding further, we must consider how our results might be affected by the possibility of misreporting. This is important because one of the underlying assumptions of the analysis above is that the size distribution of firms as observed in the Economic Census is accurate. However, since the data are self-reported, it is possible that plant managers may misreport information to Economic Census enumerators. Specifically, if the managers are aware of the increased regulatory burden that is associated with employing 10 or more workers, and if they believe that the EC enumerators will relay information to government regulatory bodies, they may wish to hide the fact that their actual employment exceeds the threshold. To see how this type of behavior might affect our results, we model it explicitly in the following subsection.

A further reason to be concerned about the possibility of misreporting is due to the fact that Economic Census enumerators were required to fill out an extra form containing the address of any establishment that reported 10 or more workers. It is conceivable that enumerators might have found it preferable to under-report the number of workers for establishments with 10 or more workers in order to avoid the extra burden of filling in the “Address Slip”. Although we do not model this type of problem explicitly in what follows, the implications are nearly identical to those of the model we do explicitly analyze.\textsuperscript{23}

\textsuperscript{23}The only difference is that higher fixed costs would replace higher marginal costs. It is, moreover, easy to show that if our estimation strategy is robust to the model of misreporting we do analyze, it is also robust to this second type of misreporting as well.
2.4.3 A Theoretical Model of Misreporting

Our model of misreporting starts with the theoretical model from Section 2.4.1, and amends it to allow firms to choose not only their true employment \(n\), but also their *reported* employment \(l\). Then, a firm with productivity \(\alpha\) faces the following profit-maximization problem:

\[
\pi(\alpha) = \max_{n,l} \alpha f(n) - wn - \tau l \cdot 1(l > 9) - F(n, l) \cdot p(n, l)
\]

where \(\alpha, f(n), w\) and \(\tau\) are all defined as they were previously. The problem is identical except that now firms pay the extra marginal cost, \(\tau\), only on their *reported* employment, and not on their true employment. Furthermore, they only pay this cost if their reported employment exceeds the threshold.\(^{24}\) There is now an incentive for firms to misreport their employment in a downward direction (i.e.: to set \(l < n\)). Counteracting this incentive is that misreporting firms may be caught by the authorities with probability \(p(n, l)\), and made subject to a fine, \(F(n, l)\). As written above, both the probability of being caught and the magnitude of the fine may in general depend on \(n\) and \(l\) in an arbitrary way. However, if one is willing to make the assumption that the expected cost of misreporting \((F \cdot p)\) is an increasing and convex function of the degree of misreporting, \(n - l\), it will be possible to use an estimation technique that will be only minimally biased by the presence of misreporting. Fortunately, based on our understanding of the context in which firms make these decisions,\(^ {25}\) we believe that this is the most reasonable assumption on the functional form of the expected cost that one could make.

One plausible way to obtain convex misreporting costs is to suppose that firms are caught with a probability that is linearly increasing in the degree of their misreporting (i.e.: \(n - l\)) and subject to a fine if caught which is also a linear function of their misreporting.

\(^{24}\)In point of fact it is most likely that firms’ answers to Economic Census enumerators have no impact on their regulatory burden, but it is possible that firms believe otherwise, and that is what is relevant.

\(^{25}\)This understanding is informed by informal interviews with small businesses in Chennai and our reading of the secondary literature.
Another possibility is that the probability of being caught is itself an increasing and convex function of the degree of misreporting and the fine if caught is fixed. In what follows we will assume the latter for clarity of exposition, but the analysis is identical for any assumption that yields convex costs of misreporting.

Specifically, suppose that misreporting firms are caught with probability \( p(n, l) = \frac{(n-l)^2}{100} \), and pay a fixed fine, \( F \), if caught. Then their profit maximization problem is:

\[
\pi(\alpha) = \max_{n,l} \alpha f(n) - wn - \tau l * 1(l > 9) - F * \frac{(n-l)^2}{100}
\]

The solution to this problem can be informally characterized as follows. The lowest productivity firms (those with \( \alpha \) below some threshold, \( \alpha_1 \)) will be unconstrained, choosing \( n \leq 9 \) and reporting truthfully (\( l = n \)). Higher productivity firms, with \( \alpha \in [\alpha_1, \alpha_2] \), will choose \( n > 9 \), exceeding the regulatory threshold, but will find it profitable to misreport their employment, setting \( l = 9 \). These firms will only appear to be “bunched” up at 9, but will in fact have higher employment. The last category of firms are those with \( \alpha > \alpha_2 \), which are productive enough to warrant hiring work forces so large that they cannot avoid detection with reasonable probability and must report \( l > 9 \). Even these firms, however, with both \( n > 9 \) and \( l > 9 \) do not find it profit-maximizing to report truthfully. They can save on their unit labor costs by shading their reported employment, and will choose \( l = n - \frac{50}{\tau} \). Note that the degree of misreporting is by a constant amount, rather than a constant proportion.\(^{26}\)

More formally, the log of the density of firms with true employment \( n \), \( \log \chi(n) \), is given by:

\[
\log \chi(n) = \begin{cases} 
\log A - \beta \log(n) & \text{if } n \in [n_{\min}, 9) \\
\log \left( \xi(n) \right) & \text{if } n \in [9, n_m(\alpha_2)] \\
0 & \text{if } n \in (n_m(\alpha_2), n_t(\alpha_2)) \\
\log A'(\tau) - \beta \log(n) & \text{if } n \geq n_t(\alpha_2)
\end{cases}
\]

\(^{26}\)This outcome is a result of the convex cost assumption.
while the log of the density of firms with reported employment \( l \), \( \log \psi(l) \) is given by:

\[
\log \psi(l) = \begin{cases} 
\log A - \beta \log(l) & \text{if } l \in [l_{\text{min}}, 9) \\
\log(\delta_l) & \text{if } l = 9 \\
0 & \text{if } n \in (9, l_t(\alpha_2)) \\
\log A'(\tau) - \beta \log(l + \frac{50}{\tau} F) & \text{if } l \geq l_t(\alpha_2)
\end{cases}
\]

where terms have been simplified and collected.\(^{27}\) Both of these densities are graphically represented in Figure 2.5 under specific values of the parameters (the true distribution, \( \chi(n) \), is represented by a thick line, and the reported distribution, \( \psi(l) \), is in blue). The key things to note are the following. First, for the range \( l \leq 9 \), the true distribution coincides with the reported/observed distribution. Second, there appears to be bunching at 9 in the reported distribution, but these firms in fact have greater than 9 workers. Third, compared to the distribution for \( n < 9 \), the true distribution and the reported distribution for \( n \gg 10 \) are downshifted (\( A'(\tau) < A \)), just as was the case in the model without misreporting.\(^{28}\) Fourth - and most significantly - the reported distribution converges to the true distribution for large \( l/n \): \( \lim_{l \to \infty} \beta \log(l + \frac{50}{\tau} F) - \beta \log(l) \to 0 \). This is due to the fact that the misreporting is by a constant amount (as noted earlier), rather than by a constant proportion.

To conclude this subsection on the theoretical implications of allowing for misreporting, we make two observations. First, misreporting may lead us to observe “bunching” in the firm size distribution when in reality there may be none. This is irrelevant for us since our estimation strategy does not rely on the bunching in any way. Second, misreporting may lead the reported/observed distribution to understate the true distribution close to the cutoff. However, because the reported distribution converges to the true distribution for large \( l/n \), misreporting is not able to induce a downshift in the reported distribution that differs from the downshift in the true distribution at large values of \( l \). Therefore, if we use an estimation strategy that focuses mostly on values far from the cutoff, our estimate of the downshift using the observed distribution is likely to reflect the real downshift and thus

\(^{27}\) Derivation to be added in an appendix.

\(^{28}\) As before, the downshift is a function of \( \tau \).
we are likely to avoid this source of bias. We develop such an estimation strategy in the following subsection.

Before proceeding, however, we should note again that the above analysis assumes that the expected costs of misreporting are strictly convex. There exist non-convex functional forms of the cost function which may lead one to observe a downshift in the reported distribution that is greater than the one in the true distribution, thus biasing any estimates of $\tau$ upwards.

2.4.4 An Empirical Strategy Robust to the Possibility of Misreporting

Since convex misreporting costs imply that misreporting will only distort the distribution of reported establishment size versus the true distribution of establishment size close to the cutoff, we estimate the model on the full distribution of establishment size. Since estimating equation 2.1 treats each establishment size as one observation, using the full distribution of establishment size will mean that the model is primarily estimated using data far from the 10-worker cutoff. However, estimating equation 2.1 on the full distribution of establishment size introduces two complications. First, we cannot perform the estimation on the empirical PMF for large firm sizes, since the empirical probability mass is truncated

---

29 The largest establishment in the 2005 EC has 22,901 workers.
at the reciprocal of the number of observations (see figure 2.6 below), while the underlying density continues to diminish in establishment size. Second, respondents appear to round their reported number of workers to the nearest multiple of 5 (see figure 2.2), a phenomenon that is more pronounced for larger establishments and that could bias our results.

Figure 2.6: Downward shift at the 10-worker threshold in the distribution of establishment size estimated on nonparametric density estimates, 2005, log scale (including all establishments). Black points = actual data; Grey = smoothed data.

To address these two problems, we first estimate the density associated with each number of workers $\chi(n)$ non-parametrically using the method of Markovitch and Krieger (2000), which addresses the econometric issues arising in nonparametric density estimation of heavy-tailed data. We then use the nonparametric density estimates as a basis for fitting the model in equation 2.1, augmented by dummy variables for having 1, 2, 8, 9 and 10 - 20 workers. Figure 2.6 depicts the strategy. The black dots show the raw data. The grey

---

30The rationale for flexibly modeling the density at 1 and 2 workers is that own account enterprises and 2-worker enterprises are likely to be household enterprises and may therefore differ fundamentally in character from their larger counterparts. The rationale for flexibly modeling the density at 8 and 9 workers is that the theory above predicts that the reported density just below the cutoff will be biased upwards by any misreporting effects. Similarly, the theory also predicts that values above - but close to - the threshold may also be biased (downwards). Therefore we also flexibly control for such values (10 to 20) as well, although doing so has only a very small effect on the estimates: as explained above, the estimates are driven mostly by observations relatively far from the threshold.
dots represent the result of the first step: nonparametric density estimates associated with each establishment size. The line shows the fit of the model in equation 2.1, augmented by the dummy variables, to the nonparametric density estimates.

Figure 2.6 above provides some evidence for the model described in section 2.4.3. The observed establishment size distribution appears to converge back to a power law with the same slope as for establishments with fewer than 10 workers, but deviates slightly from that slope at sizes just above the 10-worker cutoff. In the next section we report the results of the estimation.

2.5 Preliminary Results

In this section we apply the estimation procedure described above to the 2005 Economic Census of India and report the results. Standard errors obtained from a wild cluster bootstrap procedure with 200 replications are given in parentheses.\textsuperscript{31} In the tables below, we first report estimates for \( \tau - 1 \) at the All-India level and for a selection of States, Industries and Ownership Types. Estimates for all States, Industries and Ownership Types are reported in the Appendix. The All-India estimate on \( \tau - 1 \) is .35 and is statistically significant. This means that, on average, establishments in India that hire more than 9 workers act as though they must pay additional labor costs of 35% of the wage per additional worker. In the tables and figures it can be seen that there is substantial variation in the magnitude of our estimates of the per-worker tax by State, Industry and Ownership Type. For example, the point estimate on \( \tau - 1 \) for the State of Kerala is .14, while the estimate for Bihar, on the other hand, is .70, implying that establishments in Bihar act as though they must pay a tax of 70% of the wage for each additional worker they hire past 9 workers.

We also observe substantial differences in the size of \( \tau \) by industry: it appears that the effective tax is highest for establishments in construction and retail. As one might expect, the tax is nonexistent for establishments in the public administration sector (in fact it is

\textsuperscript{31} We cluster at the firm size level to allow for the possibility that reporting errors may be correlated by firm size.
negative, but this seems to result from the fact that the assumed power law does not fit the
distribution of establishments in this sector well). Similarly, when looking at the differences
by ownership type, we find that the estimates for $\tau$ are highest for private firms (especially
unincorporated proprietorships), and nonexistent (or negative) for government-owned firms,
where presumably the regulatory burden is less than in the private sector.

**Estimates of $\tau$ by State Using the Full Distribution of Establishment Size**

<table>
<thead>
<tr>
<th>Level</th>
<th>$\tau - 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All-India</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.347</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
</tr>
<tr>
<td><strong>By State</strong></td>
<td></td>
</tr>
<tr>
<td>Bihar</td>
<td>.693</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
</tr>
<tr>
<td>Gujarat</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
</tr>
<tr>
<td>Kerala</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
</tr>
<tr>
<td>Uttar Pradesh</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
</tr>
<tr>
<td><strong>By Industry</strong></td>
<td></td>
</tr>
<tr>
<td>Construction</td>
<td>.478</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>.268</td>
</tr>
</tbody>
</table>
Wholesale, retail  
\(0.039\)  
.637  
\(0.115\)  
Public admin., social security  
-.311  
\(0.031\)

\textbf{By Ownership Type}

\begin{tabular}{ll}
Government and PSU & -.092  \\
\multicolumn{2}{l}{\(0.028\)}  \\
Unincorporated Proprietary & .490  \\
\multicolumn{2}{l}{\(0.005\)}
\end{tabular}

2.6 Discussion and Investigation of Mechanisms

2.6.1 Interpretation of Results

Thus far we have argued that the observed downshift in the distribution of establishments with 10 or more workers is related to the existence of certain labor and industrial regulations that become binding at that point. But if the regulations are responsible for the observed effect, then differences in the substance or application of the regulations should explain (at least part of) the great variation we observe across States and Industries.\footnote{The variation across ownership types is straightforward to explain: the regulations are clearly not applied in the same way to privately owned enterprises and government enterprises. An additional explanation is that government establishments are not profit maximizing and thus would require a different motivational theory altogether to produce the observed power law distribution.} In this section we explore these dimensions of variation with the goal of reaching a deeper understanding.
regarding the causes and consequences of the costs we have tried to estimate. The regressions we do are cross-sectional and the variables used are endogenous, so the results cannot be given a causal interpretation, but we find them instructive nevertheless.

To preview our results, we do observe a correlation between our estimated costs ($\tau$) and certain measures of the substance of the regulations. Moreover, we also find a robust and independent correlation between our estimated costs and several different measures of corruption/poor state governance, suggesting that it is not only the regulations themselves but also their enforcement and application that is responsible for the high costs we estimate. We also sketch a theoretical framework of bribery and extortion which casts light on the proper interpretation of our empirical results. Finally, we present some suggestive evidence that our costs may have significant negative dynamic implications, as they are associated with lower growth in employment and productivity in registered manufacturing - and higher growth in employment in unregistered manufacturing.

### 2.6.2 $\tau$ and Corruption: Evidence from the Interstate Variation

We start by regressing our state-level estimates of $\tau$ against other established measures of the regulatory environment (see Table 2.2).\(^{33}\) These measures include the “Besley Burgess” (BB) measure of labor regulations from Aghion et al. (2008) and several measures from Dougherty (2009). The first is a measure of the number of amendments that a state government has made to the Industrial Disputes Act in either a “pro-worker” or “pro-employer” direction, as interpreted by Aghion et al. (2008), who update the measure to include amendments up to 1997.\(^{34}\) Positive values indicate more “pro-worker” amendments, which are assumed to imply a more restrictive environment for firms operating in those states. Dougherty (2009) also provides state level measures that reflect “the extent to which procedural or administrative changes have reduced transaction costs in relation

---

\(^{33}\)Note that the estimates of $\tau$ we use in all the analysis below were generated using the procedure in Section 2.4.4 that we have argued is robust to possible misreporting and non-classical measurement error.

\(^{34}\)Since there have been few state-level amendments to the IDA between 1997 and 2005, this measure should be largely the same in 2005.
to labor issues” Dougherty et al. (2014). Higher values therefore indicate an improved environment for firms. Dougherty’s measures are unique in that they cover a wide range of labor-related issues - not just the IDA. In the analysis below, we will focus on measures from Dougherty (2009) that cover reforms regarding 1) the Factories Act and 2) an overall measure of reforms. All relevant variables in our analysis have been rescaled to have mean zero and standard deviation one, with the goal of allowing comparability between regression coefficients in different specifications.

In Table 2.2, correlations are reported between $\tau$ and the three measures both by themselves and while controlling for other factors (notably state GDP per capita and the state’s share of employment in manufacturing). The Besley Burgess (BB) measure does not seem to be correlated with $\tau$ (although our power is limited by the very small number of observations) while the two measures from Dougherty (2009) are significantly correlated after applying controls (though not all the correlations are strongly significant) and have the “correct” sign: states that saw more “transaction cost reducing” reforms have lower $\tau$s. On the one hand the lack of correlation between $\tau$ and the BB measure is not surprising, as the latter capture variation only due to state amendments to the Industrial Disputes Act, which does not vary over the ten person threshold. On the other hand, if the Besley Burgess measure is meant to capture the general regulatory environment (which is how it is used in countless studies), we might well expect it to correlate with our measure of regulatory costs. That the correlation does not hold may therefore be of interest.

While the prediction regarding the correlation between $\tau$ and BB.97 may be ambiguous, that is not the case for Dougherty’s measures of transaction-cost reducing reforms related to the Factories Act. We should expect our measure of $\tau$ to correlate negatively with the latter, since the Factories Act does vary across the 10 worker threshold, and indeed we see that it does. $\tau$ is also correlated with Dougherty’s more comprehensive measure of reforms, one which aggregates reforms across all areas, although it does not appear to correlate with

---

35 In this and most of the analysis ahead, we focus on the 18 largest Indian States, for which data are most consistently available and which offer the most precise estimates of $\tau$ (the power law relationship breaks down in smaller states when there are not enough observations).
any other subcomponents (which are not depicted here).

<table>
<thead>
<tr>
<th>Table 2.2: Tau vs Other Measures of Regulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>tau</td>
</tr>
<tr>
<td>Besley-Burgess measure (regs)</td>
</tr>
<tr>
<td>(0.204)</td>
</tr>
<tr>
<td>Dougherty measure (all reforms)</td>
</tr>
<tr>
<td>(0.201)</td>
</tr>
<tr>
<td>Dougherty measure (FA reforms)</td>
</tr>
<tr>
<td>(0.187)</td>
</tr>
<tr>
<td>log of net state</td>
</tr>
<tr>
<td>(0.226)</td>
</tr>
<tr>
<td>(4.886)</td>
</tr>
<tr>
<td>share of employment in manufacturing</td>
</tr>
<tr>
<td>(6.768)</td>
</tr>
<tr>
<td>share of privately owned establishments</td>
</tr>
<tr>
<td>(1.530)</td>
</tr>
<tr>
<td>share of registered establishments</td>
</tr>
<tr>
<td>(0.204)</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

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

In addition to the above measures regarding state-level changes to the statutory, procedural and administrative aspects of the regulations, we also regress \( \tau \) against certain other measures of the labor environment. Table 2.3 reports the results of \( \tau \) regressed against per capita measures of strikes, man-days lost to strikes, lockouts and man-days lost to lockouts. One might imagine that strikes and lockouts capture relevant features of the regulatory and
labor environment,\textsuperscript{36} but we do not find them to be robustly correlated with \( \tau \).

One might also expect \( \tau \) to be correlated with aspects of the regulatory enforcement. To test this hypothesis we regress \( \tau \) against state level variables related to enforcement such as the number of inspections, convictions, and fines levied under various regulations.\textsuperscript{37} The results of the regressions for a subset of the enforcement related variables are shown in Table 2.4. In short, the only enforcement variable that is even close to being significantly correlated with \( \tau \) is the percentage of factories registered under the Factories Act that have been inspected. However, as can be seen from the table, the enforcement data are only available for a small subset of the major states, leaving very little power in the regressions. Furthermore, the regressions shown exclude Uttar Pradesh, which is a substantial outlier in the enforcement data.

\textsuperscript{36}For example, some industrial regulations explicitly undermine or support the rights of parties to engage in strikes or lockouts.

\textsuperscript{37}These data were obtained from the 2005 Indian Labour Yearbook, which we were able to attain with the generous help of Anushree Sinha and Avantika Prabhakar of NCAER.
Table 2.3: Tau vs Strikes and Lockouts

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tau</td>
<td>tau</td>
<td>tau</td>
<td>tau</td>
<td>tau</td>
<td>tau</td>
<td>tau</td>
<td>tau</td>
<td>tau</td>
</tr>
<tr>
<td>strikes per capita</td>
<td>-0.272*</td>
<td>-0.196</td>
<td>(0.145)</td>
<td>(0.148)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mandays lost due to strikes per capita</td>
<td>-0.119</td>
<td>-0.148</td>
<td>(0.157)</td>
<td>(0.159)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lockouts per capita</td>
<td>-0.0544</td>
<td>-0.0915</td>
<td>(0.146)</td>
<td>(0.143)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mandays lost due to lockouts per capita</td>
<td>-0.0527</td>
<td>-0.0995</td>
<td>(0.145)</td>
<td>(0.139)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log of net state</td>
<td>-0.400</td>
<td>-0.493*</td>
<td>-0.506**</td>
<td>-0.515**</td>
<td>(0.234)</td>
<td>(0.238)</td>
<td>(0.235)</td>
<td>(0.235)</td>
</tr>
<tr>
<td>share of employment in manufacturing</td>
<td>0.721***</td>
<td>4.352*</td>
<td>0.646***</td>
<td>5.017***</td>
<td>0.620***</td>
<td>5.191**</td>
<td>0.618***</td>
<td>5.291**</td>
</tr>
<tr>
<td>Constant</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>17</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

Only including Major Indian States

* p < 0.10, ** p < 0.05, *** p < 0.01
Table 2.4: Tau vs Enforcement of Regulations

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tau</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>percent of factories inspected</td>
<td>0.448*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.230)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>convictions under FA per factory</td>
<td>-0.0914</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.411)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prosecutions under SEA per capita</td>
<td>-0.122</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.357)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fines under SEA per capita</td>
<td>0.323</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.342)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prosecutions per inspection</td>
<td>-0.134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.254)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fines per inspection under SEA</td>
<td>2.598</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.557)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cases disposed per inspection under SEA</td>
<td>-9.616</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(16.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cases disposed per cases prosecuted under SEA</td>
<td>1.058</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.094)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log of net state domestic product pc</td>
<td>-0.0180</td>
<td>-0.648</td>
<td>-1.545**</td>
<td>-2.220**</td>
<td>-1.572**</td>
<td>-1.977**</td>
<td>-1.471**</td>
<td>-1.610**</td>
</tr>
<tr>
<td></td>
<td>(0.921)</td>
<td>(1.265)</td>
<td>(0.606)</td>
<td>(0.783)</td>
<td>(0.537)</td>
<td>(0.677)</td>
<td>(0.600)</td>
<td>(0.500)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.970</td>
<td>6.706</td>
<td>15.42**</td>
<td>22.20**</td>
<td>15.72**</td>
<td>20.08**</td>
<td>12.60</td>
<td>16.57***</td>
</tr>
<tr>
<td>Observations</td>
<td>10</td>
<td>9</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
Only including Major Indian States (except UP) for which data exist.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
To briefly summarize the results so far, when regressing $\tau$ against regulatory substance, enforcement or industrial disputes, one only observes correlations for certain specific types of measures of those phenomena. In contrast, in our remaining analysis, we will demonstrate that our measures of $\tau$ are strongly and robustly correlated with corruption, almost regardless of how it is measured. Table 2.5 reports the results of regressing $\tau$ against corruption as measured in a 2005 Transparency International (TI) Survey. Column 1 includes all states for which there is data, while the remaining columns include only the 18 largest Indian states. Column 3 adds controls for state GDP per capita, share of manufacturing in employment and some others, while Column 4 adds the aggregate measure of regulatory reform from Dougherty (2009). With no exceptions, the coefficient on the TI corruption score is consistently significant and very large in magnitude: a one standard deviation increase in a state’s corruption score is associated with a .5 standard deviation increase in $\tau$. In particular, the fact that the coefficient remains significant in Column 4 even after controlling for Dougherty’s measure of regulatory reforms suggests that the relationship between $\tau$ and corruption is at least partly independent from the relationship with the regulations themselves.

In what follows we will use the TI corruption score as our primary measure of corruption. One might be concerned, however, that the TI measure may be flawed as it is the result of individuals’ perceptions (it has been argued by some that the perception of corruption is an unreliable indicator for actual corruption). Therefore, we also regress $\tau$ against an alternative measure of corruption that is not perception based to check for robustness of the relationship between $\tau$ and corruption: Table 2.6 reports the results of $\tau$ regressed against the percent of a state’s available electricity that was lost in transmission and distribution in 2005. This variable has been used by other researchers as a proxy for corruption and poor state governance, and has the virtue of being a concrete and objective measure that does not depend on perceptions Kochhar et al. (2006). As with the TI Corruption Score,

---

38 The TI corruption measure is based on a survey of perceptions and experience regarding corruption in the public sector.
the correlations between $\tau$ and this alternative measure of corruption are significant and large in magnitude regardless of sample or controls - including, again, the addition of the Dougherty measure of regulatory reform in Column 4. To make sure that the results are not driven by the actual transmission of electricity, we control for per capita electricity available in Column 5 - which does not affect the results.

<table>
<thead>
<tr>
<th>Table 2.5: Tau vs Transparency International Corruption Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>Tau</td>
</tr>
<tr>
<td>TI Corruption Score</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>log of net state</td>
</tr>
<tr>
<td>domestic product pc</td>
</tr>
<tr>
<td>share of employment in manufacturing</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>share of privately owned establishments</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>share of registered establishments</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Dougherty measure (all reforms)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>States Included</td>
</tr>
</tbody>
</table>
* p<0.10, ** p<0.05, *** p<0.01, Standard errors in parentheses
Table 2.6: Tau vs Transmission and Distribution Losses

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>electricity</td>
<td>0.318*</td>
<td>0.648**</td>
<td>0.663**</td>
<td>0.566**</td>
<td>0.708**</td>
</tr>
<tr>
<td>transmission and distribution losses</td>
<td>(0.165)</td>
<td>(0.244)</td>
<td>(0.253)</td>
<td>(0.237)</td>
<td>(0.268)</td>
</tr>
<tr>
<td>log of net state</td>
<td>-0.477*</td>
<td>-0.536**</td>
<td>-0.459*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>domestic product pc</td>
<td>(0.222)</td>
<td>(0.205)</td>
<td>(0.229)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>share of employment</td>
<td>7.651</td>
<td>6.515</td>
<td>6.817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in manufacturing</td>
<td>(4.812)</td>
<td>(4.439)</td>
<td>(5.094)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>share of privately</td>
<td>-2.148</td>
<td>-1.440</td>
<td>-1.980</td>
<td></td>
<td></td>
</tr>
<tr>
<td>owned establishments</td>
<td>(5.677)</td>
<td>(5.200)</td>
<td>(5.824)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>share of registered</td>
<td>1.513</td>
<td>0.644</td>
<td>1.577</td>
<td></td>
<td></td>
</tr>
<tr>
<td>establishments</td>
<td>(1.307)</td>
<td>(1.284)</td>
<td>(1.343)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dougherty measure</td>
<td>-0.317*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(all reforms)</td>
<td>(0.172)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electricity</td>
<td></td>
<td>0.119</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>available (GWH)</td>
<td></td>
<td>(0.182)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-8.42e-09</td>
<td>0.643***</td>
<td>5.936</td>
<td>6.322</td>
<td>5.610</td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.163)</td>
<td>(5.746)</td>
<td>(5.253)</td>
<td>(5.910)</td>
</tr>
<tr>
<td>Observations</td>
<td>35</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>States Included</td>
<td>All</td>
<td>Major</td>
<td>Major</td>
<td>Major</td>
<td>Major</td>
</tr>
</tbody>
</table>

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

Although the state-level correlations between τ and corruption appear to be robust, the
regressions lack exogenous variation and are subject to the concern that our measures of corruption may be correlated with omitted variables that are also correlated with \( \tau \). To partially address these concerns, we attempt to take advantage of State X Industry level heterogeneity. In particular, inspired by Novosad and Asher (2014), we use 2005 World Bank Enterprise Survey (WBES) data to create an industry-level measure of “dependence on government bureaucracy”. Specifically, Indian firms in the 2005 WBES were asked how many times in a year they had an inspection or other required meeting with a government official. Averaging the firm-level responses by industry, we classify industries according to their average number of visits with officials (i.e., their dependence on government bureaucracy). If some industries have more meetings with officials, and if corruption takes place during some of these meetings, we would imagine that the costs of corruption would be highest for firms in those industries with the highest dependence on bureaucracy and in those states that have the highest levels of corruption. That is, we would expect that the interaction between industry level dependence on bureaucracy and state level corruption is positive. If found to be the case, it would be harder to argue that the result is due to the presence of omitted variables.

The hypothesis is tested in Table 2.7. To do so we generate our measures of \( \tau \) at the State X Industry level\(^{39}\) and interact each of our state level measures of corruption with a) the industry average number of visits from officials and b) the industry average duration of visits from officials. We include the interaction with average duration of visits as a placebo test: it is not clear that the duration of an inspection should be positively or negatively correlated with corruption.\(^{40}\) We then regress our State X Industry measures of \( \tau \) against the covariates including interaction terms. Our prior is that the interaction of corruption with duration of visits should be less significant than the interaction with number of visits.

\(^{39}\)Industries here are categorized according to their groupings in the World Bank Enterprise Surveys, which distinguishes 24 distinct industry categories. Examples include “auto components”, “leather and leather products”, and “food processing”.

\(^{40}\)In particular, corruption may lead to longer inspections if the process of extracting the bribe takes time, or it may lead to shorter inspections if corruption obviates the need to carry out the actual inspection.
Indeed, this is mostly what we observe. The interaction between our measures of corruption and the number of visits is at least weakly significant (at the 10% level) for one of the two measures, while the interaction between corruption and average duration of visits is never significant.

To summarize our results from these investigations, we find:

1. a correlation between $\tau$ and certain aspects of the substance of regulations as measured in Dougherty (2009),

2. a nonexistent or inconclusive relationship between $\tau$ and measures of the labor environment and enforcement of regulations, and

3. a strong and robust relationship between $\tau$ and two distinct measures of corruption.

Although none of these results can be said to be causal, we find them suggestive of a relationship between corruption and high labor costs. Next, we turn our attention to the question of how and why greater corruption would lead to higher labor costs for firms. To this end, in the following subsection we outline a simple theoretical framework to elucidate the potential connection.
Table 2.7: Tau vs. State Level Corruption Interacted with Industry Level “Dependence on Regulation” (with Industry FEs)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tau</td>
<td>0.132*</td>
<td>0.140*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0668)</td>
<td>(0.0666)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 TI Corruption Score</td>
<td></td>
<td></td>
<td>0.0583</td>
<td>0.0493</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.114)</td>
<td>(0.110)</td>
</tr>
<tr>
<td>electricity transmission and distribution losses</td>
<td></td>
<td></td>
<td>0.0583</td>
<td>0.0493</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.114)</td>
<td>(0.110)</td>
</tr>
<tr>
<td>number of inspections</td>
<td>0.127</td>
<td>0.130</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0829)</td>
<td>(0.0945)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>duration of inspections</td>
<td></td>
<td></td>
<td>-0.00610</td>
<td>-0.109</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.124)</td>
<td>(0.143)</td>
</tr>
<tr>
<td>corruption score X num inspections</td>
<td>0.101*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0545)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corruption score X duration of inspections</td>
<td></td>
<td>0.0513</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0691)</td>
<td></td>
</tr>
<tr>
<td>electricity TDLs X num of inspections</td>
<td></td>
<td>0.0112</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0848)</td>
<td></td>
</tr>
<tr>
<td>electricity TDLs X duration of inspections</td>
<td></td>
<td></td>
<td>-0.0698</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.107)</td>
<td></td>
</tr>
<tr>
<td>log of Net State Domestic Product pc</td>
<td>-0.185***</td>
<td>-0.185***</td>
<td>-0.197***</td>
<td>-0.194***</td>
</tr>
<tr>
<td></td>
<td>(0.0364)</td>
<td>(0.0374)</td>
<td>(0.0428)</td>
<td>(0.0427)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.244</td>
<td>0.0235</td>
<td>-0.235</td>
<td>-0.217</td>
</tr>
<tr>
<td></td>
<td>(0.187)</td>
<td>(0.442)</td>
<td>(0.187)</td>
<td>(0.498)</td>
</tr>
<tr>
<td>Observations</td>
<td>189</td>
<td>189</td>
<td>189</td>
<td>189</td>
</tr>
</tbody>
</table>

* p<0.10, ** p<0.05, *** p<0.01, Standard errors clustered at the State Level, Industry FEs included
A Theoretical Framework for Understanding Corruption Between Inspectors and Firms

We find it helpful to distinguish between two types of corruption that could take place between inspectors and firms: collusion and extortion. Collusion takes place when inspectors allow firms to avoid the costs of complying with regulations in exchange for bribes. However, poor state governance (which here would imply an inability to control corruption) would then lead to lower costs for firms, as greater corruption would make it easier to avoid the full costs of regulation.\footnote{See, for example, a model of corruption such as the one in Khan, Khwaja, and Olken (2014).} However, what we observed in Section 2.6.2 was a robust positive correlation between effective costs ($\tau$) and poor governance/corruption. To explain this phenomenon, we need a model of extortion. In this section, we will sketch the intuition for such a model. A fuller (but still simple) model of extortion and bribery is provided in Appendix 3.

Let us start with the observation that all firms reporting at least 10 employees fall under the jurisdiction of certain regulations. Imagine, now, that the regulations are so complex so as to make it impossible (or prohibitively costly) for any firm to be fully in compliance with all aspects of the law as written.\footnote{This does not not require much imagination. As we mentioned in Section 2.2, many of the laws have components that are antiquated, arbitrary, contradictory and confusing. That the laws may be impossible to fully comply with is suggested by some of the anecdotes we provide in Appendix 2 as well as the following observation, which we re-quote: “Rules under the Factories Act, framed in 1948, provide for white washing of factories. Distemper won’t do. Earthen pots filled with water are required. Water coolers won’t suffice. Red-painted buckets filled with sand are required. Fire extinguishers won’t do... And so on” (TeamLease Services, 2006).} Then, an inspector can, at any time, choose to subject a firm under his jurisdiction to a penalty $e$, which may include financial (e.g.: fines) and/or non-financial elements (e.g.: harassment, time needed to defend claims of violations, etc). We can think of the extent of the penalty ($e$) as a function of state governance: properly functioning governments hire and motivate inspectors to pursue substantive violations rather than minor ones, while inspectors in corrupt or dysfunctional governments can get away with threatening to impose high penalties for even minor technical violations if a bribe
is not paid (i.e.: extortion).

In such an environment, firms reporting 10 or more employees (and hence under the jurisdiction of the inspector) may face a choice between exposing themselves to the penalty, $e$, or paying a bribe, $b$. Assume that inspectors face no costs or benefits from imposing the penalty on the firm, but naturally benefit from receiving the bribe. There is thus a surplus to be had from paying/receiving the bribe $b$ and avoiding the penalty. If the inspector and firm Nash Bargain over the surplus with bargaining weights $\alpha$ and $\beta$, respectively, the problem is the following:

$$\max_b (b)^\alpha (e - b)^\beta$$

The solution is for the firm to pay a bribe $b = \frac{\alpha}{\alpha + \beta} \cdot e$. The cost born by the firm is therefore increasing in $\alpha$, the bargaining weight of the inspector, and in $e$, the maximum penalty to which the firm can be subjected. It is reasonable to imagine that this maximum level of extortion, $e$, is roughly proportional to the size of the firm, so that $e = e' \cdot n$, where $n$ is the number of workers in the firm and $e'$ is the per worker level of extortion. In that case the bribe per worker, $\frac{b}{n}$, is equal to $\frac{\alpha}{\alpha + \beta} \cdot e'$.\(^{43}\)

This framework can be embedded into the firm’s choice of true and reported employment as modeled in Section 2.4.3. In particular, the firm now faces a choice between reporting employment greater than or less than 10, where reporting less than 10 allows it to avoid the costs of bribery, and reporting greater than 10 exposes it to the bribery costs. In that framework, $\tau$ corresponds to $\frac{b}{n}$, and is therefore increasing in the bargaining power of the inspector ($\alpha$) and the corruption level of the state ($e'$). In this way, we can make sense of the empirical results above in terms of this basic framework. Again, a more fully fleshed out model that explicitly incorporates features missing here (such as an appeals process and inspector types) is provided in Appendix 3.

\(^{43}\)Again, we provide some support for the claim that bribes are proportional to the number of workers with anecdotal evidence from ipaidabribe.com in Appendix 2.
2.6.3 Possible Consequences of $\tau$

In the subsections above, we tried to argue that our estimated costs ($\tau$) are most likely due, not only to the substance of the regulations themselves, but also to high levels of corruption. In this subsection we will indicate possible consequences of high values of $\tau$. Again, the results cannot be given a causal interpretation, but we find them compelling nevertheless. In what follows we use two distinct measures of $\tau$: one which is created using all the enterprises in a state, regardless of economic sector ($\tau$) and another which is created using only the enterprises engaged in manufacturing ($\tau_{\text{manuf}}$).

Table 2.8 displays the results of employment growth in the manufacturing sector between 2010 and 2005 at the State Level regressed against our two measures of labor market distortions ($\tau$ and $\tau_{\text{manuf}}$) as well competing measures (BB and Dougherty). For each of the four measures, we observe its performance as a predictor of future employment growth in registered manufacturing as well as its correlation with employment growth in unregistered manufacturing. Interestingly, in the regressions of employment growth in registered manufacturing against $\tau$ and $\tau_{\text{manuf}}$, the coefficient on $\tau$ is negative and at least weakly significant, while the coefficient for employment growth in unregistered manufacturing is positive - significantly so in the case of $\tau_{\text{manuf}}$. This result makes sense: we should expect higher costs to negatively effect the sectors to which the costs apply - in this case the registered sector, since that is under the ambit of labor regulations while the unregistered sector is not. If these correlations reflect a causal chain, it would mean that high levels of regulator costs and corruption (as measured by $\tau$) are pushing employment from the registered to the unregistered sector.

Also included in Table 2.8 are the results of employment growth in manufacturing regressed against the BB and Dougherty measures. Neither regressor has a coefficient that is statistically significant or of a meaningful magnitude.\(^{44}\) Putting aside the considerable

\(^{44}\)One might argue that it is not quite fair to regress growth between 2010 and 2005 on a regressor that uses data only up until 1997 (as is the case for the BB measure). However, a) we have duplicated these results using from growth from 1997 to 2002 and the results are the same, and b) the Besley Burgess measure
caveat that none of these results has the virtue of exogenous variation, it would appear to be the case that our measures of labor market distortions do a better job of predicting future employment growth (or the lack thereof) than the established alternatives. This is also true when considering future growth in manufacturing *productivity* rather than employment, as shown in Table 2.9. Higher levels of $\tau$ are associated with slower growth of productivity in the *registered* manufacturing sector (less so in the unregistered sector).

Table 2.8: Manufacturing Employment Growth (2005 - 2010) vs Tau and Other Measures

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>reg manuf</td>
<td>unreg manuf</td>
<td>reg manuf</td>
<td>unreg manuf</td>
<td>reg manuf</td>
<td>unreg manuf</td>
<td>reg manuf</td>
<td>unreg manuf</td>
</tr>
<tr>
<td>tau</td>
<td>-0.0240</td>
<td>0.00197</td>
<td>(0.0176)</td>
<td>(0.0233)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tau (manuf)</td>
<td>-0.0471**</td>
<td>0.0623**</td>
<td>(0.0217)</td>
<td>(0.0256)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Besley-Burgess measure (regs)</td>
<td>-0.00525</td>
<td>0.00979</td>
<td>(0.00731)</td>
<td>(0.0142)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dougherty measure (all reforms)</td>
<td>0.0226</td>
<td>-0.0143</td>
<td>(0.0130)</td>
<td>(0.0159)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log of net state domestic product pc</td>
<td>0.00312</td>
<td>0.0189</td>
<td>0.0107</td>
<td>0.0192</td>
<td>0.00413</td>
<td>0.0140</td>
<td>0.0212</td>
</tr>
<tr>
<td>share of employment in manufacturing</td>
<td>-0.393</td>
<td>0.00558</td>
<td>-0.708**</td>
<td>0.435</td>
<td>0.0194</td>
<td>-0.559</td>
<td>-0.515*</td>
</tr>
<tr>
<td>Constant</td>
<td>0.0969</td>
<td>-0.182</td>
<td>0.0372</td>
<td>-0.229</td>
<td>0.0209</td>
<td>-0.0675</td>
<td>-0.0861</td>
</tr>
</tbody>
</table>

Observations 18 17 18 17 15 15 18 17

* $p<0.10$, ** $p<0.05$, *** $p<0.01$. Standard errors in parentheses. Only including Major Indian States

from Aghion et al. (2008) should be largely the same in 2005 due to the lack of state level reforms between 1997 and 2005.
Table 2.9: Manufacturing Productivity Growth (2005 - 2010) vs Tau and Other Measures

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tau</td>
<td>-0.0321**</td>
<td>0.00522</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0146)</td>
<td>(0.0239)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tau (manuf)</td>
<td>-0.0512**</td>
<td>-0.0567*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0181)</td>
<td>(0.0275)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Besley-Burgess</td>
<td>-0.00266</td>
<td>-0.00372</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>measure (regs)</td>
<td>(0.0122)</td>
<td>(0.0154)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dougherty measure (all reforms)</td>
<td>0.0167</td>
<td>0.00973</td>
<td>(0.0122)</td>
<td>(0.0166)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log of net state</td>
<td>-0.0160</td>
<td>-0.00902</td>
<td>-0.00478</td>
<td>-0.0120</td>
<td>-0.00455</td>
<td>-0.00723</td>
<td>0.00438</td>
<td>-0.00793</td>
</tr>
<tr>
<td>domestic product pc</td>
<td>(0.0148)</td>
<td>(0.0220)</td>
<td>(0.0121)</td>
<td>(0.0174)</td>
<td>(0.0143)</td>
<td>(0.0182)</td>
<td>(0.0145)</td>
<td>(0.0204)</td>
</tr>
<tr>
<td>share of employment in manufacturing</td>
<td>0.206</td>
<td>0.0392</td>
<td>-0.154</td>
<td>-0.349</td>
<td>0.453</td>
<td>0.526</td>
<td>0.0728</td>
<td>0.0104</td>
</tr>
<tr>
<td></td>
<td>(0.214)</td>
<td>(0.338)</td>
<td>(0.215)</td>
<td>(0.350)</td>
<td>(0.310)</td>
<td>(0.392)</td>
<td>(0.230)</td>
<td>(0.338)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.141</td>
<td>0.124</td>
<td>0.0454</td>
<td>0.198</td>
<td>-0.0208</td>
<td>0.0530</td>
<td>-0.0678</td>
<td>0.119</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
<td>(0.214)</td>
<td>(0.116)</td>
<td>(0.163)</td>
<td>(0.137)</td>
<td>(0.174)</td>
<td>(0.138)</td>
<td>(0.190)</td>
</tr>
<tr>
<td>Observations</td>
<td>18</td>
<td>17</td>
<td>18</td>
<td>17</td>
<td>15</td>
<td>15</td>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

* p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses, Only including Major Indian States

2.6.4 Inter-temporal Variation: A Final Puzzle

As we have noted, all of the analysis above uses data only from the 2005 Economic Census. However, the discontinuity observed in the 2005 data does not appear in the 1990 or 1998 ECs. Explaining the puzzling inter-temporal variation is a main goal of our future work. Perhaps the simplest explanation is that the data quality of the EC improved between 1998 and 2005. Indeed, it is intuitive that the presence of a downshift will be harder to discern upon the introduction of measurement error. This explanation has received anecdotal support from our meetings with the Directors of the state Directorates of Economics and Statistics, who largely claimed to be more confident in the results of the 2005 EC than in the results of previous rounds. One reason for this could be that the 2005 EC was the first
wave in which ICR (“intelligent character recognition”) was used to read in the raw data. This technology should have gone a long way in alleviating the considerable amount of error that comes from manual interpretation and typing.

But one might also be concerned in the other direction: that somehow the distortion observed in 2005 is anomalous and is perhaps due to something like a change in the EC’s enumeration practices. Based on our interviews with officials and enumerators in charge of collecting these data, however, we believe this is unlikely. While we have discovered that enumeration practices have changed slightly over the years, we have not discovered changes that could have produced the specific patterns we observe. For example, unlike the previous waves, the 2005 Economic Census included an “address slip” that was meant to be filled out for establishments with 10 or more workers. It is conceivable that enumerators, in an effort to avoid the extra work of filling out the address slip, preferred to misrepresent the number of workers for establishments with more than 10 workers. However, while this could reasonably explain why there are fewer 10, 11 and 12 person establishments, we find it hard to understand how this kind of phenomenon could explain why there are also fewer 30, 40 and 50 person establishments (see section 2.4.3 for more on this). Furthermore, in post-enumeration checks done in West Bengal, Bihar and Tamil Nadu, this kind of misrepresentation was not found to be in occurrence.
Nor does the culprit seem to be changes in the regulations themselves. Indeed, none of the regulations which we assume are responsible for the discontinuity in 2005 were greatly changed in the right direction between 1990 and 2005. There were reforms on the margin (reflected in the Dougherty (2009) measures), and these do seem to be correlated with our measures of \( \tau \), but most of the changes have gone in the direction of loosening regulations and thus cannot explain why \( \tau \) would have increased over time. Enforcement has also changed to some degree, but we have not yet found evidence that it can explain the variation in our data (see the interstate analysis above).

Another possible (but ultimately unlikely) explanation is that changes in the competitive environment, particularly related to the increased exposure to international markets and competition, are responsible. The period in question (1998 to 2005) saw heavy reductions in protective barriers from foreign competition - particularly through the elimination of non-tariff barriers. However, in our preliminary analysis (not yet reported) we do not find a strong link between trade liberalization and \( \tau \), and such link as exists goes in the opposite direction.

A final possibility is that changes in the availability of contract labor cause the discontinuity to show up in 2005. Indeed, there was a large speed-up in the use of contract labor over the period,\(^{45}\) however state-level changes in the fraction of contract labor in registered manufacturing are not robustly correlated with \( \tau \). The absence of such a correlation is not necessarily evidence that such a link does not exist, but it does not give support to the hypothesis either.

We have been able to share this intertemporal paradox with a number of experts regarding these issues in India but have not yet been able to find a watertight explanation. This continues to be a priority in our ongoing work.

\(^{45}\) In the registered manufacturing sector, the share of contract labor in total labor increased from 15% to 26%. Prior to this period, its growth was markedly slower.
2.7 Conclusion

Our goals in this paper are 1) to document the effect of size-based labor regulations on the misallocation of resources across firms via the employment decisions of business enterprises, 2) to estimate the net costs of the set of regulations that become binding when establishments choose to employ 10 or more workers, and 3) to shed some light on the source of these costs by demonstrating that corruption in the form of harassment bribery may play a large role in making Indian regulations costly. To the best of our knowledge, this paper is the first to provide cost estimates of regulations in India (particularly non-IDA regulations), the first to analyze the effects of regulations without using necessarily subjective evaluations of state-level amendments to labor laws, and the first to provide evidence for the link between regulatory costs and corruption.

To accomplish these tasks, we use the 2005 Economic Census of India, an uncommonly used dataset which is uniquely suited to our task because it includes the entire universe of non-farm enterprises. In our investigation, we find a significant level shift down in the natural log of the probability mass of establishments with 10 or more workers. Adapting a method from Garicano et al. (2013), we interpret this as evidence of substantial per-worker costs of operating above the 10 worker threshold. At the all-India level, we find that operating at or above the 10-worker threshold is associated with a 35% increase in the unit cost of labor as modeled. Furthermore, we observe a great deal of variation in our estimated costs by state, industry and ownership type. We estimate the highest (lowest) costs for privately owned firms (government-owned firms) and firms in construction (public administration and defense).

Exploring this variation reveals that Indian states with the highest costs also have the highest levels of corruption and poor governance (as measured through two distinct indices), and that firms in industries with high bureaucratic dependence are exposed to particularly high costs if they are also in highly corrupt states. This analysis suggests that the size of regulatory costs may have as much to do with how regulations are implemented and who
implements them, as with the content of the specific labor and industrial laws themselves. We hope that these findings will help shift the present debate away from arguments over the pro or anti-labor stance of regulations and towards arguments about clarity, bureaucracy and the proper enforcement of regulations.
Appendix 1: Full Results by State and Industry\textsuperscript{46}

Table 2.10: Estimates of Tau by State

<table>
<thead>
<tr>
<th>State</th>
<th>Tau</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andhra Pradesh</td>
<td>-.159</td>
<td>.038</td>
</tr>
<tr>
<td>Assam</td>
<td>.322</td>
<td>.041</td>
</tr>
<tr>
<td>Bihar</td>
<td>.693</td>
<td>.069</td>
</tr>
<tr>
<td>Delhi</td>
<td>.427</td>
<td>.048</td>
</tr>
<tr>
<td>Gujarat</td>
<td>.165</td>
<td>.047</td>
</tr>
<tr>
<td>Himachal Pradesh</td>
<td>-.165</td>
<td>.023</td>
</tr>
<tr>
<td>Haryana</td>
<td>.007</td>
<td>.044</td>
</tr>
<tr>
<td>Jharkhand</td>
<td>.388</td>
<td>.061</td>
</tr>
<tr>
<td>Karnataka</td>
<td>.52</td>
<td>.06</td>
</tr>
<tr>
<td>Kerala</td>
<td>.138</td>
<td>.033</td>
</tr>
<tr>
<td>Maharashtra</td>
<td>.332</td>
<td>.038</td>
</tr>
<tr>
<td>Madhya Pradesh</td>
<td>.379</td>
<td>.047</td>
</tr>
<tr>
<td>Orissa</td>
<td>.283</td>
<td>.044</td>
</tr>
<tr>
<td>Punjab</td>
<td>.096</td>
<td>.041</td>
</tr>
<tr>
<td>Rajasthan</td>
<td>.32</td>
<td>.05</td>
</tr>
<tr>
<td>Tamil Nadu</td>
<td>.397</td>
<td>.059</td>
</tr>
<tr>
<td>Uttar Pradesh</td>
<td>.502</td>
<td>.069</td>
</tr>
<tr>
<td>West Bengal</td>
<td>.151</td>
<td>.054</td>
</tr>
</tbody>
</table>

\textsuperscript{46}Standard errors generated using a wild cluster bootstrap procedure with 200 replications.
Table 2.11: Estimates of Tau by Industry

<table>
<thead>
<tr>
<th>Industry</th>
<th>Tau</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining and quarrying</td>
<td>-.042</td>
<td>.047</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>.268</td>
<td>.039</td>
</tr>
<tr>
<td>Electricity, gas and water supply</td>
<td>-.367</td>
<td>.022</td>
</tr>
<tr>
<td>Construction</td>
<td>.478</td>
<td>.047</td>
</tr>
<tr>
<td>Wholesale and retail trade</td>
<td>.637</td>
<td>.115</td>
</tr>
<tr>
<td>Hotels and restaurants</td>
<td>.468</td>
<td>.06</td>
</tr>
<tr>
<td>Transport, storage and communications</td>
<td>.334</td>
<td>.056</td>
</tr>
<tr>
<td>Financial intermediation</td>
<td>-.105</td>
<td>.044</td>
</tr>
<tr>
<td>Real estate, renting and business activities</td>
<td>.601</td>
<td>.062</td>
</tr>
<tr>
<td>Public administration and defence</td>
<td>-.311</td>
<td>.031</td>
</tr>
<tr>
<td>Education</td>
<td>-.173</td>
<td>.042</td>
</tr>
<tr>
<td>Health and social work</td>
<td>.076</td>
<td>.03</td>
</tr>
<tr>
<td>Other service activities</td>
<td>.264</td>
<td>.057</td>
</tr>
<tr>
<td>Extraterritorial organizations and bodies</td>
<td>.024</td>
<td>1.315</td>
</tr>
</tbody>
</table>
Table 2.12: Estimates of Tau by Ownership Type

<table>
<thead>
<tr>
<th>Ownership Type</th>
<th>Tau</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government and public sector undertaking</td>
<td>-.092</td>
<td>.028</td>
</tr>
<tr>
<td>Non profit institution</td>
<td>-.04</td>
<td>.038</td>
</tr>
<tr>
<td>Unincorporated proprietary</td>
<td>.43</td>
<td>.087</td>
</tr>
<tr>
<td>Unincorporated partnership</td>
<td>-.058</td>
<td>.028</td>
</tr>
<tr>
<td>Corporate non financial</td>
<td>-.197</td>
<td>.026</td>
</tr>
<tr>
<td>Corporate financial</td>
<td>-.18</td>
<td>.023</td>
</tr>
<tr>
<td>Co-operative</td>
<td>-.007</td>
<td>.022</td>
</tr>
</tbody>
</table>

Appendix 2: Anecdotal Evidence Regarding Harassment Bribery from “ipaidabribe.com”

“I am a small factory owner in Kirti Nagar Industrial Area. We follow almost all rules laid down by government for the welfare of workers. Now, even if we follow everything there is always somethings where we lack and which needs improvement. We have a factory inspector by the name of Mr.R.B.Singh (M: 9818829355). He comes to all the factories in our area, inspects them, find mistakes and then harass and blackmails us. According to him he can get our factories sealed. To avoid this, to save our time and to save the unnecessary paperwork we pay him every year. I have paid him twice in two years i.e. 10000 & 15000 and this is common with all factories. Please take a strict action against him so that he learns a lesson. I am sure he is not alone. All his colleagues are equally corrupt.”

(Reported on August 11, 2014 from New Delhi, Delhi — Report #131791)

“During the routine labor verification process by the labor department at our office, we
were advised by the consultant to pay the labor inspector a bribe to ensure that they don’t keep calling us for needless paperwork.”

(Reported on June 28, 2011 from Chennai, Tamil Nadu — Report #35064)

“The Labour Department requires a dozen odd registers to be maintained some of them which are totally outdated and pointless. E.g: Salary register, Attendance register, Leave register etc.

Our IT office has an electronic system that logs all entries/exits and leave taken. We have the records and offered to provide it to them in a printout.

Salaries are paid electronically via bank transfer.

The officer declined and said it must be maintained in a manual register!

Finally an arrangement was made where we maintain a few records manually and the rest he would overlook.

Cost of arrangement Rs 1500 twice a year even if the officer shows up only once a year for the inspection!

He is supposed to inspect twice so expects to be paid even for the time he did not show up!” (Reported on October 13, 2010 from Chennai, Tamil Nadu — Report #44950)

“Well i had gone to renew my labour license and after all the running around in the bank and the department, the signing authority asked me to pay Rs.500 for signing. When asked why 500, i was told since there are 5 employess for Rs.100 each.” (Reported on December 31, 2010 from Hyderabad, Andhra Pradesh — Report #43509)

“... in my third visit i met one of office peon in Labour office he guided me for the bribe he also investigated and advised me for bribe according to the number of Employees deployed on contract basis and for this valueble suggestion he charged me Rs. 100. Again with full confidence i went to the ALCs desk and straight away i offered him the packet which was contains the amount of Bribe Rs. 3000/- ... He issued me the license after office hours...” (Reported on March 30, 2011 from Mumbai, Maharashtra — Report #39133)
“Applying for shop & establishment [registration] & procured all documents relating to the registration. Finally inspectors are asking Rs.1000 as a bribe. If any other notice received by the company for resolving that another Rs.2000 and above, it depends on the company” (Reported on March 28, 2014 from Bangalore, Karnataka — Report #99016)

“Officer name Naveen Kumar. Mobile no. 9468104694- He is asking for a bribe of 60,000 and is saying will issue a negative report under labour laws.” (Reported on January 24, 2014 from Gurgaon, Haryana — Report #83365)

Appendix 3: Modeling Extortion (ie: Harassment Bribery)

In this Appendix we model size-based regulations in an environment where corrupt inspectors use the fact that de jure regulations are numerous, complex and burdensome in order to extort bribes from firms (ie: harassment bribery). The model aims to illustrate how corruption may lead to higher per worker costs for firms that exceed the 10 worker threshold, while being as parsimonious as possible. The set-up and timeline is described below and in Figure 2.8.

First, firms must choose their number of workers ($n \geq 10$ or $n' < 10$).\textsuperscript{47} As in Section 2.4.1, firms are characterized by a productivity parameter $\alpha$, so that firms with higher productivity would like to choose higher $n$. If firms choose $n$ greater than or equal to 10, they come under the legal purview of size-based regulations, which makes it more difficult for them to appeal extortionary practices on the part of inspectors. After choosing a level of employment, firms are randomly matched with an inspector. With probability $\kappa$, the inspector is corrupt; otherwise the inspector is honest. An honest inspector will enforce a reasonable interpretation of the spirit of the regulations if the firm has more than 10 workers. To be compliant with this “reasonable” interpretation of the regulations will cost

\textsuperscript{47}Throughout, primes will denote the values of variables on the side of the decision tree in which firms hire less than 10 workers.
the firm an amount \( F_1(n) \), which may in general depend on the number of workers in the firm. A firm with fewer than 10 workers incurs no regulatory costs if matched with an honest inspector.

If the firm is instead matched with a corrupt inspector, the inspector will threaten to report the firm for technical infractions unless it pays a bribe (which we denote \( b \) or \( b' \), depending on whether the firm has chosen \( n \geq 10 \) or \( n' < 10 \)), the value of which is determined by Nash Bargaining. The firm may choose to pay the bribe or appeal the threatened fine in court. If appealing the fine in court, the firm will win with probability \( p \) (or \( p' \)) but will incur legal fees \( (c_L) \) with certainty. If it wins the case, the firm has no further financial obligations. If the firm loses, it is obliged to pay an amount \( F_2(n) \), which we take to be much larger than \( F_1(n) \). This last assumption is tantamount to supposing that a reasonable level of compliance with regulations is not extremely costly in comparison to the punishments that could be brought by an inspector for violating the regulations - which may be a reasonable assumption in contexts where inspectors have a great amount of bargaining power and/or punishments can involve prison sentences. The assumption is also necessary for the framework to be one of extortion rather than collusion: if \( F_1(n) \) were large in comparison with \( F_2(n) \), firms would benefit from collusion and would face lower costs with corrupt inspectors than with honest ones. It is also plausible that both \( F_2(n) \) and \( F_1(n) \) are increasing functions of the number of workers, especially if we acknowledge that the full cost of any fine would include the opportunity cost of a manager’s time. We will consider the case where the total fines are directly proportional to the number of workers: \( F_i(n) = f_i \times n \). The decision tree representing the firm’s choices described above is provided in Figure 2.8.
An important assumption is that \( p' \), the probability of a firm’s winning the case when \( n' < 10 \), is much higher than \( p \), the probability of winning the case when \( n \geq 10 \). The idea is that a firm with less than 10 workers is not under the legal purview of the regulations, so any case regarding regulatory infractions brought against the firm would have no standing in court. In what follows, we will take \( p = 0 \) and \( p' = 1 \) for simplicity. As previously mentioned, if the firm meets a corrupt inspector, the value of the bribe paid to avoid going to court is determined through a process of Nash Bargaining over the surplus, where \( \alpha \) and \( \beta \) are the bargaining weights of the inspector and firm, respectively:

\[
max_b (b)^\alpha (c_L + (1 - p)F_2(n)) - b)^\beta
\]

The solution of this maximization problem is that \( b = \frac{\alpha(c_L+(1-p)F_2(n))}{\alpha+\beta} \) (and \( b' = \frac{\alpha(c_L+(1-p')F_2(n))}{\alpha+\beta} \), for firms with less than 10 workers). Given that firms meet corrupt in-
Inspectors (and thus pay bribes) with probability $\kappa$ and meet honest inspectors (and thus pay $F_1(n)$) with probability $1 - \kappa$, the expected cost for a firm with greater than 10 workers is $\kappa b + (1 - \kappa)F_1(n)$, while the expected cost for a firm with less than 10 workers is $\kappa b'$. Taking the difference and substituting in our expressions for $b$ and $b'$, we get that firms that cross the 10 worker threshold face an increase in expected costs of $\kappa \frac{\alpha}{\alpha + \beta} (p' - p)F_2(n) + (1 - \kappa)F_1(n)$.

We are interested, however, in the increase in per worker costs that firms face when exceeding the 10 worker threshold, not the increase in total costs (as discussed earlier, an increase in per worker costs is the only way to produce a downshift in the logged firm size distribution in a static model). Thus, we divide the last result by the number of workers, $n$, to get per worker costs. Before doing so, we make the further simplifications that $p = 0$, $p' = 1$, $\alpha$ and $\beta$ both equal 1 (equal bargaining weights), and that all fines are proportional to firm size ($F_i(n) = f_i * n$). Then, the increase in per worker costs for firms that exceed the 10 worker threshold is $\kappa \frac{f_2}{2} + (1 - \kappa)f_1$.

From the last result we see that if $f_2 \gg f_1$ (in particular, in this case, if $f_2 > 2f_1$), then the increase in a firm’s per worker costs for exceeding the 10 worker threshold (i.e.: what we call $\tau$ in the paper) is increasing in the proportion of corrupt inspectors, $\kappa$. Again, that we are considering a context of extortion or “harassment bribery” is implied by the assumption $f_2 \gg f_1$. It is this condition (that $f_2$ is very large) that gives corrupt inspectors the power to extract heavy bribes. We think it is a reasonable assumption given the anecdotal evidence regarding bribery we have found (some of which we present in Appendix 2). To conclude, the model above illustrates conditions that may explain the correlations we observed between corruption and $\tau$ in Section 2.6. In particular, the conclusion of the model is that firms in states with a higher proportion of corrupt inspectors (i.e: more corrupt states), face higher per worker costs for exceeding the 10 worker threshold (higher $\tau$).
Chapter 3

Distributional Impacts of Dismantling the Small-Scale Reservation Policy in India

3.1 Introduction

This paper investigates the effects of dismantling a prominent industrial regulation in India: the Small Scale Reservation Laws (SSRL). The SSRL mandated that certain goods be exclusively produced by firms maintaining less than a specified level of investment in plant and machinery. The original rationale for the SSRL was to make Indian industry to be more labor intensive by constraining the use of capital, in the hopes that it would then provide more employment to unskilled labor leaving the agricultural sector. At the height of the reservation policy, approximately 1000 goods were reserved, making up nearly 25 percent of total manufacturing output (Tewari and Wilde (2014)).

Criticism of the SSRL eventually grew, particularly after the liberalization reforms of 1991, since the laws were argued to make Indian industry less efficient and therefore less competitive with foreign producers who could now enter the Indian market. Starting in 1997, the SSRL policy was gradually dismantled, a process we refer to as “dereservation.” Each year, a number of goods were removed from the list of reserved products so that firms producing them were free to exceed the capital thresholds. By 2008, only 20 products remained reserved - although, of the ever-reserved products, these never dereserved products account for a disproportionately large number of firms.

We examine the effect of dereservation of particular goods between 2001 and 2006 on the characteristics of the cross-section of firms producing the dereserved goods. We investigate the effect of the 2001 to 2006 dereservation because in 2006 the threshold was increased from Rs. 10,000,000 (roughly $200,000) to Rs. 50,000,000 (about $1 million). Dereservation was thus a different phenomenon from 2006 onwards since the capital constraint had become
much less binding and we would expect expansion even by still-reserved goods producers. Comparing the evolution of distributions of producer characteristics for dereserved goods to never-reserved goods using the Changes-in-Changes model of Athey and Imbens (2006), we find that dereservation led to a substantial increase in the value of capital employed among producers of previously-reserved goods. We find that producers hired more employees as their scale of production increased, but less than in proportion to capital. We show suggestive evidence that the increase in the value of capital employed comes from an increase in capital intensity among establishments that had been producing dereserved goods when they were still reserved.

The magnitude of the reservation and dereservation policies have prompted substantial interest from the academic community. From a macroeconomic perspective, there has been particular interest in investigating the role reservation may have played in misallocation across firms as in Hsieh and Klenow (2009). Bollard, Klenow, and Sharma (2013) perform industry-level regression analysis and find that the fraction of output dereserved does not correlate with reallocation to more productive firms. Garcia-Santana and Pijoan-Mas (2014) theoretically examine the problem of occupational choice in a dual sector economy, with quantitative results calibrated to match the Indian reservation context. In their model, removing reservation policy can generate substantial gains in TFP. In this paper, we provide more rigorous microeconometric evidence on the effects of the policy, with the intention of building a bridge to more theory-based work.

Since beginning work on the paper, we have learned about two other working papers that investigate the process of dereservation within a difference-in-differences program evaluation framework. Tewari and Wilde (2014) look at the effect of dereservation at the industry level and find increases in productivity. In the closest paper to ours, Martin, Nataraj, and Harrison (2014) primarily investigate the within-plant effect of producing a dereserved good on various plant-level outcomes, finding small or negative effects on outcomes for firms already producing a dereserved good at the time of dereservation (i.e.: incumbents).
and positive effects on outcomes for firms that only begin production of the good after
dereservation (i.e.: entrants). However, even with respect to Martin et al. (2014), our paper
differs in methodology, objectives, and findings.

While Tewari and Wilde (2014) and Martin et al. (2014) operate within a linear, para-
metric framework, we provide the first non-parametric evidence of effects of the policy.
Non-parametric evidence is particularly important to establish a credible impact of the
deregulation given reasons to be concerned about the extent to which the capital enforce-
ment was binding, due to poor enforcement or plants’ achieving efficient scale below the
threshold (see the discussion around figure 3.2). In fact, doubts about the extent to which
the regulation was binding were a feature of policy discussions at the time of dereservation
(National Productivity Council (2009)). Further, we show in Appendix 3.9 that linear,
parametric estimates are highly sensitive to the estimation sample used, particularly with
respect to the units considered for comparison with the dereserved plants. In contrast,
our non-parametric estimates for the average effect of dereservation on capital and labor
are robust to alternative choices of comparison groups. We also focus on the effects of
dereservation alone while Tewari and Wilde (2014) and Martin et al. (2014)’s estimated
effects combine the effect of dereservation and the 2006 increase in the capital threshold for
still-reserved producers.\footnote{Martin et al. (2014) do look at longer-run effects, but using district-level variation in the fraction of reserved products in 2000, so their two sets of results are not directly comparable.}

The rest of the paper is organized as follows. Section 3.2 provides a brief explanation
of the institutional background, Section 3.3 describes the data used in our analysis, section
3.4 a descriptive analysis of reservation and dereservation, section 3.5 the empirical specifi-
cations employed to test for the effect of dereservation and section 3.6 the results. Section
3.7 provides an analysis of the effect of dereservation on plant exit. Finally, section 3.8
concludes.
3.2 Institutional Background: Small Scale Reservation in India

The definition of the Small Scale Industry (SSI) Sector has undergone several changes over time. The definition was first set out in the Industries (Development and Regulation) Act, 1951, to include establishments that do not exceed a prescribed threshold “defined in terms of investment limits in plant and machinery”. This threshold was revised on a number of occasions - usually upward, with the goal of keeping up with inflation, although the changes do not track inflation precisely. The nomenclature has also changed, so that what were previously known as “Small Scale Industries” are now referred to in government documents as “Small and Micro Enterprises”. The relevant changes in the SSI definition over the recent period are given in Table 3.1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital Limit (Millions of Rs.):</td>
<td>30</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Nomenclature:</td>
<td>“Small Scale Industry”</td>
<td>“Small Scale Industry”</td>
<td>“Small Enterprise”</td>
</tr>
</tbody>
</table>

Although the SSI sector was the focus of a large array of policies (including preferred access to government contracts and finance), product reservation was probably the most controversial of all SSI policies. The policy required all firms producing certain items to be classified as SSI enterprises (or as “Small Enterprises”, in more recent nomenclature) - and thus to maintain a level of investment in plant and machinery less than the thresholds given in Table 3.1. If an enterprise was already above the capital threshold when a good it was producing was placed on the reservation list, it could be grandfathered in and allowed to remain in production - although its capital and production levels would be capped at their current levels. Another exception to the rule was provided for enterprises that exported 50% or more of their production (National Productivity Council (2009), p 23).
The policy of reservation began in 1967, when 47 items were placed on the SSI reservation list. By 1978, this list had grown to 807 items (National Productivity Council (2009), p. 23). The exact rationale behind how and when certain items were placed on the reserved list is not entirely clear, although government reports argue that “[t]he overwhelming consideration for Reservation of an item was its suitability and feasibility for being made in the small scale sector without compromising quality aspects” (National Productivity Council (2009)). The process by which certain goods were dereserved is similarly unclear, although it seems the decisions were taken in consultation with certain members of industry. Furthermore, we read that:

“The Advisory Committee makes its recommendations for reservation/de- reservation in light of the factors like economies of scale; level of employment; possibility of encouraging and diffusing entrepreneurship in industry; prevention of concentration of economic power and any other factor which the Committee may think appropriate.” (Ministry of Micro, Small and Medium Enterprises (2013), p.40)

Therefore, the process of dereservation may not have been entirely independent of anticipated future product characteristics, although the precise rationale behind the decision to dereserve certain items rather than others at certain points in time is certainly hard to discern (Tewari and Wilde (2014)).

3.3 Data

Our main source of data on factory level outcomes is the Annual Survey of Industries (ASI). The ASI covers the organized manufacturing sector in India, where organized refers to manufacturing establishments registered with the state Directorate of Factories. The ASI is representative at the state level, with establishments employing less than 100 workers sampled from the listing maintained by the state Directorates of Factories and a complete

---

According to the Factories Act (1947), a factory is required to be registered with Directorate (and be thus classified as “organized”) if it has at least 10 workers and uses power (or if has at least 20 workers and does not use power).
enumeration of establishments employing more than 100 workers. We make use of the 2001-2008 waves of the ASI. The year of a wave refers to the “fiscal year” (so that 2001 refers to the fiscal year from April 1st, 2000 to March 31st, 2001). Establishments present in repeated waves of the survey can be tracked through a unique identifier. Establishments in the ASI may list several goods produced, each identified by ASICC code\textsuperscript{3}.

The ASI is made up of two distinct data sets. The first, called the census sector, includes plants employing 100 or more workers. These plants are surveyed every year. The second, referred to as the sample sector, includes all other plants, with each plant having a 20% probability of being sampled in a given survey year.

We constructed a mapping between the 1987 National Industrial Classification (NIC) codes used by the Ministry of Micro, Small and Medium Enterprises (MSME)\textsuperscript{4} to identify reserved products in official documents and the Annual Survey of Industries Commodity Classification (ASICC) codes used to identify products in the Indian establishment-level surveys used in the analysis and described below. We make our mapping from 1987 NIC codes to ASICC codes because the official concordance from 1987 NIC codes to 1998 NIC codes (which can be mapped through an official concordance to 2004 NIC codes) is from 3-digit 1987 NIC code to 4-digit 1998 NIC codes. The MSME Ministry identifies products by their 8-digit 1987 NIC code, so the official concordance is too coarse a mapping. Our basic procedure, similar to the approach described in Martin et al. (2014), is then to manually map 8-digit 1987 NIC codes to 5-digit 2004 NIC codes. We use an official concordance between 2004 NIC codes and ASICC codes to pick out the appropriate ASICC codes. The final mapping identifies the ASICC code for all products reserved in 1997, when the number of reserved products was greatest.

\textsuperscript{3}We have ASI data beginning with the 1999 wave, but like Martin et al. (2014), we discard the 1999 and 2001 waves due to very poor data about products produced by factories.

\textsuperscript{4}Prior to 2001, the MSME ministry was known as the Ministry of Small Scale Industries.
3.4 Descriptive analysis

In this section, we investigate the distribution of plants by a measure of their capital stock (specifically, investment in plant and machinery) for several different reservation statuses. In 2001 we compare the distributions of capital stock by plants in two categories:

- “never reserved” - producing an item that has never been on the reservation list. This is the control group in the pre-treatment period\(^5\).
- “reserved” - producing an item on the reservation list in 2001 that will be dereserved between 2001 and 2006. Firms in the reserved category are subject to the SSI constraints in 2001. This is the treatment group, before receiving treatment.

In arguing for the use of “never reserved” as a valid proxy for the changes occurring over time in the outcomes we will investigate, we point out that the distribution of factor cost ratios was quite similar across the two groups of producers in 2001 as shown in figure 3.1. Our results in this and subsequent sections are robust to restricting our analysis to producers of certain product categories\(^6\) with fractions of producers of goods to be dereserved lying in specified intervals. Setting a minimum fraction of goods to be dereserved makes it more likely that treatment and control group producers experience the same demand and supply shocks. This must be balanced against product classes where the fraction of producers to be dereserved is sufficiently large that we are concerned that dereservation would have effects on never-reserved goods producers, for example by bidding up the price of inputs. Therefore, we have experimented with providing minimum and maximum fractions of dereserved goods producers, with qualitatively similar results\(^7\). Appendix 3.10 shows the fractions of producers in each product category in 2001.

\(^5\)We can also use “already dereserved” firms as the control group. That is, firms producing goods that were dereserved prior to 2001. The results in section 3.6 are quite similar with this alternative control group.

\(^6\)We use 2- and 3-digit ASICC codes to delineate product categories.

\(^7\)Specifically, we have tried excluding producers of goods with less than 5 and 10% of producers manufacturing dereserved goods and well as excluding producers of goods with less than 5% and more than 20%.
Figure 3.1 shows kernel density plots of the distributions of the natural log of the nominal value of capital for these two categories. Surprisingly, the plants producing never-reserved goods are slightly smaller than those producing reserved goods. The vertical line shows the threshold level of capital stock in this and the subsequent figures. It is notable that both distributions are relatively smooth, particularly around the threshold. Given that firms were expected to follow different rules if operating above the threshold, one might expect some change in the shape of the distributions around the threshold. The smoothness of the distributions around the threshold and the substantial mass of plants operating above the threshold may lead observers to question whether the regulations were actually binding.
In 2006, we look at two categories:

- “never reserved” - the control group, in the post-treatment period.
- “dereserved” - producing an item that was previously on the reservation list in 2001 but has since been removed and are thus no longer subject to the SSI constraints. This is the treatment group, after receiving treatment.

Figure 3.3 shows kernel density plots of the distributions of log capital for these two categories. In contrast to figure 3.2, the distribution of capital for dereserved firms shows a bulge to the right of the capital threshold, which we will identify as the effect of dereservation after accounting for the effect of time as captured by the shift of the never-reserved distribution from 2001 to 2006. Figure 3.3 is notable for two reasons. First, the figure provides the first non-parametric evidence of an effect of the dereservation. It is difficult to explain the distortion observed in the distribution occurring right at the regulatory threshold for other reasons and the size of the distortion grows over time from 2005, when a fraction of products
was already dereserved, to 2008. Second, the disproportionate expansion of a portion of the distribution of reserved firms suggests that only some producers took advantage of the dereservation. Whether this is due to selection or information frictions is left as a question for future work.

Figure 3.3: 2006 density of log(capital)

Figure 3.4 emphasizes this point further, separating the dereserved plants into two categories:

- “incumbents” who produced a dereserved product before it was dereserved.
- “entrants” who only produce a dereserved product after it is dereserved.
We see that the bulge in the dereserved distribution in figure 3.3 is a result of the distribution of capital for incumbent firms.

Figures 3.5 - 3.7 show local linear regressions of the (capital/labor) factor cost ratio for the treatment group (reserved good producers in 2001, dereserved good producers in 2006), split in figures 3.6 and 3.7 by incumbent and entrant status. Figure 3.6 suggests that the expansion in the value of capital stock among incumbents just above the threshold was not accompanied by a concomitant increase in the wage bill. In contrast, the 2006 entrants shown in figure 3.7 display an expansion path of factor cost ratios more similar to that of reserved goods producers in 2001.
Figure 3.5: 2001, factor cost ratio (capital/labor) by log(capital) for producers of goods to be dereserved

Figure 3.6: 2006, factor cost ratio (capital/labor) by log(capital) for dereserved products - incumbents
Figure 3.7: 2006, factor cost ratio (capital/labor) by log(capital) for dereserved products - entrants

3.5 Empirical specification

We will look at the effect of dereservation on the distribution of the regulated firm attribute, capital stock. To this end, we use the Changes-in-Changes (CIC) model of Athey and Imbens (2006) (AI) to estimate the effect of dereservation on the distribution of capital stock. To capture the effect of time (from 2001 to 2006) on the distribution of capital stock, we use the evolution of the distribution of capital for never-reserved goods producers.

In our context, the AI approach is based on the assumption that, in the absence of dereservation, an establishment’s capital stock is generated by the following function.

\[ Y_t = h_t(U) \]

\( h_t(\cdot) \) is strictly increasing in \( U \), a producer-specific unobserved attribute. In the absence of dereservation, we must further assume that the distribution of \( U \) for producers of both sets of goods (dereserved from 2001 to 2006 and never-reserved) is unaffected by time. These two assumptions allow us to recover the distribution of \( Y_t \) under reservation for dereserved goods producers in period \( t' \) as \( F_{Y_t|G=d} \left( F_{Y_{t'}|G=n}^{-1} \left( F_{Y_{t'}|G=n}^{-1} (y_{t'}|G=n) | G = n) | G = d) \right) \) where \( G = \{n, d\} \).
is a random variable indexing producers of never-reserved and reserved/dereserved goods, respectively. We use this expression to compute the counterfactual distribution of producer attributes under product reservation in 2006 for dereserved products.

### 3.6 Results

Table 3.2 and figure 3.8 show the result of applying the distributional analysis described in the previous section. The observed distributions in figure 3.8 are as in figures 3.2 and 3.3: we see again that reserved firms are larger than never-reserved firms in 2001 (bold lines) and 2006 (fine lines). Dereserved (to be dereserved in 2001) establishments’ distributions of capital stock are shown with solid lines, never-reserved with dashed lines. The solid line with asterisk markers represents the counterfactual outcome for producers of dereserved goods, had the goods not been dereserved. The counterfactual is computed by assuming that the distribution of capital for reserved products would have undergone the same transformation from 2001 to 2006 as the distribution for never reserved products. The average effect of derevation is estimated at 0.2308 and is significant at the 10% level.

<table>
<thead>
<tr>
<th>log(real capital)</th>
<th>log(employment)</th>
<th>log(output)</th>
<th>log(real wage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2308</td>
<td>0.0806</td>
<td>0.0569</td>
<td>-0.0390</td>
</tr>
<tr>
<td>(0.1202)</td>
<td>(0.0746)</td>
<td>(0.1105)</td>
<td>(0.0298)</td>
</tr>
</tbody>
</table>

Note: standard errors in parentheses, based on 100 replications of the bootstrap procedure described in Athey and Imbens (2006).
We note that the effect on log(capital) is twice as large as the largest effect on capital reported in Martin et al. (2014), a fact we attribute to their effect being on capital in any year following dereservation. If the effect of dereservation on capital is increasing over time, as we show is the case in the following subsection, their effect may place more weight on smaller short-run impacts while our estimates place more weight on the medium-run impacts of the policy.

The results for other outcomes are smaller in magnitude and not statistically significant. There is weak evidence that plants expanded their use of labor along with capital, which would allay policymakers’ fears of a decrease in employment following a reallocation of
resources from smaller, more labor-intensive plants to larger, more capital-intensive plants. Given the shape of the effect on the distribution of capital, we conjecture that it may be possible to gain power to investigate the effect of the deregulation on other outcomes with this comparison group by looking at effects on plants operating just above and below the capital threshold.

3.6.1 Medium- and long-term effects comparing with “already dereserved” producers

Using “already dereserved” producers (producers of goods dereserved before 2001) as an alternative control group, table 3.3 shows that we arrive at an almost identical average effect on log(capital) by 2006. The average effect using this control group is less precisely-estimated, however. There are fewer “already dereserved” goods producers, making estimation of distributions of their characteristics less precise. Effects on employment are again similar, but effects on output with this comparison group are much larger.

Table 3.3: CIC results: 2006 impacts - dereserved 2001-2006 vs. dereserved before 2001

<table>
<thead>
<tr>
<th>log(real capital)</th>
<th>log(employment)</th>
<th>log(output)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2259</td>
<td>0.1228</td>
<td>0.3733</td>
</tr>
<tr>
<td>(0.1555)</td>
<td>(0.0865)</td>
<td>(0.1264)</td>
</tr>
</tbody>
</table>

Note: standard errors in parentheses, based on 100 replications of the bootstrap procedure described in Athey and Imbens (2006).

By 2008 (table 3.4), the impact of the 2001-2006 dereservation had become so large that it is statistically significant for all three outcomes. The evolution of the difference in the distribution of capital is shown in figures 3.9 - 3.11. The pattern of responses between capital and labor is consistent across tables 3.2, 3.3 and 3.4: capital increases by about twice as much as labor. This contrasts with the results from Martin et al. (2014), where effects on capital and labor are similar in magnitude. The difference may be due to
Martin et al. (2014)’s inclusion of the 2006-2008 dereservations, when estimating the effect of dereservation is complicated by the large increase in the threshold value of capital in 2006. Including these dereservations could mute the effect of dereservation on capital by comparing the evolution of dereserved goods producers with still-reserved goods producers who are also increasing their capital usage.

Table 3.4: CIC results: 2008 impacts - dereserved 2001-2006 vs. dereserved before 2001

<table>
<thead>
<tr>
<th></th>
<th>log(real capital)</th>
<th>log(employment)</th>
<th>log(output)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.4954</td>
<td>0.2285</td>
<td>0.5906</td>
</tr>
<tr>
<td></td>
<td>(0.2187)</td>
<td>(0.1000)</td>
<td>(0.1307)</td>
</tr>
</tbody>
</table>

Note: standard errors in parentheses, based on 100 replications of the bootstrap procedure described in Athey and Imbens (2006).
Figure 3.10: 2006 density of log(capital)

Figure 3.11: 2008 density of log(capital)
3.7 Analysis of Exit

Table 3.5 shows the total number of factories operating according to ASI estimates in 2001 and 2006. Exit rates, derived by taking the percentage change in the number of factories in 2006 after subtracting factories with opening dates after 2001, are given in the third column. We see that there were differences in exit rates between producers of dereserved products and already-dereserved products, but that exit rates were similar between dereserved and never- and always-reserved products. In other words, dereservation does not seem to be associated with higher or lower levels of exit in comparison with most of the groups that did not experience dereservation over the period.

<table>
<thead>
<tr>
<th></th>
<th>Total number in 2001</th>
<th>Total number in 2006</th>
<th>Total number in 2006 - excluding new factories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(% change)</td>
<td>(% change)</td>
</tr>
<tr>
<td>Always reserved</td>
<td>7379</td>
<td>7687</td>
<td>6482</td>
</tr>
<tr>
<td></td>
<td>(4.2)</td>
<td>(-12.2)</td>
<td></td>
</tr>
<tr>
<td>Always dereserved</td>
<td>11138</td>
<td>13398</td>
<td>10535</td>
</tr>
<tr>
<td></td>
<td>(20.3)</td>
<td>(-5.4)</td>
<td></td>
</tr>
<tr>
<td>Never reserved</td>
<td>90883</td>
<td>95064</td>
<td>77122</td>
</tr>
<tr>
<td></td>
<td>(4.6)</td>
<td>(-15.1)</td>
<td></td>
</tr>
<tr>
<td>Dereserved during (treatment)</td>
<td>2935</td>
<td>3226</td>
<td>2496</td>
</tr>
<tr>
<td></td>
<td>(9.9)</td>
<td>(-15.0)</td>
<td></td>
</tr>
</tbody>
</table>

3.8 Conclusions

In this paper, we provided non-parametric evidence of the impact of removing restrictions on the value of capital stock for producers of specific goods in India over the period from
2001 to 2006. We saw that the removal of restrictions led to a large expansion in the value of the capital stock employed by producers of the newly-unrestricted goods and that the effect grew substantially over time. In contrast to the previous literature, we find that the expansion appears to have come mostly from firms who were already producing reserved goods before dereservation. We find that employment rose due to an increase in the scale of production, but much less so than capital, as firms producing dereserved goods became more capital-intensive. The worst fear of policymakers, a fall in employment, thus did not occur over the period of our study.

The period we study, from 2001 to 2008, was also a period of high growth, however. It seems plausible that the increase in capital intensity we observe would be sustained during in a period of slower growth and would lead to slower employment growth in the previously-reserved sector than if reservation had been maintained. This is an important subject for future research on the dereservation as more data become available.

3.9 Appendix: Linear Differences in Differences - the effect of varying the estimation sample

All specifications consider two-period linear difference-in-difference specifications, with 2001 as the pre-treatment period and 2006 as the post-treatment period.
Table 3.6: Capital vs Dereservation Status

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln_fix_cap_cl_nm</td>
<td>0.0949</td>
<td>0.0804</td>
<td>0.469*</td>
<td>0.0751</td>
</tr>
<tr>
<td>(0.0711)</td>
<td>(0.0769)</td>
<td>(0.269)</td>
<td>(0.0613)</td>
<td></td>
</tr>
<tr>
<td>2006.year</td>
<td>0.105***</td>
<td>0.124***</td>
<td>-0.137</td>
<td>0.0919***</td>
</tr>
<tr>
<td>(0.0175)</td>
<td>(0.0375)</td>
<td>(0.235)</td>
<td>(0.0172)</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>13.34***</td>
<td>13.21***</td>
<td>13.16***</td>
<td>13.38***</td>
</tr>
<tr>
<td>(0.0110)</td>
<td>(0.0131)</td>
<td>(0.0220)</td>
<td>(0.00891)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>59460</td>
<td>34702</td>
<td>26212</td>
<td>57242</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

Year and firm FEs included.

1st col includes all factories, 2nd excludes never reserved,
3rd excludes still reserved and 4th excludes previously dereserved

* p < 0.10, ** p < 0.05, *** p < 0.01
Table 3.7: Capital vs Dereservation Status (interacted with readymade garments indicator)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln_fix_cap_cl_num</td>
<td>0.0912</td>
<td>0.0776</td>
<td>0.732**</td>
<td>0.0562</td>
</tr>
<tr>
<td></td>
<td>(0.0844)</td>
<td>(0.0886)</td>
<td>(0.301)</td>
<td>(0.0707)</td>
</tr>
<tr>
<td>dereserved</td>
<td>-0.215</td>
<td>-0.267</td>
<td>0.120</td>
<td>-0.226</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.200)</td>
<td>(0.441)</td>
<td>(0.192)</td>
</tr>
<tr>
<td>readymade</td>
<td>0.0299</td>
<td>0.0294</td>
<td>-0.393</td>
<td>0.0792</td>
</tr>
<tr>
<td></td>
<td>(0.149)</td>
<td>(0.149)</td>
<td>(0.252)</td>
<td>(0.141)</td>
</tr>
<tr>
<td>deresXreadymade</td>
<td>0.105***</td>
<td>0.124***</td>
<td>-0.137</td>
<td>0.0919***</td>
</tr>
<tr>
<td></td>
<td>(0.0175)</td>
<td>(0.0375)</td>
<td>(0.235)</td>
<td>(0.0172)</td>
</tr>
<tr>
<td></td>
<td>(0.0122)</td>
<td>(0.0159)</td>
<td>(0.0287)</td>
<td>(0.00966)</td>
</tr>
<tr>
<td>N</td>
<td>59460</td>
<td>34702</td>
<td>26212</td>
<td>57242</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
Year and firm FEs included.
1st col includes all factories, 2nd excludes never reserved,
3rd excludes still reserved and 4th excludes previously dereserved

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 3.8: Capital vs Dereservation Status (including interaction with entrant/incumbent status)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln_fix_cap_cl_num</td>
<td>0.0469</td>
<td>0.0386</td>
<td>0</td>
<td>-0.199</td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td>(0.157)</td>
<td>(.)</td>
<td>(0.147)</td>
</tr>
<tr>
<td>deresXent</td>
<td>0.123*</td>
<td>0.107</td>
<td>0.469*</td>
<td>0.134**</td>
</tr>
<tr>
<td></td>
<td>(0.0646)</td>
<td>(0.0711)</td>
<td>(0.269)</td>
<td>(0.0645)</td>
</tr>
<tr>
<td>2006.year</td>
<td>0.105***</td>
<td>0.122***</td>
<td>-0.137</td>
<td>0.0922***</td>
</tr>
<tr>
<td></td>
<td>(0.0175)</td>
<td>(0.0375)</td>
<td>(0.235)</td>
<td>(0.0172)</td>
</tr>
<tr>
<td>_cons</td>
<td>13.35***</td>
<td>13.22***</td>
<td>13.22***</td>
<td>13.40***</td>
</tr>
<tr>
<td></td>
<td>(0.0184)</td>
<td>(0.0277)</td>
<td>(0.00909)</td>
<td>(0.0119)</td>
</tr>
<tr>
<td>N</td>
<td>59460</td>
<td>34702</td>
<td>26212</td>
<td>57242</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
Year and firm FEs included.
1st col includes all factories, 2nd excludes never reserved,
3rd excludes still reserved and 4th excludes previously dereserved

* p < 0.10, ** p < 0.05, *** p < 0.01
Table 3.9: Capital vs Dereservation Status (including interaction with entrant/incumbent status - excluding readymade garments)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln fix cap cl num</td>
<td>0.0407</td>
<td>0.0320</td>
<td>0</td>
<td>-0.241</td>
</tr>
<tr>
<td>(0.161)</td>
<td>(0.161)</td>
<td>(.)</td>
<td>(0.157)</td>
<td></td>
</tr>
<tr>
<td>deresXent</td>
<td>0.137*</td>
<td>0.121</td>
<td>0.728**</td>
<td>0.148**</td>
</tr>
<tr>
<td>(0.0729)</td>
<td>(0.0782)</td>
<td>(0.301)</td>
<td>(0.0728)</td>
<td></td>
</tr>
<tr>
<td>2006.year</td>
<td>0.105***</td>
<td>0.124***</td>
<td>-0.137</td>
<td>0.0926***</td>
</tr>
<tr>
<td>(0.0175)</td>
<td>(0.0377)</td>
<td>(0.235)</td>
<td>(0.0172)</td>
<td></td>
</tr>
<tr>
<td>cons</td>
<td>13.34***</td>
<td>13.20***</td>
<td>13.19***</td>
<td>13.39***</td>
</tr>
<tr>
<td>(0.0179)</td>
<td>(0.0268)</td>
<td>(0.00597)</td>
<td>(0.0114)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>58489</td>
<td>33746</td>
<td>25534</td>
<td>56271</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

Year and firm FEs included.

1st col includes all factories, 2nd excludes never reserved,
3rd excludes still reserved and 4th excludes previously dereserved

* p < 0.10, ** p < 0.05, *** p < 0.01
Table 3.10: Capital vs Dereservation Status (including interaction with entrant/incumbent status - including only readymade garments)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ln_fix_cap_cl_num</td>
<td>ln_fix_cap_cl_num</td>
<td>ln_fix_cap_cl_num</td>
<td>ln_fix_cap_cl_num</td>
</tr>
<tr>
<td>deresXent</td>
<td>0.693</td>
<td>0.874**</td>
<td>0.693</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.424)</td>
<td>(0.411)</td>
<td>(. )</td>
<td>(0.424)</td>
</tr>
<tr>
<td>deresXinc</td>
<td>0.564</td>
<td>0.745**</td>
<td>0.206</td>
<td>0.564</td>
</tr>
<tr>
<td></td>
<td>(0.351)</td>
<td>(0.336)</td>
<td>(0.159)</td>
<td>(0.351)</td>
</tr>
<tr>
<td>2006.year</td>
<td>-0.366</td>
<td>-0.547*</td>
<td>0.206</td>
<td>-0.366</td>
</tr>
<tr>
<td></td>
<td>(0.330)</td>
<td>(0.313)</td>
<td>(0.159)</td>
<td>(0.330)</td>
</tr>
<tr>
<td>_cons</td>
<td>13.16***</td>
<td>13.15***</td>
<td>13.21***</td>
<td>13.16***</td>
</tr>
<tr>
<td></td>
<td>(0.0330)</td>
<td>(0.0323)</td>
<td>(0.000429)</td>
<td>(0.0330)</td>
</tr>
</tbody>
</table>

N \[ 25904 \] \[ 25889 \] \[ 25611 \] \[ 25904 \]

Standard errors in parentheses
Year and firm FEs included.

1st col includes all factories, 2nd excludes never reserved,
3rd excludes still reserved and 4th excludes previously dereserved

\* p < 0.10, \*\* p < 0.05, \*\*\* p < 0.01
### 3.10 Appendix: Fractions of plants producing in each 2-digit ASICC product class by category

<table>
<thead>
<tr>
<th>Product category</th>
<th>% of all plants producing goods in the category in the treatment group</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRUITS, VEGETABLES, CEREALS &amp; PULSES AND OTHER VEGETABLE PRODUCES LIKE LAC, GUM ETC. AND PREPARATION THEREOF</td>
<td>0.89</td>
</tr>
<tr>
<td>SALTS, SULPHER, PLASTERING MATERIALS, LIME &amp; CEMENT</td>
<td>0.07</td>
</tr>
<tr>
<td>MINERAL FUELS: OILS, PRODUCTS &amp; BY-PRODUCTS</td>
<td>14.15</td>
</tr>
<tr>
<td>INORGANIC CHEMICAL, COMPOUND OF PRECIOUS MATERIALS ETC</td>
<td>19.83</td>
</tr>
<tr>
<td>ORGANIC CHEMICALS</td>
<td>9.68</td>
</tr>
<tr>
<td>DYEING, TANNING, COLOURING, INK &amp; PAINT ETC</td>
<td>9.09</td>
</tr>
<tr>
<td>ESSENTIAL OIL, COSMETICS &amp; PERFUMES, DENTAL MATERIALS, WAX POLISHING/CLEANING MATERIALS</td>
<td>20.38</td>
</tr>
<tr>
<td>MISC CHEMICAL GOODS INCL ALBUMINOIDAL SUBSTANCES MODIFIED STARCHES (EXCL SUB-DIVISION 124) AND GLUES, ENZYMES</td>
<td>20.54</td>
</tr>
<tr>
<td>RUBBER AND MANUFACTURE OF RUBBER [THIS INCL SYNTHETIC RUBBER]</td>
<td>16.57</td>
</tr>
<tr>
<td>Description</td>
<td>Value</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>PLASTIC, PVC ARTICLES INCL PACKAGING PRODUCTS AND FOOTWEAR PLASTIC OR PVC</td>
<td>7.37</td>
</tr>
<tr>
<td>ARTICLE OF LEATHER, SADLARY AND HARNESS, TRAVEL GOODS, HAND BAGS AND SIMILAR PRODUCTS OF ANIMAL GUT</td>
<td>70.08</td>
</tr>
<tr>
<td>MISC PRODUCTS/ARTICLES OF RUBBER/PVC/LEATHER ETC</td>
<td>12.88</td>
</tr>
<tr>
<td>WOOD AND PULP OF WOOD AND PRODUCTS THEREOF EXCL FORESTRY</td>
<td>1.16</td>
</tr>
<tr>
<td>PAPER AND PAPER BOARD</td>
<td>26.07</td>
</tr>
<tr>
<td>PACKING MATERIALS AND CONTAINERS OF PAPER</td>
<td>9.06</td>
</tr>
<tr>
<td>WOOL/ANIMAL HAIR, YARN &amp; FABRICS</td>
<td>5.01</td>
</tr>
<tr>
<td>COTTON, COTTON YARN AND FABRICS</td>
<td>12.22</td>
</tr>
<tr>
<td>SYNTHETIC (MAN-MADE) AND MIXED TEXTILES</td>
<td>1.48</td>
</tr>
<tr>
<td>SPECIAL WOVEN FABRIC AND ARTICLES THEREOF</td>
<td>2.61</td>
</tr>
<tr>
<td>IRON &amp; STEEL (INCL STAINLESS STEEL) &amp; ARTICLES THEREOF</td>
<td>4.5</td>
</tr>
<tr>
<td>COPPER, NICKEL &amp; ZINC ARTICLES THEREOF</td>
<td>14.47</td>
</tr>
<tr>
<td>ALUMINIUM, TIN, LEAD &amp; OTHER BASE METALS &amp; ARTICLES THEREOF</td>
<td>19.32</td>
</tr>
<tr>
<td>MISC. MANUFACTURE OF BASE METALS, N.E.C</td>
<td>7.79</td>
</tr>
<tr>
<td>NON-ELECTRICAL MACHINE TOOLS &amp; GENERAL PURPOSE MACHINERIES AND COMPONENTS AND PARTS THEREOF</td>
<td>1.01</td>
</tr>
<tr>
<td>NON-ELECTRICAL INDUSTRY SPECIFIC EQUIPMENT/MACHINERIES INCL PARTS THEREOF</td>
<td>6.14</td>
</tr>
<tr>
<td>Description</td>
<td>Value</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>ELECTRICAL &amp; ELECTRONIC MACHINERY &amp; EQUIPMENT INCL PARTS (EXCL MEDICAL &amp; NON-CONVENTIONAL ENERGY EQUIPMENT)</td>
<td>2.39</td>
</tr>
<tr>
<td>ELECTRONICS EQUIPMENT &amp; PARTS EXCL BIO-MEDICAL EQUIPMENT</td>
<td>8.29</td>
</tr>
<tr>
<td>ROAD SURFACE VEHICLE EXCL RAILWAYS &amp; PARTS N.E.C</td>
<td>0.21</td>
</tr>
<tr>
<td>OPTICAL, PHOTOGRAPHIC, WATCH AND OTHER PRECISION EQUIPMENT, MUSICAL INSTRUMENTS AND PARTS THEOREOF</td>
<td>20.46</td>
</tr>
<tr>
<td>MISC NON CLASSIFIED MANUFACTURED ARTICLES &amp; PARTS</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>7.2</strong></td>
</tr>
</tbody>
</table>
### Appendix: example of dereservation within a 3-digit ASICC product class

Table 3.12: Product class (subdivision) 443 - Leather Footwear & Parts Thereof

<table>
<thead>
<tr>
<th>code</th>
<th>product name</th>
<th>year of dereservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>44301</td>
<td>CHAPPALS/SANDALS, LEATHER</td>
<td>2003</td>
</tr>
<tr>
<td>44302</td>
<td>FOOT WEAR, BOOT</td>
<td>never reserved</td>
</tr>
<tr>
<td>44303</td>
<td>FOOT WEAR, OTHERS, LEATHER</td>
<td>2001</td>
</tr>
<tr>
<td>44305</td>
<td>SHOE UPPER LEATHER SHEET</td>
<td>never reserved</td>
</tr>
<tr>
<td>44306</td>
<td>SHOE UPPER, LEATHER</td>
<td>2003</td>
</tr>
<tr>
<td>44311</td>
<td>SHOE SOLE, LEATHER</td>
<td>1999</td>
</tr>
<tr>
<td>44312</td>
<td>MID SOLE SHEET, LEATHER</td>
<td>never reserved</td>
</tr>
<tr>
<td>44313</td>
<td>INSOLE / OUTSOLE, LEATHER</td>
<td>never reserved</td>
</tr>
<tr>
<td>44315</td>
<td>EYE LETS, LEATHER</td>
<td>never reserved</td>
</tr>
<tr>
<td>44316</td>
<td>SHOE LINER, LEATHER</td>
<td>never reserved</td>
</tr>
<tr>
<td>44389</td>
<td>FOOTWEAR, LEATHER &amp; PARTS , N.E.C</td>
<td>never reserved</td>
</tr>
</tbody>
</table>
Bibliography


Basu, K. (2011). Why, for a Class of Bribes, the Act of Giving a Bribe should be Treated as Legal. *Mimeo*. 


Martin, L., S. Nataraj, and A. Harrison (2014). In with the Big, Out with the Small: Removing Small-Scale Reservations in India. Mimeo.


Curriculum Vitae

Michael Gechter
Boston University, Department of Economics
270 Bay State Rd
Boston MA 02215 USA
Mobile: +1-617-775-4059
Fax: +1-617-353-4449
Email: mgechter@bu.edu
Website: michaelgechter.com

Education
Ph.D., Economics, Boston University, Boston MA, May 2015 (expected)
  Dissertation Title: Three Essays in Applied Policy Evaluation in Developing Countries
  Dissertation Committee: Dilip Mookherjee (co-chair), Hiroaki Kaido (co-chair) and Kevin Lang
M.A., Political Economy, Boston University, Boston MA, 2012
B.A., Economics (Magna Cum Laude), Pomona College, Claremont CA, 2005

Fields of Interest
Development Economics, Econometrics

Fellowships and Awards
  Funded Participant, Western Economic Association International Graduate Student Workshop, 2014
  Funding for Research Assistant ($9,600), Boston University Department of Economics RA Mentor Program, 2013-2014
  Research Grant ($20,734), Weiss Family Program Fund for Research in Development Economics, 2013
  Summer Research Grant ($5,000), Boston University Department of Economics, 2013
  Department Fellowship, Boston University Department of Economics, 2011 - 2014
  Morris B. Pendleton Prize, Pomona College Department of Economics, 2005
  Phi Beta Kappa, Pomona College, 2005
  Leland M. Backstrand Memorial Award, Pomona College Department of Economics, 2004
  Student Undergraduate Research Grant, Pomona College, 2004

Work Experience
  Research Assistant for Hiroaki Kaido, Boston University, 2011 - Present
  Principal and Co-Founder, Idealistics, Inc., 2005 - 2013
Research Assistant for Tahir Andrabi, Pomona College, 2003 - 2005

Working Papers
“Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India” (Job Market Paper), January 2015.

Work in Progress
“The Effect of Small Scale Reservation Policies on Scale Economies and Firm Production in India” (with Amrit Amirapu)
“Measuring Slum Evolution Using Remote Sensing” (with Nathaniel Young)

Conference Presentations
BU-BC Econometrics Conference, Chestnut Hill MA, 2014
Northeast Universities Development Consortium (NEUDC) Conference, Boston MA, 2014
Western Economic Association International Graduate Student Workshop, Denver CO, 2014
Annual Meeting of the Eastern Economic Association, Boston MA, 2014
International Workshop on Official Data organized by Indian Statistical Institute, Kolkata, Central Statistics Office, India and International Growth Centre, Kolkata India, 2013

Invited Seminars
2014-15: Cambridge, Carlos III, CIDEx, Cornell PAM, IFPRI, ITAM, Laval, Penn State, Sciences Po
2013-14: IGIDR

Teaching Experience
Teaching Fellow, Introductory Microeconomic Analysis, Boston University Department of Economics, Spring 2011
Overall instructor rating: 4.1/5.0

Referee Experience
The Singapore Economic Review, Urban Affairs Review

Languages
Fluent in English, French and Spanish. Basic Dutch and Hindi.

Computer Skills: Stata, R, MATLAB, FORTRAN, Java, Python, SQL, \LaTeX

Other: Development Reading Group coordinator, 2011 - 2012

Citizenship: USA, Belgium
References

Professor Dilip Mookherjee
Department of Economics
Boston University
270 Bay State Rd
Boston MA 02215 USA
Phone: (617) 353-4392
Email: dilipm@bu.edu

Professor Hiroaki Kaido
Department of Economics
Boston University
270 Bay State Rd
Boston MA 02215 USA
Phone: (617) 358-5924
Email: hkaido@bu.edu

Professor Kevin Lang
Department of Economics
Boston University
270 Bay State Rd
Boston MA 02215 USA
Phone: (617) 353-5694
Email: lang@bu.edu