

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

Amrit Amirapu[†] Michael Gechter[‡]

February 9, 2017

Abstract

In this paper, we estimate the costs associated with a suite of labor regulations in India whose components have gone largely unstudied in developing countries. We take advantage of the fact that these regulations only apply to firms above a size threshold. Using distortions in the firm size distribution at the threshold together with a structural model of firm size choice, we estimate that the regulations increase firms' unit labor costs by 35%. We document a robust *positive* association between regulatory costs and exposure to corruption, which may explain why regulations appear to be so costly in developing countries.

*First version: November 22, 2014. We wish to thank Dilip Mookherjee, Arvind Subramanian, Kevin Lang, Hiroaki Kaido, Kehinde Ajayi, Sam Bazzi and Marc Rysman for input, guidance and encouragement. We also thank Sam Asher, Laveesh Bhandari, Aditya Bhattacharjea, Sharon Buteau, Areendam Chanda, Urmila Chatterjee, Nancy Chau, Bivas Chaudhuri, Bibek Debroy, Andy Foster, Francois Gourio, Jayinthe Harinath, Rana Hasan, Ravi Kanbur, Saibal Kar, Dan Keniston, Anirban Mitra, Devashish Mitra, Sandip Mitra, Paul Novosad, G. Sajeewan, Johannes Schmieder, Pronab Sen, Anushree Sinha, Sandip Sukhtankar, Shyam Sundar, John Van Reenen and numerous seminar and conference participants for many useful comments. Representatives in the state governments of BR, KL, MH, TN and WB provided invaluable assistance in understanding the data collection process for the Economic Census. We gratefully acknowledge financial support from the Weiss Family Program Fund for Research in Development Economics. Liu Li, Stefan Winata and Haiqing Zhao provided expert research assistance, supported by the BU MA-RA mentor program.

[†]School of Economics, University of Kent. Email: amrit.amirapu@gmail.com.

[‡]Department of Economics, The Pennsylvania State University. Email: mdg5396@psu.edu.

1 Introduction

Restrictive labor regulations have been blamed for some of the most significant problems faced by developing countries, including low labor force participation rates and low levels of employment in the formal sector.¹ It has even been suggested that regulations may distort the allocation of labor across firms, thus contributing to the substantially lower levels of aggregate productivity seen in developing countries (Hsieh and Klenow (2009)). What is not clear is why labor regulations should be so much costlier in a developing country setting, particularly since enforcement agencies there are typically characterized by severe resource constraints, low compliance and widespread corruption (Svensson (2005), Chatterjee and Kanbur (2013); Kanbur and Ronconi (2015)). Moreover, previous work on the subject in developing countries has focused almost exclusively on a small subset of labor regulations: namely, laws related to employment protection (e.g. firing restrictions) and minimum wages.² In actuality, labor regulations are multifaceted, encompassing many different types of employment-related laws, such as workplace safety requirements and the provision of mandated benefits (including health insurance, social security legislation, payment of gratuities, etc.). The vast majority of such labor regulations have gone almost completely unstudied in developing countries.³

In this paper, we address both of these gaps and make several further contributions to the growing literature on labor regulations in developing countries. In particular, we estimate the costs associated with a suite of labor regulations in India whose components include workplace safety regulations, social security taxes and business registration requirements.⁴

¹See, for example, Besley and Burgess (2004a); Botero, Djankov, La Porta, Lopez-De-Silanes, and Shleifer (2004); Djankov and Ramalho (2009).

²See Djankov and Ramalho (2009); Freeman (2010); Nataraj, Perez-Arce, Kumar, and Srinivasan (2014) for excellent reviews of the literature which reveal this focus.

³We are aware of only a small number of exceptions, including Botero et al. (2004), Dougherty (2009), and Dougherty, Frisanco, and Krishna (2014). However, non-minimum-wage, non-employment-protection regulations are not the main focus of any of those studies, with Botero et al. (2004) focusing primarily on the legal origins of regulation and Dougherty (2009) and Dougherty et al. (2014) principally interested in employment protection legislation. The other two exceptions of which we are aware, Gruber (1995) and Kugler and Kugler (2009), investigate the distortionary effects of payroll taxes in Chile and Colombia, respectively, but find contradictory effects.

⁴Business registration requirements are generally considered separately from labor regulations. However, in our context labor regulations intended to apply to all firms are much more likely to be enforced once

What the regulations have in common is that they only apply to firms that have hired 10 or more employees, a feature we exploit to identify the magnitude of the costs they impose on firms. Because our methodology takes advantage of this objective feature of the laws, we do not need to rely for identification on inherently subjective assessments of differences in the text of the laws across regions - a criticism which has dogged some of the best known work in the literature (see Besley and Burgess (2004b), Bhattacharjea (2009) and Fagernas (2010)).

Instead, our methodology translates observed firm behavior in response to the 10-worker threshold into estimates of the increase in unit labor costs associated with these regulations. Because our estimates are derived from firm behavior in response to actual enforcement rather than from the text of the laws we refer to our estimated labor cost increase as representing “*de facto* regulatory costs” in what follows. We find that these regulations effectively increase firms’ unit labor costs by 35%, substantially distorting economic decisions relative to a counterfactual regime without these regulations. We also apply our method to India’s most stringent, controversial piece of employment protection legislation, Chapter VB of the Industrial Disputes Act (IDA), which stipulates that any industrial establishment with more than 100 workers (in most states) must obtain prior permission from the state government before laying off workers or closing the establishment. In contrast to the substantial costs we uncover at the 10-worker threshold, we find only a small and statistically insignificant impact on unit labor costs from operating at or above the 100-worker threshold.

The next major contribution of the paper is to show that the distortionary effect of regulations depends critically on quality of governance through the extent and type of corruption present in regulatory enforcement. We distinguish between two different types of corruption in the context of regulatory enforcement: collusive and extortionary. Collusive corruption is characterized by corrupt inspectors allowing firms to avoid the *de jure* costs of abiding

enforcement agencies have records of a firm’s existence obtained through registration. This view is consistent with recent research experimentally defraying the costs of registration (de Mel, McKenzie, and Woodruff (2013); de Andrade, Bruhn, and McKenzie (2014)), which finds that informal firms behave as if registration imposes costs on them over and above the costs of registration alone.

by regulations in exchange for bribes, while extortionary corruption is characterized by corrupt inspectors who threaten to overreport regulatory infringements if they do not receive bribes. By relating differences in our estimates of *de facto* regulatory costs to differences in exposure to corruption across states and industries, we provide suggestive evidence that a) corruption is a significant determinant of regulatory costs, and b) regulatory corruption is more extortionary than collusive in nature. This may explain why regulations appear to be more costly in developing countries than in developed countries - it is not the regulations themselves that are particularly problematic, but the way in which they are enforced.

We develop our argument as follows. We begin by exhibiting the Indian establishment size distribution using data from the Economic Census of India (EC). Relatively underused until recently, the EC aims to be a complete enumeration of all non-farm establishments⁵ in India and, unlike all other Indian establishment-level datasets, it is not censored by size or restricted to include only the formal or informal sector. It is thus the only Indian dataset that permits estimation of the complete establishment size distribution - across all sizes and types of firms. We find that a power law distribution fits the data well,⁶ except for a discontinuous and proportional decrease in the density of establishments with 10 or more workers (see Figures 2 and 3). We take this distortion of the establishment size distribution at exactly the 10-worker threshold as qualitative evidence that the regulations which become binding there do affect firms' hiring decisions.

To understand and quantify the effect of the regulations on firm cost structure, we develop a simple model in which managers are endowed with heterogenous productivities and must choose their optimal employment levels. Firms that report hiring more than a threshold number of workers face higher unit labor costs due to the presence of regulations, and are

⁵The 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 primarily for expositional purposes, but also based on the observation that only a minute proportion of establishments belong to multi-establishment firms.

⁶This is not unusual. Establishment size distributions across the world have been shown to be well-fit by power law distributions (c.f. Axtell (2001); Hernández-Pérez, Angulo-Brown, and Tun (2006)).

thus smaller than they would be otherwise. [Garicano, Lelarge, and Van Reenen \(2016\)](#) (henceforth GLV) show that the magnitude of the increase in costs can be identified from characteristics of the distribution including, most importantly, the size of the downshift in density at the threshold. Our model augments GLV to allow for the possibility of strategic misreporting. That is, managers may choose to deliberately misreport their employment levels at some cost, with the goal of avoiding some or all of the additional labor costs that apply to firms above the threshold size.⁷ Fitting the model’s predicted size distribution to the one observed from the EC data, we generate an estimate of the additional labor costs that apply to firms above the 10 worker regulatory threshold that is robust to the possibility of strategic misreporting.

Although very large on average, we show that there is substantial heterogeneity in the magnitude of *de facto* regulatory costs along several dimensions including state, industry and ownership type. In the first place, we find that privately-owned establishments face the highest *de facto* regulatory costs while government-owned establishments show no significant cost increase when employing 10 or more workers. This supports our interpretation that the downshift in the distribution starting at 10 workers is due to the regulations as opposed to some other factor.⁸ Most regulations do not apply to government-owned establishments in the same way, so one would not expect their establishment size distribution to be distorted over the 10-worker threshold. We then use the state and industry level variation to explore other determinants of regulatory costs. Strikingly, we find a strong and robust positive correlation between our estimated regulatory costs and several different state-level measures of corruption.⁹ As further support for our state-level corruption results, we provide state-by-industry analysis showing that industries with higher “regulatory intensity” have higher

⁷The importance of allowing for strategic misreporting is explained in greater detail in Section 4.2. Note that “strategic misreporting” is distinct from the issue of corruption in the enforcement of labor regulations.

⁸We consider other factors in Section 4.2.

⁹These corruption measures include a subjective, perceptions-based measure 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, Kumar, Rajan, Subramanian, and Tokatlidis \(2006\)](#)).

estimated costs - especially when they are located in more corrupt states.

The link between high regulatory costs and corruption may appear surprising if one thinks of corruption as collusive, “greasing the wheels” in a highly-regulated economy by allowing firms to reduce their effective regulatory burden by bribing inspectors (e.g., [Huntington \(1968\)](#)). Our results are instead consistent with the concern that inspectors may overreport violations in order to extract greater bribes - extortionary corruption.¹⁰ We present a simple model nesting extortionary and collusive corruption which explains the patterns we see in the data. We provide additional support for our interpretation through qualitative evidence from the crowd-sourced corruption reporting website [ipaidabribe.com](#). The reports on [ipaidabribe.com](#) describe regulatory inspectors threatening to overreport violations and demanding bribes proportional to the number of workers employed in an establishment, just as in our model.

Our finding that the most contentious component of India’s employment protection legislation, Chapter VB of the IDA, does *not* have a substantial effect on unit labor costs differs from much of the earlier academic work on the subject and belies the attention the IDA has received from academics¹¹ and the business press¹² alike. We attribute this difference, first, to the fact that our methodology for estimating the impact of the legislation is very different from the strategies employed in previous papers. Specifically, our identification is based on the presence and size of distortions in the distribution of establishments at the size where Chapter VB’s firing restrictions become binding. Most previous work identifies the effect of India’s employment protection legislation based on differences in the growth of mean outcomes across states which have been coded as initiating pro-worker or pro-employer reforms to the full IDA. The coding of states into these three groups (pro-employer, pro-worker, or neutral) has been the subject of controversy in the subsequent literature (see [Fagernas \(2010\)](#);

¹⁰See [Banerjee \(1994\)](#); [Mookherjee \(1997\)](#); [Hindriks, Keen, and Muthoo \(1999\)](#); [Polinsky and Shavell \(2001\)](#); [Mishra and Mookherjee \(2013\)](#) for theoretical treatments. Empirically, [Sequeira and Djankov \(2014\)](#) and [Asher and Novosad \(2016\)](#) also provide evidence for the importance of extortionary corruption.

¹¹See, for example, [Besley and Burgess \(2004a\)](#); [Hasan, Mitra, and Ramaswamy \(2007\)](#); [Aghion, Burgess, Redding, and Zilibotti \(2008\)](#); [Adhvaryu, Chari, and Sharma \(2013\)](#); [Chaurey \(2015\)](#).

¹²E.g. [Bajaj \(2011\)](#), [Ghosh \(2016\)](#).

Bhattacharjea (2006, 2009)), and an advantage of our identification strategy is that we can side-step this controversy. Hsieh and Olken (2014), the only other paper in the literature to focus on establishment size distributions, report finding no visually striking change in the size distribution of Indian establishments at the 100-worker threshold, in accordance with our quantitative results. Another source of difference is our focus on Chapter VB, rather than the full IDA. This is partially a question of the feasibility of applying our approach (Chapter VB is size-based), but we also see focusing on Chapter VB as providing a proof of concept. If Chapter VB's complete restriction on firing is not very distortionary, it would be surprising if the law's more mild provisions are.

In addition to our contributions to the empirical literature on labor regulations, our extension of the GLV model to allow firms to strategically misreport their sizes should find applications in many other settings. Robustness to strategic misreporting in response to a size threshold is particularly crucial in developing countries because costs of such behavior are likely to be much lower than in the high-income country administrative data used by studies such as GLV. The information in the EC is self-reported - as it is in nearly all developing country datasets - and it is likely that enumerators are co-optible at relatively low cost. We show that in the presence of strategic misreporting, a naive approach to estimating GLV's model can dramatically overestimate the increase in labor costs associated with a size-based regulation. We identify firms' real responses using a reasonable theoretical restriction: that the cost of misreporting be strictly convex in the degree of misreporting. Under this assumption, misreporting can be extensive near the threshold, but becomes increasingly costly for large values. In fact, we show that the reported firm size distribution becomes arbitrarily close to the true distribution at large firm sizes, so one can minimize any bias in the estimate of regulatory costs by focusing the estimation on large firm sizes and discarding the observations close to the threshold. In our case, if one fails to account for the possibility of misreporting, the estimated increase in per-worker costs rises from 35% to 101%.

The rest of the paper is organized as follows. In the next section (Section 2), we provide

an overview of the relevant institutional details regarding Indian labor and industrial regulations. Section 3 introduces the data and provides qualitative evidence on the importance of size-based regulations using the size distribution of establishments in India. In Section 4 we describe the theoretical model and empirical strategy. Section 5 provides the main results. In Section 6, we interpret the findings, explore the multiple dimensions of variation in our results, and investigate the connection between our estimated costs and corruption. Section 7 concludes.

2 Labor Regulations in India

Most labor regulations in India only apply to establishments that are larger than a certain threshold, where size is most often measured in terms of the number of workers in the establishment. There are several thresholds at which different labor regulations start to apply, but the two most prominent thresholds occur once an establishment hires at least 10 and at least 100 workers.¹³ In most states in India, establishments that hire more than 100 workers must abide by India's most controversial piece of employment protection legislation: Chapter VB of the IDA.¹⁴ Under this regulation, establishments over the threshold must be granted government permission before closing the establishment or laying off workers. It is the IDA - of which Chapter VB is a part - that has been the subject of most academic papers on labor regulations in India. Outside of India, employment protection legislation more generally - along with minimum wage policy - has been the subject of nearly all research on labor regulations in developing countries.

In contrast, the 10 worker threshold has received far less attention from academics, even though it is extremely important due to the large number of varied regulations that start to

¹³There are other thresholds, such as at 20 workers (at which point establishments must contribute to the "Employees' Provident Fund Organisation," which operates a pension scheme for formal sector workers) and at 50 workers (at which point severance payment obligations increase under Chapter VA of the Industrial Disputes Act), but we do not analyze these thresholds because they are less contentious and do not appear to substantially distort the establishment size distribution.

¹⁴In 2005, the year to which our analysis applies, this threshold was 100 workers for all states except West Bengal, where the threshold was 50 workers.

become binding at that threshold as well as the fact that this threshold is most commonly associated with the formal/informal divide. The major regulations that start to apply once an establishment employs 10 or more workers include the following: establishments must register with the government, meet various workplace safety requirements (under the Factories Act for manufacturing establishments that use power and The Building and Other Construction Workers' Act for construction-related establishments, for example), pay insurance/social security taxes (under the Employees' State Insurance Act), distribute gratuities (under the Payment of Gratuity Act) and they must bear a greater administrative burden (under, for example, the Labor Laws Act). Other regulations are indirectly size-based, because they reference laws with size-based aspects. For example, the Maternity Benefits Act only applies to establishments designated as "factories" under the Factories Act, which means it only applies to establishments with more than 10 workers.¹⁵

In addition to - or in lieu of - the explicit costs associated with complying with the regulations, establishments with 10 or more workers may be subject to implicit costs associated with increased interaction with labor inspectors, who often 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 \(2006\)](#)).

It has been argued that the ability to extract bribes is exacerbated by the antiquated and/or arbitrary nature of certain components of the laws ([Debroy \(2013\)](#)). [TeamLease Services \(2006\)](#) provides some telling examples: "[r]ules under the Factories Act, framed in 1948,¹⁶ provide for white washing of factories. Distemper won't do. Earthen pots filled with

¹⁵Finally, there appears to be a salience effect associated with the 10 worker threshold as well: in interviews with small business owners in Chennai, we discovered that several of them appeared to believe that certain regulations (such as the Provident Fund Act) applies once you have 10 workers, when in fact they did not.

¹⁶The 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.

water are required. Water coolers won't suffice. Red-painted buckets filled with sand are required. Fire extinguishers won't do... And so on." The result of such rules is that almost all firms can be found guilty of some violation or another under the letter of the law - even if they are in compliance with the spirit of the law. Firm owners who choose not to comply with such regulations face costs if discovered and convicted.¹⁷

This kind of behavior has been referred to as "harassment bribery" (Basu (2011)). Anecdotal evidence of inspectors using the complexity, arbitrariness 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 A, which demonstrate just this kind of behavior.¹⁸ 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 for interpretation of our results in Section 6.

3 Data and the Size Distribution in India

3.1 Data

We will use the Economic Census of India (EC) as our main data source to investigate the costs associated with the regulations described in the previous section. The EC is meant to be a complete enumeration of all (formal and informal) non-farm business establishments in India at a given time. As such, it contains a very large number of units: the 2005 wave, which we will principally use, has almost 42 million observations. It is the only Indian dataset that represents the unconditional distribution of establishment size, which is essential for our analysis. Other 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 are thus unsuitable for our analysis. The

¹⁷The possible costs include fines and/or prison sentences.

¹⁸We thank Andrew Foster for this suggestion.

ASI, for example, only covers establishments 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 establishment uses no power). Therefore, the selection into the ASI varies discontinuously at precisely one of our points of interest.

Little used until recently, the EC has seen substantially more use in the literature in the past several years (e.g. [Asher and Novosad \(2015a,b, 2016\)](#); [Bertrand, Hsieh, and Tsivanidis \(2015\)](#)).¹⁹ The price to pay for uniform coverage and large sample size is that the EC does not contain very detailed information 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 are cross-sectional.

We supplement our analysis with data from a variety of other sources. From the ASI we get employment and labor 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.²⁰ Other measures of state-level regulations come from [Aghion et al. \(2008\)](#) and [Dougherty \(2009\)](#).

¹⁹[Asher and Novosad \(2016\)](#) provides a comparison of the EC and ASI, demonstrating consistency between the two sources.

²⁰We would like to thank Anushree Sinha and Avantika Prabhakar for their considerable and generous help in obtaining these data.

3.2 The Size Distribution of Establishments in India

Figure 1 shows the distribution of establishments by the number of total workers (hired and non-hired - typically family - workers) for establishments with up to 200 total workers in 2005. Perhaps the most striking feature of figure 1 is the extraordinary degree to which the distribution is right-skewed. Indeed, about half of all establishments are single person establishments, while the densities for establishments with 10 or more workers are almost imperceptible.²¹ Figure 2 shows the full distribution of establishment size frequencies according to a log scale. Each point represents one bar in the earlier histogram.

Three things are most striking about figure 2. 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 a 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 (Axtell (2001) and Hernández-Pérez et al. (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. To the best of our knowledge, ours is the first paper to document this phenomenon in India. Finally, we do not see any discernible change in the distribution at 100 workers, the relevant threshold for employment protection legislation. We will confirm this fact in our formal analysis.

Also of note from the figures above is that there appears to be a significant amount of non-classical measurement error due to rounding of establishment sizes to multiples of 5. 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]”. Our estimation procedure, described in the next section, accommodates this measurement error pattern.

²¹The densities for establishments with more than 200 workers are also imperceptible. We have omitted them only for clarity in the figure.

4 Model and Empirical Strategy

4.1 Basic Model

To interpret the downward shift from Figure 2 in economic terms, we first describe the framework developed in GLV on which our model is built. In the GLV framework, size-based regulations increase the unit labor costs of firms that exceed the size threshold, which results in a parallel downward shift in part of the theoretical firm size distribution. From the magnitude of the downshift observed in the empirical distribution, one can back out the additional labor costs imposed by the regulations.

The primitive object in the GLV framework is the distribution of managerial ability ($\alpha \sim \phi : [\underline{\alpha}, \alpha_{max}] \rightarrow \mathbb{R}$) as the primitive object, as in Lucas (1978). Firms whose managers have higher ability (α) are more productive and can profitably employ more workers. Workers are allocated to firms through a competitive labor market with a single, market clearing wage. As is common in the literature, GLV assume that the distribution of managerial ability follows a power law (e.g. $\phi(\alpha) = c_\alpha \alpha^{-\beta\alpha}$), an assumption we will maintain. This generates a power law in the theoretical firm size distribution. A firm/manager with productivity/managerial ability α faces the following profit-maximization problem:

$$\pi(\alpha) = \max_n \alpha f(n) - w\bar{\tau}n \tag{1}$$

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 $\bar{\tau}$ takes the value 1 if $n \leq N$ and $1 + \tau$ if $n > N$, where $\tau > 0$ is a proportional tax on labor. From the first order condition on this maximization problem, $\alpha = \frac{w\bar{\tau}}{f'(n)}$, 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.

One can characterize the solution as follows. First, only individuals with managerial

ability/productivity above some threshold (α_{min}) will find it profitable to manage a firm. Individuals with $\alpha < \alpha_{min}$ will be workers and receive payoff equal to w . The set of managers can be further categorized, beginning with the lowest productivity managers (those with $\alpha \in [\alpha_{min}, \alpha_1]$). One can think of these managers as unconstrained in the sense that they choose to hire fewer than N workers and thus do not fall under the purview of the size-based labor regulations. Another set of managers with slightly higher productivity (between some thresholds α_1 and α_2) maximize their profits by hiring N workers exactly to avoid the discontinuous increase in costs implied by crossing the threshold. These managers should be “bunched up” at N . By the same token, this should lead the firm size distribution to exhibit a hole or “valley”, just to the right of the threshold, because the firms that would otherwise be there are instead bunched up at N . 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).

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 strictly monotonic relationship between α and n (except for the bunching). We obtain $\chi(n)$ in closed form under the further assumption that the production function is a power function: $f(n) = n^\theta$. Then we obtain the key result that a function of the tax enters multiplicatively in the expression for the density of firm size n (for all $n > N$). 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:²²

²²For a full derivation of this result, including all omitted steps, see Appendix B.1.

$$\chi(n) = \begin{cases} \left(\frac{1-\theta}{\theta}\right)^{1-\beta} (\beta-1)n^{-\beta} & \text{if } n \in [n_{\min}, N) \\ \left(\frac{1-\theta}{\theta}\right)^{1-\beta} (N^{1-\beta} - (1+\tau)^{-\frac{\beta-1}{1-\theta}} n_u^{1-\beta}) & \text{if } n = N \\ 0 & \text{if } n \in (N, n_u) \\ \left(\frac{1-\theta}{\theta}\right)^{1-\beta} (\beta-1)(1+\tau)^{-\frac{\beta-1}{1-\theta}} n^{-\beta} & \text{if } n \geq n_u \end{cases} \quad (2)$$

where θ measures the degree of diminishing returns to scale, β represents the negative slope of the power law and τ is the implicit per worker tax. The range $[N, n_u]$ represents the “hole” in the distribution where the firms bunched up at N would otherwise be found. For n outside the range $[N, n_u]$, one can express the log of the density according to the following equation:

$$\log(\chi(n)) = \log \left[\left(\frac{1-\theta}{\theta} \right)^{1-\beta} (\beta-1) \right] - \beta \log(n) + \log((1+\tau)^{-\frac{\beta-1}{1-\theta}}) \mathbb{1}\{n > N\}, \quad (3)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. To see how τ is identified from $\chi(n)$, rewrite Equation 3 as

$$\log(\chi(n)) = \alpha - \beta \log(n) + \delta \mathbb{1}\{n > N\}. \quad (4)$$

α , β and δ can be identified by applying Equation 4 to the observed size distribution. θ is a function of α and β and is thus also identified. τ is given by

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

which is identified as long as θ and β are identified.

A researcher might therefore be tempted to choose n_u and produce a value for τ by using ordinary least squares to estimate the specification

$$\log(\chi(n)) = \alpha - \beta \log(n) + \delta \mathbb{1}\{n > N\} + \epsilon(n), \quad (5)$$

where $\epsilon(n)$ represents any deviation of the observed firm size distribution from the model coming from an idiosyncratic tendency for firms to cluster to or away from a particular size. But what if the true firm size distribution is systematically different from what is observed in the data? In the next subsection we show that when firms can misreport their size in response to the regulatory threshold, estimating Equation 5 can lead one to arrive at an upward-biased estimate of τ .²³

4.2 Strategic Misreporting

One of the underlying assumptions of the analysis above is that any deviation of the true distribution of firm size is unrelated to the regulatory threshold N . This might be a reasonable assumption when working with high quality administrative data in a developed country setting as in GLV, but the data in the Economic Census (as in most developing country datasets) are self-reported, and it is possible that establishment managers may deliberately misreport information to Economic Census enumerators. For example, if the managers are aware of the increased regulatory burden that is associated with employing 10 or more workers, and if they believe the EC enumerators will relay information to government regulatory bodies,²⁴ they may wish to hide the fact that their actual employment exceeds the threshold or more generally under-report their actual employment. In what follows, we show how misreporting generates a reported distribution that differs from the true distribution of employment and discuss the biases this difference may produce in the value of τ estimated using Equation 5 along with presenting our own solution.

Our model extends the basic GLV framework by letting firms choose not only their true

²³This conclusion also holds when using the maximum likelihood estimator from GLV or the restricted least squares specification from Appendix B of the NBER working paper version of GLV, [Garicano, Lelarge, and Van Reenen \(2013\)](#). We focus on Equation 5 for expositional simplicity and because we believe it is more likely to be implemented by researchers since it can easily be estimated using any standard regression package.

²⁴In point of fact firms' answers to Economic Census enumerators have no impact on their regulatory burden, but it is quite possible that firms believe otherwise, and that is what is relevant. If firms believe that that reporting has no effect on their regulatory burden, then there should be no incentive to misreport, and estimates based on Equation 5 would be unbiased.

employment (n), but also their *reported* employment (l). Then, a firm with productivity α faces the following profit-maximization problem:

$$\pi(\alpha) = \max_{n,l} \alpha f(n) - wn - \tau wl * \mathbb{1}\{l > N\} - F(n, l) * p(n, l) \quad (6)$$

where $\alpha, f(n), w$ and τ are all defined as they were previously. The problem is identical except that now firms pay the extra marginal cost, τw , only on their *reported* employment, and not on their true employment. Furthermore, they only pay this cost if their reported employment exceeds the threshold. 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, $F(n, l)$.²⁵ 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, but going forward we will impose the following assumption on the expected costs of misreporting:

Assumption 1. *Let $u \equiv n - l \geq 0$ denote the degree of misreporting and $M(u) = F(u) * p(u)$ denote the expected costs of misreporting. We assume that $M(u)$ is continuous, increasing, strictly convex and that $M(0) = 0$.*

We restrict u to be positive as there is no incentive for firms to over-report their employment. The assumption that misreporting costs should be convex in the degree of misreporting is both standard in the literature (e.g. [Almunia and Lopez-Rodriguez \(2015\)](#); [Kumler, Verhoogen, and Frias \(2015\)](#)) and intuitive given our understanding of the context in which Indian businesses make such decisions.²⁶ One important implication of Assumption 1 - that

²⁵Again it is the firm's *perception* that they may be caught that matters.

²⁶The intuition is that hiding larger and larger numbers of employees from enumerators or inspectors should get increasingly difficult until at some point it is impossible. This intuition is supported by our understanding of the context in which firms make such decisions, which has been informed by interviews with small businesses in Chennai. Among such enterprises it is common to hear accounts of business owners ushering employees out the back door of the establishment whenever labor inspectors arrive, but this type of behavior is clearly only possible for relatively small numbers of employees. We would like to thank Sharon Buteau and Balasekhar Sudalaimani from IFMR for helping to set up these interviews.

the extent of misreporting should be relatively lower for larger firms - finds empirical support in recent literature (e.g. Goyette and Kouame (2016)).

One plausible way to obtain strictly convex misreporting costs is to suppose that misreporting firms are caught with a probability that is itself an increasing and strictly convex function of the degree of misreporting, $n - l$, and subject to a fixed fine if caught. Another possibility is that the probability of being caught is linearly increasing in the degree of misreporting and the fine if caught is also a linear function of misreporting. But whatever the functional forms of $p(\cdot)$ and $F(\cdot)$, as long as $M(u)$ satisfies the conditions of Assumption 1, we show that for a given size x , the difference between the log of the reported density, $\psi(x)$, and the log of the true density, $\chi(x)$, becomes vanishingly small for large enough values of x .

Proposition 1. *Suppose a firm's profit maximization problem takes the form of Equation 6 and Assumption 1 holds. Then*

$$\lim_{x \rightarrow \infty} \log \chi(x) - \log \psi(x) = 0.$$

Proof. See Appendix B.3. □

To provide some intuition for this result, suppose that misreporting firms are caught with probability $p(n, l) = \min \left\{ \frac{(n-l)^2}{100}, 1 \right\}$, and pay a fixed fine, F , if caught. Then their profit maximization problem is:

$$\pi(\alpha) = \max_{n,l} \alpha f(n) - wn - \tau wl * \mathbb{1}\{l > N\} - F * \min \left\{ \frac{(n-l)^2}{100}, 1 \right\}$$

The solution to this problem can be informally characterized as follows.²⁷ The lowest productivity firms (those with α below some threshold, α_1) will be unconstrained, choosing $n \leq N$ and reporting truthfully ($l = n$). Higher productivity firms, with $\alpha \in [\alpha_1, \alpha_2]$, will choose

²⁷For simplicity we focus below on an interior solution, which is guaranteed if the parameters satisfy the following condition: $\frac{w\tau}{F} < 2$. This condition is not necessary for any of our results.

$n > N$, exceeding the regulatory threshold, but will find it profitable to misreport their employment, setting $l = N$. These firms will only *appear* to be bunched up at N , 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 completely avoid the regulation without being detected with reasonable probability and must report $l > N$. Even these firms, however, with both $n > N$ and $l > N$ 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 $l = n - \frac{50}{F}w\tau$. Note that the degree of misreporting is by a constant amount, rather than a constant proportion.²⁸

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

$$\log\chi(n) = \begin{cases} \log A - \beta \log(n) & \text{if } n \in [n_{\min}, N) \\ \log[\xi(n)] & \text{if } n \in [N, n_m(\alpha_2)] \\ - & \text{if } n \in (n_m(\alpha_2), n_t(\alpha_2)) \\ \log A - \frac{\beta-1}{1-\theta} \log(1+\tau) - \beta \log(n) & \text{if } n \geq n_t(\alpha_2) \end{cases} \quad (7)$$

while the log of the density of firms with reported employment l , $\log\psi(l)$ is given by:

$$\log\psi(l) = \begin{cases} \log A - \beta \log(l) & \text{if } l \in [l_{\min}, N) \\ \log(\delta_l) & \text{if } l = N \\ - & \text{if } l \in (N, l_t(\alpha_2)) \\ \log A - \frac{\beta-1}{1-\theta} \log(1+\tau) - \beta \log(l + \frac{50}{F}w\tau) & \text{if } l \geq l_t(\alpha_2) \end{cases}$$

where A is a function of constants and terms have been simplified and collected.²⁹

Comparing the expressions for the reported and true size distributions above, there are several points worth noting. First, for the range $l < N$, the true distribution coincides with the reported/observed distribution. Second, there appears to be bunching at N in

²⁸This will be the case for any misreporting cost function satisfying Assumption 1.

²⁹Derivation of this result can be found in Appendix B.2.

the reported distribution, but some of these firms in fact have greater than N workers. Third, compared to the distribution for $n < N$, both the true distribution *and the reported distribution* for $n \gg N$ are downshifted, and by exactly the same function of τ as in the model without misreporting (the intercepts for both distributions are $\log A - \frac{\beta-1}{1-\theta} \log(1+\tau)$ for larger firms versus $\log A$ for smaller firms). Fourth, as stated in Proposition 1, the difference between the log of the reported distribution and the log of the true distribution converges to 0 for large firms. The intuition is straightforward: the only difference in the two expressions is the constant amount $\frac{50}{F} w \tau$, the contribution of which becomes negligible at large sizes.

Proposition 1 is problematic for estimating the parameters of the model using Equation 5. In practice, Equation 5 must be estimated using data from relatively small establishments with sizes outside the range $[N, n_u]$. This is because the empirical probability of observing an establishment of a given size is truncated at $\frac{1}{\# \text{ of observations}}$, which is apparent visually in figure 2. Truncation makes the relationship between the log of the empirical probability and the log of the total number of workers nonlinear. To preserve the linear relationship, a researcher would have to omit establishment sizes large enough that truncation is not an issue.³⁰ However, Proposition 1 tells us that the misreported distribution is close to the true distribution only at large sizes and that the misreported distribution may be biased downward at establishment sizes close to a regulatory threshold. This leads to downward bias in δ (the downshift in the log density) and consequently upward bias in τ . Instead, we will develop an estimation strategy that deals with the truncation problem and allows us to focus on large firms, where the the difference between the log of the reported distribution and the log of the true distribution is close to zero.

Before proceeding, it is worth noting a second possible source of misreporting: Economic Census enumerators themselves. EC 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

³⁰The suggestion to focus on relatively smaller establishments appears in Appendix B of [Garicano et al. \(2013\)](#).

establishments with 10 or more workers in order to avoid the extra burden of filling in the “Address Slip”. However, as we show in Appendix B.4, this type of misreporting - like the previous one - only generates bias in the reported distribution for establishment sizes close to N .³¹ In fact, it is easy to show that any estimation technique that is robust to the possibility of manager-driven misreporting will also be robust to the possibility of enumerator-driven misreporting.

4.3 An Empirical Approach Robust to Strategic Misreporting

In this subsection, we develop a way of estimating Equation 5 using establishments that are least affected by the misreporting. These include establishments that are below the bunching point as well as those that are far above the size threshold. As we remarked in the previous subsection, we cannot estimate Equation 5 directly on large establishments because of truncation in the empirical probability of observing an establishment of a given size. Furthermore, as discussed in section 3, the empirical size distribution is characterized by substantial rounding to multiples of 5 workers, especially at larger sizes. Setting aside the truncation problem, OLS estimation of Equation 5 will produce downward bias in δ because sizes that are multiples of 5 are treated as single observations. Instead, their excess establishments should be distributed to nearby sizes.

To address both issues, we non-parametrically estimate the density associated with larger sizes using the method described in Markovitch and Krieger (2000) (MK). MK propose a nonparametric density estimator for heavy-tailed distributions that achieves L_1 consistency by applying a standard constant bandwidth Parzen-Rosenblatt estimator,

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right),$$

³¹The two types of misreporting can be modeled similarly. The main difference is that the marginal benefit of misreporting in the manager-driven misreporting model is replaced with a fixed benefit of misreporting in the enumerator-driven misreporting model.

to a transformation, $T : \mathbb{R} \rightarrow [0, 1]$, of the data. We use an Epanechnikov kernel and the transformation recommended by MK:

$$T(x) = \frac{2}{\pi} \arctan(x).$$

With the density estimates for the transformed data in hand, we convert back to density estimates for the original data using the transformation:

$$\hat{\psi}(l) = \hat{f}(T(l))T'(l).$$

The advantage of this approach from our perspective is that a constant bandwidth³² applied to the transformed data expands asymmetrically with respect to the original data. As we move to the right in the distribution our kernel begins to put positive weight on observations farther away, all the more so in the right of the distribution where data are scarcer. This accords with our observation that rounding in the reported distribution becomes more severe at larger sizes. We use the empirical probability for small sizes, where the establishment size distribution is better represented as a discrete variable.

We note here one difference between the model and the data. The model generates the log density of reported employment in 7. The log density is undefined for $l \in (N, l_t(\alpha_2))$ because the density of reported employment contains a hole in this region. Reporting $l \in (N, l_t(\alpha_2))$ is dominated by choosing either $l = N$ or $l \geq l_t(\alpha_2)$. However, Figures 1 and 2 clearly show that there are firms who report employing 11 workers. As in Kleven and Waseem (2013), we consider a fraction of firms to be inattentive to the threshold. That is, managers must pay a fixed cost which varies across firms to adjust their reported and actual employment in response to the threshold. In practice this would involve hiring an accountant or attorney who is knowledgeable about the text of labor regulations. Under the plausible assumption

³²Note that in this case, the bandwidth must be chosen. We cannot use cross-validation to choose the optimal bandwidth because it will recover the rounding pattern found in the data.

that the distribution of fixed costs does not vary with firm size, the fact that benefits of adjusting employment in response to the threshold rise with size means that all large firms will adjust while only some small firms will. Our approach of basing estimation primarily on firms with employment levels far from N is robust also to this rationally inattentive behavior.

We apply a modified version of Equation 5 to the log of the estimated density $\hat{\psi}(l)$ for all observed sizes. For example, when analysing the effect of regulations applying to firms employing 10 or more workers, we remove the effect of misreporting close to the threshold by adding dummy variables for size 8 and 9 and for sizes 10 - 20. The choice of 20 as the largest size for which we include a dummy is unimportant. Since Equation 5 treats each establishment size as one observation, the model is primarily estimated using data far from the 10-worker cutoff.³³ Finally, we include dummies for having 1 or 2 workers because own account and 2-worker establishments are likely to be household enterprises and may therefore differ fundamentally in character from their larger counterparts.

Figure 3 depicts the strategy. The dark grey dots show the raw data. The light grey 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 5, augmented by the dummy variables, to the nonparametric density estimates. Figure 3 above provides some evidence for the model described in section 4.2. 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.

³³The largest establishment in the 2005 EC has 22,901 workers.

5 Results

5.1 Regulations Applying to Firms Employing 10+ Workers

In this section we report the results of applying the estimation procedure described above to the 10 worker threshold in the 2005 Economic Census of India. Recall that the regulations applying to firms with 10 or more workers were of the type less-studied in the literature: workplace safety regulations, mandatory benefits and registration requirements bringing with them the threat of increased regulatory enforcement. Table 1 reports estimates for the increase in per-worker costs associated with these regulations, τ , 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 Appendix C. Standard errors, displayed beside the point estimates in parentheses, are obtained from a clustered bootstrap procedure with 200 replications. Following GLV, we cluster by industry at the 4 digit NIC code level. This allows for the possibility that differences in production technology - which could affect the firm size distribution and therefore our estimates - may be correlated by industry.³⁴ The top panel of Table 1 gives the all-India estimate of τ using our methodology. The point estimate is .35 and is significant at the $< 1\%$ level. 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.

By contrast, estimating the model without accounting for misreporting in any way yields much larger estimates. In particular, estimating Equation 5 on the size distribution omitting sizes larger than 99 workers and including the same dummy variables as in our own specification would lead us to conclude that exceeding the 10-worker threshold increases per-worker costs by 101%. This is due to a combination of rounding and the fact that the density

³⁴For robustness we have also tried alternative procedures, including a wild bootstrap and non-parametric bootstrap - both clustered at the firm size level. Clustering in this fashion allows for the possibility that reporting errors may be correlated by size and provide the most conservative approach to inference (especially the latter procedure). Under these procedures, statistical significance of our All-India estimates and estimates for most large states and industries survive (although some are only significant at the 10% level under the latter procedure).

associated with establishment sizes 21 - 99 converges only slowly back to the downshifted power law it follows at larger sizes, as predicted in our misreporting model. In other words, a “naive” estimation puts undue weight on firm sizes whose densities are biased downward by misreporting. In what follows we will focus our discussion on our misreporting-robust estimates of τ .

The lower panels of Table 1 show substantial variation in the magnitude of our misreporting-robust estimates of the per-worker tax by State, Industry and Ownership Type. For example, the point estimate on τ for the State of Kerala is .14 and is not statistically significant, while the estimate for Bihar, on the other hand, is .70 and is statistically significant at the 5% level, 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 τ by industry: it appears that the effective tax is high for establishments in manufacturing and even higher for establishments in retail and wholesale trade. Some industries have very noisy estimates, at times producing negative point estimates for τ . This is also true of some of the smaller states and ownership categories (as one can see in Appendix C), and is explained by the fact that the power law relationship can break down when there are a small number of observations in a category - as is the case for electricity, gas and water. In a few cases, negative point estimates reflect the fact that the production and market characteristics of these industries vary greatly from our model so that our model provides a poor fit of the data.³⁵ When looking at the differences by ownership type, we find that the estimates for τ are highest for private firms (particularly unincorporated proprietorships, which form by far the largest category of private firms). As one might expect, the tax is insignificant for government-owned firms, where presumably the regulatory burden is less than in the private sector.

The results above derive from the 2005 Economic Census, but we have also used data

³⁵Andhra Pradesh, the largest state to show a negative point estimate for τ , has a size distribution distorted in ways that are different from all other states, and which produces a poor fit (figure available upon request). We have concluded that this is the result of errors in data collection or recording rather than deliberate misreporting.

from the 1998 Economic Census to test whether there is inter-temporal variation in regulatory costs. Using the same empirical methodology described in Section 4.1, we estimate τ at the All-India level to be equal to .48 (.12) in the earlier data. Although somewhat larger in magnitude, this is not statistically significantly different from our 2005 estimate.³⁶ Interestingly, the downshift in the 1998 firm size distribution is not as immediately visually striking as that observed in the 2005 data, which may well reflect the fact that incentives related to misreporting - either on the part of firms or enumerators - were different in the two time periods, for example due to the Address Slip reporting requirement added in 2005. Let us reiterate, however, that our analysis in Section 4.1 shows that incentives to misreport do not affect our estimates of τ .

5.2 Employment Protection Legislation

In this subsection we report the results obtained by using our empirical strategy to test for an increase in per-worker costs for establishments that hire more than 100 workers and thus fall under the ambit of Chapter VB of the IDA, the most stringent component of India's employment protection legislation. As before, we run the test on the 2005 Economic Census and report the standard error in parentheses.³⁷ One difference in the estimation procedure is that we include dummy variables for firm sizes 1 to 20, so we are effectively comparing the distribution from 21 to 99 with that from 100 onwards. We include the dummies from 1 to 9 because we do not want to conflate the effect of the 100 worker threshold with that of the 10 worker threshold, and we include the dummies from 10 to 20 because those values will be most contaminated by misreporting - as implied by our model. The results are shown in the table below and largely conform to what the figures in Section 3 informally suggest: there is little evidence of a downshift. The implied τ is only .04 and is not statistically significant.

³⁶Relatedly, estimates of τ based on an alternative dataset comprised of the 2005/6 ASI and the 2005/6 NSSO Unorganized Manufacturing Enterprises Survey are also similar to our estimate of τ using the 2005 EC.

³⁷The standard error is obtained in an identical way as previously - from a bootstrap procedure with 200 replications, clustered at the 4 digit NIC code level.

Chapter VB of the IDA does *not* therefore appear to have an adverse effect on *the unit labor costs* of firms.³⁸

6 Discussion and Investigation of Mechanisms

In the previous section we documented considerable variation in our estimates of the costs (τ) of regulations applying to firms employing 10 or more workers across states, industries and ownership types. In this section we explore the determinants of this variation and see whether it can be explained by differences in the substance or application of regulations across states and industries.³⁹ Finding that differences in regulatory substance or enforcement do help explain the variation in τ provides further support for our assumption that the observed downshift in the distribution of establishments with 10 or more workers is indeed due to the existence of size-based regulations. Furthermore, the variation in τ helps us understand which dimensions of the regulatory regime impose the greatest costs on firms.

To preview the results, we observe a significant correlation between our estimated costs (τ) and measures of the extent to which states have reformed the role of inspectors in their regulatory regimes. This finding motivates us to look next at the link between τ and state level corruption, which we also find to be strong, robust to different measures of corruption, and independent of measures of regulatory substance. We take this finding to suggest that it is not only the regulations themselves but also their enforcement and application that is responsible for the high costs we estimate.

The finding that corruption *increases* regulatory costs may appear counter-intuitive given that some of the literature on regulations and corruption (e.g. [Khan, Khwaja, and Olken \(2015\)](#)) has emphasized the role corruption may play in *reducing* regulatory burden. In this

³⁸Note that our procedure is only capable of capturing distortions in the unit labor costs of firms, as those are the only ones that would show up as a downshift in the log firm size distribution. If the IDA imposes fixed costs, our procedure will not detect them. GLV identify fixed costs from bunching at N , but this is not possible for us because reported bunching may not reflect actual bunching, as discussed in Section 4.2.

³⁹The variation across ownership types is straightforward to explain: the regulations are clearly not applied in the same way to privately owned establishments and government enterprises.

conceptualization, firms bribe inspectors to underreport violations. However, as we show in subsection 6.3, if one allows for the possibility that inspectors can extort firms by threatening to impose large fines for technical violations, the relationship between regulatory burden and corruption becomes theoretically ambiguous and can easily be positive.⁴⁰ In Appendix D, we provide evidence that the costs we estimate may also have significant adverse effects in the longer run, as they are associated with lower future growth of employment in *registered* manufacturing - and higher future growth of employment in *unregistered* manufacturing. This evidence suggests that high effective regulatory costs may play a role in the recent trend towards informalization in the Indian economy. Before proceeding, we note that the results of the analyses we run in this section are necessarily somewhat speculative, since we do not claim to have isolated as-good-as-random variation in states' corruption environments and do not believe such variation exists. Therefore, the results in this section should be treated as suggestive, offering areas for future research to probe.

6.1 τ and Measures of Regulation

We start by regressing our state-level estimates of τ against other established measures of the regulatory environment (see Table 3).⁴¹ These include the “Besley Burgess” (BB) measure of labor regulations from Aghion et al. (2008) as well as several measures of regulatory reform from Dougherty (2009). The former is a measure of the number of amendments that a state government has made to the IDA 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 1997.⁴² Positive values indicate more “pro-worker” amendments, which are assumed to imply a more restrictive environment for firms operating in those states. Dougherty

⁴⁰As described in Section 2, India’s labor regulations are often characterized as being complex, unclear, outdated and arbitrary, so that it is difficult for firms to comply with all of their stipulations. It is thus likely that the Indian setting provides a fertile ground for extortionary corruption.

⁴¹Note that the estimates of τ we use in all the analysis below were generated using the procedure in Section 4.3.

⁴²Since there were no state-level amendments to the IDA between 1997 and 2005, this measure is appropriate for use with 2005 data.

(2009) provides state level reform indicators that reflect “the extent to which procedural or administrative changes have reduced transaction costs in relation to labor issues” by “limiting the scope of regulations, providing greater clarity in their application, or simplifying compliance procedures”.⁴³ 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.⁴⁴ In the analysis below, we will focus on an overall measure of reforms from Dougherty (2009) as well as a measure of reforms regarding the role of inspectors.⁴⁵

Going forward, 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. Furthermore, we restrict most of our subsequent analysis to include only the 18 largest states by Net State Domestic Product (NSDP), for which data are most consistently available, because it leads to the most stable samples across specifications and because our estimates of τ are most precisely measured for the biggest states.⁴⁶ Note also that in this and all further regressions in which τ is the regressand, we weight observations using analytic weights inversely proportional to the variance of our estimate of τ .⁴⁷

In Table 3, correlations are reported between τ and the other regulatory measures both by themselves and while controlling for other factors (NSDP per capita and the state’s share of

⁴³These measures are the result of surveying “a labour expert designated by the AIOE [All-India Association of Employers] or Federation of Indian Chambers of Commerce and Industry (FICCI) affiliate in the state capital” of each state, and adjusting the answers “through discussions with local union leaders, independent labour experts, employers and state labour commissioners” (Dougherty (2009)).

⁴⁴Other than the IDA, the specific areas covered in Dougherty’s index include the “Factories Act, State Shops and Commercial Establishments Acts, Contract Labour Act, the role of inspectors, the maintenance of registers, the filing of returns and union representation” (Dougherty (2009)).

⁴⁵The latter of which captures the extent to which states have reformed rules to constrain the influence of inspectors and includes such actions as limiting the number of inspector visits to one per year and requiring authorization for specific complaints.

⁴⁶Our results are robust to this choice.

⁴⁷In general, using a dependent variable that is generated with error leads to standard errors that are biased upward. Weighted least squares is a standard approach for improving precision by weighting more heavily those observations that are estimated more precisely (see Allcott (2015) for another example). Nevertheless, our conclusion does not depend on this procedure as we obtain qualitatively similar results when using unweighted regressions.

privately owned establishments). The correlation between τ and the overall index of reforms from [Dougherty \(2009\)](#) is significant at the 10% level and has the correct sign: states that saw more “transaction cost reducing” reforms have lower τ s. Yet stronger is the association between τ and the subcomponent from [Dougherty \(2009\)](#) measuring reforms related to the role of inspectors, which is significant at the 1% level and is large in magnitude: a one standard deviation increase in the degree of inspector-related reforms is associated with a .65 standard deviation decrease in τ .⁴⁸

It is reassuring to see that our measure of τ is negatively correlated with Dougherty’s measures of transaction-cost reducing reforms. We should expect this, since Dougherty’s measures includes reforms that change how firms are impacted by laws across the 10 worker threshold.⁴⁹ By contrast, the expected correlation between τ and the BB measure is more ambiguous. On the one hand one might expect no correlation between τ and the BB measure, as the latter captures variation only due to state amendments to the IDA, which does not vary over the ten person threshold. On the other hand, many studies use the BB measure to capture the general regulatory environment (e.g. [Adhvaryu et al. \(2013\)](#)) so we might well expect it to correlate with our measure of regulatory costs. As it turns out, and with the considerable caveat that our power is limited by the very small number of observations available, we find an insignificant and relatively small coefficient on the BB measure.

In addition to the above measures regarding state-level changes to the statutory, procedural and administrative aspects of the regulations, we also regress τ against certain other measures of the labor environment. [Table 9](#) in the Appendix reports the results of τ regressed against per capita measures of strikes, man-days lost to strikes, lockouts, man-days lost to lockouts and the percentage of registered factories that have been inspected. One might imagine that strikes and lockouts capture relevant features of the regulatory and labor

⁴⁸ τ is not significantly correlated with most of the 7 other subcomponent measures from [Dougherty \(2009\)](#). One notable exception is reforms related to the use of contract workers (not depicted here).

⁴⁹For example, reforms that affect the powers of inspectors certainly have a differential impact on firms above and below the threshold since firms above the threshold fall under the legal ambit of many more inspectors than firms below the threshold. See [Section 6.3](#) for a complete theoretical discussion.

environment,⁵⁰ but we do not find them to be robustly correlated with τ . We do find, echoing the results of Table 3, a robust correlation between τ and the percentage of registered factories inspected.

To briefly summarize the analysis so far, the strongest results point to a link between τ and the inspection regime. In particular, imposing reforms that constrain the powers of inspectors is correlated with lower effective regulatory costs for firms. This could be because constraining inspectors allows firms to avoid the *de jure* costs associated with following the rules, or it could be because constraining inspectors makes it harder for them to extort firms for bribes. If the latter, we should expect a strong link between τ and the corruption level of the environment. We look into this next.

6.2 τ and Corruption

The first three columns of Table 4 report the results of regressing τ against state level corruption as measured in a 2005 Transparency International (TI) Survey.⁵¹ Column 1 displays the bivariate relationship, Column 2 adds some basic controls (NSDP per capita, and the share of privately owned establishments), and Column 3 adds the measure of inspector-related 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 .6-.8 standard deviation increase in τ . Figure 5 in Appendix E, which depicts the partial residual plot associated with column 3 of Table 4, confirms that the relationship is not driven by outliers.

One might be concerned, however, that the TI measure may be flawed as it is partly the result of individuals' perceptions. To check for robustness of the relationship between τ and corruption, we regress τ against an alternative measure of corruption that is not perception-based. Columns 4-6 of Table 4 report the results of τ regressed against the (normalized)

⁵⁰For example, some industrial regulations explicitly undermine or support the rights of parties to engage in strikes or lockouts.

⁵¹The TI corruption measure is based on a survey of the perceptions and experiences regarding corruption in the public sector among 14,405 respondents (Indian households) in 20 Indian states.

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, the correlations between τ and this alternative measure of corruption are significant and large in magnitude regardless of sample or controls. 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. Again, we include a partial residual plot associated with Column 6 of Table 4 to demonstrate that the results are not driven by outliers (Figure 6 in Appendix E).

Although the state-level correlations between τ and corruption are robust, the regressions are subject to the concern that our measures of corruption may be correlated with omitted variables that also influence τ . To partially address this concern, we provide analysis in Appendix E that corroborates our results using a conceptually different source of variation by taking advantage of within-state, industry level heterogeneity in the exposure to corruption. In particular, we hypothesize that if regulations are especially costly *due to* corruption in their enforcement, then we would expect costs to be highest for those businesses that are engaged in regulation-heavy industries *when* they are located in states with high corruption. Using data from the World Bank’s 2005 Firm Analysis and Competitiveness Survey of India (FACS) to create an industry level measure of “regulatory intensity,” we test this hypothesis in Table 10 of Appendix E and find that the interaction between industry level “regulatory intensity” and state level corruption is indeed associated with higher effective regulatory costs (τ).⁵² In this analysis, bias can only arise if omitted variables exist that are correlated with the interaction between state level corruption and industry level “regulatory intensity,” which is harder to argue.

The main goal of this and the preceding tests is to demonstrate that the link between

⁵²The analysis and data construction are described more fully in Appendix E.

regulatory costs and corruption is stable and robust to very different ways of conceptualizing exposure to corruption. 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, we outline a simple theoretical framework to elucidate the potential connection in the following subsection.

6.3 A Theoretical Framework for Understanding Corruption Between Inspectors and Firms

We distinguish between two types of corruption that could take place between inspectors and firms: collusion and extortion. Collusion takes place when corrupt 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.⁵³ However, what we observed in Section 6.1 was a robust *positive* correlation between effective costs (τ) and poor governance/corruption. To explain this phenomenon, we sketch a simple model which nests both collusionary and extortionary behavior in order to illustrate how corruption might lead to *higher* per worker costs for firms that exceed the 10 worker threshold.

Before we go through the formalism of the model, it is worth pausing to flesh out the nature of the extortionary practices that firms may be exposed to when crossing the 10 worker regulatory threshold. Suppose that the regulations which apply to firms with more than 10 workers are so complex as to make it impossible (or prohibitively costly) for any firm to be fully in compliance with all aspects of the law as written. This does not require much imagination. As mentioned in Section 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 A and the descriptions of the excessive specificity of the regulations we provided in Section 2.

⁵³See, for example, a model of corruption such as the one in Khan et al. (2015).

Then, a dishonest inspector can, at any time, choose to subject a firm under his jurisdiction to a penalty, which may include financial (e.g. fines) and/or non-financial elements (e.g. harassment, time needed to defend claims of violations, prison terms). One could think of the extent of the penalty 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).

Next we describe the set-up of the model, the timeline of which is described below and in Figure 4. In the first stage, firms must choose their number of workers ($n \geq 10$ or $n' < 10$).⁵⁴ As in Section 4.1, firms are characterized by a productivity parameter α , 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 certain 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 κ , 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 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 as described above - 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

⁵⁴Throughout, primes will denote the values of variables on the side of the decision tree in which firms hire less than 10 workers.

has no further financial obligations. If the firm loses, it is obliged to pay an amount $F_2(n)$. Let us assume that $F_2(n)$ is much larger than $F_1(n)$, which is tantamount to supposing that a reasonable level of compliance with regulations is not costly in comparison to the punishments that could be brought by an inspector for violating the regulations. This may be a reasonable assumption in contexts where inspectors have a great amount of bargaining power and discretion in assigning punishments. This assumption is what makes the behavior extortionary rather than collusive: if $F_1(n)$ were large in comparison with $F_2(n)$, firms would face lower costs with corrupt inspectors than with honest ones and would be better off colluding. 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: $F_i(n) = f_i * n$. The decision tree representing the firm's choices described above is provided in Figure 4.

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 α and β 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 inspectors (and thus pay bribes) with probability κ 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 $p = 0$, $p' = 1$, α and β 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 (i.e. what we call τ in the paper) 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. τ) is increasing in the proportion of corrupt inspectors, κ . 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 anecdotal evidence regarding bribery in India (some of which we present in Appendix A). To conclude this subsection, the model above illustrates conditions that may explain the correlations we observed between corruption and τ in Section 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) will face higher per worker costs for exceeding the 10 worker threshold (higher τ) if inspectors have bargaining power and discretion in assigning heavy punishments for technical infractions.

7 Conclusion

This paper makes several contributions to the literature on labor regulations in developing countries. We provide estimates of the unit labor costs associated with a suite of regulations whose components have hitherto received little attention in the literature on developing countries. These regulations include mandatory benefits, workplace safety provisions, and reporting requirements where the literature has previously emphasized employment protection legislation and minimum wage laws. In the Indian context, we find that the costs associated with the combination of mandatory benefits, workplace safety provisions, and reporting requirements are much larger than those associated with the most stringent portion of the country’s employment protection legislation. Our results suggest that these types of regulations deserve more attention than they have received to this point.

Our results also suggest a mechanism that may explain why these regulations are so costly in a developing country context: high *de facto* regulatory costs appear to be driven by extortionary corruption on the part of inspectors. Specifically, we show that Indian states that have reformed their inspector-related regulations in a positive way face lower regulatory costs and states with the highest levels of corruption also have the highest levels of regulatory costs. Furthermore, firms in industries with high “regulatory intensity” are exposed to particularly high costs if they are also located in highly corrupt states. This analysis points to the size of regulatory costs’ having more to do with the way regulations are implemented than with the content of the specific laws themselves.

In addition to the above, our paper also makes a methodological contribution. We extend GLV’s theoretical model to allow firms to strategically misreport their sizes and simultaneously develop an empirical strategy to estimate costs from a firm size distribution under the assumptions of our model. We show that ignoring the problem of misreporting can lead to vastly over-estimating the actual costs of the regulations. We believe this contribution will find applications in other developing country settings, where the costs of strategic misreporting are typically low.

We close by noting that our analysis reveals the net costs of regulations borne by firms, but does not speak directly to the possible benefits to workers. Our results do suggest that the current regulations make it easy for inspectors to penalize firms for technical violations rather than violations of grave consequence.⁵⁵ To the extent that this is so, it is unlikely that workers derive as much protective benefit from the regulations as they might otherwise. It is difficult to arrive at more concrete conclusions, as data do not exist that would allow one to measure how workers would benefit if their employers were made to follow the spirit rather than the letter of the law. However, the results hint at an intriguing possibility: by simplifying regulations identified as costly (or by clarifying compliance and enforcement), it may be possible to reduce the costs borne by firms without diminishing effective protection for workers.

References

- Adhvaryu, A., A. Chari, and S. Sharma (2013). Firing Costs and Flexibility: Evidence from Firms' Labor Adjustments to Shocks in India. *Review of Economics and Statistics* 95(3).
- Aghion, P., R. Burgess, S. J. Redding, and F. Zilibotti (2008). The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India. *American Economic Review* 98(4), 1397–1412.
- Allcott, H. (2015). Site Selection Bias in Program Evaluation. *The Quarterly Journal of Economics* 130, 1117–1165.
- Almunia, M. and D. Lopez-Rodriguez (2015). Under the Radar: The Effects of Monitoring Firms on Tax Compliance. *Working Paper*.
- Asher, S. and P. Novosad (2015a). Market Access and Structural Transformation: Evidence from Rural Roads in India. *Working Paper*.
- Asher, S. and P. Novosad (2015b). The Impacts of Local Control over Political Institutions: Evidence from State Splitting in India. *Working Paper*.
- Asher, S. and P. Novosad (2016). Politics and Local Economic Growth: Evidence from India. *American Economic Journal: Applied Economics*, forthcoming.

⁵⁵We think of this distinction as being related to the difference between enforcing the letter versus the spirit of the law.

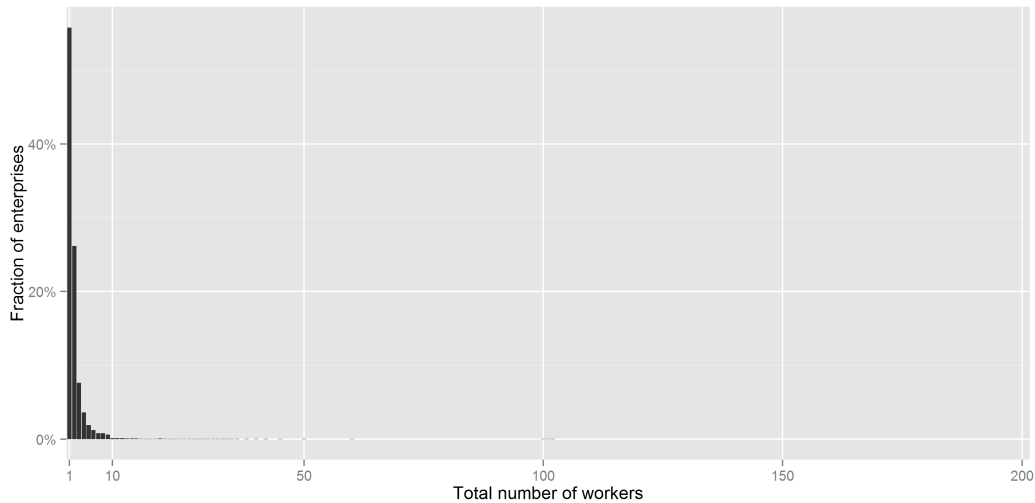
- Axtell, R. (2001). Zipf Distribution of U.S. Firm Sizes. *Science* 293(5536), 1818–1820.
- Bajaj, V. (2011, November 30). Outsourcing Giant Finds It Must Be Client, Too. *The New York Times*.
- Banerjee, A. (1994). Eliminating Corruption. In M. Quibria (Ed.), *Proceedings of the Third Annual Conference on Development Economics*. Asian Development Bank.
- Basu, K. (2011). Why, for a Class of Bribes, the Act of Giving a Bribe should be Treated as Legal. *Working Paper*.
- Bertrand, M., C.-T. Hsieh, and N. Tsivanidis (2015). Contract Labor and Firm Growth in India. *Working Paper*.
- Besley, T. and R. Burgess (2004a). Can Labor Regulation Hinder Economic Performance? Evidence from India. *The Quarterly Journal of Economics* 119(1), 91–134.
- Besley, T. and R. Burgess (2004b). Can Labor Regulation Hinder Economic Performance? Evidence from India. *The Quarterly Journal of Economics* 119(1), 91–134.
- Bhattacharjea, A. (2006). Labour Market Regulation and Industrial Performance in India: A Critical Review of the Empirical Evidence. *Indian Journal of Labour Economics* 49(2).
- Bhattacharjea, A. (2009). The Effects of Employment Protection Legislation on Indian Manufacturing. *Economic and Political Weekly XLIV*(22), 55–62.
- Botero, J. C., S. Djankov, R. La Porta, F. Lopez-De-Silanes, and A. Shleifer (2004). The Regulation of Labor. *The Quarterly Journal of Economics* 119(4), 1339–1382.
- Chatterjee, U. and R. Kanbur (2013). Regulation and Non-Compliance: Magnitudes and Patterns for India’s Factories Act. *Working Paper*.
- Chaurey, R. (2015). Labor Regulations and Contract Labor Use: Evidence from Indian Firms. *Journal of Development Economics* 114, 224–232.
- de Andrade, G. H., M. Bruhn, and D. McKenzie (2014). A Helping Hand or the Long Arm of the Law? Experimental Evidence on What Governments Can Do to Formalize Firms. *The World Bank Economic Review* (May), lhu008.
- de Mel, S., D. Mckenzie, and C. Woodruff (2013). The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka. *American Economic Journal: Applied Economics* 5(2), 122–150.
- Debroy, B. (2013). India’s Segmented Labour Markets, Inter-State Differences, and the Scope for Labour Reforms. In B. Debroy, L. Bhandari, S. Aiyar, and A. Gulati (Eds.), *Economic Freedom of the States of India*, Chapter 4. Academic Foundation.
- Djankov, S. and R. Ramalho (2009). Employment laws in developing countries. *Journal of Comparative Economics* 37(1), 3–13.

- Dougherty, S., V. Frisancho, and K. Krishna (2014). State-Level Labor Reform and Firm-Level Productivity in India. *India Policy Forum* 10.
- Dougherty, S. M. (2009). Labour Regulation and Employment Dynamics at the State Level in India. *Review of Market Integration* 1(3), 295–337.
- Fagernas, S. (2010). Labor Law, Judicial Efficiency, and Informal Employment in India. *Journal of Empirical Legal Studies* 7(2), 282–321.
- Freeman, R. B. (2010). *Labor regulations, unions, and social protection in developing countries: Market distortions or efficient institutions?* (1 ed.), Volume 5. Elsevier BV.
- Garicano, L., C. Lelarge, and J. Van Reenen (2013). Firm Size Distortions and the Productivity Distribution: Evidence from France. *NBER Working Paper* 18841.
- Garicano, L., C. Lelarge, and J. Van Reenen (2016). Firm Size Distortions and the Productivity Distribution: Evidence from France. *American Economic Review* 106(11), 3439–3479.
- Ghosh, A. (2016, February 28). Why Make in India is stumbling over our labour laws (Comment: Special to IANS). *Business Standard*.
- Goyette, J. and W. Kouame (2016). Tax Evasion, Social Infrastructures and Misallocation. *Working Paper*.
- Gruber, J. (1995). The incidence of payroll taxation: evidence from Chile. *Journal of Labor Economics* 15(53), S72–S101.
- Hasan, R., D. Mitra, and K. V. Ramaswamy (2007). Trade reforms, labor regulations, and labor-demand elasticities: Empirical evidence from India. *The Review of Economics and Statistics* 89(August), 466–481.
- Hernández-Pérez, R., F. Angulo-Brown, and D. Tun (2006). Company Size Distribution for Developing Countries. *Physica A: Statistical Mechanics and its Applications* 359, 607–618.
- Hindriks, J., M. Keen, and A. Muthoo (1999). Corruption, Extortion and Evasion. *Journal of Public Economics* 74(3), 395–430.
- Hsieh, C. T. and P. J. Klenow (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics* 124(4), 1403–1448.
- Hsieh, C.-T. and B. A. Olken (2014). The Missing "Missing Middle". *Journal of Economic Perspectives* 28(3), 89–108.
- Huntington, S. P. (1968). *Political Order in Changing Societies*. New Haven, CT: Yale University Press.
- Kanbur, R. and L. Ronconi (2015). Enforcement Matters: The Effective Regulation of Labor. *CEPR Discussion Paper No. 11098*.

- Khan, A., A. Khwaja, and B. Olken (2015). Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors. *The Quarterly Journal of Economics*, forthcoming.
- Kleven, H. J. and M. Waseem (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *Quarterly Journal of Economics* 128(2), 669–723.
- Kochhar, K., U. Kumar, R. Rajan, A. Subramanian, and I. Tokatlidis (2006). India’s Pattern of Development: What Happened, What Follows? *Journal of Monetary Economics* 53(5), 981–1019.
- Kugler, A. and M. Kugler (2009). Labor Market Effects of Payroll Taxes in Developing Countries: Evidence from Colombia. *Economic Development and Cultural Change* 57(2), 335–358.
- Kumler, T., E. Verhoogen, and J. Frias (2015). Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico. *Working Paper*.
- Lucas, R. E. J. (1978). On the size distribution of business firms. *The Bell Journal of Economics* 9(2).
- Markovitch, N. and U. Krieger (2000). Nonparametric Estimation of Long-Tailed Density Functions and its Application to the Analysis of World Wide Web Traffic. *Performance Evaluation* 42(2-3), 205–222.
- Mishra, A. and D. Mookherjee (2013). Controlling Collusion and Extortion: The Twin Faces of Corruption. *Working Paper*.
- Mookherjee, D. (1997). Incentive Reforms in Developing Country Bureaucracies. In *Annual World Bank Conference on Development Economics*. World Bank.
- Nataraj, S., F. Perez-Arce, K. B. Kumar, and S. V. Srinivasan (2014). The impact of labor market regulation on employment in low-income countries: A meta-analysis. *Journal of Economic Surveys* 28(3), 551–572.
- Polinsky, A. M. and S. Shavell (2001). Corruption and Optimal Law Enforcement. *Journal of Public Economics* 81(1), 1–24.
- Sequeira, S. and S. Djankov (2014). Corruption and firm behavior: Evidence from African Ports. *Journal of International Economics* 94, 277–294.
- Svensson, J. (2005). Eight Questions about Corruption. *Journal of Economic Perspectives* 19(3), 19–42.
- TeamLease Services (2006). India Labour Report 2006: A Ranking of Indian States by their Labour Ecosystem. Technical report.

Figures and Tables

Figure 1: 2005 Distribution of Establishment Size



Note: total number of workers is the number of workers usually working daily in an establishment. The graph is cut off at 200 total workers because the fraction of establishments of sizes greater than 200 is too small to appear in the figure. Source: 2005 Economic Census of India. Source: 2005 Economic Census of India.

Figure 2: 2005 Log-Log Distribution of Establishment Size



Note: Both axes are on a log scale. Total number of workers is the number of workers usually working daily in an establishment. Source: 2005 Economic Census of India.

Figure 3: Model Fit and Data



Note: This figure shows the fit of the model described in Section 4.3 (the black line) to the data (dark grey points). Model estimation involves non-parametric smoothing using the method described in Markovitch and Krieger (2000) with a bandwidth of 0.005 as a first step. The smoothed density estimates are shown in light grey. The second step is to fit the black line to the light grey points. Both axes in log scale. Source: 2005 Economic Census of India.

Figure 4: Decision Tree

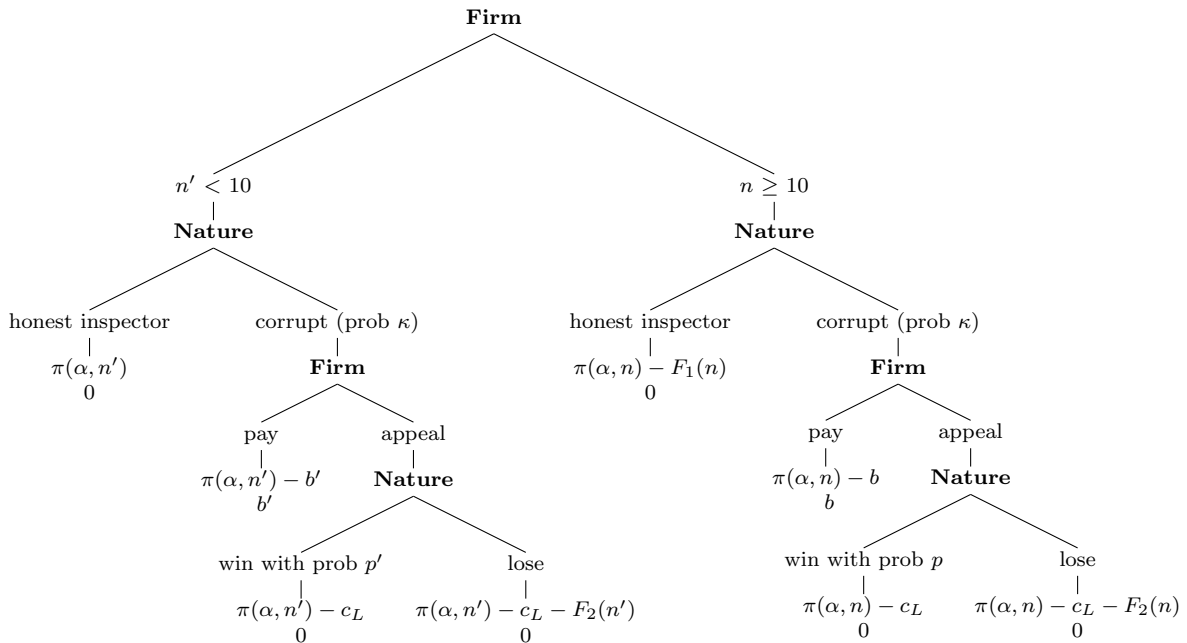


Table 1: Estimates of τ at 10 worker threshold

Level	τ	s.e.
All-India	0.347	(0.081)
By State		
Bihar	0.693	(0.302)
Gujarat	0.165	(0.151)
Kerala	0.138	(0.196)
Karnataka	0.520	(0.156)
Uttar Pradesh	0.502	(0.254)
By Industry		
Wholesale & retail trade	0.637	(0.094)
Manufacturing	0.268	(0.085)
Construction	0.478	(0.549)
Electricity, gas and water	-0.367	(0.145)
By Ownership Type		
Government and PSU	-0.092	(0.128)
Unincorporated Proprietary	0.430	(0.059)

Note: This table presents estimates of regulatory costs faced by establishments that hire 10 or more workers, using the methodology described in Section 4 with a bandwidth of 0.005. Standard errors generated using a clustered bootstrap procedure with 200 replications are presented in parentheses. Clustering is done at the 4 digit (NIC code) industry level, following [Garicano et al. \(2016\)](#). Estimates are presented for a subset of states, industries and ownership types, as well as at the All-India level. Results for all states, industries and ownership types are available in Appendix C. Source: 2005 Economic Census of India.

Table 2: Estimate of τ at 100 worker threshold

All-India	τ	s.e.
	0.044	(0.039)

Note: This table presents an estimate of regulatory costs faced by establishments that hire 100 or more workers, using the methodology described in Section 4. The standard error generated using a clustered bootstrap procedure with 200 replications is presented in parentheses. Clustering is done at the 4 digit (NIC code) industry level, following [Garicano et al. \(2016\)](#). The estimate is presented for the All-India level. Source: 2005 Economic Census of India.

Table 3: Tau vs Other Measures of Regulations

	(1)	(2)	(3)	(4)	(5)	(6)
	tau	tau	tau	tau	tau	tau
Dougherty measure (all reforms)	-0.424*	-0.431*				
	(0.212)	(0.235)				
Dougherty measure (inspector reforms)			-0.549***	-0.652***		
			(0.170)	(0.150)		
Besley-Burgess measure (regs)					0.223	0.237
					(0.182)	(0.183)
log of net state domestic product pc		-0.463*		-0.527**		-0.518**
		(0.262)		(0.204)		(0.235)
share of privately owned establishments		-0.0632		8.524		-12.80
		(7.095)		(6.707)		(8.354)
Constant	0.203	4.945	0.288*	-1.923	0.00679	16.67*
	(0.222)	(6.800)	(0.149)	(6.529)	(0.292)	(7.798)
Observations	18	18	18	18	15	15

Note: This table tests for correlations between our estimated regulatory costs (tau) and other established measures of the regulatory environment from the previous literature. Robust standard errors are reported in parentheses. Observations are weighted by the inverse variance of tau and include only the 18 largest Indian States, as measured by NSDP. Sources: Dougherty(2009); Besley and Burgess(2004); RBI.

Table 4: Tau vs State Level Measures of Corruption

	(1)	(2)	(3)	(4)	(5)	(6)
	tau	tau	tau	tau	tau	tau
TI Corruption Score	0.812**	0.806**	0.685***			
	(0.290)	(0.342)	(0.129)			
electricity losses				0.925***	0.948**	0.575***
				(0.310)	(0.326)	(0.190)
log of net state domestic product pc		-0.0667	-0.213		-0.219	-0.350**
		(0.338)	(0.285)		(0.131)	(0.133)
share of privately owned establishments		-2.323	6.335*		1.365	7.900
		(4.493)	(3.376)		(4.940)	(5.153)
Dougherty measure (inspection reforms)			-0.594***			-0.500***
			(0.0932)			(0.116)
Electricity available (GWH)					0.109	
					(0.134)	
Constant	0.334	3.075	-2.935	0.476**	1.365	-2.956
	(0.213)	(4.682)	(3.836)	(0.169)	(4.634)	(4.728)
Observations	18	18	18	18	18	18
Measure of Corruption	TI	TI	TI	TDLs	TDLs	TDLs

Note: This table reports the results of our estimated regulatory costs (tau) regressed against two different measures of corruption. Robust standard errors are reported in parentheses. Observations are weighted by the inverse variance of tau and include only the 18 largest Indian States, as measured by NSDP. Sources: Transparency International (2005); RBI; Dougherty(2009).

Appendices for Online Publication Only

A Qualitative 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. ——— (M: ———). 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 employees 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 valuable 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 ———— . Mobile no. ———— 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)

B Omitted Proofs Regarding the Theoretical Log Density of Firm Size With and Without Misreporting

B.1 Derivation of the theoretical log density without misreporting

In this part of the appendix, we include the steps omitted in section 4.1 when deriving the theoretical log density of firm size.⁵⁶ As noted earlier, the primitive of the model is the distribution of managerial ability (α), which we assume follows a power law: $\phi(\alpha) = c_\alpha \alpha^{-\beta_\alpha}$, where $c_\alpha \equiv (\beta_\alpha - 1) \underline{\alpha}^{\beta_\alpha - 1}$ and $\underline{\alpha}$ is the minimum possible value of α ⁵⁷.

⁵⁶The analysis here follows closely from that of GLV (see their Appendix B for their derivation).

⁵⁷These last two assumptions are needed to satisfy $\int_{\underline{\alpha}}^{\infty} \phi(\alpha) = 1$.

A firm with productivity α and Cobb Douglas production function ($f(n) = n^\theta$) faces the following profit-maximization problem:

$$\pi(\alpha) = \max_n \alpha n^\theta - w\bar{\tau}n$$

where again, n is the number of workers a firm employs, w is a constant wage paid to all workers, and $\bar{\tau}$ is a proportional tax on labor that takes the value 1 if $n \leq N$ and $1 + \tau$ if $n > N$, where $\tau > 0$. The resulting first order condition suggests the following general relationship between employment and productivity: $n^*(\alpha) = (\frac{\theta}{w\bar{\tau}})^{\frac{1}{1-\theta}} (\alpha)^{\frac{1}{1-\theta}}$. However, this relationship is discontinuous at N and only applies for interior solutions: some firms will find it profitable to choose the corner solution of $n(\alpha) = N$ rather than $n^*(\alpha)$.

In fact, firms can be sorted into three categories, according to their productivity, α .

1) Firms with the lowest values of α ($\alpha \in [\underline{\alpha}, \alpha_1]$) are not affected by the regulation and choose their optimal employment in an unrestricted way. In particular they choose $n^*(\alpha) = (\frac{\theta}{w})^{\frac{1}{1-\theta}} (\alpha)^{\frac{1}{1-\theta}} \leq N$. α_1 is defined such that $\pi(n^*(\alpha_1)) = \pi(N)$, where $n^*(\alpha_1) = (\frac{\theta}{w})^{\frac{1}{1-\theta}} (\alpha_1)^{\frac{1}{1-\theta}}$.

2) Firms with productivity larger than α_1 but lower than another threshold ($\alpha \in (\alpha_1, \alpha_2]$), find it optimal to choose $n^*(\alpha) = N$, rather than exceed the threshold and expose themselves to the discontinuously higher costs associated with the size-based regulation.

3) The last category includes those firms with the highest productivities ($\alpha > \alpha_2$). These firms find it optimal to exceed the threshold even though it means paying higher unit labor costs: $n^*(\alpha) = (\frac{\theta}{w(1+\tau)})^{\frac{1}{1-\theta}} (\alpha)^{\frac{1}{1-\theta}}$. α_2 is defined such that $\pi(n^*(\alpha_2)) = \pi(N)$, where $n^*(\alpha_2) = (\frac{\theta}{w(1+\tau)})^{\frac{1}{1-\theta}} (\alpha_2)^{\frac{1}{1-\theta}}$.

To summarize then, a full mapping between productivity α and firm size n is given by:

$$n(\alpha) = \begin{cases} (\frac{\theta}{w})^{\frac{1}{1-\theta}} (\alpha)^{\frac{1}{1-\theta}} \leq N & \text{if } \alpha \in [\underline{\alpha}, \alpha_1] \\ N & \text{if } \alpha \in (\alpha_1, \alpha_2] \\ (\frac{\theta}{w(1+\tau)})^{\frac{1}{1-\theta}} (\alpha)^{\frac{1}{1-\theta}} > N & \text{if } \alpha > \alpha_2 \end{cases}$$

An exact expression for the distribution of firm size, $\chi(n)$, can now be recovered as a transformation of the distribution of managerial ability, $\phi(\alpha)$, since the first-order conditions on the firms' maximization problems imply the monotonic relationship between α and n described above. Specifically, we transform $\phi(\alpha)$ into $\chi(n)$ using the change of variables formula along with the complete expression for $n(\alpha)$ above:

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

B.2 Full Derivation of the Theoretical Log Density With Misreporting

In this part of the appendix, we include the steps omitted in section 4.2 when deriving the theoretical log density of true and reported firm size in the presence of misreporting. We begin by restating the profit-maximization problem (in Equation 6) for a firm that is now allowed to choose both its true employment (n) and its *reported* employment (l):

$$\pi(\alpha) = \max_{n,l} \alpha f(n) - wn - \tau wl * \mathbb{1}\{l > 9\} - F(n, l) * p(n, l)$$

where $\alpha, f(n), w$ and τ are all defined as they were previously. As noted, the problem is similar to the case without misreporting except that now firms pay the extra marginal cost, τw , only on their *reported* employment, and not on their true employment. Furthermore, they only pay this cost if their reported employment exceeds the regulatory threshold, $N = 9$. The other new elements are $p(n, l)$, which is the probability that a misreporting firm is caught by the authorities and $F(n, l)$, which is the fine that a firm caught misreporting must pay. Therefore, firms must trade off the benefit of lower regulatory costs from under-reporting their employment against the expected cost of being caught and fined for under-reporting ($F * p$).

For now we assume a particular functional form for the misreporting costs, $F(n, l) = F$ and $p(n, l) = \min \left\{ \frac{(n-l)^2}{100}, 1 \right\}$, but we show in the next subsection of this Appendix that our main result obtains for any convex form of misreporting costs. We also reintroduce our earlier assumption that firms' production functions are power ($f(n) = n^\theta$), so that the profit maximization problem for a firm with productivity α is:

$$\pi(\alpha) = \max_{n,l} \alpha n^\theta - wn - \tau wl * \mathbb{1}\{l > 9\} - F * \frac{(n-l)^2}{100}.$$

As before, the solution to this problem looks different depending on which of three different productivity categories the firm falls into:

1) Firms with the lowest values of α ($\alpha \in [\underline{\alpha}, \alpha_1]$) are not affected by the regulation and choose their optimal employment in an unrestricted way. In particular they choose $n_1^*(\alpha) = \left(\frac{\theta}{w}\right)^{\frac{1}{1-\theta}} (\alpha)^{\frac{1}{1-\theta}} \leq N$. Because their true employment is below the regulatory threshold, they have no incentive to misreport and hence choose $l_1^*(\alpha) = n_1^*(\alpha)$. α_1 is defined such that $\pi(n_1^*(\alpha_1)) = \pi(N)$, where $n_1^*(\alpha_1) = \left(\frac{\theta}{w}\right)^{\frac{1}{1-\theta}} (\alpha_1)^{\frac{1}{1-\theta}}$.

2) Firms with productivity larger than α_1 but lower than another threshold ($\alpha \in (\alpha_1, \alpha_2]$), will choose $n > N$, exceeding the regulatory threshold, but will find it profitable to misreport

their employment, setting $l_2^*(\alpha) = N$.⁵⁸ These firms will only *appear* to be “bunched” up at 9, but will in fact have higher employment. The employment function, $n_2^*(\alpha)$, for these firms is defined implicitly from the first order condition: $\alpha\theta n_2^*(\alpha)^{\theta-1} - w - \frac{F}{50}[n_2^*(\alpha) - N] = 0$.

3) The last category includes those firms with the highest productivities ($\alpha > \alpha_2$). These firms are productive enough to warrant hiring work forces so large that they cannot choose $l = 9$ while simultaneously avoiding 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. From the first order condition on l we get: $l_3^*(\alpha) = n_3^*(\alpha) - \frac{50}{F}w\tau$. In other words, these firms can save on their unit labor costs by shading down their reported employment. Importantly, the degree of misreporting is by a constant amount that is independent of firm size ($\frac{50}{F}w\tau$).⁵⁹ These firms set their true employment according to: $n_3^*(\alpha) = (\frac{\theta}{w(1+\tau)})^{\frac{1}{1-\theta}}(\alpha)^{\frac{1}{1-\theta}}$. α_2 is defined such that $\pi(n_3^*(\alpha_2)) = \pi(n_2^*(\alpha_2))$, where $n_3^*(\alpha_2) = (\frac{\theta}{w(1+\tau)})^{\frac{1}{1-\theta}}(\alpha_2)^{\frac{1}{1-\theta}}$.

To summarize then, full mappings between productivity α and the true firm size n , as well as between productivity α and reported firm size l , are given by:

$$n^*(\alpha) = \begin{cases} (\frac{\theta}{w})^{\frac{1}{1-\theta}}(\alpha)^{\frac{1}{1-\theta}} \leq N & \text{if } \alpha \in [\underline{\alpha}, \alpha_1] \\ n_2^*(\alpha) & \text{if } \alpha \in (\alpha_1, \alpha_2] \\ (\frac{\theta}{w})^{\frac{1}{1-\theta}}(1+\tau)^{-\frac{1}{1-\theta}}(\alpha)^{\frac{1}{1-\theta}} > N & \text{if } \alpha > \alpha_2 \end{cases}$$

$$l^*(\alpha) = \begin{cases} (\frac{\theta}{w})^{\frac{1}{1-\theta}}(\alpha)^{\frac{1}{1-\theta}} \leq N & \text{if } \alpha \in [\underline{\alpha}, \alpha_1] \\ N & \text{if } \alpha \in (\alpha_1, \alpha_2] \\ (\frac{\theta}{w})^{\frac{1}{1-\theta}}(1+\tau)^{-\frac{1}{1-\theta}}(\alpha)^{\frac{1}{1-\theta}} - \frac{50}{F}w\tau > N & \text{if } \alpha > \alpha_2 \end{cases}$$

From these functions we can obtain expressions for the distributions of true and reported firm size, $\chi(n)$ and $\psi(l)$, as transformations of the distribution of managerial ability, $\phi(\alpha)$ (where $\phi(\alpha) = c_\alpha \alpha^{-\beta\alpha}$), by the change of variables formula:

$$\chi(n) = \begin{cases} c_\alpha(1-\theta)(\frac{\theta}{w})^{\frac{\beta-1}{1-\theta}}n^{-\beta} & \text{if } n \in [n_{\min}, N) \\ \left| \frac{d\alpha_2^*(n)}{dn} \right| \phi(\alpha_2^*(n)) & \text{if } n \in [N, n_2^*(\alpha_2)) \\ 0 & \text{if } n \in [n_2^*(\alpha_2), n_3^*(\alpha_2)) \\ c_\alpha(1-\theta)(\frac{\theta}{w})^{\frac{\beta-1}{1-\theta}}(1+\tau)^{-\frac{\beta-1}{1-\theta}}n^{-\beta} & \text{if } n \geq n_3^*(\alpha_2) \end{cases}$$

⁵⁸Conditional on misreporting a positive amount, setting $l = N$ is the “optimal lie” for these firms since it yields the largest benefit (by reducing firms’ regulatory burden to 0) while minimizing the misreporting costs.

⁵⁹This outcome is a result of the convex cost assumption on the misreporting function.

$$\psi(l) = \begin{cases} c_\alpha(1-\theta)\left(\frac{\theta}{w}\right)^{\frac{\beta-1}{1-\theta}}l^{-\beta} & \text{if } l \in [n_{\min}, N) \\ \int_{\alpha_1}^{\alpha_2} \phi(\alpha)d\alpha = \delta_l & \text{if } l = N \\ 0 & \text{if } l \in (N, l_3^*(\alpha_2)) \\ c_\alpha(1-\theta)\left(\frac{\theta}{w}\right)^{\frac{\beta-1}{1-\theta}}(1+\tau)^{-\frac{\beta-1}{1-\theta}}\left[l + \frac{50}{F}w\tau\right]^{-\beta} & \text{if } l \geq l_3^*(\alpha_2) \end{cases}$$

where $\beta = \theta + \beta_\alpha - \theta\beta_\alpha$ and $\alpha_2^*(n)$ is the inverse function of $n_2^*(\alpha)$, implicitly defined above. Taking the logarithm of each expression delivers the version of the distributions shown in the main text:

$$\log\chi(n) = \begin{cases} \log A - \beta\log(n) & \text{if } n \in [n_{\min}, 9) \\ \log[\xi(n)] & \text{if } n \in [9, n_m(\alpha_2)] \\ - & \text{if } n \in (n_m(\alpha_2), n_t(\alpha_2)) \\ \log A - \frac{\beta-1}{1-\theta}\log(1+\tau) - \beta\log(n) & \text{if } n \geq n_t(\alpha_2) \end{cases}$$

$$\log\psi(l) = \begin{cases} \log A - \beta\log(l) & \text{if } l \in [l_{\min}, 9) \\ \log(\delta_l) & \text{if } l = 9 \\ - & \text{if } n \in (9, l_t(\alpha_2)) \\ \log A - \frac{\beta-1}{1-\theta}\log(1+\tau) - \beta\log\left(l + \frac{50}{F}w\tau\right) & \text{if } l \geq l_t(\alpha_2) \end{cases}$$

where terms have been simplified with two substitutions.⁶⁰ These are the densities described in the main text.

B.3 Proof that Convex Misreporting Costs Imply Convergence Between True and Reported Firm Size Distributions

An important implication of the misreporting model from section 4.2 is that the difference between the log density of reported firm size, $\psi(l)$, and the log density of true firm size, $\chi(n)$, converges to 0 for large values of n, l . In this section of the appendix we show that this result does not hinge on a specific functional form for the misreporting costs, but instead requires only that the expected costs of misreporting are increasing and strictly convex in the degree of misreporting.

We replace our former expression for the expected costs of misreporting, $F(u) * p(u)$, with the simpler expression $M(u)$, since here we do not need to distinguish between the fine if caught and the probability of being caught. Given these substitutions, the problem of a firm with the option of misreporting can be written:

⁶⁰We substituted A for the expression $c_\alpha(1-\theta)\left(\frac{\theta}{w}\right)^{\frac{\beta-1}{1-\theta}}$ and $\xi(n)$ for $\left|\frac{d\alpha_2^*(n)}{dn}\right|\phi(\alpha_2^*(n))$.

$$\pi(\alpha) = \max_{n,u} \alpha f(n) - wn - \tau w(n - u) * \mathbb{1}\{n - u > 9\} - M(u)$$

which is identical to Equation 6 except for the change in variables (u for $n - l$) and the more general expression for the expected costs of misreporting. Under Assumption 1, the first order condition on u for a large firm (i.e. one whose reported employment exceeds the threshold) is: $\tau w - M'(u) = 0$. The first term denotes the benefit of increasing u by one unit (in terms of regulatory costs avoided) while the second term captures the marginal cost of u . The first term is constant, while the second starts from 0 (for $u = 0$) and increases at an increasing rate. There exists therefore some value of misreporting that satisfies the first order condition, given by $u^* = M'^{-1}(\tau w)$. Note that the optimal value for misreporting, u^* , does not depend on α . This means that, for the largest set of firms, misreporting is by the same constant amount, regardless of firm size or productivity: $l(\alpha) = n(\alpha) - u^*$.

To see that this result is all that is required for the difference between the reported distribution and the true distribution to converge to 0, consider the analysis in the previous sub-appendix, but with the more general result that $u^* = M'^{-1}(\tau w)$. Then, it is straightforward to show that $\log\chi(n) = \log A - \frac{\beta-1}{1-\theta}\log(1 + \tau) - \beta\log(n)$ and $\log\psi(l) = \log A - \frac{\beta-1}{1-\theta}\log(1 + \tau) - \beta\log(l + u^*)$ for firms above the threshold. For large values (i.e. $l = n = x \rightarrow \infty$), the difference between these two density functions goes to 0:

$$\lim_{x \rightarrow \infty} \log\chi(x) - \log\psi(x) = \beta\log(x + u^*) - \beta\log(x) = \beta\log(x(1 + \frac{u^*}{x})) - \beta\log(x) = \beta\log(1 + \frac{u^*}{x}) = 0.$$

B.4 Misreporting by Enumerators

In the text we referred to a second potential source of misreporting: not only might firms lie to enumerators about their size, enumerators themselves might lie when recording the figures reported to them by firms. One reason this might happen is 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 therefore 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’’. To show that this other source of potential misreporting is unlikely to bias our results, we consider a very simple model of enumerator misreporting. The model demonstrates that, since the cost of filling in the address slip is a fixed cost, it is not likely to lead to a ‘‘downshift’’ in the distribution, which means it is therefore unlikely to bias our estimate of τ .

The model begins with firms facing the same problem specified in Equation 1, with the same resulting distribution of the true firm size, $\chi(n)$. Then, all firms are matched with

an enumerator, who must decide how to report the size of the firm they meet. In general, the reported size, l , may or may not be equal to the true firm size, n . If an enumerator reports a size $l > 9$, they must face the burden of filling out an address slip, at cost $C > 0$. If they report $l \leq 9$, they pay no such cost. Importantly, the cost C is constant and does not depend on the size of the firm.⁶¹ Furthermore, enumerators face expected costs of misreporting, $M(u)$, where $u \equiv n - l$.⁶² $M(u)$ captures both the probability of being caught as well as the penalty faced if caught. The only assumptions we make on $M(u)$ are that it is strictly increasing in u and that $M(0) = 0$.⁶³ Then the utility maximization problem faced by an enumerator matched with a firm of size n is:

$$U(n) = \max_u -C\mathbb{1}\{n - u > 9\} - M(u)$$

For enumerators matched with firms of size $n < 9$, the optimal decision is to choose $l = n$, or $u = 0$, because there is no need to misreport. Then, their utility is maximized at 0. Enumerators matched with firms larger than 9 must decide whether to report the size truthfully and bear the address slip cost or lie in order to avoid the cost of filling out the address slip. Since $M(u)$ is increasing in u , misreporting costs will be lower than C for low values of u and higher than C for high enough values.⁶⁴ Therefore, enumerators matched with firms of an intermediate size (i.e. $n \in [9, \bar{n}]$) will find it optimal to lie by choosing $l = 9$ or $u = n - 9$, in order to avoid the fixed cost of filling out the address slip. \bar{n} is defined such that $C = M(\bar{n} - 9)$, so that an enumerator matched with a firm of size \bar{n} would be indifferent between misreporting ($l = 9$) and reporting truthfully ($l = \bar{n}$). For enumerators matched with firms of size $n > \bar{n}$, the cost of misreporting exceeds the cost of filling out the address slip, so they are better off bearing the address slip cost and reporting truthfully (i.e. setting $l = n$ or $u = 0$).

To summarize, we obtain the following relationship between reported and true employment:

$$l(n) = \begin{cases} n & \text{if } n \in [n_{min}, 9) \\ 9 & \text{if } n \in [9, \bar{n}] \\ n & \text{if } n > \bar{n} \end{cases}$$

⁶¹This reflects the fact that it is no more costly (in terms of time or hassle) to fill in an address slip for a firm of size 20 than a firm of size 200.

⁶²As before we will only consider nonnegative values of u , since there will be no incentive to over-report firm size.

⁶³The latter assumption is made only for simplicity of exposition.

⁶⁴The other possibility is that $M(u) < C \forall u$, but this case is not very interesting and is clearly not borne out by the data (it would suggest that enumerators should *always* misreport firm size to be 9).

Given this mapping between reported and actual employment, the reported firm size density is given by:

$$\psi(l) = \begin{cases} \chi(n) & \text{if } l \in [n_{\min}, 9) \\ \int_9^{\bar{n}} \chi(n) dn = \delta_e & \text{if } l = 9 \\ 0 & \text{if } l \in (9, \bar{n}] \\ \chi(n) & \text{if } l \geq \bar{n} \end{cases}$$

In other words, the enumerator misreporting will cause the reported distribution to exhibit “bunching before the threshold as well as a “valley” after the threshold, even if neither phenomena exists in the true distribution. However, the reported distribution coincides with the true distribution before the threshold and for values far above the threshold. In other words, enumerator misreporting does not cause a downshift in the reported distribution in excess of the downshift in the true distribution and hence does not bias our estimate of τ .

C Full Results by State, Industry and Ownership Type for the 10-worker Threshold

Table 5: Estimates of τ by State

<i>State</i>	<i>Tau</i>	<i>Standard Error</i>
Bihar	.693	.302
Karnataka	.52	.156
Uttar Pradesh	.502	.254
Delhi	.427	.213
Tamil Nadu	.397	.154
Jharkhand	.388	.194
Madhya Pradesh	.379	.203
Maharashtra	.332	.107
Assam	.322	.345
Rajasthan	.32	.174
Orissa	.283	.139
Gujarat	.165	.151
West Bengal	.151	.071
Kerala	.138	.196
Punjab	.096	.158
Haryana	.007	.168
Andhra Pradesh	-.159	.053
Himachal Pradesh	-.165	.159

Note: This table presents estimates of regulatory costs faced by establishments that hire 10 or more workers, using the methodology described in Section 4. Standard errors generated using a clustered bootstrap procedure with 200 replications are presented in parentheses. Clustering is done at the 4 digit (NIC code) industry level, following GLV. Estimates are presented for the 18 largest states (by NSDP). Source: 2005 Economic Census of India.

Table 6: Estimates of τ by Industry

<i>Industry</i>	<i>Tau</i>	<i>Standard Error</i>
Wholesale and retail trade	.637	.094
Real estate, renting and business activities	.601	.158
Construction	.478	.549
Hotels and restaurants	.468	.222
Transport, storage and communications	.334	.209
Manufacturing	.268	.085
Other service activities	.264	.186
Health and social work	.076	.149
Mining and quarrying	-.042	.294
Financial intermediation	-.105	.074
Education	-.173	.15
Public administration and defence	-.311	.034
Electricity, gas and water supply	-.367	.145

Note: This table presents estimates of regulatory costs faced by establishments that hire 10 or more workers, using the methodology described in Section 4. Standard errors generated using a clustered bootstrap procedure with 200 replications are presented in parentheses. Clustering is done at the 4 digit (NIC code) industry level, following GLV. Estimates are presented for all major industry categories. Source: 2005 Economic Census of India.

Table 7: Estimates of τ by Ownership Type

<i>Ownership Type</i>	<i>Tau</i>	<i>Standard Error</i>
Unincorporated proprietary	.43	.059
Co-operative	-.007	.075
Non profit institution	-.04	.095
Unincorporated partnership	-.058	.053
Government and public sector undertaking	-.092	.128
Corporate financial	-.18	.055
Corporate non financial	-.197	.05

Note: This table presents estimates of regulatory costs faced by establishments that hire 10 or more workers, using the methodology described in Section 4. Standard errors generated using a clustered bootstrap procedure with 200 replications are presented in parentheses. Clustering is done at the 4 digit (NIC code) industry level, following GLV. Estimates are presented by ownership type of the establishment. Source: 2005 Economic Census of India.

D Possible Consequences of τ

In Section 6 we argued that our estimated costs (τ) are mostly 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 τ . In what follows we use two distinct measures of τ : one which is created using all the establishments in a state, regardless of economic sector (τ) and another which is created using only the establishments engaged in manufacturing (τ_{manuf}).

Table 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 (τ and τ_{manuf}) as well alternative measures (Dougherty and BB). For each of the four measures, we observe its performance as a predictor of future employment growth in *registered* manufacturing as well as *unregistered* manufacturing. Interestingly, in the regressions of employment growth in registered manufacturing against τ_{manuf} , the coefficient on τ_{manuf} is negative and significant at the 5% level, while the coefficient for employment growth in *unregistered* manufacturing is positive and significant. This result makes sense: we should expect higher costs to negatively affect 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.⁶⁵ If these correlations reflect a causal chain, it would mean that high levels of regulatory costs and corruption (as measured by τ) are pushing employment from the registered to the unregistered sector.

Also included in Table 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.⁶⁶

⁶⁵It also makes sense that the coefficients on τ are insignificant, since τ is measured across all sectors and will be less pertinent to manufacturing performance than τ_{manuf} .

⁶⁶One might argue that it is not quite fair to regress growth between 2010 and 2005 on a regressor that uses data from 1997, as is the case for the BB measure. However, we have duplicated these results using growth from 1997 to 2002 and the results are the same. Furthermore, the Besley Burgess measure from Aghion et al. (2008) should be the same in 2005 due to the lack of state level reforms between 1997 and 2005.

Table 8: Manufacturing Employment Growth (2005 - 2010) vs Tau and Other Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	reg manuf	unreg manuf	reg manuf	unreg manuf	reg manuf	unreg manuf	reg manuf	unreg manuf
tau	-0.0240 (0.0176)	0.00197 (0.0233)						
tau (manuf)			-0.0471** (0.0217)	0.0623** (0.0256)				
Besley-Burgess measure (regs)					-0.00525 (0.00731)	0.00979 (0.0142)		
Dougherty measure (all reforms)							0.0226 (0.0130)	-0.0143 (0.0159)
log of net state domestic product pc	0.00312 (0.0178)	0.0189 (0.0214)	0.0107 (0.0145)	0.0192 (0.0161)	0.00413 (0.00863)	0.0140 (0.0168)	0.0212 (0.0154)	0.0136 (0.0195)
share of employment in manufacturing	-0.393 (0.258)	0.00558 (0.329)	-0.708** (0.258)	0.435 (0.325)	0.0194 (0.186)	-0.559 (0.362)	-0.515* (0.245)	0.0525 (0.323)
Constant	0.0969 (0.173)	-0.182 (0.209)	0.0372 (0.139)	-0.229 (0.152)	0.0209 (0.0825)	-0.0675 (0.160)	-0.0861 (0.147)	-0.131 (0.182)
Observations	18	17	18	17	15	15	18	17

Note: This table reports the results of employment growth in the registered and unregistered manufacturing sectors against several measures of the regulatory environment, including our own estimated regulatory costs (tau). Robust standard errors are reported in parentheses. Observations are unweighted and include only the 18 largest Indian States, as measured by NSDP. Sources: Besley and Burgess (2004); Dougherty(2009); RBI.

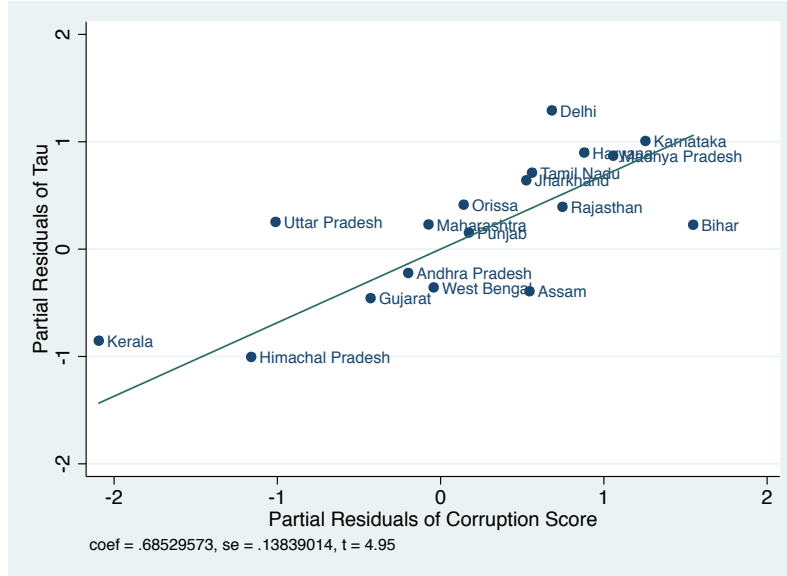
E Further Results Related to Exploration of Mechanisms

Table 9: Tau vs Strikes and Lockouts

	(1)	(2)	(3)	(4)	(5)
	tau	tau	tau	tau	tau
strikes per capita	-0.0584 (0.208)				
mandays lost due to strikes per capita		-0.00607 (0.0771)			
lockouts per capita			0.0367 (0.0789)		
mandays lost due to lockouts per capita				0.0289 (0.0750)	
percent of factories inspected					0.736** (0.290)
log of net state domestic product pc	-0.398 (0.286)	-0.387 (0.258)	-0.407 (0.264)	-0.407 (0.261)	0.547 (0.743)
Constant	4.085 (2.729)	3.946 (2.597)	4.132 (2.630)	4.138 (2.612)	-5.644 (7.681)
Observations	18	17	18	18	10

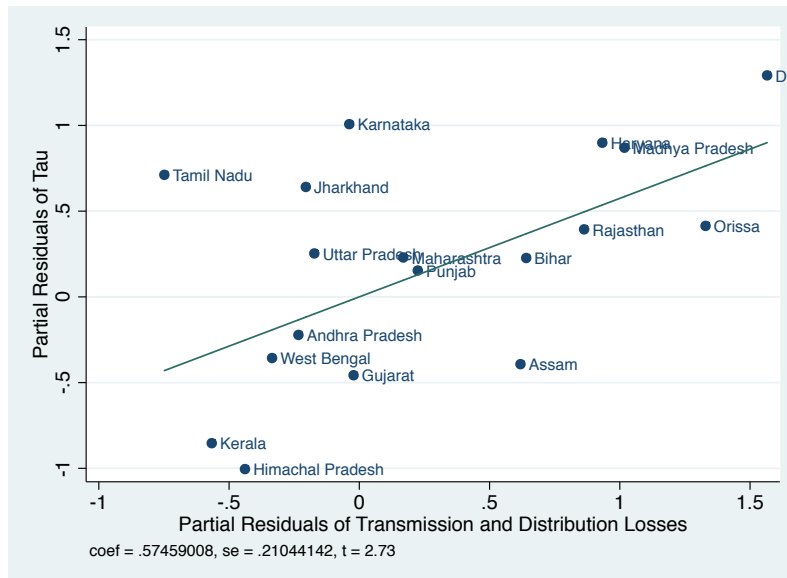
Note: This table tests for correlations between our estimated regulatory costs (τ) and other miscellaneous measures of the labor environment. Robust standard errors are reported in parentheses. Observations are weighted by the inverse variance of τ and include only the 18 largest Indian States, as measured by NSDP. Sources: Indian Labour Year Book (2005).

Figure 5: Partial Residual Plot: Tau vs Transparency International Corruption Score



Note: This figure is the graphical analogue of column 3 in Table 4. It depicts the relationship between the components of tau and the TI Corruption Score that are unexplained by the other covariates. Sources: Transparency International (2005); Dougherty(2009); RBI.

Figure 6: Partial Residual Plot: Tau vs Transmission and Distribution Losses



Note: This figure is the graphical analogue of column 6 in Table 4. It depicts the relationship between the components of tau and electricity transmission and distribution losses that are unexplained by the other covariates. Sources: Dougherty(2009); RBI.

E.1 τ and Corruption: State X Industry Analysis

In this portion of the Appendix, we explain our State X Industry analysis (described in Section 6.1) in more detail. The purpose of this analysis is to partially address concerns

that the state-level correlations between τ and corruption lack exogenous variation and may be biased if our measures of corruption are correlated with omitted variables that also influence τ . To do so, we take advantage of State X Industry level heterogeneity as an additional source of variation. We use data from the World Bank’s 2005 Firm Analysis and Competitiveness Survey of India (FACS) to create an industry level measure of the extent to which regulations are problematic, which we term “regulatory intensity”. Specifically, Indian firms in the 2005 FACS were asked whether “regulations specific to [their] industry” were problematic for their “operation and growth”. Averaging the firm-level responses by industry, we classify industries according to how likely businesses are to complain about *industry-specific* regulations. If regulations are especially costly due to corruption in their enforcement, then we would expect costs to be highest among those businesses in regulation-heavy industries *and* in states with high corruption. That is, we would expect the *interaction* between industry level “regulatory intensity” and state level corruption to be positive.

To test our hypothesis we generate our measures of τ at the State X Industry level⁶⁷ and regress those measures against interactions of state level corruption with industry level regulatory intensity. The results, shown in Table 10 with and without interaction terms, support our hypothesis. First, when excluding the interaction terms (columns 1 and 3), the main effects (state level corruption and industry level regulatory intensity) are significantly correlated with τ in the expected directions. When interaction terms are included (columns 2 and 4), their coefficients are also large and significant, suggesting that the presence of industry specific regulations is most costly when firms are located in a corrupt environment.

⁶⁷Industries here are categorized according to their groupings in the World Bank Enterprise Surveys, which distinguishes 23 distinct industry categories. Examples include “auto components”, “leather and leather products”, and “food processing”. We only generated τ for state X industry cells with a sufficient number of observations (in particular, for those with at least 40 observations in the size distribution), and were thus left with only 190 observations out of a possible 414 (23*18).

Table 10: Tau vs State Level Corruption Interacted with Industry Level “Regulatory Intensity”

	(1)	(2)	(3)	(4)
	tau	tau	tau	tau
log of net state domestic product pc	-0.348*** (0.0298)	-0.312*** (0.0311)	-0.186*** (0.0288)	-0.142*** (0.0298)
TI Corruption Score	0.0592*** (0.0140)	0.143*** (0.0291)		
electricity TDLs			0.226*** (0.0215)	0.224*** (0.0206)
Regulatory Intensity	0.259*** (0.0288)	0.314*** (0.0328)	0.0937** (0.0284)	0.103*** (0.0274)
TI Corruption Score X Regulatory Intensity		0.160** (0.0492)		
electricity TDLs X Regulatory Intensity				0.176*** (0.0435)
Constant	-0.424*** (0.00849)	-0.404*** (0.0103)	-0.534*** (0.0121)	-0.576*** (0.0156)
Observations	190	190	190	190

Note: This table reports the results of our estimated regulatory costs (tau) regressed against state level corruption, industry level regulatory intensity, and their interaction. Robust standard errors are reported in parentheses. Observations are now at the state X industry level but are still weighted by the inverse variance of tau and include only the 18 largest Indian States, as measured by NSDP. Sources: Transparency International (2005); RBI; World Bank Firm Analysis and Competitiveness Survey of India (2005).