SMALL BY CHOICE? REASSESSING THE COSTS OF SIZE-BASED REGULATIONS

KALYANI PADMAKUMAR
Department of Economics, Florida State University

Firms in many countries face stricter regulation once employment crosses a threshold. This paper asks how costly such rules are, their aggregate implications, and what trade-offs reforms entail. I study Chapter V-B of India's Industrial Disputes Act, which imposes firing restrictions and an enforcement-related wage wedge on manufacturing plants with 100 or more full-time workers. Exploiting staggered state-level reforms that raised the threshold from 100 to 300 workers, I provide causal evidence on plant responses to the change in regulatory exposure. I then combine these estimates with a dynamic model of heterogeneous plants facing labor market frictions to quantify and decompose the regulatory burden. Three findings emerge. First, the burden is overwhelmingly a firing cost. Second, full-time employment rises whether the regulation is made universal or repealed, but aggregate output and value added per worker move in opposite directions. Third, bunching-based methods understate the cost of size-based regulations when it operates primarily through firing costs, a lesson that extends beyond this setting.

KEYWORDS: Regulation, Firm Dynamics, Misallocation, Labor Market Frictions, India, Economic Development.

1. INTRODUCTION

Firms with employment exceeding a threshold level face stricter regulation than smaller firms in many countries. These size-based rules take various forms, from requiring firms to fund mandatory works councils and share profits with workers (as in France and Peru), to costly and lengthy legal procedures in the event of worker dismissals (as in India and Italy).

Kalyani Padmakumar: kpadmakumar@fsu.edu

I am profoundly grateful to Kala Krishna for her valuable insights and steadfast support. I am also deeply indebted to Fernando Parro, Shoumitro Chatterjee, James Tybout and Qi Li for their vital inputs and continuous encouragement. Special thanks to David Atkin, Ritam Chaurey, Gaurav Chiplunkar, Jonathan Eaton, Jingting Fan, Michael Gechter, Matthew Gentry, Marc Henry, Hugo Hopenhayn, Omer Koru, Vijay Krishna, Karthik Muralidharan, Nitya Pandalai-Nayar, Bradley Setzler, Daniel Xu, Stephen Yeaple, and Yingyan Zhao for their generous comments that tremendously improved this paper. All errors are mine.

Although intended to preserve existing jobs and improve working conditions, these protections typically apply only to workers hired directly by the firm on a permanent, open-ended basis.[1] Therefore, if regulatory costs are large, firms may hire fewer permanent workers and instead rely on workers hired through legally permitted alternative arrangements that fall outside the scope of these regulations.[2] If these workers are not perfect substitutes for permanent workers, firms may operate below their efficient scale with a distorted input mix, lowering aggregate output and productivity. These considerations raise several questions. How large are the costs these regulations impose on firms? What are the implications for the level and composition of employment, and for aggregate output? And what trade-offs should policymakers weigh in reforming them?

Answering these questions is challenging. Because regulatory costs are unobserved, they must be inferred from firm behavior around the threshold. A dominant approach in existing work is to infer these costs from the extent of "bunching"—the excess mass of firms that locate just below the cutoff to avoid the regulation (Guner et al., 2008, García-Santana and Pijoan-Mas, 2011, Gourio and Roys, 2014, Garicano et al., 2016, Amirapu and Gechter, 2020). The idea is that the more firms cluster below the threshold, the stronger their incentives to avoid the regulation.[3] But bunching is an imperfect diagnostic. One well-known reason is that the extent of bunching reflects not only the magnitude of regulatory costs but also other structural factors, such as how easily firms can adjust their employment (Kleven, 2016). A second reason, which this paper takes up, is that whether firms bunch depends on the nature of the regulatory cost itself, not just its magnitude. For instance, as I show, if the regulation primarily increases dismissal costs beyond the threshold, firms respond in ways that do not lead to bunching, however costly the regulation might be. Therefore, in order to measure regulatory costs credibly, one needs information beyond the observed distribution

[1] This asymmetry is a general feature of labor regulations. For instance, in the case of employment protection legislation, termination costs are significantly lower for temporary, fixed-term contracts than for permanent and open-ended contracts in virtually all OECD countries (OECD, 2026). This is also true in non-OECD countries. For instance, in India, severance pay, mandatory notice, and the requirement of government authorization for retrenchment apply only to permanent workers, and not to contract workers hired through staffing agencies (Bertrand et al., 2025).

[2] Examples include contract workers hired through staffing agencies in India, workers hired on fixed-term contracts in France, Spain, and Italy.

[3] For instance, the French firm size distribution exhibits a visible spike at 49 employees, just below the regulatory threshold of 50 employees (see Figure A.1. in the Appendix).

of firms around the threshold, such as variation in the regulation itself, together with a way to discipline the other frictions separately.

In this paper, I address these challenges by exploiting a natural experiment in the context of India's employment protection legislation. Commonly known as Chapter VB of the Industrial Disputes Act (henceforth, IDA) of 1947, this legislation applies to manufacturing plants with 100 or more full-time workers, and it imposes two distinct costs on them.[4] First, plants above the threshold must obtain government authorization before dismissing even a single worker. Securing this authorization entails considerable delays, increasing the uncertainty plants face during adverse shocks and effectively raising their firing costs. Second, government officials enforce labor regulations through frequent inspections, which gives them the power to extract bribes and increase administrative burdens on plants. Because plants incur these costs continuously and bribes are often linked to the number of workers employed, this enforcement regime raises the marginal cost of production above the threshold.

Two features of this setting make it well suited to answer the questions discussed earlier. First, beginning in 2014, several Indian states raised IDA's applicability threshold from 100 to 300 full-time workers. The staggered implementation of these reforms generates variation in regulatory costs across similarly sized-plants, within the same state, over time. Second, Indian manufacturing plants employ contract workers hired through staffing companies, who fall outside the purview of the IDA. This allows me to study how plants respond to these regulations when legally permitted alternatives are available.

My analysis proceeds in three parts. In the first part, I use the staggered reforms to provide causal evidence on how plants responded to the change in the regulatory threshold. In the second part, I build a dynamic model of heterogeneous plants facing labor market frictions, and use it to show how the two costs imposed by the IDA affect plant dynamics differently. In the third part, I combine the natural experiment and causal estimates from the first part with the structural model to quantify each cost, the aggregate implications of the regulation, and evaluate counterfactual regulatory regimes.

---

[4] Throughout the paper, "full-time workers" refers to workers hired directly by the plant on a permanent, open-ended basis—the category the IDA protects. I use the terms full-time and permanent workers interchangeably in this paper.

More specifically, in the first part, I use detailed plant-level panel data for all Indian states from 1999 to 2018 to document two key effects of the reform. First, post-policy, treated plants are 23.07% more likely to transition above the old regulatory threshold of 100 full-time workers that is no longer binding. Conversely, they are 21.68% less likely to transition above the newly binding 300 full-time worker threshold, post-policy. This suggests that the legislation distorts the plant size distribution by encouraging plants to remain small. Second, plants with fewer than 100 full-time workers in treated states reduced their reliance on contract labor and capital following the reform: the share of contract workers in production fell by 4.29% and the investment rate in plant and machinery declined by 1.3%. Since the regulation applies only to full-time workers, this indicates that plants had previously substituted toward these inputs to avoid the regulation. To rule out alternative explanations, I show that transitions and substitution patterns are unaffected around other employment thresholds. Furthermore, event studies and stacked difference-in-differences designs confirm that these significant impacts start only from the year of the reform, and these are not due to forbidden comparisons between the early-treated and late-treated states.

The above reduced-form exercise motivates three questions. First, as it has been established in Hsieh and Olken (2014) and Amirapu and Gechter (2020), there is no evidence of plants bunching at 100 full-time workers in the Indian plant size distribution (see Figure A.2. in the Appendix). Why do plants not bunch at the regulatory threshold despite the IDA being costly?[5] Second, how large are regulatory costs, and what drives them – is it primarily the delays in securing government authorization for layoffs, or the enforcement regime itself? Third, what would be the consequences of reforming the legislation, such as changing where the threshold is set? To address these questions, the second part of my analysis develops a dynamic heterogeneous firm model with multiple factors of production and labor search frictions.

The model I develop incorporates several features from existing work on heterogeneous firm dynamics models with labor market frictions – monopolistically competitive markets, plants heterogeneous in productivity which evolves over time, and an endogenous mass

[5]The absence of bunching is not because plants use other margins of adjustment. Plants could still hire full-time workers up to exactly 100 and rely on contract labor and capital beyond that point, which would generate bunching. That they do not bunch suggests a different mechanism is at work, which I explore through the quantitative model.

of plants operating in the economy (Hopenhayn, 1992, Hopenhayn and Rogerson, 1993, Melitz, 2003, Das et al., 2007, Helpman et al., 2010, Felbermayr et al., 2011, Elsby and Michaels, 2013, Acemoglu and Hawkins, 2014, Kaas and Kircher, 2015, Coşar et al., 2016, Dix-Carneiro et al., 2021). I add to this setup in two ways. First, plants employ two types of workers —full-time and contract workers. I assume they are imperfect substitutes, and I estimate their elasticity of substitution from the data. This allows plants to substitute away from full-time workers to avoid being regulated, and allows the model to account for these margins when evaluating counterfactuals. Second, motivated by the Indian institutional context, I model the adjustment of full-time employment to be costly and time-consuming. Hiring requires effort, and greater effort raises the Poisson arrival rate of workers, but high effort levels involve convex search costs. Similarly, each full-time worker separates from the plant at an exogenous Poisson rate, which plants can raise at a cost to expedite firing. Crucially, because plants choose only these rates rather than employment levels directly, the timing of worker transitions is stochastic. This formulation draws on models of network formation in international trade (Eaton et al., 2016, 2022). The regulation enters through two margins – it imposes a wedge on wage rate and firing cost once plants employ 100 or more full-time workers.

The model shows that the two regulatory wedges distort plant behavior in different ways. Suppose first that regulatory costs are dominated by the wage wedge. As plants approach the threshold from below, the expected marginal benefit of having more workers falls, reducing their search effort. The larger the wage wedge, the lower the probability that plants transition from below to above the threshold. This slower inflow into the region above the threshold shows up in the steady-state distribution as a sharp increase in the density of plants below the threshold and a sharp decrease in plant density above the threshold—the canonical bunching signature.

Suppose instead that regulatory costs are dominated by the wedge on firing costs. Then the contrast between plant density on either side of the threshold is weaker than under the wage wedge. This is because the firing wedge reduces both upward and downward transitions around the threshold. For plants above the threshold, a higher firing wedge makes them more likely to hoard labor and stop firing actively altogether, since the marginal benefit from dismissing a worker no longer justifies the cost. For plants below the threshold, optimal search effort decreases, since a worker who is costly to fire in the future is worth

less to have today. Thus, the firing wedge reduces transitions in both directions, causing plant density to rise on both sides of the threshold and attenuating the bunching signature. The model therefore suggests that a regulation dominated by the firing cost wedge can be very costly while producing little bunching.

In the third part, I estimate the model on plants in treated states using the simulated method of moments procedure, assuming that their pre-reform size distribution was in steady state. The causal estimates from the first part of the paper serve as key moments. I use the reform-induced change in contract labor share to identify the elasticity of substitution between full-time and contract workers. The reform-induced change in plant transitions around 100 full-time workers, together with the density of plants on either side of the threshold, identifies the two regulatory costs. The estimates show that the firing cost accounts for the major part of the regulatory burden. Over and above the baseline firing costs that all plants face, the IDA imposes an additional firing cost equal to 1.66 times the annual wage for each worker dismissed. The wedge on their wage rate is only 1.54%. Therefore, the absence of bunching at 100 full-time workers in the Indian plant size distribution does not imply that the legislation is cheap; it is instead what we would expect when the regulatory burden is dominated by firing costs.

What would be the consequences of reforming this legislation? Although the estimated regulatory costs are large, the answer depends on several countervailing forces. First, plants have been substituting towards contract workers to circumvent the legislation, so the impact of reforms depend on the estimated elasticity of substitution between full-time and contract workers. Second, since the regulatory burden is overwhelmingly a firing cost, the net impact of a reduction in firing costs on employment is ambiguous: while plants may hire more aggressively in response to positive shocks, they may also fire more readily after negative ones. Third, general equilibrium effects, such as those from increased firm entry, could amplify or dampen the overall impact on output.

After accounting for these channels, the estimated model implies that removing the IDA would raise aggregate output by 1.56%, full-time employment by 11.79%, share of full-time employment by 11.54%, and value added per worker by 1.33%. That full-time employment increases indicates that the increase in hiring dominates the increase in separations. That plants readily substitute toward full-time workers once the regulation is lifted suggests that their earlier reliance on contract labor was largely a response to the regula-

tion rather than to underlying production needs. With lower firing costs, plants that had been hoarding full-time labor release them, while more productive plants search more intensively for workers, improving the allocation of full-time labor across plants, causing aggregate output and value-added per worker to rise. Raising the threshold from 100 to 300 full-time workers, the reform that was actually enacted, delivers roughly half of these gains.

This paper contributes primarily to two strands of literature. First, it relates to the literature on size-dependent labor regulations and the measurement of their costs. A large body of work has quantified the distortions caused by size-dependent labor regulations (Schivardi and Torrini, 2008, Guner et al., 2008, García-Santana and Pijoan-Mas, 2011, Gourio and Roys, 2014, Garicano et al., 2016, Amirapu and Gechter, 2020). These papers recover regulatory costs by comparing the observed firm size distribution to a counterfactual one—the distribution absent the regulation—using moments such as the excess mass of firms just below the threshold or their probability of transitioning above it. The principal advantage of this approach is that it exploits the variation the regulation itself creates across firms of different sizes; no further variation across time or space is required. The difficulty, as discussed earlier, is that regulatory costs and the other frictions firms face in adjusting employment must be recovered from the same firm size distribution – they compete to explain the same features of the distribution, and how that competition is resolved determines the estimated costs. This difficulty is well recognized in the literature on responses to tax kinks and notches, where the recommended solution is additional variation in the discontinuity (Kleven, 2016).

I contribute to our understanding of these regulations by combining differences in regulatory exposure across plants of different sizes with quasi-experimental variation in the location of the threshold itself. This setting delivers size distributions under two regulatory regimes for the same set of plants, generated by the same underlying adjustment frictions but different regulatory costs. I use this variation to provide causal evidence on how plants respond to a change in the regulatory threshold. I then combine the quasi-experimental variation with a structural model to separately estimate regulatory costs and other adjustment frictions. Finally, I show when the observed size distribution can and cannot reveal the cost of a size-contingent regulation. A regulation that acts like a tax on wages leaves the canonical bunching signature: excess mass below the threshold and missing mass above it,

with the contrast growing in the size of the wedge. A regulation that imposes a wedge on firing costs attenuates this signature, even when the regulation is costly, because it affects transitions in both directions across the threshold. Thus, when firing costs dominate the regulatory burden, or when a regulation combines a large wage wedge with a large firing cost, a bunching-based estimator will understate its cost. Nimier-David et al. (2026) is a related recent paper that uses quasi-experimental variation in the regulatory threshold to identify the causal effects of size-based profit-sharing requirements on firms and workers in France.

Second, it relates to the literature on misallocation in developing economies. A large literature has quantified the importance of misallocation for aggregate outcomes (Restuccia and Rogerson, 2008, Hsieh and Klenow, 2009, 2014, Hsieh and Olken, 2014, Restuccia and Rogerson, 2017, David and Venkateswaran, 2019, Baqaee and Farhi, 2020). Subsequent work has identified several sources of these distortions in Indian manufacturing, including capital market frictions (Banerjee and Duflo, 2014, Bau and Matray, 2023), labor market regulations (Besley and Burgess, 2004, Aghion et al., 2008, Chaurey, 2015), land market distortions (Duranton et al., 2015, Sood, 2020), entry and exit barriers (Bollard et al., 2013, Chari, 2011, Chatterjee et al., 2025), and product market regulations (Sivadasan, 2009, Alfaro and Chari, 2014).

Among labor regulations, Chapter V-B of the IDA has received the most attention, and the literature on it has moved in two stages. The first, beginning with the seminal work of Besley and Burgess (2004), measures the costs of the IDA using state-level indices of regulatory stringency constructed from the letter of the law (Hasan et al., 2007, Ahsan and Pagés, 2009). For instance, Besley and Burgess (2004) classifies states as pro-worker or pro-employer by aggregating state-level amendments to the IDA between 1958 and 1992, and compares plant outcomes across state groups. However, several concerns have been raised about how these amendments were interpreted, weighted, and aggregated, as well as about cross-state differences in enforcement arising from differences in administrative capacity (Bhattacharjea, 2006, 2021). More recent work instead focuses on plants' ability to use contract workers, who fall outside the Act, as an alternative to regulated full-time employment (Chaurey, 2015, Chiplunkar et al., 2024, Bertrand et al., 2025). Bertrand et al. (2025) shows that the improved access to contract workers relieved part of the IDA's burden, narrowing the gap in value added per worker between large and small plants and ac-

counting for a meaningful share of the increase in manufacturing productivity. Chiplunkar et al. (2024) shows that restricting access to contract labor reduced plants' propensity to operate formally, lowering aggregate productivity and welfare. Both Bertrand et al. (2025) and Chiplunkar et al. (2024) take the IDA as given and ask what the availability of an alternative to it has been worth.

I contribute to this literature by providing causal evidence on how the IDA affects plant outcomes. The threshold reforms allow me to do this without relying on state-level regulatory indices or cross-state differences in enforcement. I further ask what burden remains of the IDA given the availability of contract workers, what drives these costs, and what the aggregate implications of reforming the IDA would be. This requires accounting for the specific ways in which the IDA affects plants: it applies only when plants cross a size threshold, imposes a wedge on the wage rate, and raises the cost and uncertainty of adjusting full-time employment over and above the adjustment frictions plants already face. I therefore build and estimate a dynamic heterogeneous-plant model with labor market frictions that captures these features explicitly. In the model, adjusting employment is a costly, time-consuming, and uncertain process, and the IDA introduces an additional layer of costs and uncertainty when plants cross the regulatory threshold. I estimate the regulatory costs separately from the adjustment frictions, and decompose the regulatory burden into the costs associated with the enforcement regime and the firing restriction. This decomposition matters for evaluating reforms to the IDA because the two costs affect plants through different margins, and which cost dominates determines the trade-offs policymakers face. For instance, if the IDA is primarily a firing cost, as it turns out to be, reforming it would change both hiring and separations, and the net effect on employment would depend on which force dominates and may not be favorable.

The rest of the paper is organized as follows. In section 2, I discuss the IDA in greater detail and highlight the two channels through which it imposes costs on plants. Section 3 describes the data source that I use for my analysis. In Section 4, I present causal evidence on how plants responded to the change in IDA's threshold. Section 5 describes the quantitative model. In subsections 5.1 to 5.4, I lay out the model setup and equilibrium conditions. In subsection 5.5, I demonstrate the mechanisms through which size-based rules affect plant dynamics. Sections 6 and 7 discuss how I take the model to the data, model fit, parameter estimates, and the counterfactual results. Section 8 concludes.

2. INSTITUTIONAL BACKGROUND

A complex web of 463 Acts, 32542 compliances, and 3048 filings governs employer-employee relationships in Indian industries. Among these, the piece of legislation that has attracted the most attention in policy circles is Chapter VB of the Industrial Disputes Act (IDA) of 1947. Chapter VB of the IDA mandates that industrial plants employing more than 100 workmen obtain permission from a government authority before laying off even one worker. IDA defines a workman as "any person employed in any industry, to do any skilled or unskilled manual or clerical work, for hire or reward." This definition includes full-time or permanent workers hired directly by the plant for production. However, it excludes managerial staff and contract workers, as the latter are hired through staffing companies rather than directly by the plant.

The relevant authority is typically a labor court or industrial tribunal whose primary role is to adjudicate disputes between employers and employees. The labor court grants or refuses permission after conducting an inquiry in which both the employer and the affected worker present their case. Once permission is secured, the employer must pay the affected worker severance of 15 days' wages per year of service. If either party is dissatisfied with the verdict, they may appeal to the High Court or the Supreme Court.

It is widely argued that the IDA constrains plants from laying off workers even during adverse shocks, effectively raising firing costs for plants above the 100 full-time worker threshold. The opportunity cost of following the due process laid down by these rules is particularly high given India's weak state capacity. As of 1998, there existed only one labor court for every 437,868 workers. The number of cases pending before labor courts exceeds 100,000 as of 2020, with 35% unresolved for over a year and 37% pending for more than three years (Table A1 in the Appendix).[6] This large pendency of cases has been attributed to three factors: the absence of affected parties at hearings, frequent adjournments sought by parties to file documents, and parties approaching higher courts to challenge orders of reference or preliminary rulings (Debroy and Bhandari, 2008).

In addition to the costs arising from judicial delays, the manner in which the legislation is enforced also imposes costs on plants. Government officials monitor compliance through

[6]The number of labor-related cases pending before the High Court and Supreme Court also runs into the hundreds of thousands, as shown in Table A2 in the Appendix.

frequent inspections of plant premises, and inspectors have the authority — and, in practice, the discretion — to extract bribes and increase the administrative burden on plant owners. Amirapu and Gechter (2020) documents several instances of inspectors exploiting the arbitrary and antiquated nature of these laws to extract bribes. Importantly, plant owners incur these costs continuously, not only during adverse shocks, and the bribes they pay typically increase with the number of workers (Debroy and Bhandari, 2008, Amirapu and Gechter, 2020).

Although the magnitude of regulatory costs cannot be directly observed in the data, forward-looking plants would be expected to account for these costs when deciding whether to enter the market or expand beyond 100 full-time workers. These costs should also be reflected in their demand for other inputs, such as contract workers and capital.[7] To infer whether the regulation is costly, however, one requires an exogenous variation in regulatory costs.

Given this objective, a series of policy changes implemented in several Indian states acquire significance. In 2014, the state of Rajasthan raised the IDA's applicability threshold from 100 to 300 full-time workers, thereby relaxing regulatory costs for plants employing between 100 and 300 full-time workers. Between 2015 and 2017, several other states followed suit and implemented the same reform. Thus, while the threshold stood at 100 full-time workers for all states until 2013, by 2017 a subset of states had raised it to 300 while many others retained the original threshold.[8] This staggered implementation across states generates variation in regulatory costs that I exploit to test whether plants responded to the reform. Section 3 describes the data, and I present my analysis in the subsequent sections.

3. DATA SOURCE

In this paper, I use the Annual Survey of Industries (ASI) dataset, published by the Ministry of Statistics and Programme Implementation, Government of India. The unit of

[7] Plants hire contract workers through staffing companies on a temporary contractual basis. These workers are not full-time workers of the plant, earn approximately 30% less than full-time workers, and do not fall under the purview of the IDA (Bertrand et al., 2025).

[8] Figure A3 in the Appendix lists the treated states and their year of treatment. India follows a federal system of governance, with a Central Government for the entire country and a separate government for each state. The Constitution of India confers the authority to amend this regulation on both the Central and State Governments.

enumeration in ASI is the plant-level, and the sampling frame includes all registered manufacturing plants in the country.[9] By law, all plants employing more than ten workers (if using power) or more than twenty workers (if not using power) must be registered.

Plants in the sampling frame belong either to the census sector or to the sample sector. The census sector includes all plants employing more than 100 total workers, as well as plants located in the six least industrially developed states, regardless of employment. The remaining plants belong to the sample sector. Since the grouping of plants into census and sample sectors depends on employment, which varies over time, this classification is updated periodically. The plants to be surveyed each year are drawn from the sampling frame as follows. Plants belonging to the census sector are surveyed yearly until the next revision occurs. Plants in the sample sector within each state are arranged into groups based on their 4-digit industry classification, and one-fifth of plants are selected from each (state, 4-digit industry) combination using stratified circular systematic sampling.

This paper uses the plant-level panel version of ASI from 1999 to 2018. ASI has information about the state in which each plant operates, and the analysis includes manufacturing units from all major Indian states. ASI reports detailed plant-level information on industry classification, employment and wages by worker type (full-time workers, contract workers, and managerial staff), output and profitability measures (sales, expenditure on materials, profits) and balance sheet information (the book value of fixed assets, current assets, current liabilities).

The ASI is better suited for this analysis than two other commonly used datasets on Indian manufacturing – the Economic Census and the Centre for Monitoring Indian Economy (CMIE) Prowess Database. Although the Economic Census covers the universe of all manufacturing plants in India, it publishes their data only once every five years. Hence, it is not possible to track individual plants over time and learn about their dynamics. The CMIE Prowess Database primarily contains balance sheet information for large firms. However, only 10% of firms provide information on the number of workers they employ. Moreover, Indian labor laws apply at the plant level rather than the firm level, making firm-level datasets such as CMIE Prowess less suitable for this analysis.

[9] I use the terms establishment, factory, and plant interchangeably in this paper.

4. REDUCED-FORM EVIDENCE

4.1. *Policy Impact on Plant Size Transitions*

This section examines whether the policy change affected plant transitions around the thresholds of 100 and 300 full-time workers. Specifically, for plants employing at most 100 (or 300) full-time workers in a given period, I analyze whether they are more (or less) likely to exceed 100 (or 300) full-time workers in the subsequent period after the policy is implemented.

The intuition is as follows. Before the policy change, if the regulation imposed significant costs on plants, crossing the 100 full-time worker threshold would have led to a steep increase in their labor costs. The policy change removed this steep increase at 100 workers and reintroduced it at the 300-worker threshold. Hence, if the regulation was binding for plants to begin with, one would expect the policy to increase their likelihood of crossing the 100-worker threshold and reduce their likelihood of crossing the 300-worker threshold. Since there were no such steep increases in costs at other employment thresholds, one should not observe significant effects at other levels.

I test this hypothesis formally as follows. Since the policy impact on plant size transitions is likely to vary with plant size, I divide the distribution of full-time workers, l^f, into five bins. The variable Bin_{ijst} takes values from 1 to 5, depending on the number of full-time workers employed by plant i in industry j, state s, and year t.

$$Bin_{ijst} = \begin{cases} 1 & \text{if } l^f_{ijst} \in [0, 50] \\ 2 & \text{if } l^f_{ijst} \in (50, 100] \\ 3 & \text{if } l^f_{ijst} \in (100, 200] \\ 4 & \text{if } l^f_{ijst} \in (200, 300] \\ 5 & \text{if } l^f_{ijst} > 300 \end{cases} \tag{1}$$

The dependent variable Y_{ijst} – defined in equation (2) – captures transitions of plants across these size bins. For example, if plant i employed between 51 and 100 full-time workers in period $t-1$ ($Bin_{ijst-1} = 2$), Y_{ijst} equals one if the plant employs more than 100 workers in period t, and zero otherwise. Similarly, if the plant employed between 201 and 300 workers in period $t-1$ ($Bin_{ijst-1} = 4$), Y_{ijst} equals one if it exceeds 300 workers in period t. I do

not distinguish between plants that remain in the same bin and those that move down, as my focus is exclusively on upward transitions. Note the dependent variable is defined only for plants that are observed in both t and $t-1$.

$$Y_{ijst} = \begin{cases} 1 & \text{if } Bin_{ijst} > Bin_{ijst-1} \\ 0 & \text{if } Bin_{ijst} \leq Bin_{ijst-1} \end{cases} \tag{2}$$

Using these definitions of Bin_{ijst} and Y_{ijst}, I estimate the following specification:

$$Y_{ijst} = \delta \, \mathbb{1}[Post_{st}] + \sum_{k=2}^{5} \beta_k^{post} \, \mathbb{1}[Post_{st}] \, \mathbb{1}[Bin_{ijst-1} = k] + \alpha_{jt} + \theta_{sk} + \lambda_{kt} + u_{ijst} \tag{3}$$

Here, k denotes the lagged size bin, with $k \in \{1, 2, 3, 4, 5\}$. Bin 1 is the base category, so all effects are measured relative to this group. $\mathbb{1}[Post_{st}]$ is a dummy variable that takes the value one if state s implemented the policy in year t or earlier, and zero otherwise. To minimize omitted variable bias, I incorporate a rich set of fixed effects in the regression specification. Industry-year fixed effects α_{jt} control for time-varying industry-specific shocks. State-by-lagged size bin fixed effects θ_{sk} account for time-invariant differences across states in the propensity to transition from each bin. Lagged size bin-by-year fixed effects λ_{kt} capture aggregate shocks in transition probabilities that vary by plant size. Standard errors have been clustered at the state level.

With these controls, δ captures the policy impact on upward transitions from the base category. β_k^{post} captures the differential impact of the policy on upward transitions from bin k relative to the base category. $\delta + \beta_k^{post}$ captures the total policy impact on upward transitions for plants in bin k. These coefficients of interest are identified by comparing transitions within the same state and lagged size bin pre- and post policy, after accounting for time-varying industry-specific and size-specific shocks.

Consistent with the economic intuition described above, if the regulation imposed significant costs on plants, the policy should have differential effects on plant transitions from bins 2 and 4. Specifically, transitions from Bin 2 (51-100 workers) should increase relative to the base category, as the regulatory burden at the 100-worker threshold is removed. Conversely, transitions from Bin 4 (201-300 workers) should decrease relative to the base

category, due to the new threshold at 300 workers. Bins 1 and 3 serve as placebo tests – the policy change should have no significant impact on upward transitions from these bins, since no discontinuities in labor costs exist around the 50 and 200-worker thresholds.

Figure 1 reports estimates of $\delta + \beta_k^{post}$ for each bin k from equation (3). Following the policy, there is a 3.89 percentage point (pp) increase in the probability with which plants in the treated states transition from 51-100 full-time workers to above 100 full-time workers. In the pre-policy period, plants in this size range had a 16.86% probability of transitioning above 100 workers in the subsequent period, so the estimated effect represents a 23.07% increase. Conversely, there is a 3.43 pp (21.68%) decrease in the probability with which plants transition from 201-300 full-time workers to above 300 post-policy. There is no statistically significant impact of the policy on transitions around any other threshold, suggesting these results are not due to reporting errors or other shocks in treated states.[10]

4.2. *Robustness Checks*

In this section, I address several concerns regarding the validity of the results in Figure 1 and rule out alternative explanations for the observed patterns in transition probabilities.

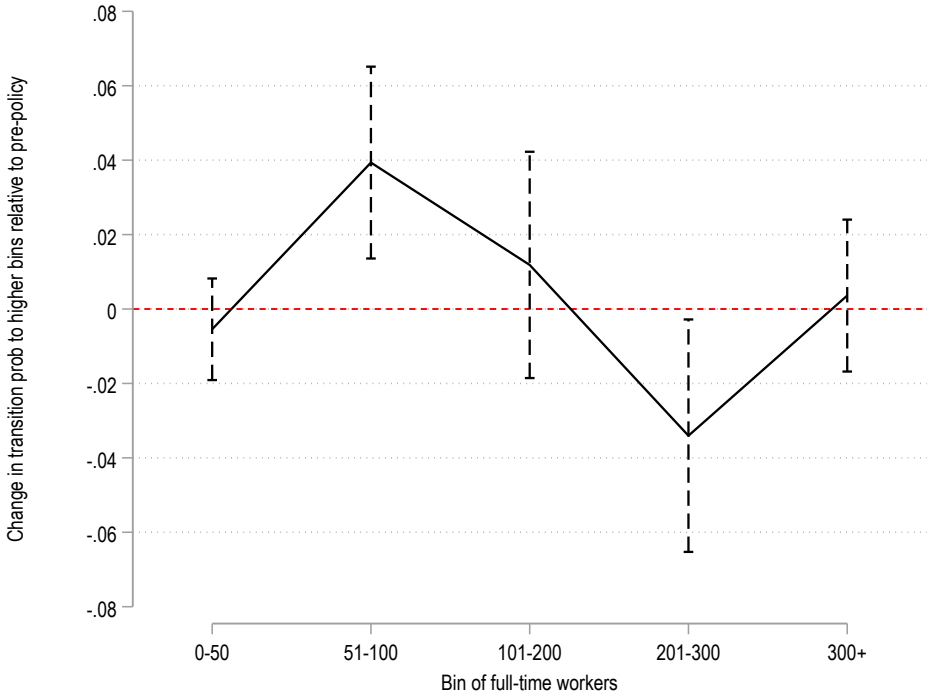
4.2.1. *Robustness to Choice of Bins*

One concern is that the plant size bins used in Figure 1 may be too broad to precisely capture the policy's impact. When bins span a large range of employment levels, the estimated coefficients could pick up the effects of other contemporaneous shocks that are unrelated to the policy change.

To address this, I re-estimate specification (3) using finer bins that provide more granularity around 100 full-time workers. Specifically, I divide the distribution of full-time workers into the following intervals: 0–20, 21–40, 41–60, 61–80, 81–100, 101–200, 201–300, and 300+. I then define bins based on these finer intervals. The dependent variable cap-

[10] One concern is whether plants were strategically under-reporting their full-time workforce before the reform to avoid the costs associated with the 100-worker threshold, and that the policy change simply induced more truthful reporting without any real changes in employment levels. As I show in Section 4.3, plants responded to the policy change by adjusting multiple margins beyond full-time employment, including reduced utilization of contract workers. This pattern of substitution toward full-time workers is what we would expect if plants faced real relaxation in regulatory costs. If plants had always been employing workers above the threshold but merely misreporting their numbers, we would not observe these complementary adjustments on other margins.

FIGURE 1.—Policy Impact on Transitions of Plants



Note: Each observation is at the plant-year level. Plants are assigned to size bins based on their full-time employment: 0–50, 51–100, 101–200, 201–300, and 300+. The figure plots the estimated policy impact on the probability that a plant in a given bin (horizontal axis) transitions to a higher bin in the following year (specification (3)). Bins 51–100 and 201–300 are policy-relevant, while the remaining bins serve as placebos. The sample includes all major Indian states and covers 1999–2018. Results are also reported in column 1 of Table I.

tures transitions across these finer bins, and is defined as in (2) – it equals one if a plant transitions to a higher bin and zero otherwise.

The empirical specification is similar to the one in equation (3), but with eight bins, resulting in more fixed effects and placebo thresholds:

$$Y_{ijst} = \delta \, \mathbb{1}[Post_{st}] + \sum_{k=2}^{8} \beta_k^{post} \, \mathbb{1}[Post_{st}] \, \mathbb{1}[Bin_{ijst-1} = k] + \alpha_{jt} + \theta_{sk} + \lambda_{kt} + u_{ijst} \quad (4)$$

Table A.3 in the Appendix reports the policy impact on upward transitions from each bin. Following the policy change, there is a 3.96 percentage point (13.58%) increase in the probability with which plants in the treated states transition from 81-100 full-time workers to 100+ full-time workers. Conversely, there is a 3.82 percentage point (24.14%) decrease

in the probability with which plants transition from 201-300 full-time workers to 300+ full-time workers post-policy. Specification (4) allows me to test the policy impact at multiple placebo thresholds – 20, 40, 60, 80, and 200 full-time workers – and there is no statistically significant impact on transitions around any of these thresholds.

The robustness to finer bins confirms that the baseline results are not driven by the choice of bins, and the policy effects are concentrated in the bins immediately adjacent to the 100 and 300 full-time worker thresholds. Therefore, throughout the rest of the paper, I retain the broader bins used in specification (1) to maintain adequate sample sizes in each bin and preserve statistical power, particularly near the thresholds where plant density may be lower.

4.2.2. *Event Studies*

A second concern from Figure 1 is whether the differential effects for plants near regulatory thresholds – relative to the base category – represent genuine policy responses or a mere continuation of pre-existing patterns. In other words, if the transition probability of bin k relative to bin 1 was already changing systematically before policy, the estimated β_k^{post} from specification (3) would be biased by that underlying trend.

To address this concern, I estimate an event study specification that replaces the single post-policy indicator in equation (3) with a full set of year-relative-to-policy indicators. Specifically, I estimate:

$$\begin{aligned} Y_{ijst} = &\sum_{\tau \neq -1} \delta_\tau \mathbb{1}[t - T_s = \tau] + \sum_{k=2}^{5} \sum_{\tau \neq -1} \beta_{k\tau} \mathbb{1}[t - T_s = \tau] \mathbb{1}[Bin_{ijst-1} = k] \\ &+ \alpha_{jt} + \theta_{sk} + \lambda_{kt} + u_{ijst} \end{aligned} \tag{5}$$

Here, Bin_{ijst} follows the definition in (1), Y_{ijst} is defined in equation (2), and T_s denotes the year in which state s implemented the policy. Event time is indexed by τ, where $\tau = 0$ denotes the year of policy implementation, and negative (positive) values correspond to years before (after) implementation. The omitted category is the year immediately preceding policy implementation ($\tau = -1$). The fixed effects are the same as in the baseline specification.

Under specification (5), the estimated coefficient $\beta_{k\tau}$ captures how the gap in transition probabilities between bin k and bin 1 changed at event time τ relative to $\tau = -1$. The identifying assumption is that, absent the policy, the gap would have remained constant over time. This implies $\beta_{k\tau}$ must be zero for all pre-policy periods.

Figures A.4, A.5 and A.6 in the Appendix plot the estimated effects $\beta_{k\tau}$ for bins 2, 3, and 4, respectively.[11] Three key findings emerge. First, for Bin 2 (51 to 100 workers), the estimated coefficients are close to zero and statistically insignificant in all pre-policy years, indicating stable relative transition probabilities prior to the policy. From the policy year, the coefficients become positive and statistically significant, implying that Bin 2 plants became more likely to transition upward relative to Bin 1 following implementation.

Second, the coefficients for Bin 4 (201–300 workers) are likewise close to zero and insignificant prior to the policy. They turn negative and statistically significant since the policy year, indicating a decline in the likelihood of upward transitions for Bin 4 plants relative to Bin 1. Third, the coefficients for Bin 3 – the placebo bin – remain close to zero and statistically insignificant throughout the entire event window, both before and after the policy change.

The absence of pre-trends in the policy-relevant bins, combined with the precise timing of the effects at the reform year, strongly supports the causal interpretation of the baseline results.

4.2.3. *Stacked Difference-in-Differences*

Recent advances in the difference-in-differences literature have shown that in settings with staggered treatment timing, the standard two-way fixed effects (TWFE) estimator can produce biased estimates when treatment effects are heterogeneous across units or over time (Goodman-Bacon, 2021, De Chaisemartin and d'Haultfoeuille, 2020, Sun and Abraham, 2021). This bias arises because the TWFE estimator implicitly uses already-treated units as controls for newly treated units – the so-called "forbidden comparisons".

To verify if the results in Figure 1 are driven by such forbidden comparisons, I implement a stacked difference-in-differences design following Cengiz et al. (2019). In my setting, states adopted the policy in four distinct years – 2014, 2015, 2016, and 2017 – so I divide

[11] Bin 5 is excluded as plants in this bin cannot transition to a higher employment category by construction.

the treated states into four cohorts based on their adoption year. For each cohort, I create a dataset that includes all plants from states in that cohort together with all plants from the never-treated states. I then stack the four cohort-specific datasets into a single dataset, and estimate the following specification on the stacked data:

$$Y_{ijstc} = \delta \, \mathbb{1}[Post_{tc}] + \sum_{k=2}^{5} \beta_k^{post} \, \mathbb{1}[Post_{tc}] \, \mathbb{1}[Bin_{ijst-1} = k] + \alpha_{jt} + \theta_{sk} + \lambda_{kt} + \gamma_{ct} + u_{ijstc} \tag{6}$$

where c denotes the cohort dataset. The treatment indicator $\mathbb{1}[Post_{tc}]$ equals one if year t is at or after the adoption year for cohort c. Besides the fixed-effects from the baseline specification, specification (6) also includes cohort-by-year fixed effects (γ_{ct}). The inclusion of cohort-by-year fixed effects ensures that coefficients are identified by comparing the treated units with only the never-treated units in each cohort, thereby blocking "forbidden comparisons".

Column 1 of Table A.4 in the Appendix reports the results from estimating specification (6). The estimated policy effects remain quantitatively similar to the baseline findings: a statistically significant (3.5 pp) increase in the probability of transitioning from 51-100 to 100+ full-time workers, a statistically significant (4.1 pp) decrease in the probability of transitioning from 201-300 to 300+ full-time workers, and no significant effects for placebo bins.

4.3. *Policy Impact on Substitution Patterns*

The above results show that the policy change impacted plants' size transitions around the old and new regulatory thresholds. Did it also impact plants' usage of other factors of production on which the regulation does not apply? Understanding these substitution patterns is crucial for assessing the full impact of the regulation, as plants may respond to regulations not only by adjusting their full-time workforce but also by reallocating resources towards other inputs.

Contract workers represent a particularly relevant margin of adjustment in the Indian context. As described earlier, contract workers are hired by plants through staffing companies on a temporary contractual basis. Critically, they are not considered full-time workers of the plant and therefore do not fall under the purview of chapter VB of the Industrial

Disputes Act (Bertrand et al., 2025). This regulatory distinction creates an incentive for plants to substitute between full-time and contract workers. Similarly, capital represents another potential margin of adjustment if plants can substitute machinery for labor to avoid crossing regulatory thresholds.

In this section, I examine whether the policy change affected plants' substitution patterns. The policy impact on substitution patterns would vary based on plant size. Before the policy change, plants near the 100-worker threshold faced regulatory costs when expanding their full-time workforce. To the extent that full-time workers, contract workers, and capital are substitutable, these plants would have relied on contract labor and capital to avoid crossing the threshold. Following the policy change, we expect plants near the 100-worker threshold to reduce their reliance on these other factors, as the regulatory cost of expanding full-time employment is removed.

To test this hypothesis, I estimate the same difference-in-differences regression as in specification (3), replacing the transition probability outcome variable with measures of factor substitution. Specifically, I estimate:

$$Y_{ijst} = \delta \, \mathbb{1}[Post_{st}] + \sum_{k=2}^{5} \beta_k^{post} \, \mathbb{1}[Post_{st}] \, \mathbb{1}[Bin_{ijst-1} = k] + \alpha_{jt} + \theta_{sk} + \lambda_{kt} + u_{ijst} \quad (7)$$

where Y_{ijst} is either the contract labor share or logarithm of investment rate in plant and machinery of plant i in industry j, state s, and year t. The contract labor share is defined as the ratio of contract workers to total non-managerial workers. The investment rate in plant and machinery is defined as actual additions to plant and machinery divided by the net closing value of plant and machinery. Bin_{ijst}, $Post_{st}$, and fixed effects are defined as in specification (3).

Columns 2 and 3 of table I show the results from estimating specification (7). Following the policy change, contract labor share and investment rate in plant & machinery decreased significantly by 2 percentage points (4.29%) and 1.3% for plants near the 100-worker threshold (Bin 2). These patterns are consistent with plants substituting back toward full-time workers when regulatory costs were removed at 100 full-time workers. There is no statistically significant impact of the policy on these outcomes around the 50, 200 and 300 worker thresholds.

TABLE I

POLICY IMPACT ON TRANSITIONS AND SUBSTITUTION PATTERNS

	(1) Transition probabilities	(2) Share of contract labor	(3) Log(Investment rate)
$\mathbb{1}[Post_{st}]$	-0.005	0.005	0.004
	(0.007)	(0.012)	(0.006)
$\mathbb{1}[Post_{st}] \times l^f_{ijst-1} \in (50, 100]$	0.044***	-0.025*	-0.013*
	(0.009)	(0.013)	(0.007)
$\mathbb{1}[Post_{st}] \times l^f_{ijst-1} \in (100, 200]$	0.017	0.004	-0.003
	(0.013)	(0.02)	(0.007)
$\mathbb{1}[Post_{st}] \times l^f_{ijst-1} \in (200, 300]$	-0.029**	0.002	-0.004
	(0.014)	(0.025)	(0.012)
$\mathbb{1}[Post_{st}] \times l^f_{ijst-1} > 300$	0.007	0.009	-0.003
	(0.005)	(0.022)	(0.008)
Observations	198703	105988	210387
Industry × Year FE	Yes	Yes	Yes
State × Lagged size-bin FE	Yes	Yes	Yes
Lagged size-bin × Year FE	Yes	Yes	Yes

Notes: Each observation is at the plant-year level. Plants are assigned to size bins based on their full-time employment: 0–50, 51–100, 101–200, 201–300, and 300+. In column 1, the dependent variable equals one if a plant transitions to a higher bin in the following year (specification (3)). In column 2, the dependent variable is the contract labor share, defined as the ratio of contract workers to total non-managerial workers. In column 3, the dependent variable is the log investment rate in plant and machinery, defined as actual additions divided by the net closing value. Column 1 corresponds to the estimates reported in Figure 1. Columns 2 and 3 report estimates from specification (7). Post is a state-year indicator equal to one if the state had raised the regulatory threshold to 300 full-time workers by that year. Bin 0–50 is the omitted category. All specifications include industry-by-year, state-by-lagged size bin, and lagged size bin-by-year fixed effects. The sample includes all major Indian states and covers 1999–2018. Robust standard errors clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To summarize, this section has provided comprehensive evidence on how plants responded to the policy change along multiple dimensions. Plants with less than or equal to 100 full-time workers in any given year responded in two ways to the reform. First, their likelihood of employing above 100 workers in the subsequent year increased significantly, as documented in the previous subsection. Second, their share of contract workers and investment rate in plant and machinery in the subsequent year decreased significantly after the policy. These results establish that IDA significantly impacts not only the full-time

employment growth of plants but also their broader input allocation decisions. However, several important questions remain unanswered. First, it is unclear why plants do not exhibit bunching behavior around the regulatory threshold. Second, a comprehensive policy evaluation requires quantifying the magnitude of regulatory costs and understanding their implications. To address these questions, the next section develops and estimates a quantitative model that explicitly captures how size-based regulations affect plant behavior.

5. MODEL

I develop a dynamic model of heterogeneous plants that employ full-time workers, contract workers, and materials, and face costs of adjusting full-time employment. Plants enter by paying a sunk cost, draw productivity from a distribution, experience idiosyncratic productivity shocks over time, and exit exogenously. Given their current productivity and full-time employment, plants choose contract labor and materials at each instant to maximize short-run profits, while dynamically adjusting full-time employment through costly hiring and separation decisions. Because these adjustments are time-consuming and costly, plants gradually expand or contract their full-time workforce rather than adjusting instantaneously.

In this framework, I incorporate size-based labor regulations that increase both the marginal cost of production and the cost of firing once the full-time employment of a plant exceeds a legal threshold. The model captures how these regulations affect plant dynamics – entry, growth, and substitution between full-time and contract workers.

I proceed as follows. I first describe the model setup and then define the equilibrium conditions of the model. I then use the model to highlight the channels through which the regulation impacts plant outcomes. I present results from estimating this model on Indian data in Section 6.

5.1. *Setup*

I assume that time is continuous and plants in the economy discount the future at rate $r \geq 0$. The economy consists of an endogenous mass of active plants. Plants are engaged in monopolistic competition, and each plant produces a unique variety. Hence, the total mass of active plants at any instant determines the mass of available varieties in the economy at that instant.

Potential entrants in this economy can choose to become active by paying an entry cost worth c_e. Once they become active, they start their operations by drawing a productivity ϵ from distribution $f(\epsilon)$. Newly active plants begin operations with zero full-time workers. I now describe the production technology and hiring process of active plants.

5.1.1. *Production Technology*

Let q denote the output of a plant. Output q is produced using full-time workers l_f, contract workers l_c, and intermediate inputs m. I allow for substitutability between full-time workers and contract workers by aggregating them using a CES function. Output from this CES function is then combined with intermediate inputs (m) through a Cobb-Douglas function to produce output q.

More specifically, full-time and contract workers are combined into a composite labor input L using the CES function in (8) with elasticity of substitution $\nu > 1$.

$$L = (l_f^{\frac{\nu-1}{\nu}} + l_c^{\frac{\nu-1}{\nu}})^{\frac{\nu}{\nu-1}} \tag{8}$$

The resulting labor composite L is combined with intermediate inputs m using the Cobb-Douglas production function in (9) to produce output q. In equation (9), $\alpha_M \in (0,1)$ and $\alpha_L \in (0,1)$ denote the share of materials and labor in output respectively, and ϵ represents plant productivity. The production function in equation (9) is analogous to the ones used in Krusell et al. (2000) and Parro (2013).

$$q = \epsilon L^{\alpha_L} m^{\alpha_M} \quad \text{with} \quad \alpha_L + \alpha_M < 1 \tag{9}$$

For a plant that produces output q as per equation (9), its revenue in equilibrium is given by equation (10). In (10), $\mu > 1$ is the constant elasticity of substitution across the different varieties available in the market, I stands for aggregate expenditure, and P is the Dixit-Stiglitz price index.

$$R(q) = q^{\frac{\mu-1}{\mu}} \left(\frac{I}{P}\right)^{\frac{1}{\mu}} P \tag{10}$$

5.1.2. *Productivity Dynamics*

Plant-level productivity, ϵ, follows an exogenous, first-order Markov process in continuous time. More specifically, following Shimer (2005), the logarithm of plant-level productivity $z = \log(\epsilon)$ lies on the following grid with $2n_\epsilon + 1$ points spaced $\Delta_\epsilon > 0$ apart and centered around $\bar{z}$:

$$z \in \{\bar{z} - n_\epsilon \Delta_\epsilon, \bar{z} - (n_\epsilon - 1)\Delta_\epsilon, \ \ldots, \ \bar{z}, \ \ldots, \bar{z} + (n_\epsilon - 1)\Delta_\epsilon, \bar{z} + \ n_\epsilon \Delta_\epsilon\}.$$

Productivity shocks arrive at the Poisson rate $\lambda_\epsilon > 0$. When a shock occurs, the new log-productivity z' moves up or down by one grid point. Let $p(z)$ denote the probability of an upward jump conditional on the current state z. Then:

$$\begin{aligned} p(z) &= \mathbb{P}(z' = z + \Delta_\epsilon \mid z) = \frac{1}{2}\left(1 - \frac{z - \bar{z}}{n_\epsilon \Delta_\epsilon}\right) \\ 1 - p(z) &= \mathbb{P}(z' = z - \Delta_\epsilon \mid z) = \frac{1}{2}\left(1 + \frac{z - \bar{z}}{n_\epsilon \Delta_\epsilon}\right) \end{aligned} \tag{11}$$

A transition in log productivity from z to $z + \Delta_\epsilon$ induces a corresponding transition in productivity levels from ϵ to $\epsilon \ \exp(\Delta_\epsilon)$. The probability $p(z)$ decreases as z increases above $\bar{z}$, ensuring mean reversion towards $\bar{z}$. In other words, for plants with larger values of current productivity z, their probability of receiving an even higher productivity shock is lower. Hence, the parameters of the productivity process can be summarized by $(n_\epsilon, \Delta_\epsilon, \lambda_\epsilon, \bar{z})$.

5.1.3. *Hiring Technology*

Plants face different costs while hiring full-time workers, contract workers, and materials, as I explain below. Suppose the regulatory environment is such that plants fall under the purview of labor regulations if their full-time workforce l_f exceeds a threshold $\ell^* > 0$. Given this regulatory environment, the total wage bill of a plant with l_f full-time workers is given by:

$$\begin{cases} w_f l_f & \text{if } l_f \leq \ell^* \\ w_f (1 + \tau) l_f & \text{if } l_f > \ell^* \end{cases} \tag{12}$$

Here, w_f is the wage rate per full-time worker that the plant takes as given, and $\tau \geq 0$ captures the wedge above the wage rate imposed by the regulation. As argued in Section 1, $\tau \geq 0$ can be interpreted as the costs that plants face because of frequent inspections to check for compliance with the IDA.[12]

In addition to the wages that they have to pay full-time workers, the process of hiring them is also costly and time-consuming. In order to hire workers, a plant with productivity ϵ and l_f full-time workers has to choose an effort level or search intensity defined by $\sigma(\epsilon, l_f) \geq 0$. New full-time workers then arrive at a Poisson rate equal to $\sigma(\epsilon, l_f) \geq 0$: the higher the search intensity, the shorter the expected time until the next worker arrives.

However, plants cannot always choose high levels of search intensity because it is costly. Let $C^h(\sigma, l_f)$ denote the cost function associated with search intensity σ for a plant with l_f full-time workers. I assume the following functional form for $C^h(\sigma, l_f)$ following Coşar et al. (2016) and Eaton et al. (2022):

$$C^h(\sigma, l_f) = \frac{1}{2} \frac{\sigma^2}{(l_f + 1)^{\phi_h}} \tag{13}$$

Note the cost function $C^h(\sigma, l_f)$ is twice continuously differentiable, strictly increasing, and strictly convex in σ, with $\lim_{\sigma \to \infty} C^h_\sigma(\sigma, l_f) = \infty$. $\phi_h \in [0, 2]$ captures the extent of economies of scale in hiring. If $\phi_h = 0$, there are no economies of scale in hiring, and the cost of a given search effort is the same for all plants. On the other hand, if $\phi_h = 2$, then the cost of a given search effort per worker is the same for all plants.

The total wage bill of a plant with l_c contract workers is $w_c l_c$, where w_c is the wage rate per contract worker. Analogously, the total expenditure incurred by a plant on materials is given by $r_m m$. Plants take w_c and r_m as given. Unlike full-time workers, plants do not face any other costs while hiring contract workers and materials.[13]

[12] The wage rate w_f is taken as exogenous in the model. Formal manufacturing accounts for only 15-20% of India's GDP, so wages are largely determined outside the sector.

[13] The assumption of no adjustment costs for contract workers is reasonable given the Indian institutional context. As described earlier, plants hire these workers through staffing companies, so plants do not incur any hiring costs per se while employing them. Similarly, contract workers are not protected by labor laws, so there are no costs associated with firing them.

5.1.4. *Firing Costs*

Plants face firing costs only for full-time workers. Each full-time worker gets separated from the plant exogenously at rate $\delta_0 > 0$. Plants can expedite separations by choosing a higher separation rate $\delta \geq \delta_0$. Assuming these separation rates are independent across full-time workers, separation shocks arrive for a plant with l_f full-time workers at Poisson rate δl_f. Higher δ reduces the expected time until worker separation.

However, plants cannot always choose high δ because there are costs associated with it. Let $C^f(\delta, l_f)$ denote the cost of a plant with l_f full-time workers that chooses separation rate δ. I assume the following functional form for $C^f(\delta, l_f)$.[14]

$$C^f(\delta, l_f) = \begin{cases} \frac{1}{2}\frac{(\delta l_f - \delta_0 l_f)^2}{(l_f)^{\phi_f}} & \text{if } l_f \leq \ell^* \\ \frac{1}{2}\frac{(\delta l_f - \delta_0 l_f)^2}{(l_f)^{\phi_f}} + \gamma(\delta l_f - \delta_0 l_f) & \text{if } l_f > \ell^* \end{cases} \tag{14}$$

Here, $\ell^* > 0$ is the threshold beyond which regulations apply to plants, $\phi_f \in [1, 2]$ captures economies of scale in firing, and $\gamma \geq 0$ captures the wedge on firing costs imposed by the regulation. $\gamma \geq 0$ can be interpreted as the costs that plants face because of the uncertainty that they face when they are hit by an adverse shock. In contrast to full-time workers, I assume plants can adjust contract workers and materials without firing costs, making these variable inputs that adjust flexibly to current productivity ϵ and the stock of full-time workers l_f.

5.1.5. *Plant Exit*

Every plant exits at an exogenous Poisson rate $s > 0$. I assume that an exiting plant receives a scrap value of 0.[15]

[14] The baseline component of the separation cost mirrors the search cost specification in equation (13): quadratic in the excess flow rate and scaled by plant size.

[15] I do not model plant exit as an endogenous decision. As noted in Chatterjee et al. (2025), exiting is difficult in India, taking 4.3 years on average, and many manufacturing plants in the economy remain dormant while waiting to exit. Therefore, for the purposes of this paper, I abstract from endogenous exit and focus on how the IDA affects the behavior of entrants and incumbents.

5.2. *Characterization of the Incumbent's Problem*

At any given instant, an incumbent plant's state is characterized by its productivity ϵ and stock of full-time workers l_f. I assume the Poisson processes governing worker arrivals, worker separations, plant exit, and plant-level productivity shocks are independent of each other. The plant's value function $V(l_f, \epsilon)$ satisfies the following Hamilton-Jacobi-Bellman equation in continuous time.

$$\begin{aligned}(r+\sigma+\delta l_f+s+\lambda_\epsilon)V(l_f,\epsilon) = \max_{\sigma\geq 0,\delta\geq\delta_0}\Big\{&\Pi(l_f,\epsilon)-w_f(1+\tau\mathbb{I}\{l_f>\ell^*\})l_f-C^h(\sigma,l_f)\\ &-C^f(\delta,l_f)+\delta l_f V(l_f-1,\epsilon)+\sigma V(l_f+1,\epsilon)\\ &+\lambda_\epsilon p(\epsilon)V(l_f,\epsilon^+)+\lambda_\epsilon(1-p(\epsilon))V(l_f,\epsilon^-)\Big\},\end{aligned} \tag{15}$$

where $\epsilon^+ = \epsilon\exp(\Delta_\epsilon)$ and $\epsilon^- = \epsilon\exp(-\Delta_\epsilon)$ are the new productivity levels after positive and negative productivity shocks, following the transition probability matrix in (11).[16] The flow profit $\Pi(l_f, \epsilon)$ is determined by optimizing over variable inputs:

$$\Pi(l_f,\epsilon) = \max_{l_c,m\geq 0}\left\{R(q(\epsilon,l_f,l_c,m))-w_c l_c - r_m m\right\} \tag{16}$$

where $R(q(\epsilon, l_f, l_c, m))$ denotes plant-level revenues from (10). Equation (15) can be interpreted as follows. Given productivity level ϵ and l_f full-time workers, a plant earns a flow payoff of $\Pi(l_f,\epsilon) - w_f(1+\tau\mathbb{I}\{l_f>\ell^*\})l_f - C^h(\sigma,l_f) - C^f(\delta,l_f)$ until the next event occurs. $\Pi(l_f, \epsilon)$ denotes the flow profits earned by the plant and is pinned down by the optimal values of l_c and m that maximize the right hand side of equation (16). Since there are no adjustment costs for l_c and m, their optimal values are such that their marginal revenue product equals their respective factor price.

The time until the next event depends on the Poisson arrival rates s, δl_f, σ and λ_ϵ. The higher these arrival rates, the faster the next event will occur, on average. There are four possible events that could occur next. Plant destruction occurs with probability $\frac{s}{s+\delta l_f+\sigma+\lambda_\epsilon}$, yielding zero scrap value. Worker separation occurs with probability $\frac{\delta l_f}{s+\delta l_f+\sigma+\lambda_\epsilon}$, reducing

[16] Following (11), transition probabilities are defined over log productivity, $z = \log(\epsilon)$. For notational convenience, I denote it by $p(\epsilon)$ rather than $p(\log(\epsilon))$.

employment to $l_f - 1$ with continuation value $V(l_f - 1, \epsilon)$. Worker arrival occurs with probability $\frac{\sigma}{s+\delta l_f+\sigma+\lambda_\epsilon}$, increasing employment to $l_f + 1$ with continuation value $V(l_f + 1, \epsilon)$. Finally, a productivity shock occurs with probability $\frac{\lambda_\epsilon}{s+\delta l_f+\sigma+\lambda_\epsilon}$, changing productivity to ϵ^+ with probability $p(\epsilon)$ or to ϵ^- with probability $1 - p(\epsilon)$.

While plants take s and λ_ϵ as given, they can expedite the arrival and separation of workers by choosing σ and δ, respectively. The optimal search intensity of a plant with productivity ϵ and l_f full-time workers is captured by the policy function $\sigma(l_f, \epsilon)$, and it satisfies the following first order condition:

$$\sigma(l_f, \epsilon) = (l_f + 1)^{\phi_h} \max\left\{0,\ V(l_f + 1, \epsilon) - V(l_f, \epsilon)\right\} \tag{17}$$

Equation (17) suggests that the optimal search effort plants put towards hiring depends on their marginal cost of effort and their expected marginal benefit from search. The latter includes the increase in present value of lifetime profits from having more workers, captured by $V(l_f + 1, \epsilon) - V(l_f, \epsilon)$.

Analogously, the optimal separation rate chosen by a plant with productivity ϵ and l_f full-time workers, $\delta(l_f, \epsilon)$, depends on the marginal firing cost and the expected marginal benefit from firing, and is given in Equation (18).

$$\delta(l_f, \epsilon) = \delta_0 + l_f^{\phi_f - 1} \max\left\{0,\ V(l_f - 1, \epsilon) - V(l_f, \epsilon) - \gamma\, \mathbb{I}\{l_f > \ell^*\}\right\} \tag{18}$$

5.3. *Characterization of the Entrant's Problem*

At each instant, a large mass of potential entrants considers entering the market. Entry requires paying a sunk cost of $c_e > 0$. Upon paying this cost, a plant draws its initial productivity ϵ from the ergodic distribution $f(\epsilon)$ of the Markov process for $\log(\epsilon)$ defined in equation (11). A potential entrant will enter only if their expected value of entry is at least as high as c_e. Therefore, in a stationary equilibrium with positive entry, the free entry condition should hold with equality at all times:[17]

$$\sum_{\epsilon \in \mathcal{E}} V(0, \epsilon) f(\epsilon) = c_e, \tag{19}$$

[17] Since plants exit at Poisson rate $s > 0$, a stationary equilibrium requires positive entry.

where $\mathcal{E} = \{e^{\bar{z}-n_\epsilon\Delta_\epsilon}, \ldots, e^{\bar{z}+n_\epsilon\Delta_\epsilon}\}$ denotes the discrete productivity grid, and $V(0,\epsilon)$ is defined in (15). The left-hand side of equation (19) represents the expected value of entry.

The free-entry condition pins down the equilibrium Dixit-Stiglitz price index. Suppose the expected value of entry exceeds the sunk entry cost. Then the mass of entrants increases, leading to a larger mass of active plants in the economy. This reduces the aggregate price index, making entry less valuable for potential entrants. The price index falls until the expected value of entry equals the sunk entry cost, restoring equilibrium.

5.4. *Steady State Plant Size Distribution*

Let $\bar{M}$ denote the total mass of plants in the economy. In steady state, how is this mass distributed across states (l_f, ϵ)? Let $M(l_f, \epsilon)$ be the mass of plants with l_f workers and productivity ϵ, where $\sum_{l_f,\epsilon} M(l_f,\epsilon) = \bar{M}$. The evolution of $M(l_f,\epsilon)$ over an infinitesimal time interval, $\dot{M}(l_f,\epsilon)$, depends on inflows to and outflows from state (l_f,ϵ). These transitions occur due to productivity shocks, plant exit, hiring, firing, and plant entry. In steady state, net flows into each state are zero, which yields a system of equations that can be solved for the steady-state plant size distribution, conditional on $\bar{M}$.

More formally, the law of motion describing $M(l_f,\epsilon)$, for $l_f \in \{1,2,...\}$ and $\epsilon \in \mathcal{E}$ is specified in Equation (20). Inflows into state (l_f,ϵ) comprise the following plants: (1) those from state (l_f-1,ϵ) that get a worker with Poisson rate $\sigma(l_f-1,\epsilon)$, (2) those from state (l_f+1,ϵ) that lose a worker due to separation at Poisson rate $\delta(l_f+1,\epsilon)(l_f+1)$, (3) those from state (l_f,ϵ^-) that receive a higher productivity shock at Poisson rate $\lambda_\epsilon p(\epsilon^-)$, and (4) those from state (l_f,ϵ^+) that receive an adverse productivity shock with Poisson rate $\lambda_\epsilon(1-p(\epsilon^+))$.[18] Similarly, outflows from state (l_f,ϵ) includes plants from state (l_f,ϵ) that either get a worker, lose a worker, receive a productivity shock or get destroyed with Poisson rates $\sigma(l_f,\epsilon), \delta(l_f,\epsilon)l_f, \lambda_\epsilon$ and s, respectively.

$$\begin{aligned}\dot{M}(l_f,\epsilon) =& M(l_f-1,\epsilon)\sigma(l_f-1,\epsilon) + M(l_f+1,\epsilon)\delta(l_f+1,\epsilon)(l_f+1) \\ &+ M(l_f,\epsilon^-)\lambda_\epsilon p(\epsilon^-) + M(l_f,\epsilon^+)\lambda_\epsilon(1-p(\epsilon^+)) \\ &- M(l_f,\epsilon)\left[\sigma(l_f,\epsilon) + \delta(l_f,\epsilon)l_f + \lambda_\epsilon + s\right]\end{aligned} \tag{20}$$

[18] As before, $\epsilon^+ = \epsilon\exp(\Delta_\epsilon)$ and $\epsilon^- = \epsilon\exp(-\Delta_\epsilon)$ denote productivity levels adjacent to productivity ϵ in the productivity grid.

Similarly, the mass of plants with 0 workers and productivity ϵ evolves over time according to Equation (21). In equation (21), $s\bar{M}f(\epsilon)$ denotes the mass of entrants that receive a productivity draw ϵ from the ergodic distribution.

$$\begin{aligned}\dot{M}(0,\epsilon) =& M(1,\epsilon)\delta(1,\epsilon) + M(0,\epsilon^-)\lambda_\epsilon p(\epsilon^-) + M(0,\epsilon^+)\lambda_\epsilon(1-p(\epsilon^+)) + s\bar{M}f(\epsilon) \\ &- M(0,\epsilon)\left[\sigma(0,\epsilon) + \lambda_\epsilon + s\right]\end{aligned} \tag{21}$$

In steady state, $\dot{M}(l_f,\epsilon) = 0$ for all l_f and ϵ. This yields a system of equations that can be solved for the steady-state distribution. For non-zero employment levels ($l_f \geq 1$):

$$M(l_f,\epsilon) = \left\{ \frac{\begin{array}{c} M(l_f-1,\epsilon)\sigma(l_f-1,\epsilon) + M(l_f+1,\epsilon)\delta(l_f+1,\epsilon)(l_f+1) \\ + M(l_f,\epsilon^-)\lambda_\epsilon p(\epsilon^-) + M(l_f,\epsilon^+)\lambda_\epsilon(1-p(\epsilon^+)) \end{array}}{s + \delta(l_f,\epsilon)l_f + \sigma(l_f,\epsilon) + \lambda_\epsilon} \right\} \tag{22}$$

For plants with zero workers:

$$M(0,\epsilon) = \frac{M(1,\epsilon)\delta(1,\epsilon) + s\bar{M}f(\epsilon) + M(0,\epsilon^-)\lambda_\epsilon p(\epsilon^-) + M(0,\epsilon^+)\lambda_\epsilon(1-p(\epsilon^+))}{s + \sigma(0,\epsilon) + \lambda_\epsilon} \tag{23}$$

The system of equations characterized by equations (22) to (23) for all $\epsilon \in \mathcal{E}$ imply the following. Given any (l_f,ϵ) pair, if plants lower their search effort, then they will spend more time in (l_f,ϵ) on average. As a result, M(l_f,ϵ) at steady state will be higher.

Finally, in equilibrium, aggregate revenue must equal aggregate expenditure. The total revenue for all plants is $\sum_{l_f,\epsilon} M(l_f,\epsilon)R(l_f,\epsilon)$, where $R(l_f,\epsilon)$ denotes the revenue of a plant in state (l_f,ϵ) after optimally choosing variable inputs. Setting this equal to aggregate expenditure I yields:

$$\sum_{l_f,\epsilon} M(l_f,\epsilon)R(l_f,\epsilon) = I \tag{24}$$

This market-clearing condition, combined with the free-entry condition (19), determines both the total mass of plants $\bar{M}$ and the price index P in equilibrium.

5.5. *Discussion of Mechanisms*

Motivated by the Indian institutional context, I model the regulation as imposing two distinct costs on plants above the threshold: a wedge τ on the wage bill and a wedge γ on firing costs (see equations (12) and (14)). In this subsection, I use the model to show that the two wedges leave different imprints on the steady-state density of plants around the regulatory threshold. This difference allows me to estimate them separately in Section 6.

5.5.1. *Impact of wage wedge τ on density of plants around the threshold*

What happens to the steady-state distribution of plants around ℓ^* workers as we vary τ, holding all else fixed? Consider plants with fewer than ℓ^* workers. As τ increases, their expected marginal benefit from hiring additional workers decreases, so their optimal search effort falls as per equation (17). Hence, plants below ℓ^* spend more time on average in their current states, and fewer plants transition from below to above the regulatory threshold. With all else equal, this implies that the steady-state density of plants below the regulatory threshold increases, and the steady-state density of plants above the regulatory threshold falls.

I illustrate this with a numerical example. Fixing $\ell^* = 100$ workers and setting all parameters at the values in Table A.5, I solve for the steady-state distribution using equations (22)–(23). I then steadily increase τ, holding the remaining parameters fixed, and recompute the steady-state distribution for each value. Figure 2a reports the resulting change in the density of plants in a band below the threshold (91–99 workers) and a band above it (101–110 workers), relative to the benchmark. As τ rises, the density below the threshold increases monotonically while the density above it falls, consistent with the intuition above.

5.5.2. *Impact of firing cost wedge γ on density of plants around the threshold*

What happens to the steady-state distribution of plants around ℓ^* workers as we vary γ, holding all else fixed? Consider first plants with more than ℓ^* workers. As γ increases, a wider range of plants stop firing actively altogether, as their expected marginal benefit from firing a worker no longer exceeds the wedge they must pay to do so.[19] At the same time,

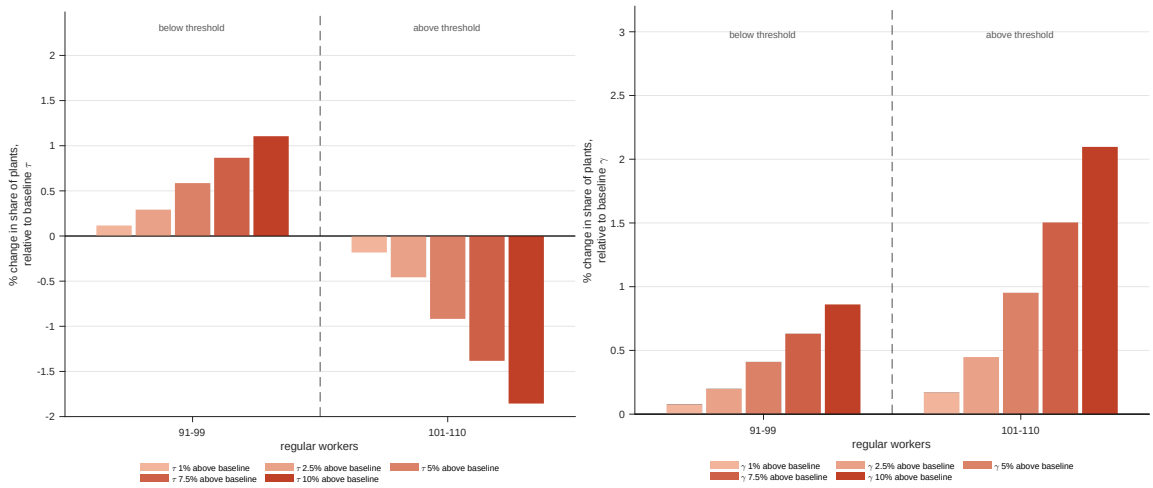
[19] A wider range of plants choose $\delta = \delta_0$.

plants with more than ℓ^* workers reduce their optimal search effort, since the option to fire an additional worker in future is now more costly to exercise. Together, these mean that plants above ℓ^* spend more time on average in their current states. With everything else remaining the same, the steady-state density of plants just above the regulatory threshold increases. The larger the firing cost wedge, the greater this increase.

Now consider plants with fewer than ℓ^* workers. Here two forces work in opposite directions. As γ increases, fewer plants transition from below to above ℓ^* workers, as the option to fire a worker in the future is more costly to exercise above the threshold. At the same time, fewer plants transition from above the threshold to below it, since plants above ℓ^* are now retained there. The steady-state density of plants just below the regulatory threshold therefore reflects both fewer outflows to and fewer inflows from higher employment levels, and the net effect depends on which force dominates.

I illustrate this with a numerical example. Fixing $\ell^* = 100$ workers and setting all parameters at the values in Table A.5, I solve for the steady-state distribution using equations (22)–(23). I then steadily increase γ, holding the remaining parameters fixed, and recompute the steady-state distribution for each value. Figure 2b reports the resulting change in the density of plants in the same two bands, relative to the benchmark. As γ rises, the density of plants above the threshold increases monotonically, and by more than the density below it.

To summarize, holding all else equal, a wage wedge lowers transition probabilities from below to above the regulatory threshold. This produces the canonical signature of a size-based regulation: excess mass below the threshold and missing mass above it. On the other hand, a firing cost wedge lowers both upward and downward transitions around the regulatory threshold, thereby attenuating or even reversing the bunching signature. So if firing costs dominate the regulatory burden, or if the regulation combines a high wage wedge with a high firing cost, a bunching-based estimator would understate its cost. Transitions around the threshold together with plant density on either side of it, provided that other frictions are disciplined separately, therefore allow me to disentangle the nature of the regulatory costs.



(a) Firm Size Distribution and Wage Wedge τ (b) Firm Size Distribution and Firing Wedge γ

FIGURE 2.—Impact of Regulatory Costs on Steady-State Plant Size Distribution

Notes: The height of each bar is the percentage change in the share of plants in each employment band, relative to the baseline. The baseline is one in which all parameters, including both wedges, take the values reported in Table A.5. Panel (a) raises the wage wedge τ above its baseline value while holding the firing wedge γ and all other parameters fixed. Panel (b) raises γ above its baseline value while holding τ and all other parameters fixed. In each case, I recompute the policy functions and the steady-state plant size distribution from equations (22)–(23). The regulatory threshold is held at 100 full-time workers throughout.

6. ESTIMATION PROCEDURE

In order to quantify the magnitude of regulatory costs, I fit the above framework to plant-level data from the treated states. Recall that the regulatory threshold was 100 full-time workers prior to the reform. The policy reform that was implemented in these states increased the regulatory threshold from 100 to 300 full-time workers. I assume that the plant size distribution of treated states during the year before the reform was in steady state. The causal impacts shown in Section 4 are informative about how transition probabilities and the contract labor share of plants in the treated states around 100 full-time workers changed because of the policy reform. I exploit these causal impacts to identify key parameters in my model. The reform-induced variation in the contract labor share is informative about the elasticity of substitution between full-time and contract workers. The reform-induced variation in transition probabilities around the threshold, together with the density of plants on either side of it, pins down the magnitude of the regulatory costs. The detailed estimation procedure is described below.

6.1. *Parameters Not Estimated*

Several parameters are not identified by the model, so I either take these from external sources or calibrate them. Following Chatterjee et al. (2025), I set the annual Poisson exit rate s to 0.03. Following De Loecker et al. (2016), I set the elasticity of substitution between different varieties μ to 3.94. The rate of time preference r is set to 0.05, corresponding to an annual interest rate of 5%. I also normalize the price of materials r_m to 1.

The material and labor shares in output, α_M and α_L, are calibrated using all-India data. For each industry, I first calculate the median share of deflated material expenditure in deflated revenue, and the median share of deflated total wage bill in deflated revenue. These medians directly map to $\alpha_M \left(\frac{\mu-1}{\mu}\right)$ and $\alpha_L \left(\frac{\mu-1}{\mu}\right)$, respectively. Dividing each median share by $\frac{\mu-1}{\mu}$ yields the industry-specific values of α_M and α_L. Because the model has a single sector, I then average these values across industries to obtain the values used in estimation.

The full-time and contract wage rates, w_f and w_c, are calibrated from the treated states in the year before the reform. Specifically, w_f is computed as the median of the deflated wage bill for full-time workers divided by the number of full-time workers, while w_c is computed as the median of the deflated wage bill for contract workers divided by the number of contract workers. The aggregate expenditure in treated states I is set to be equal to the sum of deflated revenues for all plants in the treated states during the year before the reform. Finally, the Dixit-Stiglitz price index P for the year before reform is normalized to 1.

6.2. *Parameters to be Estimated*

The remaining parameters to be estimated are collected in Ω below:

$$\Omega = (\upsilon, \bar{z}, \Delta_\epsilon, n_\epsilon, \lambda_\epsilon, \tau, \gamma, \phi_h, \delta_0, \phi_f)$$

I proceed sequentially as follows. As the first step, I estimate the elasticity of substitution between full-time and contract workers (υ). In the second step, I compute plant productivities and estimate parameters of the productivity evolution process ($\bar{z}, \Delta_\epsilon, n_\epsilon, \lambda_\epsilon$). In the third step, I estimate the remaining parameters using simulated method of moments ($\tau, \gamma, \phi_h, \delta_0, \phi_f$). I describe each step in more detail below.

6.2.1. *Elasticity of Substitution between Workers*

I first estimate the elasticity of substitution between full-time and contract workers (v). Recall from (16) that flow profits of a plant with productivity ϵ and l_f full-time workers are given by:

$$\Pi(l_f, \epsilon) = \max_{l_c, m \geq 0} \left\{ R(q(\epsilon, l_f, l_c, m)) - w_c l_c - r_m m \right\} \tag{25}$$

I take the first order conditions of (25) with respect to l_c and m, and combine them to obtain equation (26). Here, l_c^*, m^* denote optimal contract labor and materials, at which the marginal revenue product of each input equals its factor price.

$$\log\left[\frac{\alpha_L(\frac{\mu-1}{\mu})R(q(\epsilon, l_f, l_c^*, m^*)) - w_c l_c^*}{w_c l_c^*}\right] = \left(\frac{v-1}{v}\right)\log\left[\frac{l_f}{l_c^*}\right] \tag{26}$$

The left-hand side of equation (26) measures revenue – adjusted for markup and payments to variable inputs – relative to the contract-labor wage bill. The right-hand side is the ratio of full-time to contract workers in equilibrium. Identification of v comes from comparing the extent to which the left-hand side of equation (26) responds to reform-induced variation in equilibrium $\frac{l_f}{l_c^*}$ around the regulatory threshold. When full-time and contract labor are highly substitutable, plants respond to a relaxation of size-based regulations primarily by substituting away from contract labor towards full-time workers. This leads to a sharp decline in contract-labor expenditure $w_c l_c^*$, while the labor composite, output, and revenues change only modestly. As a result, a given percentage increase in equilibrium $\frac{l_f}{l_c^*}$ will be accompanied by a large increase in the left-hand side of equation (26) when v is high. In contrast, when substitution between full-time and contract labor is limited, the policy-induced increase in full-time employment leads to little adjustment in contract labor, and contract-labor expenditures $w_c l_c^*$ remain largely unchanged. The increase in full-time labor raises the effective labor composite, but revenue net of markup and payments to variable inputs increases only modestly. Consequently, the left-hand side responds only weakly to changes in equilibrium $\frac{l_f}{l_c^*}$ when v is low.

Equation (27) is the empirical counterpart of equation (26), which I estimate using two-stage least squares (2SLS).

$$\log\left[\frac{\alpha_L(\frac{\mu-1}{\mu})R(q(\epsilon,l_f,l_c^*,m^*))-w_c l_c^*}{w_c l_c^*}\right]_{ijst} = \left(\frac{v-1}{v}\right)\log\left[\frac{l_f}{l_c^*}\right]_{ijst} + \alpha_{jt}+\theta_{sk}+\lambda_{kt}+u_{ijst} \tag{27}$$

Here, i indexes plant, j industry, s state, and t time. α_{jt} are industry$\times$year fixed effects that control for time-varying industry-specific shocks, θ_{sk} are state $\times$ lagged size-bin fixed effects that account for time-invariant differences across state-lagged size bins, λ_{kt} are lagged size bin $\times$ year fixed effects that capture aggregate shocks by plant size. In equation (27), $\log(l_f/l_c^*)$ is treated as an endogenous regressor and is instrumented using the policy-induced variation in the regulatory threshold in the treated states. The first-stage regression follows the same specification as in equation (7). The fitted values from the first-stage regression are used in equation (27) to recover $\hat{v}$. Note the estimating equation includes the same fixed effects structure as the first stage, ensuring that identification of v is driven exclusively by reform-induced variation in labor composition around the regulatory threshold.[20]

6.2.2. *Productivity and Productivity Evolution Parameters*

Using the estimated parameters of the revenue function (v, α_L, α_M, μ), and observed data on deflated revenues and deflated inputs, I recover plant-level productivity ϵ as a residual from the revenue function. Rearranging equation (10) yields equation (28) for the logarithm of plant productivity, where i indexes plants and t time.[21]

$$\log(\epsilon_{it}) = \frac{\mu}{\mu-1}\log\left(\frac{R_{it}}{P_t}\right) - \frac{1}{\mu-1}\log(\frac{I_t}{P_t}) - \alpha_M\log(m_{it}) - \frac{\alpha_L v}{v-1}\log\left(l_{f,it}^{\frac{v-1}{v}} + l_{c,it}^{\frac{v-1}{v}}\right) \tag{28}$$

[20] Although equation (27) links observable plant characteristics to the elasticity of substitution between worker types, estimating this via ordinary least squares (OLS) would yield biased estimates of v. For instance, there could be unobserved plant-level productivity or demand shocks that affect both equilibrium $\frac{l_f}{l_c^*}$ and the left-hand side of equation (27). I estimate equation (27) using 2SLS to address this endogeneity concern.

[21] All nominal variables are expressed in real terms as described earlier – in particular, m_{it} denotes deflated material expenditure.

I then estimate the parameters of the stochastic process governing plant-level productivity. As described in Section 5.1.2, the logarithm of productivity $z_{it} \equiv \log \epsilon_{it}$ follows an exogenous, first-order Markov process on a grid of $2n_\epsilon + 1$ points spaced $\Delta_\epsilon > 0$ apart. Shocks arrive at Poisson rate $\lambda_\epsilon > 0$, and conditional on a shock, z moves up or down one grid point with mean-reverting probabilities given in equation (11). In continuous time, this stochastic jump process can be expressed as:

$$dz_t = -\rho\,(z_t - \bar{z})\,dt + \varsigma\,dx_t, \tag{29}$$

where $\rho = \frac{\lambda_\epsilon}{n_\epsilon}$ is the speed of mean reversion, $\varsigma = \Delta_\epsilon\sqrt{\lambda_\epsilon}$ is the instantaneous volatility, $\bar{z}$ is the long-run mean of log productivity, and dx_t is a martingale innovation with zero mean and variance dt. Following the argument in Shimer (2005) and Eaton et al. (2022), as the productivity grid gets finer – formally, by rescaling the grid spacing and shock arrival rate appropriately – the innovations converge to a Brownian motion, and the above stochastic jump process converges to the Ornstein–Uhlenbeck (O-U) process in equation (30).

$$dz_t = -\rho\,(z_t - \bar{z})\,dt + \varsigma\,dB_t, \qquad dB_t \sim \mathcal{N}(0, dt), \tag{30}$$

where B_t is a standard Brownian motion. The mapping between the discrete-grid parameters and the O-U parameters remains $\rho = \lambda_\epsilon/n_\epsilon$ and $\varsigma = \Delta_\epsilon\sqrt{\lambda_\epsilon}$; the long-run mean $\bar{z}$ is unchanged. Derivations of the convergence and the parameter mapping are provided in Appendix II.

My data are annual, so I observe plant productivity at discrete yearly intervals. Therefore, I estimate the following discrete-time counterpart of the O-U process using annual log-productivities at the plant level $\{z_{it}\}$:

$$z_{it+1} = (1 - e^{-\rho})\bar{z} + e^{-\rho} z_{it} + \xi_{it+1}, \qquad \xi_{it+1} \sim \mathcal{N}\left(0, \frac{\varsigma^2}{2\rho}(1 - e^{-2\rho})\right) \tag{31}$$

From equation (31), I estimate the persistence $e^{-\hat{\rho}}$ and the innovation variance $\frac{\hat{\varsigma}^2}{2\hat{\rho}}(1 - e^{-2\hat{\rho}})$ by ordinary least squares. The long-run mean $\bar{z}$ is computed from the estimated intercept of equation (31). I then fix the number of grid points n_ϵ, guided by numerical convenience and the desired coverage of the productivity distribution. Given $\hat{\rho}$ and $\hat{\varsigma}^2$, the discrete-grid

parameters are determined as $\hat{\lambda}_\epsilon = n_\epsilon \hat{\rho}$ and $\hat{\Delta}_\epsilon = \hat{\varsigma}/\sqrt{\hat{\lambda}_\epsilon}$. This procedure ensures that the exogenous productivity dynamics in the model reproduce the observed annual persistence and volatility of plant-level productivity in the data.

6.2.3. *Estimation of Remaining Parameters*

The third step involves estimating the remaining policy invariant parameters and the magnitude of regulatory costs given in $\Omega = \{\tau, \gamma, \phi_h, \delta_0, \phi_f\}$ using Simulated Method of Moments (SMM). Ω is chosen to minimize the following objective function:

$$G(\Omega) = [\bar{m} - m(\Omega)]' W [\bar{m} - m(\Omega)] \tag{32}$$

Here, $\bar{m}$ is the vector of empirical moments, $m(\Omega)$ are the model counterparts of $\bar{m}$, and W is a diagonal weighting matrix with diagonal elements equal to the inverse of the squared empirical moments.[22] I construct three sets of moment conditions. The first set targets features of the pre-reform steady-state plant size distribution and plant dynamics. The second targets the corresponding features of the post-reform distribution. The third targets the reform-induced changes in plant dynamics estimated in Section 4.

Specifically, for a given parameter vector Ω and regulatory threshold ℓ^* at 100 full-time workers, I compute the steady-state plant size distribution using equations (22)–(23). From this distribution, I compute the moments reported in Panel A of Table II. To capture how hiring behavior varies with plant size, I compare the average employment growth rate among expanding plants across the first, second, and third quartiles of the plant-size distribution. To capture how firing behavior varies with plant size, I compute the mean contraction rate among contracting plants with fewer than 100 workers, the mean contraction rate among contracting plants with 101–200 workers, and the slope of the contraction rate on log plant size.

To capture plant dynamics around the regulatory threshold prior to the reform, I use: (1) the mass of plants at exactly 100 workers relative to the mass at 101 workers; (2) the mean mass of plants with 101–105 workers relative to those with 91–99 workers; and (3) the

[22] The moments I target differ substantially in scale, as Table II shows. This weighting places equal weight on the percentage deviation of each model moment from its empirical counterpart, and makes the objective function independent of the units in which the moments are measured.

mean mass of plants with 96–100 workers relative to those with 92–95 workers. Finally, to characterize the right tail of the distribution, I compute the log-odds of the share of plants employing fewer than 507 full-time workers, which corresponds to the 99th percentile of plant size in the data. These model moments are matched with their empirical counterparts from the pre-reform years to form the first set of moment conditions.

I then consider an alternative regime in which the regulatory threshold is raised from 100 to 300 full-time workers. I recompute the policy functions and the steady-state plant size distribution under this regime, allowing the price index to adjust so that the free entry condition (19) continues to hold. From the post-reform distribution, I compute two of the measures described above: the mass of plants at exactly 100 workers relative to the mass at 101, and the mean mass of plants with 96–100 workers relative to those with 92–95. These form the second set of moment conditions, reported in Panel B of Table II. Comparing the two regimes, I also compute the change in the probability that plants transition from below to above 100 full-time workers, and the change in the contract labor share of plants near 100 full-time workers. These are matched with their reduced-form counterparts estimated in Section 4 to form the third set, reported in Panel C.

6.3. *Identification and Model Fit*

While all parameters in Ω are estimated jointly and every moment responds to more than one parameter, certain moments are particularly informative about particular parameters.

The economies-of-scale parameter in hiring, ϕ_h, governs how the marginal cost of search differs between large and small plants for the same search effort. It therefore determines whether larger plants find it relatively easier or harder to expand than smaller ones. The ratios of average employment growth rates among expanding plants across quartiles of the plant-size distribution are informative about ϕ_h. The right tail of the plant size distribution is also informative about ϕ_h. I summarize the right tail by the log-odds of the share of plants employing fewer than 507 full-time workers, the 99th percentile of plant size in the pre-reform data.

Analogously, the economies-of-scale parameter in firing, ϕ_f, governs how the marginal cost of expediting separations differs between large and small plants. The slope of the contraction rate on log plant size is therefore directly informative about ϕ_f.

The two regulatory wedges, τ and γ, are identified jointly from plant behavior around 100 full-time workers before and after the reform. As discussed in Section 5.5, both wedges reduce the probability with which plants transition from below to above 100 full-time workers prior to the reform. What distinguishes the two is the density of plants on either side of the threshold: higher τ builds up mass below 100 workers and depletes it above, whereas higher γ builds up mass above 100 workers. Therefore, plant density on either side of 100 full-time workers both before and after the reform, together with the reform-induced change in transition probabilities and in contract labor share around 100 workers, is informative about the two wedges.

Finally, the exogenous separation rate δ_0 absorbs the residual component of mean contraction rates and threshold moments, once ϕ_h, ϕ_f, τ and γ are disciplined by the above moments.

Table II lists the set of moments used to estimate the parameters in Ω. Model predictions match their data counterparts closely. Panels A and B show that the model captures plant dynamics well both before and after the reform. Panel C shows that it reproduces the reform-induced changes estimated in Section 4.2.

6.4. *Parameter Estimates*

Table III reports the parameter estimates. The elasticity of substitution between full-time and contract workers $\hat{v}$ is estimated to be 2.08, suggesting moderate substitutability between the two types of workers. The estimated productivity process has an annual autocorrelation of $e^{-\hat{\rho}} = 0.868$ and mean log productivity $\bar{z} = -2.6$. In continuous time, this corresponds to a mean-reversion rate $\hat{\rho} = 0.141$ and volatility $\hat{\varsigma} = 0.41$. For the numerical solution, I set $n_\epsilon = 20$, implying 41 grid points for productivity. This implies an annual shock arrival rate $\hat{\lambda}_\epsilon = 2.82$ and grid spacing $\hat{\Delta}_\epsilon = 0.244$.

The estimated economies-of-scale parameters imply that larger plants adjust employment more readily than smaller ones, on both the hiring and firing margins. Turning to regulatory costs, the estimates are substantial. Over and above the baseline firing costs that all plants face, the IDA imposes an additional firing cost equal to 1.657 times the annual wage for each worker fired. The wedge on their wage rate is 1.54%.

TABLE II

TARGETED MOMENTS: MODEL VERSUS DATA

Moments	Model	Data
Panel A: Pre-reform steady state		
Hiring behavior by plant size		
Expansion growth rate: quartile 1 / quartile 2	1.608	1.550
Expansion growth rate: quartile 2 / quartile 3	1.154	1.160
Firing behavior by plant size		
Mean contraction rate, plants with <100 workers	−0.416	−0.377
Mean contraction rate, plants with 101–200 workers	−0.271	−0.300
Slope of contraction rate on log plant size	0.028	0.029
Dynamics around the regulatory threshold		
Mass at 100 workers / mass at 101 workers	1.233	1.234
Mean mass at 101–105 / mean mass at 91–99	1.059	1.106
Mean mass at 96–100 / mean mass at 92–95	1.066	1.076
Right tail of the distribution		
Share of plants below 507 workers (log-odds)	5.103	4.595
Panel B: Post-reform steady state		
Mass at 100 workers / mass at 101 workers	1.032	0.986
Mean mass at 96–100 / mean mass at 92–95	0.971	0.975
Panel C: Change induced by the reform		
Transition prob. from below to above 100 workers (pp)	5.4	5
Contract labor share around 100 workers (% change)	−4.2	−4.29

Notes: Data moments are computed from ASI plant-level data for the treated states. Panel A reports moments from the years before the reform, Panel B from the years after the reform, and Panel C the reform-induced changes estimated in Section 4. The 507-worker cutoff in Panel A corresponds to the 99th percentile of plant size in the pre-reform data. Transition probabilities in Panel C refer to plants with 81–100 full-time workers moving to 100 or more workers, corresponding to the finer bin specification in Section 4.2. All data moments in panels A and B are first computed within each 2-digit industry and then averaged across industries, weighting each industry by its share of total revenue in the pre-reform years.

7. COUNTERFACTUAL EXPERIMENTS

Having estimated the magnitude of regulatory costs, I now use the model to quantify their aggregate implications. I consider two sets of counterfactual experiments. The first

TABLE III

PARAMETER ESTIMATES

Parameter	Description	Estimate
Panel A: Estimated by 2SLS		
υ	Elasticity of substitution between worker types	2.08
Panel B: Estimated from the productivity process		
ρ	Mean reversion rate of the O-U process	0.141
ς	Instantaneous volatility of the O-U process	0.410
Δ_ϵ	Spacing of the log productivity grid	0.244
λ_ϵ	Annual arrival rate of productivity shocks	2.820
Panel C: Estimated by SMM		
ϕ_h	Economies of scale in hiring	0.917
ϕ_f	Economies of scale in firing	1.347
δ_0	Exogenous separation rate	0.081
τ	Wedge on wage rate imposed by regulation	1.54%
γ/w_f	Firing cost per worker dismissed, in annual wages	1.657

Notes: Panel A reports the elasticity of substitution between full-time and contract workers, estimated by two-stage least squares, as described in Section 6.2.1. Panel B reports parameters of the productivity process as described in Section 6.2.2. Panel C reports parameters estimated by simulated method of moments as described in Section 6.2.3.

set holds the regulatory threshold at 100 full-time workers and considers two cases: removing the wage wedge τ alone, and removing both regulatory wedges (τ and γ). The wage wedge arises from the enforcement regime rather than from the letter of the law itself, so removing it corresponds to eliminating the rents extracted through inspections while leaving the restriction on dismissals in place. Removing both wedges corresponds to repealing the regulation altogether. The second set holds both wedges at their estimated values in Table III and varies the threshold ℓ^*. This speaks to the design of the regulation rather than its intensity.

Table IV reports the first set of results. Removing the wage wedge alone raises aggregate output by 0.13% and full-time employment by 1.65%, but leaves value added per worker essentially unchanged. Removing both wedges raises output by 1.56%, full-time employment by 11.79%, and value added per worker by 1.33%. Therefore, the regulatory burden

in this setting is overwhelmingly a firing cost: eliminating the costs of corrupt enforcement, while leaving the constraint on dismissals intact, recovers less than a tenth of the available output gains.

Multiple forces are at work here. Removing the firing wedge lowers the cost of firing workers, so separations rise. It also increases the value of hiring additional workers, so hiring rises as well. That full-time employment increases indicates that the increase in hiring dominates the increase in separations. This impacts aggregate output and value added per worker through two channels. First, workers get reallocated across plants: those that were hoarding labor release it, and those that search more intensely absorb it, so workers move toward plants that use them more productively. Second, more plants enter, since the value of entry rises.

TABLE IV

COUNTERFACTUAL EFFECTS OF REMOVING REGULATORY WEDGES

Outcome	(A) $\ell^* = 100, \tau = 0$, γ fixed	(B) $\ell^* = 100, \tau = 0, \gamma = 0$
Aggregate output	$+0.13\%$	$+1.56\%$
Full-time employment	$+1.65\%$	$+11.79\%$
Share of full-time employment	$+1.47\%$	$+11.54\%$
Value added per worker	-0.05%	$+1.33\%$

Notes: All entries are percentage changes relative to the baseline equilibrium. The baseline equilibrium is one in which the regulatory threshold is set at 100 full-time workers, and both regulatory wedges take their estimated values from Table III. Column (A) removes the wage wedge τ arising from the enforcement regime while retaining the firing wedge γ. Column (B) removes both wedges. The regulatory threshold is held at 100 full-time workers throughout. In each experiment, I recompute the policy functions and the steady-state plant size distribution, allowing the price index to adjust so that the free entry condition (19) continues to hold. Aggregate output is measured as I/P, where P is the equilibrium price index, since aggregate expenditure is fixed and market clearing (24) holds in every regime. Value added per worker is real value-added divided by total number of workers, comprising both full-time and contract workers.

Table V reports the results from varying the regulatory threshold, holding both wedges at their estimated values. Aggregate output and value added per worker are monotone in ℓ^*. Relative to a threshold of 100 full-time workers, extending the regulation to all plants lowers output by 1.14%, tightening the threshold to 50 workers lowers it by 0.48%, relaxing it to 300 workers raises it by 0.82%, and removing the regulation entirely raises it by 1.56%. The reform that actually took place in the treated states, which relaxed the threshold to 300

full-time workers, therefore delivers roughly half of the gains available from full repeal. The mechanism behind the output response is the same one at work in the first set of experiments.

Aggregate full-time employment, by contrast, is not monotone. It rises by 2.78% when the regulation is extended to all plants (column A) and by 11.79% when the regulation is removed entirely (column D), but these two increases have very different sources. When the regulation binds universally, a wide range of plants stop firing actively altogether, and employment rises because separations fall by more than hires. Output falls by 1.14% and value added per worker by 2.22% at the same time, so the additional workers are retained at plants that would otherwise release them. When the regulation is removed altogether, employment rises alongside output and value added per worker, reflecting reallocation rather than retention. Reading the employment response in isolation would thus give a misleading account of the regulation's effects, and it is the joint behavior of output and employment that distinguishes the two cases.

TABLE V

COUNTERFACTUAL EFFECTS OF VARYING THE REGULATORY THRESHOLD

Outcome	(A) $\ell^* = 0$	(B) $\ell^* = 50$	(C) $\ell^* = 300$	(D) No threshold
Aggregate output	-1.14%	-0.48%	$+0.82\%$	$+1.56\%$
Full-time employment	$+2.78\%$	-1.03%	$+4.55\%$	$+11.79\%$
Share of full-time employment	$+1.66\%$	-1.17%	$+4.54\%$	$+11.54\%$
Value added per worker	-2.22%	-0.62%	$+0.81\%$	$+1.33\%$

Notes: All entries are percentage changes relative to the baseline equilibrium. The baseline equilibrium is one in which the regulatory threshold is set at 100 full-time workers and both wedges take their estimated values from Table III. Column (A) extends the regulation to all plants; column (B) tightens the threshold to 50 full-time workers; column (C) relaxes it to 300 full-time workers, corresponding to the reform enacted in the treated states; and column (D) removes the regulation entirely. In each experiment, both wedges are fixed at their estimated values. Aggregate output is measured as I/P, where P is the equilibrium price index, since aggregate expenditure is fixed and market clearing (24) holds in every regime. Value added per worker is real value-added divided by total headcount, comprising both full-time and contract workers.

8. CONCLUSION

Firms in many countries face stricter regulation once their employment crosses a threshold. This paper asks how costly such rules are, what they imply for overall employment

and output, and what trade-offs policymakers face in reforming them. I answer these questions in the context of Chapter V-B of India's Industrial Disputes Act of 1947 (IDA), which applies to manufacturing plants with 100 or more full-time workers. The IDA imposes two distinct costs: a restriction on dismissal requiring government authorization, and an enforcement regime of frequent inspections whose burden rises with the full-time workforce. To measure these costs, I exploit a set of reforms beginning in 2014, in which several Indian states raised the IDA's applicability threshold from 100 to 300 full-time workers. My analysis has three parts. First, I use the staggered reforms to provide causal evidence on how plants responded to the change in regulatory threshold. Second, I build a dynamic model of heterogeneous plants facing labor market frictions and use it to show how the two costs imposed by the IDA affect plant dynamics differently. Third, I combine the natural experiment and causal estimates with the structural model to quantify each cost, assess the aggregate implications of the regulation, and evaluate counterfactual regulatory regimes.

I find the burden to be overwhelmingly a firing cost. Decomposing the regulatory burden in this way is possible because the two costs leave different imprints on the plant size distribution. A wage wedge lowers the probability that plants transition from below to above the threshold, producing the canonical signature of bunching: excess mass below and missing mass above the threshold. A firing wedge lowers transitions in both directions, since plants above the threshold hoard labor amidst negative productivity shocks while plants below it search less intensively for workers who will be costly to let go of later. It therefore attenuates the bunching signature, however costly the underlying regulation may be. This yields a broader implication that extends beyond this setting: wherever a size-based rule operates primarily through dismissal costs, a bunching-based estimator will understate its cost.

How do these plant-level distortions translate into aggregate outcomes? The answer depends importantly on how broadly the regulation applies. Extending it to all plants reduces output and value added per worker, even as full-time employment and its share in total employment rise. This is because a wide range of plants stop actively firing altogether, so workers are retained at plants that would otherwise release them, weakening the alignment between full-time employment and plant productivity, and lowering aggregate output and value added per worker. Repealing the regulation has the opposite effect: plants that had been hoarding full-time labor release it, while more productive plants search more in-

tensively for workers, improving the allocation of full-time labor across plants. Aggregate output rises by 1.56%, value added per worker by 1.33%, full-time employment by 11.79%, and share of full-time employment by 11.54%. The 100-to-300-worker threshold reform, which I study using staggered state-level implementation, recovers roughly half of these gains, raising output by 0.82% and value added per worker by 0.81%. These estimates provide a quantitative assessment of the threshold change subsequently adopted nationally in India's Industrial Relations Code, 2020. All of these results, however, are measured from the plant side. Whether reforming the IDA leaves workers better off, through the composition of employment or the security of the jobs it was intended to protect, is a question I leave for future research.

REFERENCES

ACEMOGLU, DARON AND WILLIAM B HAWKINS (2014): "Search with multi-worker firms," *Theoretical Economics*, 9 (3), 583–628. [5]

AGHION, PHILIPPE, ROBIN BURGESS, STEPHEN J REDDING, AND FABRIZIO ZILIBOTTI (2008): "The unequal effects of liberalization: Evidence from dismantling the License Raj in India," *American Economic Review*, 98 (4), 1397–1412. [8]

AHSAN, AHMAD AND CARMEN PAGÉS (2009): "Are all labor regulations equal? Evidence from Indian manufacturing," *Journal of Comparative Economics*, 37 (1), 62–75. [8]

ALFARO, LAURA AND ANUSHA CHARI (2014): "Deregulation, misallocation, and size: Evidence from India," *The Journal of Law and Economics*, 57 (4), 897–936. [8]

AMIRAPU, AMRIT AND MICHAEL GECHTER (2020): "Labor regulations and the cost of corruption: Evidence from the Indian firm size distribution," *Review of Economics and Statistics*, 102 (1), 34–48. [2, 4, 7, 11]

BANERJEE, ABHIJIT V AND ESTHER DUFLO (2014): "Do firms want to borrow more? Testing credit constraints using a directed lending program," *Review of Economic Studies*, 81 (2), 572–607. [8]

BAQAEE, DAVID REZZA AND EMMANUEL FARHI (2020): "Productivity and misallocation in general equilibrium," *The quarterly journal of economics*, 135 (1), 105–163. [8]

BAU, NATALIE AND ADRIEN MATRAY (2023): "Misallocation and capital market integration: Evidence from India," *Econometrica*, 91 (1), 67–106. [8]

BERTRAND, MARIANNE, CHANG-TAI HSIEH, AND NICK TSIVANIDIS (2025): "Contract labor and establishment growth in India," *Econometrica*, 93 (4), 1411–1448. [2, 8, 9, 11, 20]

BESLEY, TIMOTHY AND ROBIN BURGESS (2004): "Can labor regulation hinder economic performance? Evidence from India," *The Quarterly journal of economics*, 119 (1), 91–134. [8]

BHATTACHARJEA, ADITYA (2006): "Labour market regulation and industrial performance in India: A critical review of the empirical evidence," *Indian Journal of Labour Economics*, 49 (2), 211–232. [8]

——— (2021): "Labour market flexibility in Indian manufacturing: A critical survey of the literature," *International Labour Review*, 160 (2), 197–217. [8]

BOLLARD, ALBERT, PETER J KLENOW, AND GUNJAN SHARMA (2013): "Indias mysterious manufacturing miracle," *Review of Economic Dynamics*, 16 (1), 59–85. [8]

CENGIZ, DORUK, ARINDRAJIT DUBE, ATTILA LINDNER, AND BEN ZIPPERER (2019): "The effect of minimum wages on low-wage jobs," *The Quarterly Journal of Economics*, 134 (3), 1405–1454. [18]

CHARI, AMALAVOYAL V (2011): "Identifying the aggregate productivity effects of entry and size restrictions: An empirical analysis of license reform in India," *American Economic Journal: Economic Policy*, 3 (2), 66–96. [8]

CHATTERJEE, SHOUMITRO, KALA KRISHNA, KALYANI PADMAKUMAR, AND YINGYAN ZHAO (2025): "No Country for Dying Firms: Evidence from India," Tech. rep., National Bureau of Economic Research. [8, 26, 34]

CHAUREY, RITAM (2015): "Labor regulations and contract labor use: Evidence from Indian firms," *Journal of Development Economics*, 114, 224–232. [8]

CHIPLUNKAR, GAURAV, RITAM CHAUREY, AND VIDHYA SOUNDARARAJAN (2024): "Employment Protection Legislation and Informality: Theory and Evidence from India," Tech. Rep. 4204033, SSRN. [8, 9]

COŞAR, A KEREM, NEZIH GUNER, AND JAMES TYBOUT (2016): "Firm dynamics, job turnover, and wage distributions in an open economy," *American Economic Review*, 106 (3), 625–63. [5, 25]

DAS, SANGHAMITRA, MARK J ROBERTS, AND JAMES R TYBOUT (2007): "Market entry costs, producer heterogeneity, and export dynamics," *Econometrica*, 75 (3), 837–873. [5]

DAVID, JOEL M AND VENKY VENKATESWARAN (2019): "The sources of capital misallocation," *American Economic Review*, 109 (7), 2531–2567. [8]

DE CHAISEMARTIN, CLÉMENT AND XAVIER D'HAULTFOEUILLE (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American economic review*, 110 (9), 2964–2996. [18]

DE LOECKER, JAN, PINELOPI K GOLDBERG, AMIT K KHANDELWAL, AND NINA PAVCNIK (2016): "Prices, markups, and trade reform," *Econometrica*, 84 (2), 445–510. [34]

DEBROY, BIBEK AND LAVEESH BHANDARI (2008): *India Labour Report 2008*, Teamlease Services. [10, 11]

DIX-CARNEIRO, RAFAEL, PINELOPI K GOLDBERG, COSTAS MEGHIR, AND GABRIEL ULYSSEA (2021): "Trade and informality in the presence of labor market frictions and regulations," Tech. rep., National Bureau of Economic Research. [5]

DURANTON, GILLES, EJAZ GHANI, ARTI GROVER GOSWAMI, AND WILLIAM KERR (2015): "The misallocation of land and other factors of production in India," Policy research working paper, The World Bank. [8]

EATON, JONATHAN, DAVID JINKINS, JAMES TYBOUT, AND DANIEL XU (2016): "Two-sided search in international markets," in *2016 Annual Meeting of the Society for Economic Dynamics*. [5]

EATON, JONATHAN, DAVID JINKINS, JAMES R TYBOUT, AND DANIEL XU (2022): "Two-sided search in international markets," Tech. rep., National Bureau of Economic Research. [5, 25, 37]

ELSBY, MICHAEL W L AND RYAN MICHAELS (2013): "Marginal jobs, heterogeneous firms, and unemployment flows," *American Economic Journal: Macroeconomics*, 5 (1), 1–48. [5]

FELBERMAYR, GABRIEL, JULIEN PRAT, AND HANS-JÖRG SCHMERER (2011): "Globalization and labor market outcomes: Wage bargaining, search frictions, and firm heterogeneity," *Journal of Economic theory*, 146 (1), 39–73. [5]

GARCÍA-SANTANA, MANUEL AND JOSEP PIJOAN-MAS (2011): "Small scale reservation laws and the misallocation of talent," Discussion Paper DP8242, CEPR. [2, 7]

GARICANO, LUIS, CLAIRE LELARGE, AND JOHN VAN REENEN (2016): "Firm size distortions and the productivity distribution: Evidence from France," *American Economic Review*, 106 (11), 3439–79. [2, 7]

GOODMAN-BACON, ANDREW (2021): "Difference-in-differences with variation in treatment timing," *Journal of econometrics*, 225 (2), 254–277. [18]

GOURIO, FRANÇOIS AND NICOLAS ROYS (2014): "Size-dependent regulations, firm size distribution, and reallocation," *Quantitative Economics*, 5 (2), 377–416. [2, 7, 1]

GUNER, NEZIH, GUSTAVO VENTURA, AND YI XU (2008): "Macroeconomic implications of size-dependent policies," *Review of economic Dynamics*, 11 (4), 721–744. [2, 7]

HASAN, RANA, DEVASHISH MITRA, AND KRISHNARAJAPET V RAMASWAMY (2007): "Trade reforms, labor regulations, and labor-demand elasticities: Empirical evidence from India," *The Review of Economics and Statistics*, 89 (3), 466–481. [8]

HELPMAN, ELHANAN, OLEG ITSKHOKI, AND STEPHEN REDDING (2010): "Inequality and unemployment in a global economy," *Econometrica*, 78 (4), 1239–1283. [5]

HOPENHAYN, HUGO AND RICHARD ROGERSON (1993): "Job turnover and policy evaluation: A general equilibrium analysis," *Journal of political Economy*, 101 (5), 915–938. [5]

HOPENHAYN, HUGO A (1992): "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica: Journal of the Econometric Society*, 1127–1150. [5]

HSIEH, CHANG-TAI AND PETER J KLENOW (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124 (4), 1403–1448. [8]

——— (2014): "The life cycle of plants in India and Mexico," *The Quarterly Journal of Economics*, 129 (3), 1035–1084. [8]

HSIEH, CHANG-TAI AND BENJAMIN A OLKEN (2014): "The missing" missing middle"," *Journal of Economic Perspectives*, 28 (3), 89–108. [4, 8, 1]

KAAS, LEO AND PHILIPP KIRCHER (2015): "Efficient firm dynamics in a frictional labor market," *American Economic Review*, 105 (10), 3030–60. [5]

KLEVEN, HENRIK JACOBSEN (2016): "Bunching," *Annual Review of Economics*, 8 (1), 435–464. [2, 7]

KRUSELL, PER, LEE E OHANIAN, JOSÉ-VÍCTOR RÍOS-RULL, AND GIOVANNI L VIOLANTE (2000): "Capital-skill complementarity and inequality: A macroeconomic analysis," *Econometrica*, 68 (5), 1029–1053. [23]

MELITZ, MARC J (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *econometrica*, 71 (6), 1695–1725. [5]

NIMIER-DAVID, ELIO, DAVID SRAER, AND DAVID THESMAR (2026): "The effects of mandatory profit-sharing on workers and firms: evidence from France," *The Quarterly Journal of Economics*, 141 (3), 2205–2267. [8]

OECD (2026): *OECD Employment Outlook 2026: Geographic Disparities in Jobs and Incomes*, Paris: OECD Publishing. [2]

PARRO, FERNANDO (2013): "Capital-skill complementarity and the skill premium in a quantitative model of trade," *American Economic Journal: Macroeconomics*, 5 (2), 72–117. [23]

RESTUCCIA, DIEGO AND RICHARD ROGERSON (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 11 (4), 707–720. [8]

——— (2017): "The causes and costs of misallocation," *Journal of Economic Perspectives*, 31 (3), 151–174. [8]

SCHIVARDI, FABIANO AND ROBERTO TORRINI (2008): "Identifying the effects of firing restrictions through size-contingent differences in regulation," *Labour Economics*, 15 (3), 482–511. [7]

SHIMER, ROBERT (2005): "The cyclical behavior of equilibrium unemployment and vacancies," *American economic review*, 95 (1), 25–49. [24, 37, 8]

SIVADASAN, JAGADEESH (2009): "Barriers to competition and productivity: Evidence from India," *The BE Journal of Economic Analysis & Policy*, 9 (1). [8]

SOOD, ARADHYA (2020): "Land market frictions in developing countries: Evidence from manufacturing firms in India," Working Paper. [8]

SUN, LIYANG AND SARAH ABRAHAM (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics*, 225 (2), 175–199. [18]

Online Appendix Material for

Small By Choice? Reassessing the Costs of Size-Based Regulations

by Kalyani Padmakumar

I: Institutional Background and Reduced-Form Evidence

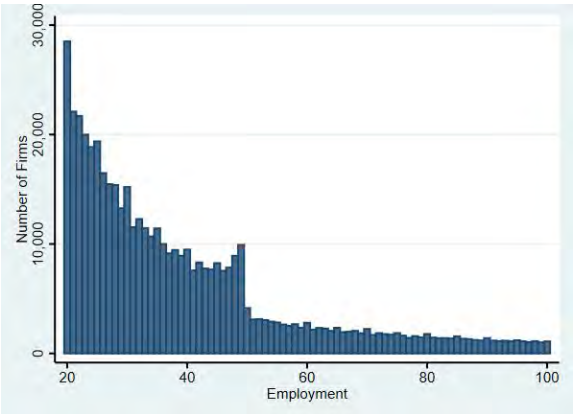


FIGURE A.1.—French firm size distribution
Source: Gourio and Roys (2014)

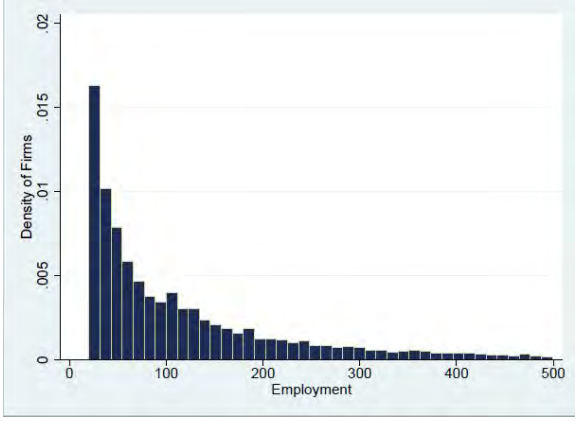


FIGURE A.2.—Indian plant size distribution
Source: Hsieh and Olken (2014)

TABLE A.1

PERCENTAGE DISTRIBUTION OF CASES PENDING BEFORE LABOR COURTS IN INDIA AS OF OCTOBER 2020

	All India
0 to 1 year	35.24%
1 to 3 years	27.75%
3 to 5 years	15.78%
5 to 10 years	13.38%
10 to 20 years	6.90%
20 to 30 years	0.93%
30+ years	0.02%
Total	146660

Source: National Judicial Data Grid, Shyam Sundar (2020)

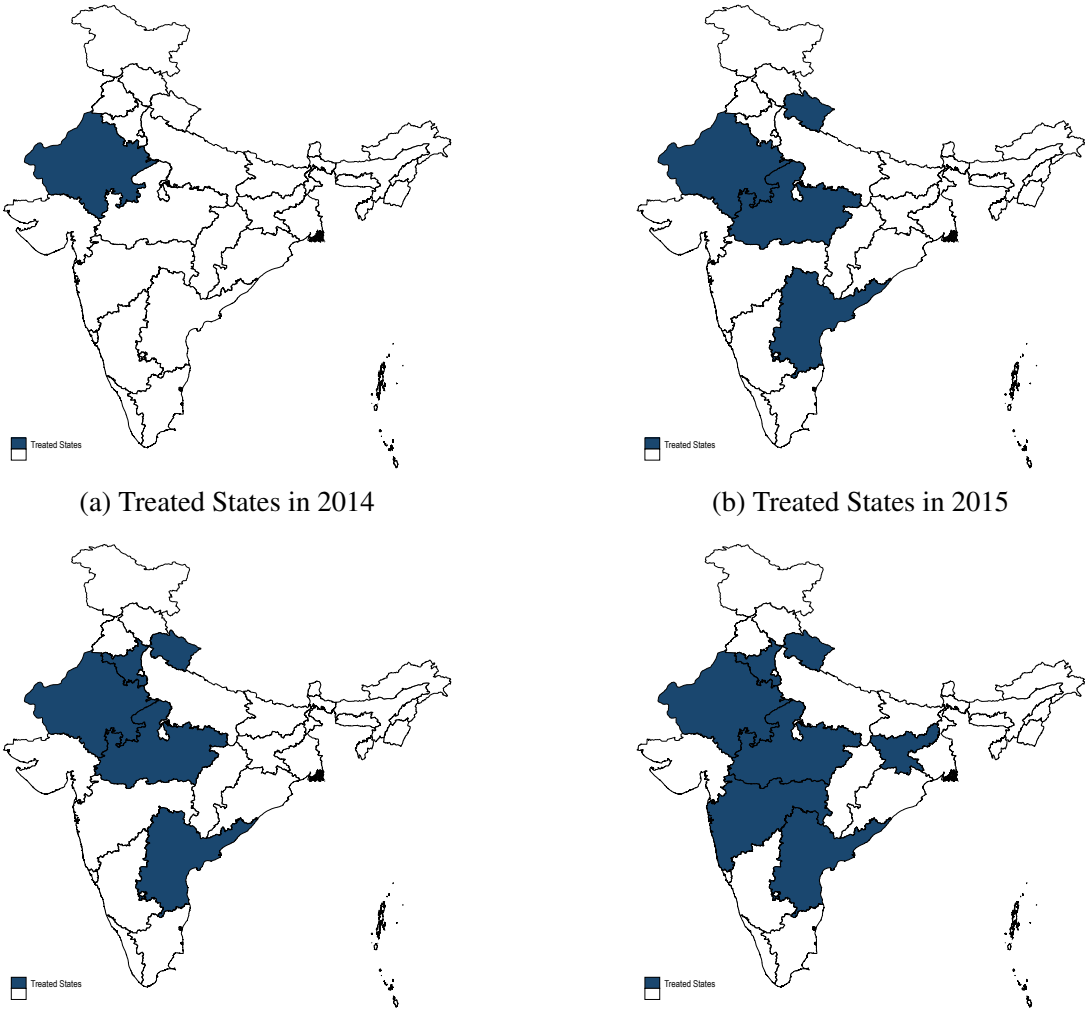
TABLE A.2

PENDENCY OF LABOR CASES IN HIGH COURTS AS OF DECEMBER 2013

	Pendency of Cases as on Dec 2013	% Share
Allahabad	10,43,398	23.38%
Madras	5,57,479	12.49%
Bombay	3,49,837	7.84%
Rajasthan	3,07,640	6.89%
Calcutta	2,80,006	6.27%
Punjab-Haryana	2,62,760	5.89%
Madhya Pradesh	2,61,611	5.86%
Andhra Pradesh	2,32,459	5.21%
Orissa	2,06,822	4.63%
Karnataka	1,96,972	4.41%
Kerala	1,32,159	2.96%
Patna	1,32,155	2.96%
Jammu-Kashmir	93,038	2.08%
Gujarat	91,953	2.06%
Jharkhand	72,958	1.63%
Delhi	64,652	1.45%
Himachal Pradesh	60,073	1.35%
Chhattisgarh	44,139	0.99%
Guwahati	40,912	0.92%
Uttarakhand	20,686	0.46%
Tripura+Manipur+Meghalaya	10,876	0.24%
Sikkim	120	0.00%
Total	44,62,705	100%

Source: Lok Sabha Unstarred Question No. 812 (November 24, 2014); Shyam Sundar (2020)

FIGURE A.3.—List of treated states and their year of treatment



(a) Treated States in 2014

(b) Treated States in 2015

(c) Treated States in 2016

(d) Treated States in 2017

Notes: Each panel shows the set of Indian states (in blue) that had raised the IDA's applicability threshold from 100 to 300 full-time workers by the indicated year. Rajasthan was the first state to implement the reform in 2014, followed by Madhya Pradesh, Andhra Pradesh, and Uttarakhand in 2015, Haryana in 2016, and Maharashtra and Jharkhand in 2017. Unshaded states retained the original threshold of 100 full-time workers throughout the sample period.

TABLE A.3

POLICY IMPACT ON TRANSITIONS (ROBUSTNESS CHECK WITH FINER BINS)

	(1) Transition probabilities
$\mathbb{1}[Post_{st}]$	-0.01
	(0.013)
$\mathbb{1}[Post_{st}] \times l^f_{isjt-1} \in (20, 40]$	0.005
	(0.015)
$\mathbb{1}[Post_{st}] \times l^f_{isjt-1} \in (40, 60]$	-0.005
	(0.020)
$\mathbb{1}[Post_{st}] \times l^f_{isjt-1} \in (60, 80]$	0.010
	(0.020)
$\mathbb{1}[Post_{st}] \times l^f_{isjt-1} \in (80, 100]$	0.050***
	(0.015)
$\mathbb{1}[Post_{st}] \times l^f_{isjt-1} \in (100, 200]$	0.021
	(0.012)
$\mathbb{1}[Post_{st}] \times l^f_{isjt-1} \in (200, 300]$	-0.028**
	(0.012)
$\mathbb{1}[Post_{st}] \times l^f_{isjt-1} > 300$	0.010*
	(0.005)
Observations	198703
Industry × Year FE	Yes
State × Lagged size-bin FE	Yes
Lagged size-bin × Year FE	Yes

Notes: Each observation is at the plant-year level. Plants are assigned to finer size bins based on their full-time employment: 0–20, 21–40, 41–60, 61–80, 81–100, 101–200, 201–300, and 300+. The dependent variable equals one if a plant transitions to a higher bin in the following year. The table reports coefficients from specification (4). Post is a state-year indicator equal to one if the state had raised the regulatory threshold to 300 full-time workers by that year. Bin 0–20 is the omitted category. Bins 81–100 and 201–300 are policy-relevant, while the remaining bins serve as placebos. The sample includes all major Indian states and covers 1999–2018. Robust standard errors clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Event Study Graphs: Policy Impact on Transitions of Plants

FIGURE A.4.—Bin: 51-100 workers

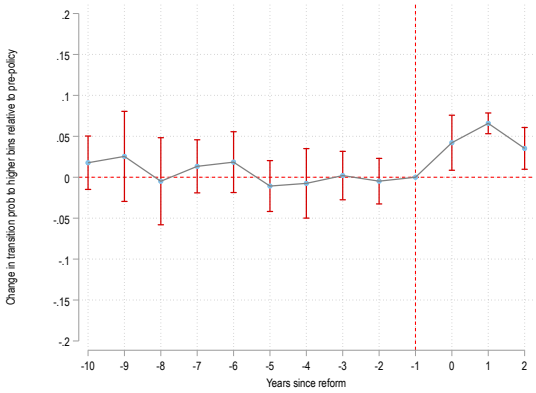


FIGURE A.5.—Bin: 101-200 workers

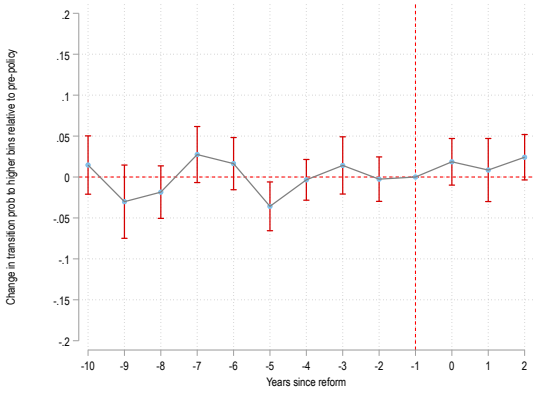
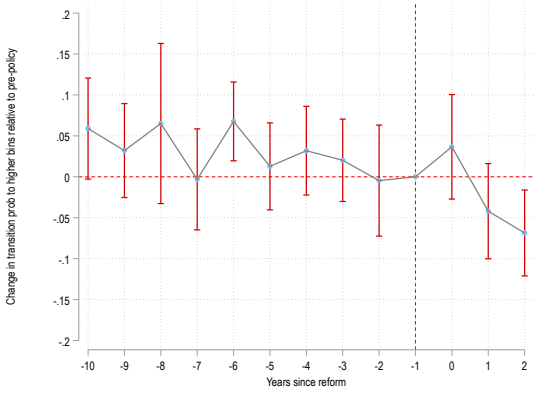


FIGURE A.6.—Bin: 201-300 workers



Note: Each figure plots estimated coefficients and 95% confidence intervals from the event study specification (5). The horizontal axis measures event time relative to the year of policy implementation ($\tau = 0$). The omitted period is $\tau = -1$, and $\tau = -10$ groups all periods ten or more years before the reform. Each figure above captures the change in the gap in upward transition probabilities between the indicated bin and Bin 1 (0–50 workers) at event time τ, relative to $\tau = -1$. Figure A.4 reports estimates for Bin 2 (51–100 workers), figure A.5 for Bin 3 (101–200 workers), and figure A.6 for Bin 4 (201–300 workers). The sample includes all major Indian states and covers 1999–2018.

TABLE A.4

POLICY IMPACT ON TRANSITIONS AND SUBSTITUTION PATTERNS: STACKED DIFFERENCE-IN-DIFFERENCES

	(1) Transition probabilities	(2) Share of contract labor	(3) Log(Investment rate)
$\mathbb{1}[Post_{st}]$	-0.006	0.012	0.006
	(0.007)	(0.013)	(0.006)
$\mathbb{1}[Post_{st}] \times l^f_{ijst-1} \in (50, 100]$	0.041***	-0.026*	-0.013
	(0.010)	(0.015)	(0.008)
$\mathbb{1}[Post_{st}] \times l^f_{ijst-1} \in (100, 200]$	0.015	0.001	-0.004
	(0.014)	(0.024)	(0.008)
$\mathbb{1}[Post_{st}] \times l^f_{ijst-1} \in (200, 300]$	-0.034**	-0.004	-0.004
	(0.015)	(0.027)	(0.013)
$\mathbb{1}[Post_{st}] \times l^f_{ijst-1} > 300$	0.007	0.002	-0.003
	(0.005)	(0.025)	(0.009)
Observations	602961	304415	636446
Industry × Year FE	Yes	Yes	Yes
State × Lagged size-bin FE	Yes	Yes	Yes
Lagged size-bin × Year FE	Yes	Yes	Yes
Cohort × Year FE	Yes	Yes	Yes

Notes: Each observation is at the plant-year-cohort level. Treated states are divided into four cohorts based on their year of policy adoption (2014, 2015, 2016, 2017). Each cohort dataset includes plants from that cohort's treated states together with plants from never-treated states. The four datasets are stacked, and the specification includes cohort-by-year fixed effects to ensure comparisons are made only between treated and never-treated states within each cohort. In column 1, the dependent variable equals one if a plant transitions to a higher size bin in the following year. In column 2, the dependent variable is the contract labor share. In column 3, the dependent variable is the log investment rate in plant and machinery. Post is a cohort-year indicator equal to one if year t is at or after the adoption year for cohort c. Bin 0–50 is the omitted category. All specifications include industry-by-year, state-by-lagged size bin, lagged size bin-by-year, and cohort-by-year fixed effects. The sample includes all major Indian states and covers 1999–2018. Robust standard errors clustered at the state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

APPENDIX: II: MODEL

A: Convergence of the Discrete Productivity Process to the Ornstein–Uhlenbeck Process

This section provides details on the convergence of the discrete stochastic process for log productivity described in Section 5.1.2 to the continuous-time Ornstein–Uhlenbeck (O-U) process stated in equation (30). It then derives the discrete-time representation used for estimation in equation (31) and characterizes the stationary distribution.

A.1 Conditional Moments of the Discrete Process

Recall from Section 5.1.2 that log productivity $z = \log(\epsilon)$ lies on a grid with $2n_\epsilon + 1$ points spaced Δ_ϵ apart, centered at $\bar{z}$. Shocks arrive at Poisson rate λ_ϵ, and conditional on a shock, z moves up or down by Δ_ϵ with probabilities $p(z)$ and $1 - p(z)$ as defined in equation (11). Within an infinitesimal interval dt, the change dz satisfies:

$$dz = \begin{cases} 0 & \text{with probability } 1 - \lambda_\epsilon\, dt, \\ +\Delta_\epsilon & \text{with probability } \lambda_\epsilon\, dt \cdot \frac{1}{2}\left(1 - \frac{z-\bar{z}}{n_\epsilon \Delta_\epsilon}\right), \\ -\Delta_\epsilon & \text{with probability } \lambda_\epsilon\, dt \cdot \frac{1}{2}\left(1 + \frac{z-\bar{z}}{n_\epsilon \Delta_\epsilon}\right) \end{cases}$$

Computing the conditional mean:

$$\begin{aligned} \mathbb{E}[dz \mid z] &= \Delta_\epsilon \cdot \lambda_\epsilon\, dt \cdot \tfrac{1}{2}\left(1 - \tfrac{z-\bar{z}}{n_\epsilon \Delta_\epsilon}\right) - \Delta_\epsilon \cdot \lambda_\epsilon\, dt \cdot \tfrac{1}{2}\left(1 + \tfrac{z-\bar{z}}{n_\epsilon \Delta_\epsilon}\right) \\ &= -\frac{\lambda_\epsilon}{n_\epsilon}(z - \bar{z})\, dt \;=\; -\theta(z - \bar{z})\, dt, \end{aligned} \tag{A.33}$$

where $\theta \equiv \lambda_\epsilon / n_\epsilon$. For the conditional variance:

$$\begin{aligned} \text{Var}[dz \mid z] &= \mathbb{E}[dz^2 \mid z] - \{\mathbb{E}[dz \mid z]\}^2 \\ &= \Delta_\epsilon^2 \lambda_\epsilon\, dt \;=\; \varsigma^2\, dt, \end{aligned} \tag{A.34}$$

where $\varsigma^2 \equiv \Delta_\epsilon^2 \lambda_\epsilon$ and the $(dt)^2$ term vanishes for infinitesimal dt. The change in log productivity can therefore be written as:

$$dz = -\theta(z - \bar{z})\, dt + \varsigma\, dx, \qquad dx \equiv \frac{dz - \mathbb{E}[dz \mid z]}{\varsigma}, \tag{A.35}$$

where dx is a zero-mean, unit-variance (per dt) innovation.

A.2 Convergence to the O-U Process

On the discrete grid, dx in (A.35) has jumps and is therefore not Gaussian. Following Shimer (2005), consider refining the grid by replacing the parameters $(\Delta_\epsilon, \lambda_\epsilon, n_\epsilon)$ with $(\Delta_\epsilon\sqrt{\delta},\ \lambda_\epsilon/\delta,\ n_\epsilon/\delta)$ for $\delta > 0$. Under this reparameterization, the conditional mean and variance are invariant:

$$\theta = \frac{\lambda_\epsilon/\delta}{n_\epsilon/\delta} = \frac{\lambda_\epsilon}{n_\epsilon}, \qquad \varsigma^2 = (\Delta_\epsilon\sqrt{\delta})^2 \cdot \frac{\lambda_\epsilon}{\delta} = \Delta_\epsilon^2 \lambda_\epsilon.$$

As $\delta \to 0$, the step size $\Delta_\epsilon\sqrt{\delta} \to 0$, the arrival rate $\lambda_\epsilon/\delta \to \infty$, and the number of shocks within any interval dt grows without bound. Each shock contributes a vanishingly small increment, so dz becomes the sum of a large number of independent, small, mean-zero shocks. Since $\mathbb{E}[dz \mid z]$ and $\mathrm{Var}[dz \mid z]$ remain finite and constant across δ, the Lindeberg condition is satisfied: for every $\xi > 0$, each individual shock size $|dz_i| = \Delta_\epsilon\sqrt{\delta}$ satisfies $|dz_i|/\varsigma \to 0$ as $\delta \to 0$. By the Lindeberg–Feller central limit theorem for triangular arrays, the standardized innovation converges in distribution:

$$dx \xrightarrow{d} \mathcal{N}(0, dt) \quad \text{as } \delta \to 0.$$

Substituting into (A.35) and writing $dx = dB_t$ where B_t is a standard Brownian motion yields the O-U process:

$$dz_t = -\theta(z_t - \bar{z})\, dt + \varsigma\, dB_t, \qquad dB_t \sim \mathcal{N}(0, dt). \tag{A.36}$$

A.3 Discrete-Time Representation

To derive the discrete-time counterpart used for estimation in equation (31), multiply both sides of the O-U process (A.36) by the integrating factor $e^{\theta t}$:

$$e^{\theta t}\,dz_t + \theta e^{\theta t} z_t\,dt = \theta\bar{z}\,e^{\theta t}\,dt + \varsigma\,e^{\theta t}\,dB_t \implies d\Big[e^{\theta t} z_t\Big] = \theta\bar{z}\,e^{\theta t}\,dt + \varsigma\,e^{\theta t}\,dB_t$$

Integrating from t to $t+\Delta t$:

$$e^{\theta(t+\Delta t)} z_{t+\Delta t} - e^{\theta t} z_t = \bar{z}\left[e^{\theta(t+\Delta t)} - e^{\theta t}\right] + \varsigma \int_t^{t+\Delta t} e^{\theta s}\,dB_s$$

Dividing through by $e^{\theta(t+\Delta t)}$:

$$z_{t+\Delta t} = e^{-\theta\Delta t} z_t + \bar{z}\Big(1 - e^{-\theta\Delta t}\Big) + \xi_{t+\Delta t}, \tag{A.37}$$

where the innovation $\xi_{t+\Delta t} \equiv \varsigma\, e^{-\theta(t+\Delta t)} \int_t^{t+\Delta t} e^{\theta s}\,dB_s$ is an Itô integral with respect to Brownian motion. Since $\mathbb{E}\Big[\int_t^{t+\Delta t} e^{\theta s}\,dB_s\Big] = 0$, it follows that $\mathbb{E}[\xi_{t+\Delta t}] = 0$. By the Itô isometry,

$$\text{Var}[\xi_{t+\Delta t}] = \varsigma^2\, e^{-2\theta(t+\Delta t)} \int_t^{t+\Delta t} e^{2\theta s}\,ds = \varsigma^2\, e^{-2\theta\Delta t} \cdot \frac{e^{2\theta\Delta t} - 1}{2\theta} = \frac{\varsigma^2}{2\theta}\Big(1 - e^{-2\theta\Delta t}\Big) \tag{A.38}$$

For annual data ($\Delta t = 1$), equation (A.37) becomes:

$$z_{t+1} = (1 - e^{-\theta})\,\bar{z} + e^{-\theta}\, z_t + \xi_{t+1}, \qquad \xi_{t+1} \sim \mathcal{N}\left(0,\, \frac{\varsigma^2}{2\theta}\left(1 - e^{-2\theta}\right)\right), \tag{A.39}$$

which corresponds to equation (31) in the main text. I estimate this by OLS to recover the persistence $e^{-\hat{\theta}}$ and the innovation variance $\hat{\sigma}^2_\xi = \frac{\hat{\varsigma}^2}{2\hat{\theta}}(1 - e^{-2\hat{\theta}})$. The long-run mean $\bar{z}$ is recovered from the intercept. Given $\hat{\theta}$ and $\hat{\varsigma}^2$, the discrete-grid parameters follow by inverting $\theta = \lambda_\epsilon / n_\epsilon$ and $\varsigma^2 = \Delta_\epsilon^2 \lambda_\epsilon$:

$$\hat{\lambda}_\epsilon = n_\epsilon \hat{\theta}, \qquad \hat{\Delta}_\epsilon = \frac{\hat{\varsigma}}{\sqrt{\hat{\lambda}_\epsilon}}.$$

A.4 Stationary Distribution

Let $p(z,t)$ denote the probability density of log productivity taking value z at time t. The O-U process in (A.36) has an associated Fokker–Planck equation governing the evolution of the density $p(z,t)$. This density evolves over time according to:

$$\frac{\partial p(z,t)}{\partial t} = -\frac{\partial}{\partial z}[-\theta(z-\bar{z})\,p(z,t)] + \frac{\varsigma^2}{2}\frac{\partial^2 p(z,t)}{\partial z^2}$$

In steady state, $\partial p/\partial t = 0$. Let $p(z)$ denote the stationary density. Setting the time derivative to zero and integrating once with respect to z:

$$\theta(z-\bar{z})\,p(z) + \frac{\varsigma^2}{2}\frac{\partial p(z)}{\partial z} = C$$

Since $p(z)$ and $p'(z)$ must vanish as $z \to \pm\infty$, the constant $C = 0$. This yields the first-order ODE:

$$\frac{\partial p(z)}{p(z)} = -\frac{2\theta(z-\bar{z})}{\varsigma^2}\,\partial z$$

Integrating:

$$\log p(z) = -\frac{\theta(z-\bar{z})^2}{\varsigma^2} + D \qquad \Longrightarrow \qquad p(z) = A\exp\left(-\frac{\theta(z-\bar{z})^2}{\varsigma^2}\right),$$

where A is a normalizing constant. Requiring $\int_{-\infty}^{\infty} p(z)\,dz = 1$ and evaluating the Gaussian integral gives $A = \sqrt{\theta/(\pi\varsigma^2)}$. Therefore, the stationary distribution is:

$$p(z) = \frac{1}{\sqrt{2\pi(\varsigma^2/2\theta)}}\exp\left(-\frac{(z-\bar{z})^2}{2(\varsigma^2/2\theta)}\right), \tag{A.40}$$

which is Gaussian with mean $\bar{z}$ and variance $\varsigma^2/(2\theta)$. This is the ergodic distribution $f(\epsilon)$ from which entering plants draw their initial productivity, as described in Section 5.3.

TABLE A.5

PARAMETER VALUES USED FOR MODEL SIMULATIONS

Parameter	Description	Value
r	Rate of time preference	0.05
s	Annual plant exit rate	0.03
α_L	Labor share in production	0.075
α_M	Materials share in production	0.627
μ	Elasticity of substitution across varieties	3.94
ν	Elasticity of substitution, full-time vs contract labor	2
w_f	Annual wage rate per full-time worker	1.25
w_c	Annual wage rate per contract worker	1
I	Aggregate expenditure	14,075,569
$\bar{z}$	Mean of log plant productivity	-2.6
n_ϵ	Number of grid points either side of $\bar{z}$	20
Δ_ϵ	Spacing of the log productivity grid	0.244
λ_ϵ	Annual arrival rate of productivity shocks	2.82
ℓ^*	Regulatory threshold (full-time workers)	100
τ	Wage wedge above the threshold	0.0154
γ	Firing cost wedge above the threshold	2.0713
ϕ_h	Economies of scale in hiring	0.9167
ϕ_f	Economies of scale in firing	1.3466
δ_0	Annual exogenous worker separation rate	0.0807