Preliminary Draft

How Trade and Local Labor Market Institutions Shape Informality: Evidence from Indian Microdata*

Maria Bas†

Maxime Berrou ‡

Pamela Bombarda§

August 26, 2025

Abstract

This paper examines the effects of India's trade liberalization since the 1990s on firm and labor informality. Relying on different sets of data, we apply a difference-in-differences methodology, exploiting exogenous variation in industry-level tariffs to assess how tariff changes have shaped informality. Our findings indicate that import competition leads to higher informality within industries that undergo substantial tariff reductions. However, the presence of lower-cost foreign intermediates promotes formal employment, effectively offsetting the adverse impacts of import competition. Notably, reductions in input tariffs resulted in a net increase in formal employment ranging from 6.5 to 15 percentage points. Crucially, these effects vary across states and depend on the labor market institutions considered. Industries that enjoyed access to cheaper foreign intermediates located in states with pro-worker labor market institutions experienced an increase in formal employment relative to those in pro-employer states.

Keywords: informal and formal employment, trade liberalization, household and firm data.

---

*We would like to thank Johannes Boehm, Mohammad Hoseini, Pierre-Louis Vézina and participants of CITP Nottingham 2024, CY University ThEMA, Paris 1 IELM, ETSG Athens 2024, and King's Economics Brownbag seminar for helpful comments and suggestions. We are also grateful to the Ministry of Statistics and Programme Implementation (MOSPI) staff for their support with data access. All errors are our own.

†University of Paris 1, Centre d'Economie de la Sorbonne (CES). maria.bas@univ-paris1.fr

‡CY Université, THEMA (UMR CNRS 8184), France. maxime.berrou@cyu.fr

§CY Université, THEMA (UMR CNRS 8184), France. pamela.bombarda@cyu.fr

1 Introduction

Over the past few decades, trade liberalization has played a central role in the structural reforms of developing countries. However, these countries are yet facing critical challenges, related to rising inequality, insufficient social security systems, and widespread informal markets. It is essential to examine how these institutional and market characteristics interact with trade reforms, as they can produce effects that significantly diverge from those seen in developed countries (Atkin and Khandelwal, 2020; Atkin et al., 2022; and Atkin and Donaldson, 2022). Developing countries are often characterized by domestic distortions that hinder efficient resource allocation and economic growth. Among these distortions, such as rigid labor market regulations, limited enforcement capacity or credit market imperfections, some are closely tied to a defining characteristic of these economies: the widespread prevalence of informality. Indeed, informal firms in developing countries contribute to nearly half of economic activity (La Porta and Shleifer, 2014), and the share of employment in the informal sector has been steadily increasing over recent decades. In India, for instance, 90% of the workforce lacks access to social insurance, and 85% of employment in the manufacturing sector is informal (Mehrotra, 2019). Despite rapid economic growth following the structural reforms of the early 1990s, India remains a striking example of the persistence of informality.

The aim of this study is to analyze the impact of India's trade liberalization on a critical aspect of the labor market: the informal sector. Focusing on the period of major trade reforms between 1990 and 2010, we disentangle the distinct effects of trade liberalization—such as increased import competition, access to cheaper foreign intermediates, and expanded market access—on informality. Furthermore, recognizing that trade reforms may have varying effects depending on local labor institutions, this study also seeks to explore the interaction between trade liberalization and labor market policies. The regional variation in labor market institutions offers a fertile ground for empirical analysis, enriching our understanding of the relationship between trade policy and informality. Our contribution to the literature is threefold. First, we provide a detailed account of informality patterns in India. By leveraging both firm-level and individual-level data, we document that: (i) the informal sector constitutes a significant portion of the economy, accounting for approximately 15% of total economic activity, (ii) while the share of informal employment began to decline following the trade reforms of 1991, it remained notably high throughout the period under study, and (iii) changes in overall formality are primarily driven by shifts in formality within industries, rather than by reallocation across sectors. Second, by leveraging multiple datasets, we provide causal evidence on the different channels of trade reforms. Third, we investigate how ow differences in local labor market institutions, such as state-level labor regulations, union strength, enforcement of labor regulation, and changes in employment protection legislation, mediate the impact of trade liberalization on informality.

In the decades after independence, India maintained a highly regulated economy with minimal external trade. While initial steps toward liberalization began in 1976, they were limited and sector-specific. A major shift began in the 1990s when the Indian government relaxed its control over investment and import licensing, paving the way for substantial trade liberalization policies that continued into the early 2000s. This sequence of reforms, well-documented, and largely exogenous in nature, makes India a particularly compelling setting for studying the impact of trade on informality. Moreover, India's institutional diversity, particularly in local labor market regulations, allows for assessing how trade policies affect regions differently. The varying enforcement capabilities and regulatory frameworks across Indian states may contribute to differing levels of informality.

To evaluate the impact of these reforms on informality, we draw on four datasets that provide a comprehensive picture of the Indian economy: (i) formal manufacturing firm-level data from the Annual Survey of Industries (ASI), (ii) informal manufacturing firm-level data from the Unorganized Manufacturing Surveys (UMES), (iii) worker-level data from the Employment Unemployment Survey (NSSEU), and (iv) economic census data, which encompasses the full range of economic units in India. These datasets help overcome common challenges faced by other studies on informality in low-income settings, particularly the lack of detailed information on informal firms. Furthermore, they allow us to validate our findings by using alternative measures of informality. Following the literature (Maloney, 1999, Bosch and Maloney, 2010, Bosch et al., 2012, Hasan et al., 2007 among others), we classify informal firms as those registered in UMES, while formal firms are those in ASI. In contrast, using NSSEU data, informal employees are defined as those without access to retirement funds or social security benefits, which are legally mandated for formal labor contracts. Then, to capture the role of trade policies, we combine these data with tariffs data.

Our analysis focuses on two major Indian trading partners: the European Union (EU25) and the United States, which together accounted for 40% to 47% of India's total trade in 2000.[1] The period under analysis is marked by significant tariff reductions, with the average applied tariff rates falling from 65% to 15%. This decline reflects a concerted effort to simplify and rationalize India's tariff structure. To capture the different effects of trade liberalization, we use four-digit product-level import and export tariff data from the World Integrated Trade Solution (WITS).[2] We then construct a measure of input tariffs by combining these import tariffs with Indian Input-Output tables. A key econometric concern when analyzing the effects of trade policy is the potential endogeneity of tariff changes, as trade policy may be shaped by underlying economic conditions. In the case of India, however, tariff reductions during the period 1990–2000 can be viewed as largely exogenous. The initial wave of liberalization in 1991 was triggered by a severe balance of payments crisis, which necessitated intervention from the International Monetary Fund (IMF) and imposed structural adjustment requirements. In contrast, trade reforms in the 2000s

[1] As a robustness check, we also extended the set of India's main trading partners to China, Japan and Singapore.
[2] Throughout the analysis, we also refer to import tariffs as output tariffs.

were not driven by an immediate economic crisis, which in theory could have allowed the Indian government more discretion in setting trade policy. Nevertheless, Batra (2022) argues that tariff adjustments during this later period were still heavily shaped by international commitments and ongoing liberalization agendas, limiting the scope for policy manipulation. Consistent with this interpretation, our data reveal no systematic relationship between initial four-digit tariff levels and subsequent tariff changes across the two decades studied, supporting the view that tariff reductions were not endogenously determined by sector-specific conditions.

By leveraging this plausibly exogenous variation in tariffs across industries over time, we first investigate the relationship between trade liberalization and changes in firms' demand for formal labor using aggregated firm-level data at the industry and state level. Our identification strategy takes advantage of the exogenous and heterogeneous reduction in tariffs across 4-digit industries over two periods, as well as the variation in the share of formal firms across industries and states over time. Our estimations account for three key channels through which international trade can influence informality: market access, import competition, and access to foreign technology. A potential concern for identification is that industry-level, time-varying factors unrelated to trade policy may have influenced firms' decisions to hire formal versus informal labor. To account for this, we include industry-specific linear trends, constructed by interacting the total employment level in each four-digit industry in the initial year with year fixed effects. To control for time varying unobservable shocks related to domestic policies at the state level (like VAT implementation in the 2000s), we also include state-year fixed effect. Our findings indicate that reductions in output tariffs, which reflect increased import competition, led to a decline in the share of formal employment within industries. Sectors experiencing larger output tariff cuts saw a more pronounced reduction in formal employment relative to those facing smaller cuts. In contrast, reductions in input tariffs, capturing improved access to foreign intermediate goods, had the opposite effect: they increased the share of formal employment, with stronger gains in industries more exposed to input tariff liberalization. Importantly, the negative impact of increased import competition on formality is fully offset by the positive effect of input tariff reductions. Quantitatively, the combined effect of trade liberalization led to a net increase in the formal employment share, ranging from 6.5 to 15 percentage points, depending on the decade under consideration. To reinforce our industry-by-state level findings, we present an extension of the analysis to the individual level using NSSEU data, as documented in the online appendix.

To examine whether the effects of trade liberalization on the reallocation of firms and workers between the formal and informal sectors vary depending on the local institutional environment, we consider four distinct dimensions of labor market institutions. First, we draw on a state-level index of labor regulations that captures variation in hiring and firing costs across Indian states. This index enables us to classify states as either pro-worker or pro-employer, providing a useful proxy for the rigidity or flexibility of local labor laws. Second, we assess the role of labor union presence by distinguishing between states with high and low levels of unionization, measured by the share of unionized workers in our working population. Third, we examine enforcement capacity by calculating the distance from each district to the nearest regional labor office, which is responsible for monitoring and enforcing labor regulations.[3] Finally, we exploit a 2003 policy reform in Andhra Pradesh that increased penalties for formal firms avoiding employment protection provisions by hiring informal workers. Our main findings show that industries benefiting from access to cheaper foreign intermediates experienced a greater increase in formal employment in states with pro-worker labor institutions compared to those in pro-employer states. This suggests that the gains from input trade liberalization were more likely to translate into formal job creation where institutional conditions favored worker protections.

To validate our findings and address potential limitations in our analysis, we conduct a series of robustness checks. First, we account for the influence of other contemporaneous domestic reforms, such as the gradual reduction of restrictions on foreign direct investment (FDI) and the dismantling of the License Raj during the 1990s, which could have independently influenced firms' decisions regarding formal versus informal employment. Second, to account for potential measurement errors in our definition of informality, we use alternative definitions of formal employment. Finally, we relax the assumption of perfect labor mobility by adopting a local labor market approach, constructing a district-level weighted average of tariffs to capture localized exposure to trade liberalization. Across all specifications, our results remain robust and consistent with the benchmark estimates.

Finally, we examine the extent to which the reallocation of workers toward the formal sector, driven by input trade liberalization, contributes to aggregate productivity gains.[4] Following McCaig and Pavcnik (2018), we start by computing the labor productivity gap between the formal and the informal sector using the ratio of the average revenue product of labor. We estimate the gap to be 36.6 in the 1990s and 17.2 in the 2000s. After accounting for heterogeneity in workers' human capital, measurement error and differences in output-labor elasticity across sectors, the gap narrows to 12.3 and 5.81. These results are validated by a wage-based approach that finds similar results (13.9 and 5.43). We then estimate that 9.9% of workers were reallocated to the formal sector during the first decade of liberalization, with this figure dropping to 4.7% in the subsequent decade. Our results are in line with other works in India, and suggest that the adjusted labor productivity gap there is particularly large compared to other low-income countries, even after those adjustments (McCaig and Pavcnik, 2018; La Porta and Shleifer, 2014; Nataraj, 2011). This indicates substantial potential aggregate productivity gains following liberalization.

Related Literature. We first contribute to a literature seeking to establish a causal link between international trade and informality. Those studies have found empirical evidence of three channels through which trade reforms can affect the reallocation of firms and workers between the formal

[3]Greater distance is interpreted as indicative of weaker enforcement capacity.

[4]Recent contributions include Banerjee and Duflo (2005), Restuccia and Rogerson (2008), La Porta and Shleifer (2008, 2014), Hsieh and Klenow (2009a), McMillan and Rodrik (2011), McCaig and Pavcnik (2018), among others.

and informal sectors. The first channel involves import competition, which increases informality. An early contribution to that literature by Goldberg and Pavcnik, 2003 provides evidence of a positive relationship between trade reforms and informality in Colombia. Ben Yahmed and Bombarda (2020) find similar results in Mexico. Dix-Carneiro and Kovak (2019), using Brazilian data, adopt a local labor market approach. They find that formally employed workers located in regions more exposed to import competition experienced a reduction in the probability of being employed in a formal job relative to those in less exposed regions. The second channel between trade and informality is access to higher-quality foreign inputs, which decreases informality. Bas and Bombarda (2023) show that input-trade liberalization is the main channel that affects the reallocation of workers from informal to formal manufacturing employment in Mexico after the NAFTA agreement. The last channel is the expansion of market access and foreign demand shocks. McCaig and Pavcnik (2018) show that the U.S.-Vietnam bilateral trade agreement, implemented in 2001, increased export opportunities in Vietnam, which reallocated labor away from household work. We add to that literature by studying India, where studies on informality are scarce, and by providing evidence that those channels are at play in India following unilateral trade liberalization.

This study also relates to works seeking to evaluate how domestic regulations and institutions may impede development. Low-income economies are characterized by pervasive domestic distortions. In the Indian context, a substantial body of evidence suggests that labor regulations have hindered the growth of the formal sector and limited its capacity to effectively respond to shocks (Besley and Burgess, 2004; Hasan et al., 2007; Aghion et al., 2008; Panagariya, 2007, among others). Additionally, the causes of informality have been debated in the literature, which highlights the central role of regulations. Some authors argue that inefficient and costly regulations prevents firms from expanding and becoming formal (De Soto, 1989). Others believe that informal firms opt for not registering and paying regulatory costs and that they compete unfairly with formal firms (Levy, 2008). Recent research has sought to investigate how the effects of trade reforms depends on the size of those regulations (Atkin and Donaldson, 2022). Our contribution to this literature is to examine how the size of the informal sector responds to interactions between trade liberalization and labor market institutions specific to the Indian context, including hiring and firing costs, unionization, labor law enforcement, and employment protection legislation.

This paper is organized as follows. Section 2 describes the main channels through which trade liberalization might affect informality. Sections 3 and 4 presents the institutional setting of the study and the main data sources used. Sections 5 describes the identification strategy, and section 6 describes the main results. Section A.2 further explores the role of labor market regulations , and section 9 develops a local labor market approach. Section 10 discusses the labor productivity gap and quantifies reallocations between sectors. Section 11 concludes.

2 Theoretical motivation

The goal of this section is to present the conceptual framework to guide our empirical analysis. There is a fast-growing theoretical literature linking trade and informality. Those models have helped rationalize the three empirical links between trade and informality described above by defining mechanisms through which trade impacts the labor demand of formal and informal firms. Most of those works share the same characteristics: they are general equilibrium trade models with heterogeneous firms, featuring intra-industry trade à la Melitz (2003) and two coexisting sectors (formal and informal). The model by Dix-Carneiro et al. (2024) is particularly rich. They develop a general equilibrium trade model of heterogeneous firms with search and matching frictions based on Cosar et al. (2016), Meghir et al. (2015) and Ulyssea (2018). Their model features the three channels linking trade and informality, and several types of domestic labor market institutions. The remainder of the discussion in this section presents the main theoretical insights from Dix-Carneiro et al. (2024) as well as other models that are relevant to our study.

We follow the current literature, such as Ulyssea (2018) and Dix-Carneiro et al. (2024), and posit that informality stems from the existence of regulations imposed on firms, such as taxes and labor laws. The government enforces those regulations, but only imperfectly, leading to incentives for some firms to evade them and operate informally. Detection by the government is possible but not certain; if it happens, it is costly for informal firms (in the form of closures or fines). For firms choosing under which status to operate, this leads to a trade-off between paying costly regulations (formality) and risking detection (informality). Larger firms have a greater probability of detection than smaller firms, and informality is therefore a size-dependent distortions.

We now turn to the three types of mechanisms linking trade and informality. First, import competition reduces domestic market shares of formal firms competing with foreign ones, which leads to the exit of the least efficient formal firms, thereby increasing the share of informal firms. In this setting, the decline in formal employment leads to both and increase in unemployment and in informal employment.[5]

Second, reductions in international trade barriers may reduce cost of intermediates good in two distinct way: by allowing formal firms to source the same inputs from abroad at a lower costs (Dix-Carneiro et al., 2024), or to have access to a more productive technology (Bas and Bombarda, 2023). In the theoretical setting proposed by Dix-Carneiro et al. (2024), a reduction in trade barriers affecting intermediate inputs has ambiguous effects on informality. On the one hand, access to cheaper inputs tends to make all firms more productive creating incentives for the most efficient informal firms to become formal. On the other hand, input tariff cuts by reducing marginal costs increase profitability and can also lead to entry of low productivity firms in the informal sector. Moreover, in this framework, the lower cost of inputs can promote exports by increasing

[5]This channel was already present in the models developed by Kovak (2013) and (Dix-Carneiro and Kovak, 2017, 2019), which had showed that the presence of a large informal sector acted as a buffer to displaced workers due to import competition.

firms' profitability generating incentives for productive informal firms to grow and formalize. The net effect of trade on informality will depend on the values of the parameters. Another factor potentially at play in this channel is worker heterogeneity and the potential complementarity between the worker characteristics and intermediate inputs. Bas and Bombarda (2023) develop a theoretical framework based on the extension of the Melitz (2003) proposed by Ulyssea (2018) in order to highlight the skilled-biased foreign input mechanism through which trade can affect informality. Their model predicts that input-tariff cuts reduce the relative unit costs of formal firms vis-à-vis informal ones. This reduces the cutoff threshold required to become a formal firm, and thereby increases formal employment. In their framework, where foreign inputs are skilled-biased, the reallocation effect of input-trade liberalization will be more pronounced among skilled workers.

Third, expansion of market access abroad provide new opportunities for exporting firms in the domestic formal sector. Those firms that export are able to expand the foreign demand for their products, and therefore they need to increase their (formal) labor demand. This might induce a reallocation of workers from the informal to the formal sector. At the same time, the increase in export profitability due to trade variable costs reductions might create incentives for productive informal firms to become formal (Dix-Carneiro et al., 2024).

This paper also investigates if these channels vary depending on the type of domestic labor market institutions. In the setting developed by Dix-Carneiro et al. (2024), labor market institutions are a crucial determinant of the size of the informal sector. By influencing the cost of operating formally, these institutions can affect how formal firms adjust their labor demand in response to trade liberalization, with important implications for informality.

3 Indian Trade Reforms

3.1 Twenty years of trade reforms

In the decades following independence, the Indian economy was highly regulated. Policymakers sought to attain self-sufficiency and to minimize dependence on external trade. To reach this goal, exchanges with the rest of the world were tightly controlled. The primary tool to limit import flows was restrictive licensing requirements: only goods listed on the Open General License (OGL)—a positive list—were allowed for imports, and importers were required to prove that the goods would be used by them directly, rather than being resold (Panagariya, 2004). Additionally, the average tariff rate on imports was prohibitively high, averaging 87%, and non-tariff barriers accounted for approximately 90% of the value-added in manufacturing (Hasan et al., 2007). Despite some limited reforms, this system remained in place until 1991.[6]

[6]Though early steps toward liberalization began in 1976 and continued through the 1980s, these reforms were gradual and sector-specific, failing to dismantle the broader system of controls and restrictions. As Panagariya

The era of economic liberalization started suddenly in 1991. Sweeping reforms were initiated against the backdrop of a severe balance of payments crisis, with foreign exchange reserves plummeting to levels barely sufficient to cover two weeks of imports. The twin deficits—fiscal and current account—had reached unsustainable levels, exacerbated by the external shocks of the Gulf War and the collapse of the Soviet Union, India's largest trading partner at the time (Topalova and Khandelwal, 2011). To deal with those problems, India requested financial support from the IMF, which was granted conditionally on macroeconomic stabilization policies and structural reforms. These reforms primarily targeted key structural barriers to economic growth, focusing on industrial licensing, import restrictions, the financial sector, the tax system, and trade policy. These measures aimed at dismantling the long-standing constraints of the license raj, reducing bureaucratic hurdles, and liberalizing trade to align India with global economic practices.

The 1991 reforms marked a shift in India's economic model, moving from an inward-focused to a globally integrated approach. The import licensing system was abolished, allowing for easier trade, with tariffs becoming the primary mechanism for regulating global trade participation (Hasan et al., 2007). Over the next two decades, import tariffs were significantly reduced. By 1992, average import tariffs were around 65%. Over the next decade, reductions were made as part of India's liberalization, stabilizing at 35% by the late 1990s. Tariff liberalization accelerated in the 2000s, particularly around 2004-2005, driven by India's WTO commitments. During this period, tariffs fell from 35% to 15%, and in 2007-2008, they dropped further to approximately 12%. After the 2008 financial crisis, the government took a more cautious approach, balancing liberalization with domestic industry protection, and tariffs remained stable for the rest of the decade. In contrast, export tariffs remained relatively unchanged, as India's developed trade partners had already little trade barriers on Indian goods.[7]

Tariff reductions were accompanied by a simplification of India's tariff schedule, cutting the number of slabs and peak duties. As shown in Table A.1, the standard deviation of output tariffs dropped from 35 in 1990 to 20 in 2010, and the maximum tariff rate fell from 241% to 150%.[8] Tariff liberalization, spanning two decades, targeted different sectors. In the early 1990s, cuts focused on capital goods and essential inputs, allowing Indian firms to source new foreign inputs that were previously unavailable (Goldberg et al., 2010; Topalova and Khandelwal, 2011). On the other hand, consumer goods remained regulated until around 1997-1998, after which they too were liberalized. Figure A.3 illustrates this by plotting average tariff reductions across two-digit industries. For example, the "office, accounting, and computing machinery" sector experienced immediate tariff cuts in 1991, while large domestic industries like textiles, initially more protected, were liberalized in later phases.

(2004) notes, significant liberalization only began in the 1990s.

[7]Figure A.1 in Appendix A.1.

[8]India did not target specific partners for liberalization or enter major trade agreements during this period, limiting partner-specific reforms. As Batra (2022) writes, "India's FTAs are limited to shallow integration provisions."

3.2 Data on Trade Barriers

The tariff data used in our analysis is sourced from the World Integrated Trade Solution (WITS). We focus on the EU25 and the United States because they are India's two main trading partners, accounting for 40% and 47% of India's import and export shares in 2000, respectively.[9] Among the three types of tariffs reported by WITS, we use the effectively applied tariff rates (AHS), which reflect the actual tariffs imposed after accounting for preferential trade agreements. These rates are available at the ISIC Rev. 3 level, which corresponds to the 4-digit NIC-1998 classification used in India.

Import tariffs at the product level (also referred to as output tariffs) are calculated as a simple average across India's two main trading partners, EU25 and the United States. In contrast, export tariffs on Indian products are constructed as a weighted average of the tariffs levied by these partners. The weights reflect each partner's share in India's total exports for a given industry and are fixed at the beginning of the period. This approach ensures that changes in export tariffs are not driven by potentially endogenous shifts in trading partners' preferences or demand over time. It is also important to note that during this period, India signed very few trade agreements, and those it did enter were mostly with smaller economies. These agreements were generally limited in scope and lacked the depth and comprehensive coverage typical of broader free trade agreements.[10] India's relatively low engagement in preferential trade agreements minimizes measurement errors highlighted by Teti (2023).

To build our measure of input tariff, we follow the standard practice in the literature and compute the input tariff faced by final good producer in industry j as a weighted average of the output tariffs, τ_{kt}, for every period. This yields the following measure:

$$\text{input tariff}_{jt} = \sum_k w_{kj}^{1998} \times \tau_{kt}, \quad \text{where} \quad w_{kj}^{1998} = \frac{\text{input}_{kj}^{1998}}{\sum_k \text{input}_{kj}^{1998}}. \tag{1}$$

In line with Amiti and Konings (2021), the weights, w_{kj}^{1998}, represent the cost share of industry k in the production of a good in industry j. These weights are computed using the Indian Input-Output table for 1998 for a total of 51 manufacturing input sectors.[11]

[9] Data from the Observatory of Economic Complexity.

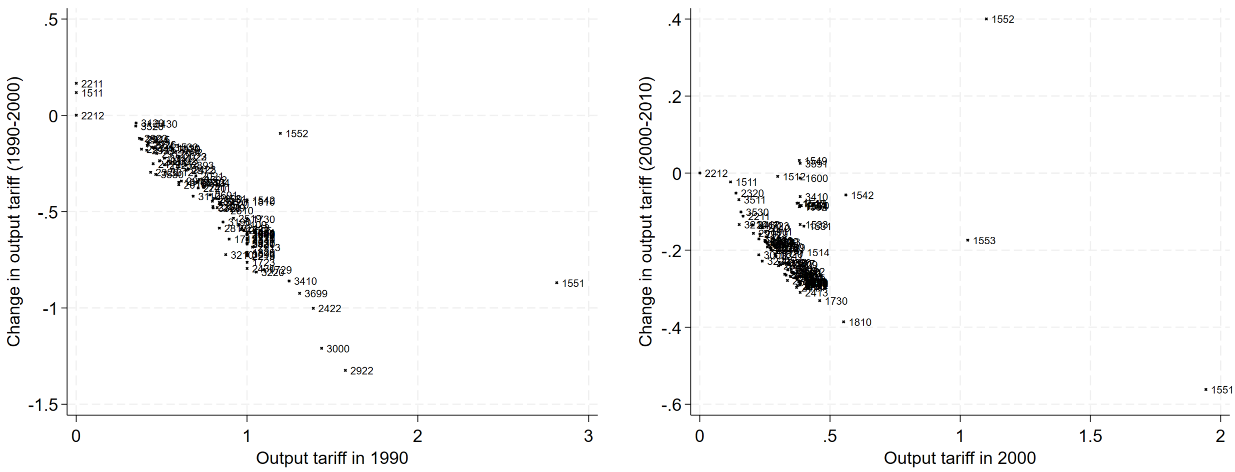
[10] India's bilateral trade agreements during this period include agreements with Sri Lanka (effective in 2000), Singapore (2005), Thailand (2006), Chile (2007), and South Korea (2010). Two multilateral agreements were also implemented: the South Asia Free Trade Agreement in 2006 and the India-Mercosur Preferential Trade Agreement in 2009.

[11] The correlation between the output and the input tariffs is 0.61, between the output and the export tariff is 0.10 and between the input and the export tariffs is -0.03.

3.3 Exogeneity of Trade Policy

Tariff liberalization in India can be considered largely exogenous, particularly during the period 1990-2000. This exogeneity can be understood by examining the institutional and economic context of the two key liberalization waves in 1991 and 2004. The 1991 liberalization was triggered by a severe balance of payments crisis, which required India to seek assistance from the the IMF. As part of the IMF's conditionality for financial support, trade policy reforms were externally imposed, rather than being the result of domestic policy deliberation. In contrast, the 2000s represent a period when trade reforms were not prompted by an immediate economic crisis, theoretically allowing for greater discretion by the Indian government in shaping trade policy. Nonetheless, even during this period, policy adjustments were largely constrained by international commitments. Specifically, by 2005, India was implementing its commitments under the Uruguay Round agreements, which required reductions in bound tariff rates. Moreover, as noted in the WTO's 2007 Trade Policy Review (WTO, 2007), the government had publicly committed to aligning India's applied tariff rates on non-agricultural goods with those of ASEAN economies.

Figure 1: Output tariff changes and initial tariff level, by decade



(a) 1990-1999 (b) 2000-2010

Source: Authors' calculation based on WITS. 1511 Manufacture of meat products; 1520 Manufacture of dairy products; 1542 Manufacture of sugar; 2211 and 2212 Publishing of books, brochures, musical books, newspapers, and other publications; 1551, 1552, and 1553 Manufacture of alcoholic beverages.

To validate these arguments, we analyze Indian output tariff data with respect to EU and US trading partners to determine whether a systematic relationship exists between initial 4-digit tariff levels and subsequent tariff changes over the two decades under study. Panel (a) in Figure 1 reveals a negative relationship for the 1990s: industries with higher tariff levels in 1990 experienced larger tariff reductions. This pattern aligns with India's deliberate strategy to liberalize heavily protected sectors to enhance competitiveness. The only notable outliers are industries related to alcoholic beverages, which, while still liberalized, underwent more modest reductions. Panel (b) reveals a similar trend for the 2000s, although the relationship is less precise. Higher initial tariffs continued to be associated with larger reductions; however, some industries maintained relatively elevated tariff levels by the end of the decade. A closer analysis highlights that many of the industries with minimal tariff cuts, or even increased protection, were concentrated in food products and beverages, reflecting India's historically protectionist stance toward agriculture and food-related industries. Overall, Figure 1 suggests that tariff reductions were widespread across the Indian economy, with only a few exceptions concentrated in specific sectors. This suggests that endogeneity concerns are largely limited to these protected industries, while the broader pattern of liberalization reflects India's commitment to reducing trade barriers and integrating into the global economy. Figure A.4 in Appendix A further confirms that a similar pattern holds for input tariffs.

Table 1: Decline in trade barriers and pre-reform industry characteristics

	(1) Industry size$_j$	(2) Formal employment share$_j$	(3) Capital-labor ratio$_j$	(4) Firms 3+ years share$_j$	(5) ln Output$_j$
Panel A: First wave of trade liberalization (1990s)					
$\Delta\tau_{O,j}$	0.009	-0.078	-0.077	0.054	0.183
	(0.007)	(0.122)	(0.088)	(0.056)	(0.623)
$\Delta\tau_{I,j}$	0.028	0.292	-0.122	0.037	2.597
	(0.023)	(0.229)	(0.185)	(0.128)	(1.658)
$\Delta\tau_{X,j}$	0.006	0.862**	0.219	-0.082	3.173
	(0.024)	(0.361)	(0.180)	(0.307)	(3.285)
2-industry FE	Yes	Yes	Yes	Yes	Yes
Observations	74	91	91	91	91
R-squared	0.503	0.638	0.344	0.492	0.347
Panel B: Second wave of trade liberalization (2000s)					
$\Delta\tau_{O,j}$	0.005	0.621**	0.488	-0.035	-1.036
	(0.011)	(0.262)	(0.394)	(0.113)	(2.360)
$\Delta\tau_{I,j}$	0.041*	0.279	-0.031	0.180	2.096
	(0.024)	(0.385)	(0.458)	(0.222)	(3.714)
$\Delta\tau_{X,j}$	-0.077	-0.915	0.374	0.193	-2.219
	(0.054)	(1.350)	(1.616)	(0.361)	(10.359)
2-industry FE	Yes	Yes	Yes	Yes	Yes
Observations	84	105	105	105	105
R-squared	0.491	0.634	0.401	0.317	0.385

Notes: This table presents regressions of changes in 4-digit industry j tariff variations on industry j initial characteristics. Panel A uses 1990 characteristics and 1990–1999 tariff changes; Panel B uses 2000 characteristics and 2000–2010 changes. Industry size$_j$ is employment in industry j over total manufacturing employment. Formal employment share$_j$ is ASI employment over ASI+UMES employment. Firms 3+ years share$_j$ is the share of firms older than 3 years. Output is measured as log gross sales. All regressions are weighted by the square root of industry employment. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A related concern is that industries with the highest initial tariff levels, which experienced the greatest tariff reductions, may systematically differ in their underlying characteristics from industries with lower initial tariffs. For example, larger or more formal industries could have been more heavily protected from import competition. This could complicate the causal interpretation of our regression estimates, as trade policy may have been influenced by unobserved factors such as the size of the informal sector or political considerations, raising the possibility of reverse causality. To address this concern, we follow Topalova and Khandelwal (2011) and examine whether initial industry characteristics are systematically related to tariff changes. For each decade, we regress the initial industry characteristics on the tariff change, as presented in Table 1. These characteristics include: (i) the relative size of the industry (measured as industry employment over total manufacturing employment), (ii) the share of employment in formal firms, (iii) the capital-labor ratio, (iv) the share of firms older than three years, and (v) output, measured by gross sales value. In the first decade of liberalization, we find no significant relationship between these characteristics and subsequent changes in either output or input tariffs. In the second decade, only two significant correlations emerge: industry size for input tariffs, and the share of formal employment for output tariffs. Overall, these results suggest that tariff changes were largely exogenous with respect to initial industry characteristics, particularly during the 1990s.

4 Indian Labor Market and Informality

4.1 Labor market institutions

Labor costs in India vary significantly across regions, reflecting the heterogeneity of local labor market institutions. This section aims to document the main features that contribute to this variation.[12]

Besley and Burgess (2004) provide the first attempt to measure geographic heterogeneity in labor market regulations in India. Their approach exploits the fact that labor regulations are governed by a nationwide piece of legislation, the Industrial Disputes Act (IDA). Since 1947, it has served as the key framework aimed at protecting workers in the organized sector from exploitation (Besley and Burgess, 2004). However, states possess the authority to amend central labor legislation and are responsible for its enforcement. Over times, states have deviated from the central framework by passing additional laws, leading to significant regulatory divergence across states. Besley and Burgess (2004) thus compile and categorize each amendment made by states over the period 1947-1992. Each amendment is either coded as neutral (0), pro-worker (+1), or pro-employer (-1). The amendments are then cumulated for each state. In short, Besley and Burgess (2004) define a measure of labor regulations which varies et the state-time level.

A vast literature has intended to refine the classification proposed by Besley and Burgess (2004).[13] To account for those critiques, we follow Gupta et al. (2009), Chaurey (2015) and

[12] The discussion is limited to aspects of labor market regulation that we believe are most relevant in light of the framework described in section 2.

[13] Subsequent works focusing on Indian labor market regulations have debated whether the amendments that Besley and Burgess (2004) use are truly capturing meaningful changes in states' regulations. They have argued

Chakraborty et al. (2024), among others, and adopt a time-invariant classification of state regulations.[14] More specifically, we rely on the classification used in Chakraborty et al. (2024), which categorizes the states of Gujarat, Maharashtra, Orissa, West Bengal as "pro-worker"; Andhra Pradesh, Karnataka, Rajasthan, Tamil Nadu, and Uttar Pradesh as "pro-employer"; and Assam, Bihar, Haryana, Jammu and Kashmir, Punjab, Kerala and Madhya Pradesh as "neutral". Since we are mostly interested in the effect of pro-worker regulations relative to other types of regulations, we further classify states in two categories: those with pro-worker labor market regulations (subsequently referred to as "pro-worker") and those with either neutral or pro-employer regulations ("pro-employer").

Labor laws do not always translate into higher labor costs in developing economies (Almeida and Carneiro (2012); Ulyssea (2020)). This may be the case for two reasons. First, the state may be unable to enforce existing regulations (for instance, due to a lack of inspectors in a given location). Second, the state's regulations may be designed in such a way that permits firms to legally exploit loopholes. This implies that two firms operating under similar regulatory environment may not have the same labor costs if one firm respects those regulations while the other is able to avoid them. In India, public authorities have taken measures to ensure labor regulations play their role, and set up a Chief Labor Commissioner in 1945. To keep up with the country's industrial development and the complex regulatory system, has set up a network of regional offices was set up over the years, most recently in 2005 (see map A.6). The role of those offices is to ensure enforcement of labor laws. Some states have also punctually taken steps to make it harder for firms to avoid regulations. For instance, the state of Andhra Pradesh introduced a reform in 2003 to limit the hiring of contract workers, which provided a legal way for formal firms to hire workers who are in effect informal.

Lastly, Indian workers have organized to increase their bargaining power relative to their employers and to ensure that they respect regulations. In most cases, negotiations between unions and firms are carried out at a fairly decentralized level. The Trade Unions Act (1926) allows any group of seven workers within a firm to form a union, allowing negotiations to take place directly within the firm (Ahsan et al., 2017). That institutional setting has lead to the creation of large number of unions in India, potentially more than 100,000 (Bamber et al., 2011). This institutional setting has two consequences in light of our study. First, it has induced important variation in unionization rates across the country. Map A.7 shows the average share of workers within a district reporting belonging to a union. Unionization shares range from less than 5% to more than 25%,

that some states are classified on the basis of just one or two amendments, or that the categorization of those amendments as neutral, pro-worker or pro-employer is sometimes not obvious. As an illustration, Bhattacharjea (2006) writes that "Gujarat is designated as pro-worker (a strange characterization, to which I return below) because of a solitary amendment which it passed in 1973, allowing for a penalty of Rs 50 a day on employers for not nominating representatives to firm-level joint management councils."

[14]Adopting a time-invariant classification has several advantages. First, the Besley and Burgess (2004) index stops in 1992 and thus does not cover the period of our analysis. Second, it reduces the risk of measurement error due to authors' potentially arbitrary classification of state amendments Gupta et al. (2009).

with the South and the North-East the most unionized areas. Second, the Trade Unions Act defines the framework through which bargaining can take place between workers and formal firms. In practice, informal workers do not benefit from this text, which does not define how unionization can operate in the informal economy (Routh, 2015). For that reason, informal workers are rarely unionized and it is difficult for them to organize to improve their bargaining position relative to their employers.

4.2 Data on firms and workers

We now present the datasets on firms and workers. Table A.3 in Appendix A provides a detailed description of the different datasets used. Data on formal firms is sourced from the *Annual Survey of Industries* (ASI).[15] The ASI is a repeated cross-section that provides comprehensive coverage of manufacturing establishments registered under the Industrial Disputes Act.[16] Specifically, it includes establishments with 10 or more workers that use power electricity, as well as establishments with at least 20 workers that do not use power electricity. Then, we use data from the *Unorganized Manufacturing Surveys* (UMES), which focus exclusively on unorganized manufacturing establishments. UMES is designed to represent all manufacturing enterprises not covered by the ASI.[17] Similar to the ASI, the UMES provides establishment-level information on key variables such as labor, wages, fixed assets, energy use, and sales. Following the methodology outlined in previous studies (Hoseini and Briand, 2020 and Chakraborty et al., 2021 among others), we aggregate both ASI and UMES datasets to create a panel dataset at the state-industry level.[18] We restrict the analysis to years for which there is both ASI and UMES information: 1990, 1994, 2000, 2005 and 2010.

Next, we consider additional data source that provides worker-level information and enables us to distinguish between formal and informal individuals. Specifically, we use the National Sample Survey Organisation's *Employment and Unemployment Surveys* (NSSEU). This survey offers detailed information on individuals' work status, as well as their personal and household characteristics. The first NSSEU survey was conducted in 1983, with subsequent rounds carried out periodically in the following years. We use rounds conducted in 1983, 1993-1994, 1999–2000, 2004–2005, and 2009–2010. Since the variables included in the NSSEU vary between rounds, we restrict our analysis to variables consistently reported across all rounds. These nationwide surveys

[15] Another commonly used firm-level dataset is *PROWESS*, but it is not suitable for our research purposes due to its lack of representativeness and non-systematic information on labor.

[16] A panel version is available for large firms starting in 1998, but our analysis does not require following firms over time.

[17] The unorganised manufacturing sector has roughly one-third share in the total contribution by the manufacturing sector in the GDP, see Appendix A. More information is available in MOSPI (2012).

[18] To proceed, we make use of the sampling weights provided by both surveys. Those weights can be used to recover information about the universe of manufacturing firms. For each sampled establishment that appears in the surveys, they indicate how many other establishments possess similar characteristics. More information is provided in Appendix A.

sample approximately 70,000 rural and 45,000 urban households per round. As NSSEU data do not track individuals over time, we employ methods to estimate individual fixed effects using independent repeated cross-sectional data. More information is provided in appendix A.

Finally, we incorporate data from the 1990 and 1998 Economic Censuses of India to build the employment shares for the district level tariff used in section 9, and to create additional control variables.[19]

4.3 Measuring Informality

By its very nature, the informal economy is challenging to study, particularly in the context of developing countries. The term "informal economy" encompasses all economic activities undertaken by economic units that are either not regulated by formal legal frameworks, or are inadequately protected under such arrangements. Consequently, empirical studies on informality encounter numerous challenges. For example, economic activities conducted beyond the state's oversight are often excluded from national statistics, making any comprehensive analysis difficult. Fortunately, Indian statistical authorities design their firm and labor force surveys in a way that enables the identification of informal activities. We now detail how we use the data to assign a formality status to both firms and workers, and what our preferred and alternative measures are.

First, we identify firm informality. We follow the literature and assign it based on the firm's registration status (Ulyssea, 2018). Specifically, we rely on the two types of firm-level data described above: we classify all firms in the ASI database as formal and those in the UMES database as informal.[20] This choice is justified by the fact that establishments included in the ASI are registered under the Factories Act (FA), which was enacted in 1948. The FA serves as a key legislative framework for regulating the safety, health, and welfare of employees, with inspections by government officials ensuring compliance. As the FA primarily applies to formal establishments, firms listed in the ASI database can generally be classified as formal entities.[21] Differently, firms in UMES are not registered with FA and thus do not bear the costs associated with labor regulations and inspections; as such, we classify them as informal. This approach aligns with previous studies using those datasets to build a state-industry panel (Chakraborty et al., 2021; Hoseini and Briand, 2020). We also employ an alternative measure of informality that is not based on the division between the ASI and UMES datasets. Specifically, we use a question asked to ASI firms regarding whether they report contributing to employee provident funds. The alternative mea-

[19] Although Economic Censuses are available for the period 1990–2010 and cover all economic units in India, they do not allow for distinguishing between the formal and informal sectors, making them unsuitable for our core analysis (Amirapu and Gechter, 2019).

[20] Indian statistical agencies generally avoid using the term informal due to its imprecision. Instead, they prefer the term unorganized, which has a specific definition: broadly, non-farm establishments with fewer than 10 workers. In practice, unorganized firms are often informal.

[21] Note that this measurement captures the extensive margin of informality. Although ASI firms are formal, they may employ informal workers, as documented by the ILO Mehrotra, 2019.

sure of informality treats all firms that allocate a non-zero monetary amount to their employee's provident funds as formal, and informal if they do not.[22]

We then turn to identifying labor informality. Empirical studies typically assign a formality status to workers by looking for variables indicating access to specific benefits or protections that workers are generally entitled to.[23] In line with those works, we use the NSSEU survey to classify workers as formal or informal by exploiting three variables. Our preferred measure is based on information on access to retirement benefits. According to Indian labor regulations, all firms with 20 or more employees are required to contribute to an employee provident fund. Workers can then claim the fund as a lump-sum transfer of money at the time they retire. [24] We classify a worker as formal if they report that their employer contributes to a provident fund; otherwise, the worker is classified as informal. Since the survey only asks this question to wage workers, we assign a value of 0 to all self-employed individuals. This approach is appropriate because all self-employed workers in the survey are individuals operating their own businesses, and are typically very small. Unfortunately, this information is only reported from the 1999–2000 onwards.

To be able to extend our analysis to the previous decade, we propose two alternative measures of informality. The first measure is based on the status the worker has. There are two types of wage workers in India: regular worker, who are directly employed by the firm, and non-regular workers, who are employed through successive short-term contracts. As regular wage workers are much more likely to receive benefits than non-regular wage workers, this status provides a proxy for more stable, formal employment. The NSSEU survey indicates what category of wage employment the workers belong to, and permits us to define formal workers as those who are regular workers, and informal otherwise. Finally, we simply use the wage workers and the self-employed dummies, classifying the latter as informal. The correlation between the two measures is quite high and positive, indicating those alternative definitions do a good job at classifying survey workers in either sector.[25]

Some facts. We now turn to discussing some facts about formal and informal workers and firms in India. Evidence from both firm-level and worker-level data indicates that the informal sector constitutes a substantial portion of the economy in several developing countries (La Porta and Shleifer, 2014). We discuss those issues Table A.4 provides aggregated nationwide statistics ASI and UMES firms for 2005. It reveals that only 1% of firms and 19% of workers are employed

[22]It is important to note that this question is not included in the UMES survey. Consequently, under this alternative definition, all UMES firms are still classified as informal. This difference between this alternative measure of firm informality and the previous one is that here some ASI firms will be considered informal.

[23]For example, Bas and Bombarda (2023) classify formal workers as wage earners with social security coverage, while Ulyssea (2020) defines formality as having a formal labor contract.

[24]In India, labor regulations often exempt small firms, meaning that operating informally does not necessarily imply illegality. For our analysis, the critical factor is that formality imposes costs on firms, rather than whether informality is unlawful. Informality can still be relatively costly if informal firms face challenges that formal firms do not, as documented in Section 4.

[25]The correlation between the preferred benefit-based definition formality and the definition based on being a regular wage worker is 0.67. The correlation between the benefit-based and the wage worker dummies is 0.54.

in registered firms. Furthermore, Table A.4 highlights significant disparities between ASI and UMES. Indeed, ASI firms employ more than twice as many workers as UMES firms, hold 82% of the total capital stock, generate 64% of manufacturing output, and account for 61% of total emoluments (including wages, bonuses, and benefits).

These aggregate findings are driven by underlying micro-level characteristics, which are relatively constant over time (see table A.5 for descriptive statistics of both ASI and UMES in 1990, 2000 and 2010). Starting with firm informality, we first document that formal firms differ from informal firms along a number of dimensions. Table 2 highlights these differences by reporting formal firms' premia from bivariate OLS regressions. Each column represents a possible dependent variable, and the explanatory variable is a dummy variable indicating whether the firm comes from the ASI dataset. Since the dependent variable data are in logarithms, the coefficients can be interpreted as percentages. For example, over the observed period, formal firms employed, on average, 245% more workers than informal firms. Additionally, formal firms demonstrated higher productivity, earning more, possessing greater capital in fixed assets, and achieving higher sales. These performance differences can be attributed to several characteristics of the informal market. While informal firms benefit from not being bound by regulatory compliance, they remain small and encounter significant challenges, as highlighted in Table A.6. Specifically, 7% of informal firms reported being unable to access electricity, 25% cited frequent power outages, and nearly half struggled with insufficient capital. Additionally, around 25% reported difficulties in effectively marketing their products. Relative to formal firms, informal firms often face challenges when conducting their business, for instance having a limited access to financing (La Porta and Shleifer, 2014).

Table 2: Formality Premia

Dependent variables	Labor	Earnings per Worker	Capital per Worker	Sales per Worker
Formal firms' premia	2.455***	4.570***	0.411**	3.521***
	(0.115)	(0.276)	(0.195)	(0.314)
4-industry × year FE	Yes	Yes	Yes	Yes
State × year FE	Yes	Yes	Yes	Yes
Observations	825,340	699,314	749,312	589,816
R-squared	0.689	0.560	0.365	0.543

Notes: Calculations are based on ASI and UMES for 1990, 1994, 2000, 2005, and 2010. All variables are in logs. *Formal firms's premia* is a binary variable equal to 1 if the firm is from the ASI dataset and 0 if it is from the UMES dataset. *Labor* is the natural log of the total number of employees working for the firm. *Earnings per Worker* is the natural log of total earnings paid by the firm over the last year divided by the total number of employees. *Capital per Worker* is the natural log of the total fixed asset value divided by the total number of employees. *Sales per Worker* is the natural log of the firm's total sales over the last year divided by the total number of employees.

We now turn to labor informality. Table A.7 analyzes informality using individual-level data from the NSSEU. In 2005, only 10% of workers across all sectors were classified as formal. The data indicate that informal workers are predominantly younger, have lower levels of education, are more likely to be female, and are overrepresented among the illiterate population. These workers are also more likely to reside in rural areas and are significantly less likely to be employed in large firms. Similar levels of informality are observed within the manufacturing sector. These findings are consistent with other studies on the Indian labor force. For example, Mehrotra (2019) estimates a formality share of 10.4% for 2004–2005.[26]

5 Identification Strategy

This section outlines the identification strategy used to evaluate the impact of Indian trade reforms on the allocation of economic activity between the formal and informal sectors. Specifically, we examine how trade liberalization, through three distinct channels, affects firms' demand. The analysis is based on firm-level data from the ASI and UMES surveys, aggregated at the 4-digit industry and state level. The study focuses on two distinct periods of trade reform: 1990-2000 and 2000-2010. Our identification strategy exploits the exogenous and heterogeneous reduction in tariffs across 4-digit industries over the two periods and the variation of the share of formal firms across industries and states over time. We estimate following equation:

$$FS_{jst} = \alpha + \beta_1 \tau_{O,jt} + \beta_2 \tau_{I,jt} + \beta_3 \tau_{X,jt} + \gamma_{js} + \mu_{st} + \delta_{jt} + Trend_{jt} + \varepsilon_{jst} \tag{2}$$

where FS_{jst} denotes the share of formal employment in 4-digit industry j and state s at time t. It is calculated as the ratio of employment in formal firms relative to total employment (the sum of employment in both formal and informal firms) within the same industry-state-year cell. In our preferred measure, FS_{jst} is the share of firms belonging to the ASI dataset over all firms (ASI and UMES) within each industry-state cell.[27]

Equation (2) accounts for three key channels through which international trade can influence informality: import competition, access to foreign technology, and market access. Specifically, $\tau_{O,jt}$ is the output tariff at the 4-digit ISIC level, capturing the import competition channel. This coefficient is expected to be positive, as lower tariffs intensify import competition, potentially driving some firms out of the market and increasing the prevalence of informality within domestic industries. $\tau_{I,jt}$ is the input tariff, reflecting the cost of imported intermediates. A decline in input tariffs reduces the relative price of foreign inputs, which can enhance firm productivity and incentivize formalization. Therefore, we expect the coefficient on input tariffs to be negative. Our

[26] The share of the labor force employed in the informal sector ranges from 35 percent in Chile to 80 percent in Bolivia and Peru, with Mexico reaching 55% (Perry et al., 2007). In contrast, South Asia, including India, exhibits some of the highest levels of informality globally. Over 75% of the total non-agricultural labor force in South Asia is informal, compared to 65% in Latin America and 45% in the Middle East and North Africa (Bussolo and Sharma, 2022).

[27] As a robustness check, we also use an alternative definition of informality based on whether firms in the ASI and UMES datasets report making contributions to employee provident funds.

specification also control for the export tariff, $\tau_{X,jt}$, to account for foreign demand shocks. A reduction in export tariffs can expand market access for domestic firms, boosting labor demand, particularly among more productive, formal-sector firms, and promoting a reallocation of workers from the informal to the formal sector.

To address time-invariant characteristics specific to 3-digit industries within each state, we include industry-state fixed effects, γ_{js}. These fixed effects absorb persistent differences in how industries operate across states, capturing, for instance, local institutional environments, infrastructure quality, or historical specialization patterns. This adjustment is particularly important in a large and heterogeneous country like India, where geographic variation contributes substantially to cross-sectional differences in economic activity and industrial structure. To control for time-varying factors at the state level, we use state-year fixed effects, μ_{st}. This is a crucial adjustment, as many contemporaneous reforms in India were implemented at the state level, consistent with the country's federal structure. These fixed effects capture the influence of state-level policies that may affect the size of the informal sector. For example, the introduction of the Value Added Tax (VAT) in the early 2000s, which promoted formalization, as well as other labor market policies, such as minimum wage laws and employment restrictions on firing, which can influence firms' hiring decisions (see Hoseini and Briand, 2020 and Soundararajan, 2019, among others). Additionally, we include industry-year fixed effects, δ_{jt}, which control for aggregate time-varying shocks within 2-digit industries. These fixed effects absorb any shocks that are common to all firms within a given 2-digit industry and year, such as national subsidies targeting specific sectors, sector-wide technological changes or global demand shifts.

A remaining identification concern is that firms' choices between formal and informal labor may also be influenced by other industry-specific, time-varying factors unrelated to trade policy, such as pre-existing trends in growth or structural transformation. To address this, we include industry-specific initial size trends, $Trend_{jt}$, constructed by interacting total employment in the 4-digit industry at the beginning of the period with year dummies. This allows us to flexibly control for differential trends in formalization across industries that may be correlated with initial industry size but unrelated to tariff changes. In doing so, we compare industries exposed to different tariff reductions but following similar pre-reform employment trajectories. Thus, we compare industry-state cells exposed to varying degrees of trade liberalization, while ensuring they are otherwise similar in observable characteristics. Finally, because tariff variation occurs at the 4-digit industry level and error terms may be correlated within industries, we cluster standard errors at this level in all regressions.

6 Baseline Results

To distinguish the liberalization effect over the two decades, Table 3 presents the estimation results for Equation (2), distinguishing between the two waves of trade liberalization.[28] Panel A presents the results for the first wave of trade reform (1990-2000), using data from the 1990, 1994, and 2000 decades. In contrast, Panel B displays the estimates for the second wave (2000-2010), based on data from the 2000, 2005, and 2010 decades.

Columns (1) and (4) show that the coefficient on output tariffs is positive and significant, indicating that industries facing larger cuts in output tariffs have experienced a relatively greater decrease in the share of formal workers compared to industries with smaller output tariff reductions. Conversely, the coefficient on input tariffs is negative and significant, suggesting that industries facing larger cuts in input tariffs have seen a relatively greater increase in the share of formal workers. These findings account for several potential confounding factors. First, to control for time-varying shocks at the state level, such as changes in minimum wage laws or modifications to the tax system that could influence firms' hiring decisions, we include state-year fixed effects. Additionally, to capture heterogeneity across industry-location pairs, we include state-industry fixed effects. Lastly, to control for broader industry-specific reforms, we incorporate 2-digit industry-year fixed effects. To account for the fact that exporting firms may have benefited from expanded market access abroad, potentially increasing their revenue and facilitating a higher share of formalization, as observed for Vietnam (McCaig and Pavcnik, 2018), all regressions in Table 3 include export tariffs as a control. However, as illustrated in Table A.1, the variation in export tariffs during this period is minimal, which limits its ability to be effectively identified.

Columns (2) and (5) account for systematic differences in industry trajectories over time related to their initial characteristics by including industry-specific trends based on initial size. These results are consistent with the framework outlined in Section 2. The estimates in column (2) indicate that, for industries experiencing the average reduction in output tariffs (45 percentage points), the share of formal employment declined by approximately 6 percentage points. In contrast, a similar average reduction in input tariffs (46 percentage points) is associated with an increase of about 21 percentage points in the share of formal employment. Similar patterns emerge in Panel B, where the average reductions in output and input tariffs were 20 and 23 percentage points, respectively. Over the period 2000–2010, these tariff changes resulted in a net increase of 6.5 percentage points in the share of formal employment.

Overall, these findings indicate that input-trade liberalization more than offsets the adverse effects of import competition on the share of formal employment. These results remain robust across alternative specifications. In columns (3) and (6), we use an alternative definition of informality based on whether firms in both the ASI and UMES samples report contributing to the employees' provident fund. Specifically, the dependent variable is defined as the share of

[28] In section 7, we also present results based on a long-run approach, covering the period from 1990 to 2010.

total employment in firms that declare positive expenditures on the provident fund, relative to total employment in the industry-state. The estimates remain robust and stable when using this alternative measure of informality.

Table 3: The effects of trade liberalization on the share of formal workers

	Panel A: 1990-2000			**Panel B: 2000-2010**		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variables	Share of formal workers in industry j and state s					
	Registered	Registered	Provident Fund	Registered	Registered	Provident fund
$\tau_{O,jt}$	0.166*	0.142*	0.130*	0.228**	0.263**	0.244**
	(0.084)	(0.075)	(0.067)	(0.109)	(0.116)	(0.102)
$\tau_{I,jt}$	-0.527***	-0.468***	-0.415***	-0.478***	-0.503***	-0.491***
	(0.060)	(0.052)	(0.050)	(0.070)	(0.086)	(0.074)
Export τ_{jt}	Yes	Yes	Yes	Yes	Yes	Yes
State × year FE	Yes	Yes	Yes	Yes	Yes	Yes
State × 3-industry FE	Yes	Yes	Yes	Yes	Yes	Yes
2-industry × year FE	Yes	Yes	Yes	Yes	Yes	Yes
Trends$_{jt}$	No	Yes	Yes	No	Yes	Yes
Observations	1,934	1,934	1,934	2,019	2,019	2,019
R-squared	0.845	0.860	0.863	0.772	0.783	0.790

Notes: OLS estimation split by decade. Panel A is restricted to rounds 1990, 1994 and 2000, and Panel B to rounds 2000, 2005 and 2010. We use two definition of formal employment. *Registered* is the share of workers employed in firms belonging to the ASI dataset over total employment (in ASI and UMES) in the industry-state. *Provident Fund* is the share of workers employed in firms reporting having positive expenses for their employees' provident fund over total employment (in ASI and UMES) in the industry-state. Tariffs are measured at the industry j at time t. *Trends*$_{jt}$ accounts for industry dynamics at four-digit, constructed interacting initial employment level for each decade with year dummies. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To understand the channels behind these results, we investigate the role of capital intensity. If foreign intermediate inputs and capital are complementary in the production process, then industries that are initially more capital intensive should benefit more from input trade liberalization. To assess this mechanism, we interact the tariff measures with capital intensity, constructed as the initial capital-to-labor ratio in a given industry and decade. The results presented in Table A.8 support this intuition: the estimated coefficient on input tariffs is larger in absolute value for industries that are initially more capital intensive, suggesting that these industries experience a greater expansion in the share of formal employment in response to increased access to foreign inputs.

7 Robustness and Sensitivity Analysis

This section presents a series of robustness and sensitivity checks. First, we consider alternative measures to capture the informality dimension. Second, we show that the results remain robust when using the share of formal firms as an outcome. Third, we control for other potentially confounding policies implemented during the same period, such as FDI liberalization and industrial delicensing. Fourth, we re-estimate the baseline results without splitting the sample by decade.

Finally, we adopt a long-run approach by estimating a stacked regression specification.

Share of formal firms. In the baseline results, we find evidence of reallocations of workers between sectors at the state-industry level. A related question is whether there were also reallocations of firms between the formal and the informal sector due to trade liberalization, the extensive margin of informality. To investigate this, we estimate equation (2) using the share of formal firms as dependent variable instead of the share of formal workers in the state-industry. Table B.2 shows that output tariff cuts induce a reallocation of firms from formal to informal sector by reducing the share of formal firms, while input tariff reductions generate an increase in the share of formal firms. This is in line with the conceptual framework presented in section 2, in which trade modifies the relative costs of formalization and my However this effects are smaller in magnitude compare to the effects of tariff cuts on the share of formal workers.

Contemporaneous policies. As detailed in section 3, tariff reductions are only one dimension of a more general move towards liberalization of the Indian economy. This is a strong concern in the 1990s of the analysis, when the bulk of the reforms was taking place. Several of those liberalization reforms increased the incentives of firms to operate formally by reducing the distortions impeding the growth and profit opportunities of the formal sector. Our estimates may pick up the effect of those reforms. While the specification of equation (2) accounts for policy changes at the state-level, the two-digit industry fixed effect might fail at accurately controlling for other reforms if they had a differential effect on industries within a two-digit category. To limit those issues, we explicitly control for two of the most import reforms of that period. Relying on data from Aghion et al. (2008), we control for the relaxation of restrictions on foreign direct investment as well as of the end of licensing, which regulated entry and production the formal sectors for many industries. Results presented in Table B.3 show that our results are robust even when accounting for those reform.

Estimating both decades together. In the baseline results, we separate the analysis by decade to exploit the two waves of liberalization (1991 and 2004-2005). This allow us to argue that the effects of trade reforms hold across liberalization periods. However, one could argue that sample splitting could induce a bias when studying the second decade, even under plausibly exogenous tariff reductions, because industry-states were already liberalized in the first year (2000). Indeed, the year 2000 enters as the last year of the first decade sample, and the first year of the second decade sample. To account for that, we re-estimate equation (2) and keep the whole sample for the period 1990-2010 together. Table B.1 shows that our results are robust across the entire liberalization period.

8 The role of labor market institutions

Labor regulations significantly influence how trade shocks are absorbed within the economy. This section explores the interplay between trade liberalization and labor market policy, focusing on how labor market institutions shape the relationship between trade and informality. By influencing the cost of operating formally, these institutions can affect how formal firms adjust their labor demand in response to trade liberalization, with significant implications for informality levels. We consider four distinct types of labor market institutions: state-level labor regulations, labor union presence in each state, enforcement of labor regulation, and changes in employment protection legislation.[29]

Labor market regulations. Labor regulations in India are based on a central legislation, the Industrial Disputes Act of 1947. However, over time, individual states have passed amendments and deviated from this national framework. We follow the most recent classification for labor market regulation, where we classify each state as pro-worker or pro-employer.[30] Using this measure, we examine the complementarity between trade and labor market regulations in driving the reallocation of firms and workers between the formal and informal sectors. Specifically, we want to evaluate the differential effect of trade liberalization depending on the type of labor market regulations.[31] We expect that in pro-worker states, the effects of output tariff reductions (import competition) on informality to be smaller compared to pro-employer states. This is because, in more regulated pro-worker states, where hiring and firing costs are greater, firms are less responsive to changes in trade policies. The effect of input tariff reduction in pro-worker states is less straightforward. On the one hand, we would expect formal firms in pro-worker states to hire relatively less formal workers because they may anticipate that those newly-hired workers could be harder to fire later on. On the other hand, our data show that firms in pro-worker states are larger (and possibly, on average more profitable). This implies that firms in pro-worker states can better afford to bear the labor costs associated to hiring registered workers. Thereby, the share of formal firms would be larger in pro-worker states and so the effects of input tariff cuts should be greater than in pro-employer states.

To assess the interplay between trade shocks and labor market regulation, we follow Aghion et al. (2008) and estimate the following regression:

$$FS_{jst} = \alpha + \beta_1 \tau_{O,jt} + \beta_2 \tau_{I,jt} + \beta_3 \tau_{O,jt} \times \text{pro-W}_s + \beta_4 \tau_{I,jt} \times \text{pro-W}_s + \gamma_{js} + \mu_{st} + \delta_{jt} + Trend_{jt} + \varepsilon_{jst} \quad (3)$$

[29]Statistics on these different labor market institutions are provided in Table A.2.

[30]Neutral states are coded as pro-employer. As explained in section 4.1, Besley and Burgess (2004) proposed an initial classification of labor regulations based on the period 1958-1992. However, given our time frame of 1990–2010, their index is not applicable. We therefore adopt an alternative classification, used in several more recent studies such as Gupta et al. (2009), Chaurey (2015), and Chakraborty et al. (2024). Unlike the Besley and Burgess (2004) measure, this classification is time-invariant, which solves the potential concern related to the fact that state labor regulations are responding to changes in trade openness.

[31]Table **??** in Appendix A indicates that firms in pro-worker states tend to be larger, produce more, and pay higher wages, making them better able to adapt and formalize when input tariffs decrease.

where FS_{jst} denotes the share of formal employees in 4-digit industry j and state s at time t. pro-W_s represents a dummy variable which takes the value 1 every time a state is a pro-worker state. Differently, we assign a 0 if a state is either pro-employer or neutral. All other controls are described in equation (2). The coefficients of interest, β_3 and β_4 are identified by the combination of industry-year variation in tariffs interacted with state-year variation in labor regulation. Table C.1 presents the results. Columns (1), (2) and (3) show the results on the full sample. Results in columns (2) and (3) suggest that the impact of input tariff cuts on the share of formal employment is greater in pro-worker states. The input tariff–labor regulation interaction coefficient is negative and significant, indicating that a reduction in input tariffs allowed industries in states with pro-worker regulations to increase formal employment relative to industries in pro-employer states. This result could be due to the fact that pro-worker states have on average larger formal firms, better able to absorb the increased costs associated with pro-worker regulations. Indeed, Panagariya (2007) explains that hiring and firing costs discourage firm entry in pro-worker states. In the last two columns, we split the sample between pro-worker (column 4) and pro-employer (column 5) states. Our estimates suggest that the negative effect of import competition on the share of formal workers plays a role only in pro-worker states. On the contrary, the effect of input tariffs cuts on the share of formal employment continues to be larger in pro-worker states.

Labor Unions. Next, we explore the specific role of unionization. The strength of unions in a state might affect the bargaining process between firms and workers. Following the implementation of NAFTA in the United States, Cristea and Lopresti (2024) find that states which had taken provisions to weaken unions experienced lower employment growth in manufacturing. Evidence is more mixed in developing economies. Data limitations prevent Goldberg and Pavcnik (2005) from examining the role of unions in the context of Colombia's trade reforms. However, they argue that unions in Colombia are weak and, as a result, are unlikely to significantly alter the impact of these trade reforms. In contrast, unions in India have historically been strong and may therefore have influenced the impact of trade reforms on informality. Firms located in states with stronger unions may opt for lower levels of formal employment to avoid union-related constraints, particularly following large import competition shocks that reduce their domestic market share. Differently, following a positive trade shock, which improves access to better foreign inputs, formal firms in states with stronger unions may be more willing to share part of the surplus generated through foreign technology upgrading.

We evaluate these predictions by estimating a variation of equation (3), where tariffs are interacted with a measure of unionization. We calculate the share of unionized workers as the ratio of workers who report union membership to the total number of workers (both formal and informal) in a state, using the sampling weights provided by the survey. To avoid endogeneity issues, we use information on unionization in 1993.[32] We classify a state as having high unionization if its share of

[32]Unfortunately, data on union membership were not collected prior to 1993.

unionized workers in 1993 is above the sample median; 0 otherwise. Next, we interact output and input tariffs with our measure of unionization. The results, presented in columns (1) to (3) of Table C.2, indicate that the negative effect of output tariff reductions on the share of formal employment is more pronounced in states with a higher share of unionized workers. Specifically, the coefficient attached to the interaction between output tariffs and the high unionization dummy is positive and significant. The reduction in output tariffs prompted industries in states with stronger unions to decrease formal employment more relative to industries in states with weaker unions. This result seems at odds with previous studies documenting how strong unions mitigate employment losses in developed economies, where there is no informal sector. In India, the informal sector is large and informal workers are rarely unionized. In a context of greater import competition, we may expect formal firms to try to bypass unions by hiring informal workers not benefiting from collective bargaining. These findings are confirmed in column (4), which uses an alternative measure of unionization based on historical unionization rates prior to 1977, as proposed by Aghion et al. (2008). The last two columns (5) and (6) show that our results are also robust when we split the sample between high and low-unionized states. The effects of both import competition and access to foreign inputs on the share of formal employment are stronger in states with a higher share of unionized individuals (column 5).

Enforcement. This part examines how the enforcement of labor regulations influences the labor market effects of trade. Rigid labor market regulations may lead to poorer labor market outcomes and potentially exacerbate employment losses from trade shocks, especially in countries with a large informal sector. Unfortunately, we do not directly observe enforcement in our data. To overcome this limitation, we build on the approach of Ponczek and Ulyssea (2022) and measure the enforcement of labor regulations using the distance between a worker's district and the nearest "Chief Labor Commissioner Office". Those regional labor offices were established to enhance the enforcement of labor laws. Individuals located in districts closer to the labor office are likely to benefit from better enforcement of these regulations. In such districts, we expect the negative effects of import competition on informality to be weaker.

To match study this mechanism, we need to rely on a more detailed geographical information not available in the state-industry panel. We therefore rely on individual-level data from the NSSEU for the years 1999, 2004, and 2009, which provides disaggregated information at the district level. We estimate a linear probability model in line with equation (1) to assess the likelihood of being formally employed. Our key variable of interest is the interaction term between output (or input) tariffs and an indicator variable, "far from labor office." This variable takes the value of 1 if the district is above the median distance from the state's labor office, and 0 otherwise. Results are presented in Table C.3. Workers in industries exposed to relatively large output tariff reductions and located in districts far from labor offices are less likely to be formally employed than workers in industries with similar tariff reductions and located in districts close to the labor offices. In districts far from labor offices, it may be easier for firms to dismiss workers, as labor laws are harder to enforce and the probability of detection, and penalty, is lower. The last two columns control for the differential impact of trade liberalization based on the distance to the state capital, to ensure that our results are not picking up the effect of remoteness. Column (5) further includes unobservable shocks that vary across 3-digit industries and states over time, incorporating 3-digit industry-state fixed effects. The main results remain robust after including these controls.

Employment protection legislation. Finally, we investigate how employment protection legislation that penalizes firms for circumventing labor laws can alter the effect of trade liberalization in contexts with high levels of informality. As described above, labor laws in India increase the cost of hiring formal labor, especially in pro-worker states. To bring down those labor costs, firms circumvent labor laws by hiring workers indirectly through contractors (Chaurey, 2015). In the 1990s, as the economy liberalized, formal firms started exploiting a loophole in the state's labor regulations and hired contract labor instead of regular workers. Unlike workers on the payroll, IDA provisions do not apply to contract workers, who do not enjoy benefits nor social security and can therefore be considered as informal (Chaurey et al., 2023).

To study how stricter enforcement of employment protection legislation affects how firms exploit the informality margin as a response to trade policy changes, we leverage a reform that took place in Andhra Pradesh in 2003. The goal of the reform was to strengthen employment protection in the state by forcing formal sector firms to comply with EPL, or face penalties. The reform can therefore be interpreted as a sudden increase in the cost of exploiting the intensive margin of informality. Because the reform affects hiring decisions, we do not expect the import competition channel to play a role here. On the contrary, the increase in formal labor demand due to access to foreign inputs may be relatively smaller in post-reform Andhra Pradesh. Unable to hire contract workers at a low cost, some formal firms may not hire as many workers as they would otherwise have. Using the state-industry panel, we estimate an equation similar to equation (3). Tariffs are interacted with a treated dummy taking value 1 for observations in Andhra Pradesh after 2003, and 0 otherwise. Results are presented table C.4. Columns (1) and (2) interact the output (input) tariffs with a treatment variable, and results are not significant. In column (3), with both interactions, the coefficient on the input tariff and reform indicator is positive and significant, showing that a reduction in input tariffs made industries located in post-reform Andhra Pradesh decrease formal employment relative to industries in non-reformed states. This means that the reform mitigated the increase in the labor demand of formal firms due to input tariff cuts. It suggests that part of the increase in employment by formal firms after the reduction in input tariffs was driven by the hiring of contract workers. This finding is in line with those of Chaurey et al. (2023) who find that firms in locations with stricter enforcement of employment protection legislation that keep hiring workers hire them formally, but that many firms opt for not hiring workers. Column (4) is a robustness check with a similar specification but an alternative measure of state-industry formality, and those results are similar.

9 Local labor market approach

Our within industry-level analysis, which compares changes in probability of finding a formal job across individuals in industries with differing levels of trade exposure, is not appropriate to identifying reallocation across sectors. If workers have limited mobility between local labor markets in response to trade-induced labor market shocks, as shown by Topalova (2010) and Autor et al. (2013), the main adjustment mechanism will involve the reallocation of workers from contracting industries to other sectors within the same local labor market.[33] The NSSEU dataset, spanning more than a decade and encompassing nearly all Indian districts, with both urban and rural areas, allows us to examine the localized effects of trade liberalization.[34]

A growing body of recent research on the local labor market effects of trade reforms underscores that trade liberalization, driven by import competition, impacts workers differently depending on their geographical location. This variation stems from the unique industrial compositions of different regions (e.g., Topalova, 2010; McCaig, 2011; Kovak, 2013; Autor et al., 2013; Hakobyan and McLaren, 2016; Acemoglu et al., 2016; Dix-Carneiro and Kovak, 2017; Dix-Carneiro and Kovak, 2019). These studies incorporate the regional industrial structure to assess how trade liberalization affects local labor markets, accounting for differences in industrial specialization across regions.

To account for regional variations in the impacts of trade, we exploit the NSSEU survey, specifically the information regarding the district where each individual report resides.[35] Following Topalova (2005), the trade policy measures used are the district-level tariff, calculated as the 1998 employment-weighted average applied ad valorem tariff at time t. In line with the existing literature, we compute a weighted average of tariffs at the district level, using the industrial labor distribution at the beginning of the period as weights. Following Topalova, 2010 and Kovak, 2013, we calculate weighted averages for input, output, and export tariffs, considering both tradable and non-tradable industries active in each district. The weights are based on employment shares at the district-industry level. To mitigate potential endogeneity concerns stemming from changes in the industry composition within a district over time, we use fixed weights derived from the employment structure of each district in the first year it enters the survey. The district-level weighted average tariff for inputs, outputs, or exports is calculated as follows:

$$\tau_{dt}^{1998} = \sum_j \frac{\text{Emp}_{dj}^{1998}}{\sum_j \text{Emp}_{dj}^{1998}} \tau_{jt} \tag{4}$$

where Emp_{dj}^{1998} represents employment in sector j operating in district d. To calculate employ-

[33]For reallocation across sectors refer to Topalova (2010) section C.

[34]We initially focus on the decade 2000-2010, and are currently extending the analysis to the entire liberalization period (1990-2010), using the definition of informality related to having a non-periodic contract. We also intend to explore the role of labor market regulations, which was highlighted in section A.2.

[35]New districts are created by splitting with existing districts throughout the period, so we use 1983 district borders in all analyses. The total number of districts is 414 and remains stable over time.

ment shares for constructing local tariffs and other industry-level and local-level variables, we use economic censuses from 1990 and 1998.[36] Figure A.5 illustrates the geographical distribution of input and export tariffs, highlighting the spatial heterogeneity across regions. The most liberalized areas, where local tariffs experienced the largest reductions, are concentrated in districts near major cities such as Mumbai, Delhi, Hyderabad, Bangalore, and Kolkata. Coastal regions and areas along the Ganges River also appear to have been significantly impacted, as well as southern states like Kerala and Tamil Nadu. In contrast, districts in the Deccan Plateau, central India, and the Northeast were less affected, likely due to their smaller and less productive manufacturing sectors.

We use these local-level measures of exposure to openness to assess their effect on formal employment by estimating the following equation:

$$F_{ijdt} = \alpha + \beta_1 \times \tau_{O,dt} + \beta_2 \times \tau_{I,dt} + \beta_3 \times \tau_{X,dt} + \alpha_{jd} + \alpha_{st} + \alpha_i + \gamma X_i + \varepsilon_{ijdt} \quad (5)$$

Similarly to equation (1), F_{ijdt} is an indicator variable that takes the value one if individual i is employed as a formal worker in industry j, district m, at time t, and zero if the individual works as an informal employee. To account for trade openness, we control for three channels at the district level: output tariffs,$\tau_{O,dt}$, input tariffs, $\tau_{INP,dt}$, and export tariffs, $\tau_{X,dt}$. To account for potential differences in time-invariant characteristics across industries in different districts, we include sector-district fixed effects, γ_{jd}. To account for state-level reforms (such as VAT implementation or changes to minimum wage laws), local election outcomes (e.g., the 2004 national and local elections that brought the Congress Party back to power), and the Maoist insurgency—a low-intensity conflict affecting Eastern India between 2000 and 2009 (Couttenier et al., 2023)—we include state-year fixed effects, μ_{dt}. Finally, we include a trend in the share of tradable industries at the beginning of the period to ensure that our results are not influenced by factors unrelated to trade policy changes—such as increasing international integration driven by place-based policies like special economic zones. The results are presented in Table 4. As in the within-industry analysis, greater import competition decreases the probability of working formally, whereas input tariff reductions increase it.

[36] Due to the survey design and the structure of the sampling weights, the NSSEU data cannot be aggregated at the district-sector level in a representative manner. To overcome this limitation, we use the Economic Censuses, which cover the entire universe of establishments in India and can therefore be aggregated at the district-sector level.

Table 4: Local Labor Markets (NSSEU)

	(1)	(2)	(3)	(4)
VARIABLES	Formal	Formal	Formal	Formal
$\tau_{O,dt}$	0.142*		0.303**	0.250**
	(0.082)		(0.135)	(0.124)
$\tau_{I,dt}$		-0.193	-0.572**	-0.635**
		(0.166)	(0.278)	(0.281)
Trend_{dt}^{1998}				-0.005**
				(0.002)
Industry $\times$ district FE	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes
Year $\times$ state FE	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes
Export τ_{dt}	Yes	Yes	Yes	Yes
Observations	533,852	533,852	533,852	533,852
R-squared	0.705	0.705	0.705	0.706

Notes: LPM estimation between 1999 and 2010 (all sectors). Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Tariffs in district d at time t. Individual characteristics include age, square of age, years of education, household size and urban location. Pseudo-individual FE includes gender, religion, literacy, birth-year and district. Heteroskedasticity-robust standard errors clustered by district are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

10 Aggregate Productivity and Labor Reallocation

A large informal sector can lead to significant resource misallocation at the macroeconomic level. Table 2 highlights differences between formal and informal firms, showing the existence of a formality premium. In this section, we aim to quantify the impact of trade policies on the Indian economy. Building on the macroeconomic development accounting literature, in this section we assess how trade liberalization affects aggregate labor productivity in India's manufacturing sector. Following McCaig and Pavcnik (2018), we first estimate the labor productivity gap, adjusting for potential biases and measurement issues, and then we assess the extent of resource reallocation between the formal and informal sectors. We depart from their approach by taking a long-run perspective, allowing us to assess trends in the size of the productivity gap and to compare the potential aggregate effects of the trade reforms over two liberalization episodes, which we analyze separately.

Labor Productivity Gap. We use a Cobb-Douglas production function for both formal and informal sectors, such that $Y = AK^{\alpha_s}L^{1-\alpha_s}$, where A is total factor productivity, K is capital, L is labor, α_s and $(1-\alpha_s)$ represent the output elasticity of capital and labor respectively. The subscript $s \in \{f, i\}$ distinguishes between the formal, f, and informal, i, sectors. Assuming homogeneous labor and perfectly competitive markets, wages are equal to the marginal revenue product of labor (MRPL), which implies:

$$w_s = \text{MRPL}_s = (1 - \alpha_s)\text{ARPL}_s \tag{6}$$

where $ARLP_s$ denotes the average revenue product of labor in sector s. As a result, the gap in marginal revenue products of labor across sectors is proportional to the observed gap in average revenue products, adjusted for differences in output elasticities of labor:

$$\frac{w_f}{w_i} = \frac{\text{MRPL}_f}{\text{MRPL}_i} = \frac{(1 - \alpha_f)\text{ARPL}_f}{(1 - \alpha_i)\text{ARPL}_i} \tag{7}$$

Following the methodologies adopted in development accounting research (e.g., Caselli, 2005; Gollin et al., 2014), we use two alternative measures to assess productivity gaps across formal and informal sectors: wages and revenue per worker. Using ASI and UMES datasets, we retrieve information on wages, revenues, and labor for the years 1994 and 2005. To construct the revenue-based measure for the formal sector, we calculate the average revenue product of labor (ARPL) as total revenue divided by total employment of ASI firms in a sector. For the informal sector, we proceed similarly but using UMES information. Similarly, the wage-based measure is constructed using total annual earnings divided by total employment in each sector.[37] As we later account for various compositional and measurement differences, we refer to this initial estimate as the unadjusted productivity gap.

Estimates of the labor productivity gap between the formal and informal sectors by liberalization decade are reported in Table 5, with Panel A corresponding to 1994 and Panel B to the 2005. Columns (1) and (2) show estimates of the gap across all manufacturing industries. Columns (3) to (6) repeat the analysis for different subsamples to ensure the results are not driven by specific industries or locations. Specifically, columns (3) and (4) focus on the textile and apparel industries, while columns (5) and (6) restrict the sample to Maharashtra, India's richer largest state in terms of GDP. Three key findings emerge from Table 5. First, the labor productivity gap between the formal and informal sectors is consistently large across decades, subsamples, and various methodological adjustments. The unadjusted measure in column (1) of panel B indicates a productivity gap of 16 in favor of formal sector, which is broadly consistent with findings in the recent literature.[38] Second, there is a substantial decline in the gap over time. Considering all sectors, the unadjusted gap fell by more than half, from 34.1 to 16. This trend likely reflects the

[37]Total annual earnings encompass wages/salaries as well as additional payments, such as bonuses and welfare expenses made by the employer for wage workers.

[38]Nataraj (2011) find a gap of 12.4 for Indian manufacturing over a similar period, and McCaig and Pavcnik (2018) find a gap of 9 for Vietnam. These differences highlight cross-country variation in the productivity gap. Compared to India, Vietnam has a smaller informal sector, higher GDP per capita, and a lower level of extreme poverty, which may help explain its narrower gap.

structural transformation of the Indian economy following liberalization reforms in the early 1990s, such as industrial delicensing and tariff reductions. These policies encouraged the entry of small formal firms, which are captured in ASI data and contributed to lowering average and marginal productivity in the formal sector—thereby narrowing the gap. Finally, consistent with McCaig and Pavcnik (2018), the wage-based approach consistently yields a much smaller productivity gap than the revenue-based approach. This difference arises because wage-based estimates are less sensitive to the measurement errors that affect revenue-based measures. We examine these issues in more detail in the next section.

The unadjusted gaps reported in Table 5 may not accurately reflect the true productivity difference between the formal and the informal sectors. Three potential sources of bias could affect these estimates: (i) differences in workforce characteristics across sectors, (ii) measurement error in the data, and (iii) sector-specific output elasticity of labor. We begin by addressing the first concern, i.e. differences in workforce composition. Controlling for differences in human capital is important, as workers in the formal sector tend to be more educated and more productive (Gollin et al., 2014; McCaig and Pavcnik, 2018). Failing to account for this would mechanically inflate the estimated gap. Thus, following Gollin et al. (2014), we implement a two-step procedure. First, we estimate the average years of education for workers in each sector.[39] Next, following the approach in Gollin et al. (2014), we adjust for human capital by applying a standard Mincer-style correction. Specifically, we multiply the average years of schooling in each sector by the estimated return to education from Banerjee and Duflo (2005), and use the resulting ratio to scale our productivity measures. This yields the human capital–adjusted productivity gaps reported in the second row of Table 5. As expected, the adjustment narrows the gap: in manufacturing, it falls from 34.1 to 27.7 in the 1990s, and from 16.0 to 12.9 in the 2000s. These results suggest that differences in human capital account for roughly 20% of the observed productivity gap.

The second adjustment addresses potential measurement errors in the ASI and UMES surveys. While major mismeasurement is unlikely—since both datasets are designed to be complementary and representative, and their methodologies remained stable between 1990 and 2010, micro-level evidence suggests that data from the informal sector may still suffer from inaccuracies.[40] For instance, De Mel et al. (2009) find that informal firms in Sri Lanka underreport revenues by up to 30%, while data on time use collected by Fafchamps et al. (2014) show that Ghanaian microenterprise workers' effective working hours account for only 89% of total hours spent in their

[39] Since ASI and UMES do not have information on education, we turn to NSSEU dataset. To match NSSEU data with ASI and UMES, we exploit the fact that NSSEU provides individual-level information on the industry, size, and power usage of the firm where the worker is employed—key criteria that define also ASI and UMES surveys. This allows us to use NSSEU data to compute the average education level for firms in ASI and UMES respectively. For instance, to measure average education of ASI firms, we proceed by considering all individuals who report being employed in firms with more than 20 workers without power or more than 10 workers with power. We apply a similar approach to construct the average education level for the informal firms (UMES).

[40] As discussed in Section 4.2, both ASI and UMES are constructed to offer nationally representative data and are methodologically consistent over time.

business. Following McCaig and Pavcnik (2018), we adjust the ARPL ratio by multiplying it by 0.7 (to account for underreported revenue) and by 0.89 (to correct for overestimated labor supply). The resulting adjusted productivity gaps, shown in row 3 of Table 5, fall to 17.2 in the 1990s and 8.09 in the 2000s.[41]

Up to this point, the analysis has implicitly assumed equal output-labor elasticities across sectors, that is, equation (7) is estimated under the assumption that $\alpha_f = \alpha_i$. However, this is unlikely to hold in practice. The informal sector, being more labor-intensive, likely exhibits a higher output elasticity of labor than the formal sector. Ignoring this difference may lead to an overstatement of the productivity gap. To address this, we again follow McCaig and Pavcnik (2018) and apply a correction based on an elasticity ratio of 1.5 between the informal and formal sectors. This value is drawn from Restrepo-Echavarria (2014), who estimate labor elasticities of 1.0 for the informal sector and 0.68 for the formal sector. Incorporating this adjustment further narrows the productivity gap in manufacturing—to 11.5 in the 1990s and 5.39 in the 2000s.

This analysis of the labor productivity gap yields three main takeaways. First, about two-thirds of the unadjusted labor productivity gap can be explained by differences in education, measurement error, and output-labor elasticities. After adjusting for these factors, the gap in manufacturing falls to 11.5 (revenue-based) and 13.9 (wage-based) in the 1990s, and to around 5.4 in both measures in the 2000s. For comparison, McCaig and Pavcnik (2018) report even smaller gaps for Vietnam. Second, despite the adjustments, the labor productivity gap in India remains sizable, suggesting that labor reallocation from the informal to the formal sector could yield substantial efficiency gains. Third, the decline over time is robust across different subsamples and methods. Equation (7) neglects the potential role of sector-specific distortions, such as taxes, labor laws or unequal access to capital, that may drive a wedge between the ARPL and the wage received by workers (Hsieh and Klenow, 2009b). While the adjustments made do not account for such distortions, the similarity between the adjusted estimates from both estimation approaches reassures bias induced by those distortions is likely minimal.

Reallocation across Sectors. Next, we assess the contribution of labor reallocation between the formal and informal sectors to aggregate productivity growth. To do this, we estimate the share of workers who transitioned from informal to formal employment during the two waves of trade liberalization, using the estimated effects of output and input tariff reductions reported in Table 3. Specifically, we multiply these coefficients by the corresponding decadal changes in tariffs and weight the resulting effects by initial industry employment shares. We then average across industries. This yields the following expression:

$$\text{Reallocation to Formal} = \sum_j s_j \left(\hat{\beta}_1 \times \Delta\tau_{Oj} - \hat{\beta}_2 \times \Delta\tau_{Ij} \right) \tag{8}$$

[41] This adjustment pertains to revenues and cannot be applied to wage-based measures.

where s_j denotes the share of manufacturing employment in industry j at the beginning of each decade (1990 or 2000), $\Delta\tau_{Oj}$ and $\Delta\tau_{Ij}$ represent the change in output and input tariffs for industry j over the corresponding decade, and $\hat{\beta}_1$ and $\hat{\beta}_2$ are the estimated coefficients from columns (2) and (5) of Table 3. Notice that equation (8) accounts for the effective rate of protection, i.e. the net effect of trade liberalization considering both input and export tariffs.[42]

Results of this reallocation are reported in the fifth row of Table 5.The estimated share of workers moving from the informal to the formal sector is positive and economically meaningful: approximately 15% in the 1990s, and 5% in the 2000s. This suggests that the increase in labor demand by formal firms—driven by improved access to foreign inputs—more than offset the opposing effect of import competition. Given the sizable productivity gap between sectors, these reallocations imply substantial potential gains in aggregate productivity. Moreover, the changing magnitude of the gap over time has important implications. The productivity impact of trade-induced reallocation depends on the initial gap: the larger the gap, the greater the potential gains from shifting labor toward the formal sector.

We explore these measures to account for the heterogeneity across sectors. Table D.1 shows results for industries that experienced above- and below-median tariff cuts. As expected, labor reallocation is more pronounced in sectors with larger tariff reductions, reported in columns (3)–(4), than in sectors with smaller reductions, in columns (5)–(6). The adjusted productivity gap is larger in more protected industries, suggesting potential productivity gains from further liberalization. We also investigate the role of labor market regulations. Table D.2 splits the analysis between pro-worker and pro-employer state. While the magnitude of worker reallocations does not vary significantly across regulatory regimes, the adjusted productivity gap is slightly larger in pro-employer states. This finding is consistent with the notion that pro-worker regulations raise the cost of formality to a point where firms near the productivity threshold opt to remain informal, which increases average productivity of the informal sector and reduces the productivity gap. The implications of Table D.2 are ambiguous. On the one hand, a larger productivity gap suggests greater efficiency gains from reallocation, implying that pro-employer states may enjoy larger increases in aggregate productivity than pro-worker states. On the other hand, this may not be the case if the additional formal labor demand in pro-employer states is driven by low-productivity formal firms, rather than high-productivity ones.

[42]As discussed in Section 3, the export tariffs declined by less than 5 percentage points over the 20 year period. Therefore, we do not include them in the computation of the net effect.

Table 5: Productivity Gap and Reallocation to Formal

	All		Textile		Maharashtra	
	Revenue based (1)	Wage based (2)	Revenue based (3)	Wage based (4)	Revenue based (5)	Wage based (6)
Panel A						
Unadjusted	34.1	17.2	28.9	10.7	22.0	9.18
Adjusted by human capital	27.7	13.9	24.2	8.97	14.7	6.16
+ measurement error in revenue and time worked	17.2		15.1		9.20	
+ differences in output-labor elasticity	11.5		10.0		6.13	
Share of workers reallocated in the 1990s	0.146	0.146	0.195	0.195	0.145	0.145
Initial share of workers in the formal sector	0.209	0.209	0.175	0.175	0.353	0.353
Panel B						
Unadjusted	16.0	6.68	13.8	7.51	7.57	2.33
Adjusted by human capital	12.9	5.42	11.6	6.28	5.08	1.56
+ measurement error in revenue and time worked	8.09		7.23		3.16	
+ differences in output-labor elasticity	5.39		4.82		2.11	
Share of workers reallocated in the 2000s	0.050	0.050	0.065	0.065	0.055	0.055
Initial share of workers in the formal sector	0.178	0.178	0.124	0.124	0.283	0.283

Notes: Panel A is based on 1994 data, and panel B on 2005 data. Formal sector information comes from ASI, informal sector information from UMES. The sample comprises all manufacturing industries in columns (1) and (2), textile and apparel industries in columns (3) and (4), and all manufacturing industries in the state of Maharashtra in columns (5) and (6). The productivity gap reported in columns (1), (3), and (5) is based on the average revenue product of labor, plus subsequent adjustments. The average revenue product of labor is the ratio of total revenue divided by total employment within each sector. The labor productivity gap reported in columns (2), (4), and (6) is based on the ratio of total annual earnings divided by total employment in each sector, plus subsequent adjustments. Human capital information comes from the NSSEU survey for round 1999-2000.

11 Conclusion

This study provides causal evidence that trade liberalization reallocated workers between the formal and informal sectors between 1990 and 2010 in India. We find that the import competition and access to foreign inputs drive this reallocation process. The identification strategy exploits exogenous variations of output and input tariffs. We first rely on an aggregated state-industry panel and regress the share of employment of formal firms over total employment (of formal and informal firms) over two decades of trade reforms. While import competition reduces the share of formal employment and increasing informality relatively more, the positive effect of access to foreign inputs on the share of formal workers more than compensates the negative effect of import competition. Those results are confirmed by an analysis at the individual level, and are driven by within-industry reallocations. In addition, we find that domestic labor market institutions modify the way trade policy impacts formality. We consider a wide range of labor market institutions and find that heterogeneity in hiring and firing restrictions, unionization, enforcement and employment protection legislations all induce a differential effect of trade reforms on informality. In particular, we find that the formalization due to greater access to foreign inputs is relatively stronger in pro-worker states, and mitigated in states where there are penalties to hiring contract labor. Moreover, strong union presence and low enforcement both exaggerate the reallocations towards the informal sector.

Our analysis based on a development accounting framework reveals that the labor productivity gap between the formal and the informal sector is particularly large, relative to Latin American and South East Asian countries. Over the two decades of trade liberalization, the gap decreased and the overall effect on formalization is positive. We estimate that in manufacturing, 14.6% of manufacturing workers became more formal, and 5% in the following decade. Those results are economically substantial: out of a manufacturing workforce of approximately 24 million in 1990, they imply that 3.5 million workers formalized in the 1990s, and an additional 1.2 million in the following decade.

Our findings also have clear policy implications. First, when liberalizing trade in developing economies, the effect of access to new varieties of foreign inputs on formalization should not be underestimated. Second, domestic labor market institutions induce a differential impact of trade reforms in contexts where informality is widespread. In large countries such as India, characterized by diverse local institutions, identifying and measuring accurately the most relevant ones is especially important. Third, the timing of the reforms matters: the productivity gap between sectors varies over time, and so do the potential productivity gains from trade. Finally, researchers have long known that a country cannot skip the stage of industrialization on its development path (Panagariya, 2004). It is therefore crucial to better understand how to design policies to strengthen manufacturing performance, and ultimately economic and social development.

References

Acemoglu, D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016): "Import Competition and the Great US Employment Sag of the 2000s," Journal of Labor Economics, 34, S141–S198.

Aghion, P., R. Burgess, S. J. Redding, and F. Zilibotti (2008): "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India," American Economic Review, 98, 1397–1412.

Ahsan, R. N., A. Ghosh, and D. Mitra (2017): "International trade and unionization: Evidence from India," Canadian Journal of Economics, 50, 398–425.

Almeida, R. and P. Carneiro (2012): "Enforcement of Labor Regulation and Informality," American Economic Journal: Applied Economics.

Amirapu, A. and M. Gechter (2019): "Labor Regulations and the Cost of Corruption: Evidence from the Indian Firm Size Distribution," The Review of Economics and Statistics, 102, 34–48.

Amiti, M. and J. Konings (2021): "Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia," American Economic Review, 97(5), 1611–1638.

Atkin, D. and D. Donaldson (2022): "The role of trade in economic development," in Handbook of International Economics: International Trade, Volume 5, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Elsevier, vol. 5 of Handbook of International Economics, 1–59.

Atkin, D., A. Khandelwal, L. Boudreau, R. Dix-Carneiro, I. Manelici, P. Medina, B. McCaig, A. Morjaria, L. Pascali, B. Rijkers, and M. Startz (2022): "International Trade," VoxDevLit, 4, 1–45.

Atkin, D. and A. K. Khandelwal (2020): "How Distortions Alter the Impacts of International Trade in Developing Countries," Annual Review of Economics, 12, 213–238.

Attanasio, O., P. Goldberg, and N. Pavcnik (2004): "Trade reforms and wage inequality in Colombia," Journal of Development Economics, 74, 331–366.

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," American Economic Review, 103, 2121–68.

Bamber, G. J., R. D. Lansbury, and N. Wailes (2011): "Employment relations in India," in International Comparative Employment Relations: Globalization and Change, ed. by V. C. S. Ratnam and A. Verma, Thousand Oaks, CA: Sage Publications.

BANERJEE, A. V. AND E. DUFLO (2005): "Growth Theory through the Lens of Development Economics," in Handbook of Economic Growth, Vol. 1A, Amsterdam: North-Holland, 473–552.

BAS, M. AND P. BOMBARDA (2023): "Foreign Technology and Informal Employment: Evidence from Mexico," Working paper.

BATRA, A. (2022): India's Trade Policy in the 21st Century, Routledge.

BEN YAHMED, S. AND P. BOMBARDA (2020): "Gender, Informal Employment and Trade Liberalization in Mexico," World Bank Economic Review, 34(2), 259–283.

BESLEY, T. AND R. BURGESS (2004): "Can Labor Regulation Hinder Economic Performance? Evidence from India," The Quarterly Journal of Economics, 119, 91–134.

BHATTACHARJEA, A. (2006): "Labour Market Regulation and Industrial Performance in India: A Critical Review of the Empirical Evidence," The Indian Journal of Labour Economics, 49, 211–32.

BOSCH, M., E. GOÑI-PACCHIONI, AND W. MALONEY (2012): "Trade liberalization, labor reforms and formal–informal employment dynamics," Labour Economics, 19, 653–667, special Section on: Price, Wage and Employment Adjustments in 2007-2008 and Some Inferences for the Current European Crisis.

BOSCH, M. AND W. F. MALONEY (2010): "Comparative analysis of labor market dynamics using Markov processes: An application to informality," Labour Economics, 17, 621–631.

BUSSOLO, M. AND S. SHARMA (2022): "Hidden Potential RethinkingInformalityinSouthAsia," Tech. rep., South Asia Development Forum.

CASELLI, F. (2005): "Accounting for Cross-Country Income Differences," in Handbook of Economic Growth, ed. by P. Aghion and S. Durlauf, Elsevier, vol. 1A of Handbook of Economic Growth, 679–741.

CHAKRABORTY, P., D. MITRA, AND A. SUNDARAM (2024): "Import competition, labor market regulations, and firm outsourcing," Journal of development Economics, 168.

CHAKRABORTY, P., R. SINGH, AND V. SOUNDARARAJAN (2021): "Import Competition, Formalization, and the Role of Contract Labor," Working Paper.

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

CHAUREY, R., G. CHIPLUNKAR, AND V. SOUNDARARAJAN (2023): "Employment Protection Legislation and Informality: Theory and Evidence from India," Working Paper.

Cosar, A. K., N. Guner, and J. Tybout (2016): "Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy," American Economic Review.

Couttenier, M., N. Monnet, and L. Piemontese (2023): "The Economic Costs of Conflict: A Production Network Approach," Working Paper.

Cristea, A. D. and J. Lopresti (2024): "Organized Labor When Things Go South: Unions and the Labor Market Consequences of NAFTA," Working Paper.

De Mel, S., D. McKenzie, and C. Woodruff (2009): "Measuring microenterprise profits: Must we ask how the sausage is made?" Journal of Development Economics, 88, 19–31.

De Soto, H. (1989): The Other Path: The Invisible Revolution in the Third World, Harper Row.

Dix-Carneiro, R., P. K. Goldberg, C. Meghir, and G. Ulyssea (2024): "Trade and Domestic Distortions: the Case of Informality," Working Paper.

Dix-Carneiro, R. and B. Kovak (2019): "Margins of labor market adjustment to trade," Journal of International Economics, 117, 125–142.

Dix-Carneiro, R. and B. K. Kovak (2017): "Trade Liberalization and Regional Dynamics," American Economic Review, 107, 2908–46.

Fafchamps, M., , D. McKenzie, S. Quinn, and C. Woodruff (2014): "Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana," Journal of Development Economics, 106, 211–226.

Goldberg, P. and N. Pavcnik (2003): "The Response of the Informal Sector to trade Liberalization," Journal of Development Economics, 72(2), 463–496.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010): "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India*," The Quarterly Journal of Economics, 125, 1727–1767.

Goldberg, P. K. and N. Pavcnik (2005): "Trade, wages, and the political economy of trade protection: evidence from the Colombian trade reforms," Journal of International Economics, 66, 75 – 105.

Gollin, D., D. Lagakos, and M. E. Waugh (2014): "The Agricultural Productivity Gap," Quarterly Journal of Economics, 939–93.

Guillerm, M. (2017): "Pseudo-panel methods and an example of application to Household Wealth data," Economie et Statistique / Economics and Statistics, 491-492, 109–130.

Gupta, P., R. Hasan, and U. Kumar (2009): "Big Reforms but Small Payoffs: Explaining the Weak Record of Growth in Indian Manufacturing," India Policy Forum.

Hakobyan, S. and J. McLaren (2016): "Looking for Local Labor Market Effects of NAFTA," The Review of Economics and Statistics, 98, 728–741.

Hasan, R., D. Mitra, and K. Ramaswamy (2007): "Trade Reforms, Labor Regulations, and Labor Demand Elasticities: Empirical Evidence from India," The Review of Economics and Statistics, 89, 466–481.

Hoseini, M. and O. Briand (2020): "Production efficiency and self-enforcement in value-added tax: Evidence from state-level reform in India," Journal of Development Economics.

Hsieh, C.-T. and P. J. Klenow (2009a): "Misallocation and Manufacturing TFP in China and India*," The Quarterly Journal of Economics, 124, 1403–1448.

——— (2009b): "Misallocation and Manufacturing TFP in India and China," The Quarterly Journal of Economics, 124.

Imbert, C. and J. Papp (2015): "Labor Market Effects of Social Programs: Evidence from India's Employment Guarantee," American Economic Journal: Applied Economics, 7, 33–63.

Kovak, B. K. (2013): "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?" The American Economic Review, 103, 1960–1976.

La Porta, R. and A. Shleifer (2008): "The Unofficial Economy and Economic Development," Brookings Papers on Economic Activity, 275–363.

——— (2014): "Informality and Development," Journal of Economic Perspectives, 28(3), 109–126.

Levy, S. (2008): Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico, Brookings Institution Press.

Maloney, W. F. (1999): "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico," World Bank Economic Review, 13, 275–302.

McCaig, B. (2011): "Exporting out of poverty: Provincial poverty in Vietnam and U.S. market access," Journal of International Economics, 85, 102–113.

McCaig, B. and N. Pavcnik (2018): "Export Markets and Labor Allocation in a Low-Income," American Economic Review, 7, 1899–1941.

McMillan, M. S. and D. Rodrik (2011): "Globalization, Structural Change and Productivity Growth," National Bureau of Economic Research Working Paper 17143.

Meghir, C., R. Narita, and J.-M. Robin (2015): "Wages and Informality in Developing Countries," American Economic Review, 105.

Mehrotra, S. (2019): "Informal Employment Trends in the Indian Economy: Persistent informality, but growing positive development," Tech. rep., International Labour Organization.

Melitz, M. J. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," Econometrica, 71, 1695–1725.

MOSPI (2012): IHSN Study Report, National Sample Survey Office, M/o Statistcs and Programme Implementation(MOSPI), Government of India (GOI), Vienna, Austria.

Nataraj, S. (2011): "The impact of trade liberalization on productivity: Evidence from India's formal and informal manufacturing sectors," Journal of International Economics, 85, 292–301.

Panagariya, A. (2004): "India's Trade Reform," Tech. rep., Columbia University.

——— (2007): "Why India Lags Behind China and How It Can Bridge the Gap," The World Economy, 30, 229–248.

Perry, G., W. Maloney, P. Fajnzylber, A. Mason, and J. Saavedra-Chanduvi (2007): "Informality: exit and exclusion," World Bank Latin America and Caribean Studies.

Ponczek, V. and G. Ulyssea (2022): "Enforcement of Labor Regulation and the Labor Market Effects of Trade: Evidence from Brazil," the Economic Journal, Volume 132(641).

Restrepo-Echavarria, P. (2014): "Macroeconomic Volatility: The Role of the Informal Economy," European Economic Review, 70, 454–469.

Restuccia, D. and R. Rogerson (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," Review of Economic Dynamics, 2, 707–720.

Routh, S. (2015): "Informal Workers' Aggregation in India: An Evolving Model of Collective Action," Sociologie et sociétés, 47, 177–200.

Soundararajan, V. (2019): "Heterogeneous effects of imperfectly enforced minimum wages in low-wage labor markets," Journal of Development Economics.

Srivastava, R. (2016): "Structural change and non-standard forms of employment in India," Tech. rep., International Labour Organization.

Teti, F. (2023): "Missing Tariffs," Working paper.

Topalova, P. (2005): "Trade Liberalization, Poverty, and Inequality: Evidence from Indian Districts," Working Paper 11614, National Bureau of Economic Research.

——— (2010): "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," American Economic Journal: Applied Economics, 2, 1–41.

Topalova, P. and A. Khandelwal (2011): "Trade Liberalization and Firm Productivity: The Case of India," The Review of Economics and Statistics, 93, 995–1009.

Ulyssea, G. (2018): "Firms, Informality, and Development: Theory and Evidence from Brazil," American Economic Review, 108, 2015–47.

——— (2020): "Informality: Causes and Consequences for Development," Annual Review of Economics, 12, 525–46.

Wacziarg, R. and J. S. Wallack (2004): "Trade liberalization and intersectoral labor movements," Journal of International Economics, 64.

WTO (2007): Trade Policy Review, World Trade Organization, Geneva, Switzerland.

Appendices

A Descriptive Statistics

This section provides additional details about the datasets used.

A.1 Description of Tariffs

Figure A.1: Average Indian industry tariff level, 1990-2010

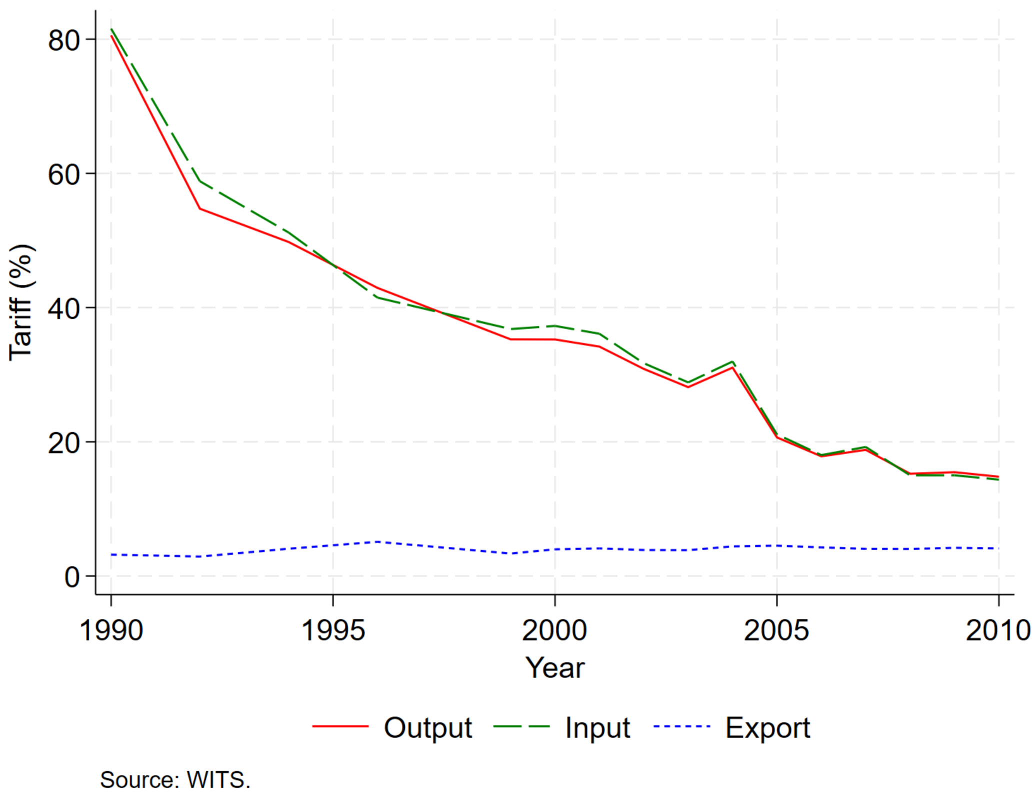


Table A.1: Description of tariffs for 1990, 2000 and 2010

	Mean	SD	Min.	Max.	Obs.
1990					
$\tau_{O,jt}$	0.81	0.35	0.00	2.81	104
$\tau_{I,jt}$	0.82	0.37	0.04	1.80	104
$\tau_{X,jt}$	0.03	0.06	0.00	0.42	104
2000					
$\tau_{O,jt}$	0.35	0.20	0.00	1.94	115
$\tau_{I,jt}$	0.37	0.20	0.02	0.81	115
$\tau_{X,jt}$	0.04	0.18	0.00	1.75	115
2010					
$\tau_{O,jt}$	0.15	0.20	0.00	1.50	116
$\tau_{I,jt}$	0.14	0.14	0.01	0.50	116
$\tau_{X,jt}$	0.04	0.17	0.00	1.75	116

Notes: $\tau_{O,jt}$, $\tau_{I,jt}$ and $\tau_{X,jt}$ are the output, input and export tariffs for industry j at time t.

Figure A.2: Average Indian input tariff for the main trade partners of India, 1999-2010

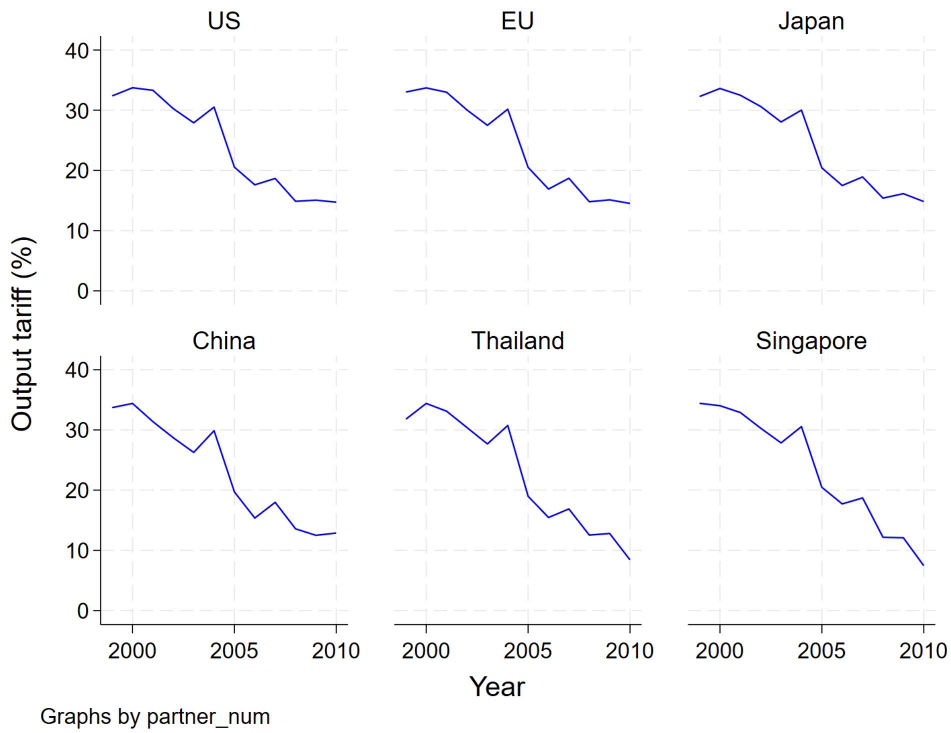


Figure A.3: Average Indian tariff for select two-digit industries, 1999-2010

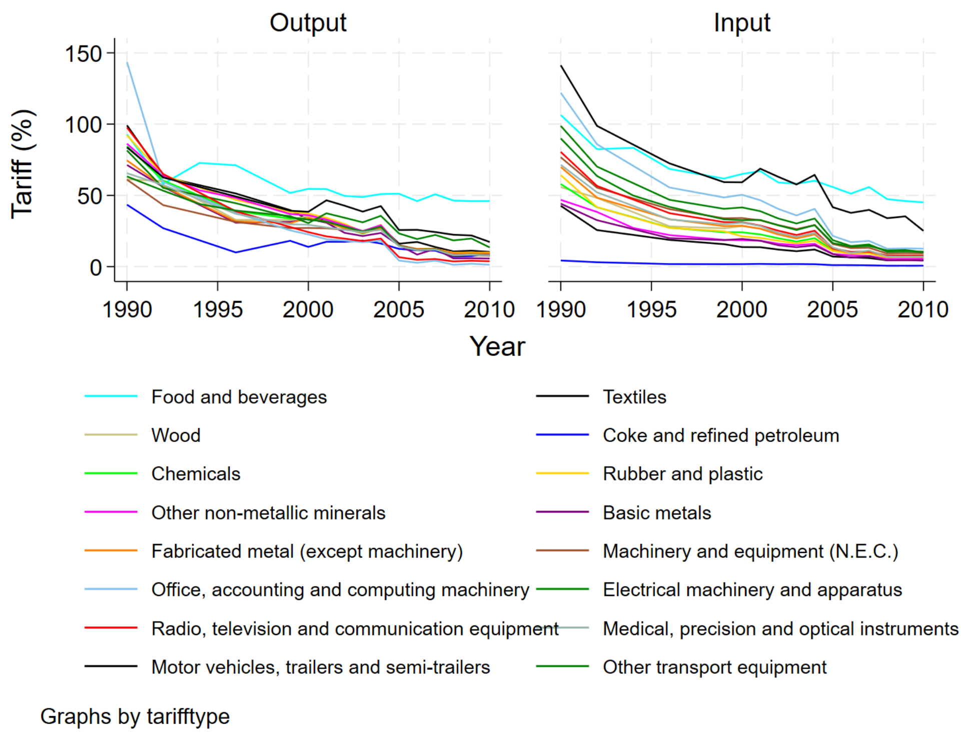
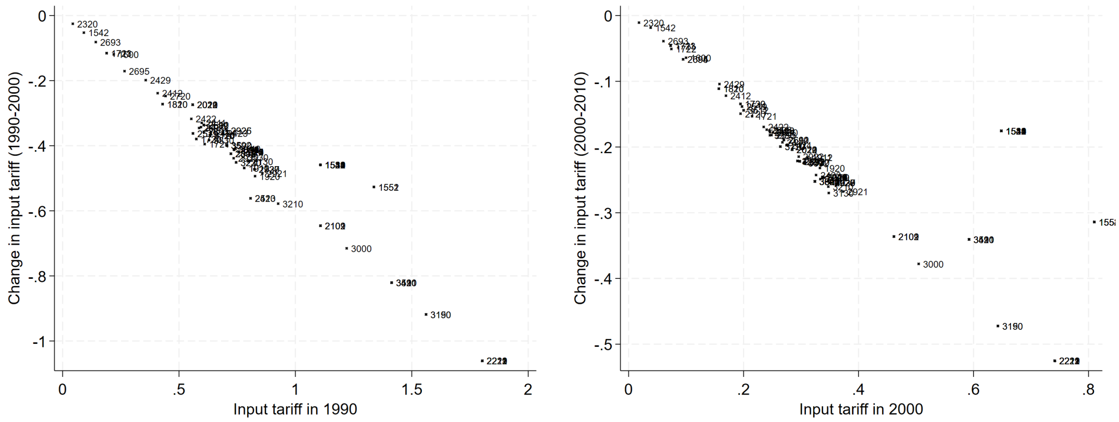


Figure A.4: Input tariff changes and initial tariff level, by decade

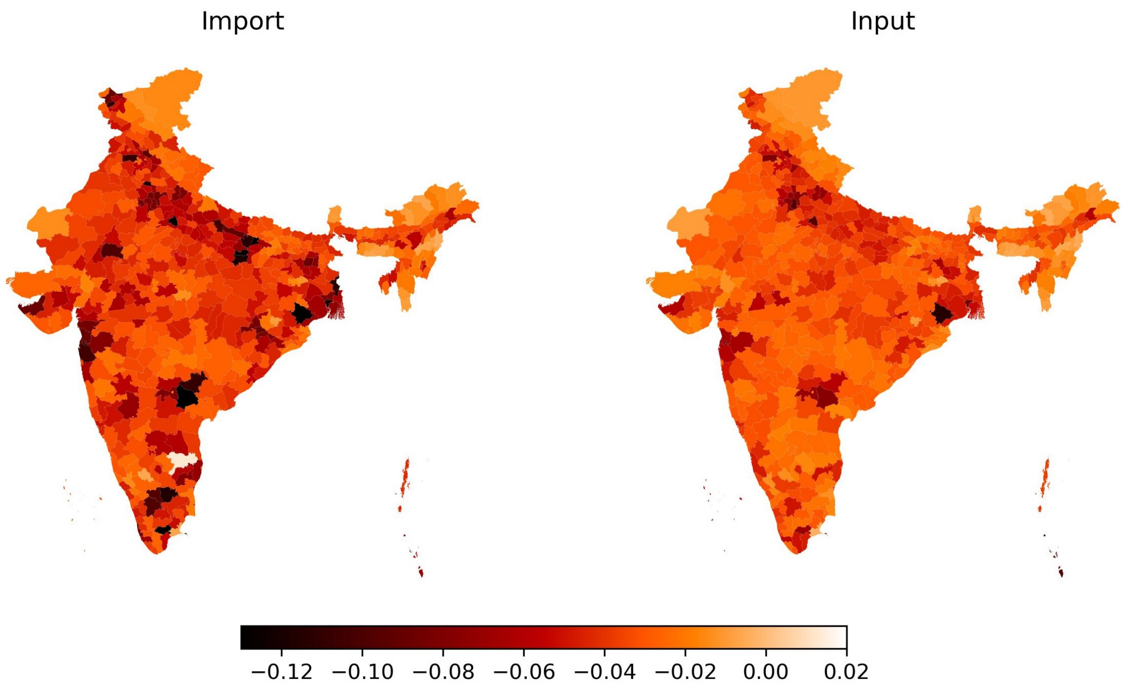


(a) 1990-1999

(b) 2000-2010

Source: Authors' calculation based on WITS. 1511 Manufacture of meat products; 1520 Manufacture of dairy products; 1542 Manufacture of sugar; 1551, 1552, and 1553 Manufacture of alcoholic beverages.

Figure A.5: Average district tariff change between 1999 and 2010



A.2 Statistics on Labor Market Institutions

Table A.2: Labor market regulations and industry-state characteristics

	Mean	SD	Min	Max
Labor market regulations				
Pro-worker$_s$	0.32	0.47	0.00	1.00
Share of unionized labor$_s^{1993}$	0.13	0.08	0.03	0.37
Regional office$_{st}$	0.84	0.37	0.00	1.00
Industry-state characteristics				
Share of ASI labor$_{jst}$	0.33	0.31	0.00	1.00
Share of contract labor in ASI firms$_{jst}$	0.13	0.16	0.00	0.88
Share of ASI firms$_{jst}$	0.10	0.18	0.00	0.96
Share of ASI capital$_{jst}$	0.52	0.35	0.00	1.00
Share of ASI output$_{jst}$	0.07	0.24	-0.88	5.51
Observations	828			

Notes: Calculations based on the state-industry panel in 2000. Only manufacturing.

Figure A.6: Location of the regional labor offices

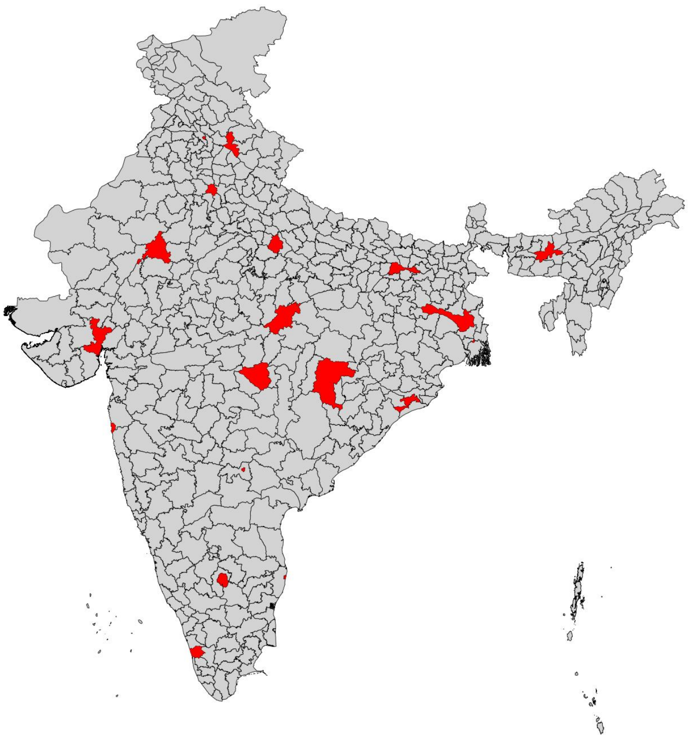
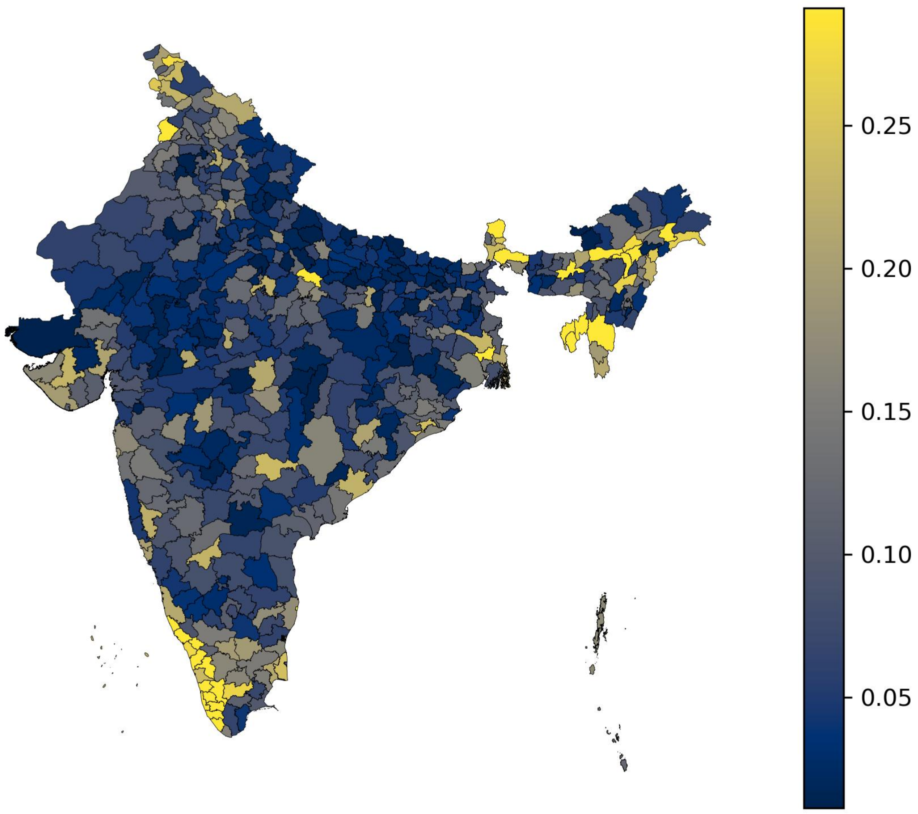


Figure A.7: Unionization shares by district in 2009



A.3 ASI and UMES

The ASI has been collected annually since 1960. It is based on accounting years starting on 1st April and ending on 31st March. We assign the year as the start of the accounting year. For instance, the accounting year starting on 1st April 1994 and ending on 31st March 1995 is assigned the year 1994. Despite some changes over time, the structure of ASI is consistent enough to recover key firm-level information over those two decades. We recover information about location (state and district), urban or rural status, year of initial operation, 5-digit industry, number of employees, wages and contribution to benefits, fixed assets, gross sales value. UMES is collected every five years, starting in 1990. For some years, they also provide information about services. In that case, we restrict the data to manufacturing establishments only. For later years, UMES data offer valuable information about the reality of operating informally in India, which we use in section 5 for the description of the institutional setting. We restrict data collection to the years for which corresponding informal data was available: 1990, 1994, 2000, 2005 and 2010.[43]

We combine ASI and UMES following Hoseini and Briand (2020) and Chakraborty et al. (2021) to created state-industry level panels. This implies using sampling weights provided by the surveys. Those weights are provided in both surveys and can be used to estimate the approximate number of establishments with similar characteristics including for the first stage units of the sampling frame. For ASI, this is straightforward: the frame is designed to ensure accurate coverage of all industries that are present in a given state, so even small industries are accounted for.[44] For UMES, the sample is not designed separately for industries within a district, so the weights should be representative at the district level only. This implies that, in theory, industries with few firms within a district might not be accounted for if they are too small. We argue that this issue does not threaten our approach in practice, for two main reasons. First, as argued by Hoseini and Briand (2020), UMES focus on informality. Unlike formal firms, informal firms exist in large numbers, so random sampling within a district makes it unlikely that informal firms in an industry would be missed. Second, we aggregate the results at the state-industry, so even if sampling was biased in one district to underestimate a particular industry, there is no reason to believe that the same bias would hold in another district in the same state.

We then merge those two datasets by state-industry-year. We create a balanced panel in which we only keep state-industry variables that are present for all the years and provide information about both ASI and UMES. We refer to the resulting panel as the ASI-UMES panel. As argued above, the ASI-UMES panel should cover all manufacturing establishments in the state-industry.

[43] To harmonize the NIC categories across surveys over time, we use concordance tables provided by the Ministry of Statistics and Programme Implementation.

[44] Importantly, industries are representative at the 4-digit level (NIC-1998 classification), the same level of precision as our explanatory variables, tariffs.

Table A.3: Summary of data

Dataset	Source	Description	Years
Annual Survey of Industries (ASI)	MOSPI	Survey data on formal manufacturing firms. Contains information on employment, revenue, expenditures, intermediate inputs and capital stock.	1990, 1994, 2000, 2005, 2010
Unorganized Manufacturing Survey (UMES)	NSSO	Survey data on informal manufacturing firms. Contains information on employment, revenue, expenditures and capital stock.	1990, 1994, 2000-2001, 2005-2006, 2010-2011
Employment Unemployment Survey (NSSEU)	NSSO	Survey data on employment. Contains information on household and personal characteristics, activity status during the previous year and week, earnings and benefits.	1987-1988, 1993-1994, 1999-2000, 2004-2005, 2009-2010.
Economic Census (EC)	CSO	All units engaged in the production or distribution of goods or services other than for the sole purpose of own consumption.	1990, 1998

Table A.4: Relative size of ASI and UMES

	ASI	UMES	$\frac{ASI}{ASI+UMES}$
Firms	117.63	17,172.69	0.01
Labor	8,515.59	36,781.56	0.19
Capital	495,410.95	110,641.65	0.82
Output	1,636,564.26	891,986.07	0.65
Emoluments	78,950.49	49,853.69	0.61

Notes: Calculation based on ASI and UMES data for 2005 using sampling weights. ASI and UMES are only for manufacturing industries. Firms and labor, for *ASI* and *UMES* expressed in 1000's, while capital, output, and emoluments are reported in 10,000,000's.

Table A.5: Description of the firm data (ASI and UMES)

	1990		2000		2010	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Annual Survey of Industries (ASI)						
Labor	107.44	408.91	151.50	627.84	173.15	661.35
Capital per labor	97217.84	330778.48	438839.54	12727037.46	578406.00	4873631.00
Emoluments per sales	118.96	22620.70	2.08	89.52	0.35	9.20
Plus 3 years	0.86	0.35	0.87	0.33	0.86	0.35
Importer	.	.	0.15	0.36	0.16	0.37
More than 3 years	0.72	0.45	0.80	0.40	0.79	0.40
Observations	45716		29533		33273	
Panel B: Unorganized manufacturing surveys (UMES)						
Labor	1.69	1.08	3.04	4.87	3.08	8.75
Capital per labor	40632.03	275501.45	40439.12	82352.61	94703.56	567592.27
Emoluments per sales	0.08	0.91	0.09	1.01	0.11	2.69
Plus 3 years	0.75	0.43	0.90	0.30	0.76	0.43
Registered	0.16	0.37	0.22	0.41	0.26	0.44
Observations	89516		199904		86384	

Notes: Only manufacturing.

Table A.6: Challenges Faced by UMES Firms

	Mean / SD
No electricity	0.07
	(0.25)
Power cuts	0.25
	(0.43)
Shortage of capital	0.47
	(0.50)
Non-availability of raw materials	0.12
	(0.33)
Problems with marketing	0.21
	(0.41)
Observations	84,262

Notes: Calculations are based on micro-level UMES data for 2005. All variables are binary, taking the value 1 for "yes."

A.4 NSSEU

Worker-level data is sourced from the National Sample Survey Organisation's Employment and Unemployment Surveys (NSSEU). It contains information on household and individual characteristics, usual and subsidiary activity during the previous year (including employer's characteristics for wage workers), and time use during a reference week. A key advantages of the NSSEU is its extensive coverage, encompassing both urban and rural areas as well as the majority of Indian districts. To account for oversampling of certain households, the NSSEU provides sampling weights, ensuring that the data is representative at the local level Imbert and Papp (2015).[45] [46]

Table A.7: Description of Worker Level Data

	All Industries			**Manufacturing**		
	All	Informal	Formal	All	Informal	Formal
Age	35.71	35.09	40.97	33.82	33.46	36.97
	(11.91)	(11.99)	(9.70)	(11.44)	(11.48)	(10.55)
Woman	0.27	0.28	0.17	0.27	0.28	0.14
	(0.44)	(0.45)	(0.37)	(0.44)	(0.45)	(0.34)
Not Literate	0.28	0.31	0.04	0.24	0.26	0.08
	(0.45)	(0.46)	(0.20)	(0.43)	(0.44)	(0.28)
High School	0.11	0.10	0.16	0.12	0.11	0.17
	(0.31)	(0.31)	(0.37)	(0.32)	(0.31)	(0.38)
Urban	0.32	0.29	0.58	0.51	0.49	0.72
	(0.47)	(0.46)	(0.49)	(0.50)	(0.50)	(0.45)
Formal	0.10	0.00	1.00	0.10	0.00	1.00
	(0.31)	(0.00)	(0.00)	(0.30)	(0.00)	(0.00)
Wage Worker	0.42	0.35	1.00	0.43	0.37	1.00
	(0.49)	(0.48)	(0.00)	(0.50)	(0.48)	(0.00)
Firm Size						
1-5 Workers	0.64	0.75	0.14	0.61	0.68	0.03
	(0.48)	(0.43)	(0.35)	(0.49)	(0.47)	(0.16)
6-9 Workers	0.08	0.08	0.09	0.08	0.09	0.03
	(0.27)	(0.26)	(0.28)	(0.27)	(0.28)	(0.16)
10-19 Workers	0.07	0.05	0.13	0.07	0.07	0.06
	(0.25)	(0.22)	(0.34)	(0.25)	(0.25)	(0.24)
20+ Workers	0.13	0.06	0.45	0.17	0.10	0.74
	(0.34)	(0.23)	(0.50)	(0.38)	(0.30)	(0.44)
Observations	217,377	194,726	22,651	24,528	22,041	2,487

Calculations are based on the NSSEU data for 2004–2005. The definition of informality is based on retirement benefits. *All Industries* includes primary, manufacturing, and services industries. *Manufacturing* is a subsample of 4-digit NIC-1998 codes between 1500 and 4000. *All*, *Formal*, and *Informal* categories represent all workers, formal workers, and informal workers, respectively.

[45] The definition of "local" varies by survey year due to changes in the NSSEU's sampling design. Up to and including round 55 (1999–2000), the data was representative at the district level in rural areas and at the region level in urban areas. From subsequent rounds onward, representativity was extended to the district level for both rural and urban areas.

[46] To harmonize the NSSEU classifications across surveys over time we use concordance tables provided by the Ministry of Statistics and Programme Implementation.

A.5 Additional Results

The role of capital intensity. The goal of this table is to study the interactions between input tariff and capital intensity. Industry-states are classified as high capital intensive if their capital-labor ratio is greater than the median, and as low capital intensive otherwise.

Table A.8: Industry-state-level regressions: heterogeneity by capital intensity

	Panel A: 1990-2000		**Panel B: 2000-2010**	
	(1)	(2)	(3)	(4)
VARIABLES	Registered	Provident fund	Registered	Provident fund
$\tau_{O,jt}$	0.148**	0.132*	0.238**	0.221***
	(0.074)	(0.067)	(0.092)	(0.079)
$\tau_{I,jt}$ × Low c_j	-0.429***	-0.401***	-0.425***	-0.421***
	(0.045)	(0.048)	(0.066)	(0.060)
$\tau_{I,jt}$ × High c_j	-0.519***	-0.434***	-0.665***	-0.637***
	(0.074)	(0.078)	(0.087)	(0.083)
State × year FE	Yes	Yes	Yes	Yes
2-industry × year FE	Yes	Yes	Yes	Yes
State × 3-industry FE	Yes	Yes	Yes	Yes
Trends_{jt}	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes
Observations	1,934	1,934	2,019	2,019
R-squared	0.861	0.863	0.788	0.794

Notes: OLS estimation split by decade. Panel A is restricted to rounds 1990, 1994 and 2000, and Panel B to rounds 2000, 2005 and 2010. Only manufacturing sectors. We use two definition of formal employment. *Registered* is the share of workers employed in firms belonging to the ASI dataset over total employment (in ASI and UMES) in the industry-state. *Provident Fund* is the share of workers employed in firms reporting having positive expenses for their employees' provident fund over total employment (in ASI and UMES) in the industry-state. Tariffs are measured at the industry j at time t. Low and High c_j indicate industries with below and above median capital intensity, respectively, where capital intensity is defined as the initial capital-to-labor ratio in a given industry and decade. $\textit{Trends}_{jt}$ accounts for industry dynamics at four-digit, constructed interacting initial employment level for each decade with year dummies. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Additional Results and Robustness Checks

This section presents robustness checks using ASI and UMES data.

Estimation without splitting the sample by decade. This table follows the same specification as table , but does not split the results by decade.

Table B.1: Industry-state-level regressions: both decades

Within-industry analysis

VARIABLES	(1) Registered	(2) Registered	(3) Provident Fund
$\tau_{O,jt}$	0.172**	0.170**	0.164**
	(0.075)	(0.078)	(0.069)
$\tau_{I,jt}$	-0.558***	-0.525***	-0.492***
	(0.050)	(0.047)	(0.045)
State × year FE	Yes	Yes	Yes
2-industry × year FE	Yes	Yes	Yes
State × 3-industry FE	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes
Trend$_{jt}$	No	Yes	Yes
Observations	3,291	3,291	3,291
R-squared	0.795	0.804	0.812

Notes: OLS estimation using 1990, 1994, 2000, 2005 and 2010. *Registered* is the share of firms belonging to the ASI dataset over all firms (ASI and UMES) in the industry-state. *Provident Fund* is the share of firms reporting having positive expenses for their employees' provident fund, over all firms (ASI and UMES) in the industry-state. Tariffs in industry j at time t. Only manufacturing sectors. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

PUT A SIMILAR TABLE USING LONG RUN DIFFERENCES

The share of formal firms. The dependent variable is the share of formal (ASI) firms over total (ASI and UMES) firms.

Table B.2: Industry-state-level regressions: share of formal firms

Within-industry analysis			
	(1)	(2)	(3)
VARIABLES	Registered	Registered	Provident fund
$\tau_{O,jt}$	0.061**	0.057**	0.051**
	(0.027)	(0.025)	(0.020)
$\tau_{I,jt}$	-0.086***	-0.037*	-0.031*
	(0.020)	(0.021)	(0.017)
State × year FE	Yes	Yes	Yes
2-industry × year FE	Yes	Yes	Yes
State × 3-industry FE	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes
Observations	3,291	3,291	3,291
R-squared	0.589	0.617	0.622

Notes: OLS estimation by decades using 1990, 1994, 2000 , 2005 and 2010. *Registered* is the share of firms belonging to the ASI dataset over all firms (ASI and UMES) in the industry-state. *Provident Fund* is the share of firms reporting having positive expenses for their employees' provident fund, over all firms (ASI and UMES) in the industry-state. Tariffs in industry j at time t. $Trends_{jt}$ is four-digit industry number of firms interacted with year. Only manufacturing sectors. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The role of other reforms. This table focuses explicitly controls for role of contemporaneous industrial reforms. The sample is restricted to the years 1990, 1994 and 2000. The two policies considered are FDI liberalization and industry delicensing (Aghion et al., 2008).

Table B.3: Industry-state-level regressions: the role of other reforms

Within-industry analysis			
	(1)	(2)	(3)
VARIABLES	Registered	Registered	Provident fund
$\tau_{O,jt}$	0.142*	0.123	0.111
	(0.075)	(0.077)	(0.071)
$\tau_{I,jt}$	-0.468***	-0.492***	-0.436***
	(0.052)	(0.049)	(0.056)
FDI reform_{jt}		0.212***	0.184***
		(0.064)	(0.063)
Delicensed_{jt}		-0.044*	-0.042*
		(0.024)	(0.022)
State × year FE	Yes	Yes	Yes
2-industry × year FE	Yes	Yes	Yes
State × 3-industry FE	Yes	Yes	Yes
Trend_{jt}	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes
Observations	1,934	1,778	1,778
R-squared	0.860	0.866	0.868

Notes: OLS estimation by decades using 1990, 1994 and 2000. *Registered* is the share of firms belonging to the ASI dataset over all firms (ASI and UMES) in the industry-state. *Provident Fund* is the share of firms reporting having positive expenses for their employees' provident fund, over all firms (ASI and UMES) in the industry-state. Tariffs in industry j at time t. $Trends_{jt}$ is four-digit industry labor interacted with year. FDI $\text{reform}_j t$ is the share of 6-digit HS code within an industry that were automatically open for FDI after 1991. $\text{Delicensed}_j t$ is a binary variable taking value one if the industry had been delicensed by year t. Only manufacturing sectors. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C The role of labor market institutions

Classification of state-level regulations: We follow the most recent classification for labor market regulation, where we classify each state as pro-worker or pro-employer (Besley and Burgess, 2004, Gupta et al. (2009), Chakraborty et al. (2024) among others).

Table C.1: Trade liberalization and labor market regulations

Dependent variables	Share of formal workers in industry j and state s				
	Full sample			**High-LMI**	**Low-LMI**
	(1)	(2)	(3)	(4)	(5)
$\tau_{O,jt}$	0.209***	0.276***	0.222***	0.527	0.220***
	(0.078)	(0.105)	(0.083)	(0.360)	(0.083)
$\tau_{I,jt}$	-0.489***	-0.398***	-0.404***	-0.944***	-0.360***
	(0.135)	(0.129)	(0.127)	(0.164)	(0.135)
$\tau_{O,jt} \times$ pro-W$_s$	0.304		0.284		
	(0.191)		(0.175)		
$\tau_{I,jt} \times$ pro-W$_s$		-0.401*	-0.385**		
		(0.210)	(0.191)		
Observations	1,972	1,972	1,972	637	1,329
R-squared	0.775	0.776	0.777	0.802	0.771
State $\times$ year FE	Yes	Yes	Yes	Yes	Yes
2-industry $\times$ year FE	Yes	Yes	Yes	Yes	Yes
3-industry $\times$ state FE	Yes	Yes	Yes	Yes	Yes
Industry trends	Yes	Yes	Yes	Yes	Yes
Export tariff	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimation. The dependent variable is the share of ASI-employed labor over total labor in industry j, state s and time t in all regressions. The aggregated firm panel for the years 2000, 2005 and 2010 is used for columns (1) and (2), a sample restricted to high-LMI states in column (3) and to low-LMI states in column (4). *High LMI* is a subsample of states with pro-worker labor laws, and *low LMI* is a subsample of states with pro-employer and neutral laws. Only manufacturing sectors. All regressions include 4-digit industry employment trends and control for export tariffs. Heteroskedasticity-robust standard errors clustered by 4-digit industries-state are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Labor Unions: here we explore the specific role of the degree of unionization in a state.

Table C.2: Trade liberalization and unionization

Dependent variables	Share of formal workers in industry j and state s					
	Full sample				**High-union**	**Low-union**
	(1)	(2)	(3)	(4)	(5)	(6)
$\tau_{O,jt}$	0.203***	0.273***	0.213***	0.212***	0.517*	0.216***
	(0.077)	(0.104)	(0.081)	(0.079)	(0.299)	(0.077)
$\tau_{I,jt}$	-0.488***	-0.411***	-0.418***	-0.430***	-0.774***	-0.386***
	(0.136)	(0.136)	(0.133)	(0.134)	(0.215)	(0.144)
$\tau_{O,jt}$ × High unionization$_s$	0.268*		0.249*			
	(0.156)		(0.144)			
$\tau_{I,jt}$ × High unionization$_s$		-0.271	-0.251			
		(0.225)	(0.212)			
$\tau_{O,jt}$ × High historical unionization$_s$				0.250*		
				(0.146)		
$\tau_{I,jt}$ × High historical unionization$_s$				-0.223		
				(0.220)		
State × year FE	Yes	Yes	Yes	Yes	Yes	Yes
2-industry × year FE	Yes	Yes	Yes	Yes	Yes	Yes
State × 3-industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Export tariff	Yes	Yes	Yes	Yes	Yes	Yes
Industry trend	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,290	2,290	2,290	1,936	1,126	1,158
R-squared	0.774	0.774	0.775	0.776	0.785	0.780

Notes: OLS estimation by decades using 1990, 1994, 2000, 2005 and 2010. Results reported in columns 1 to 4 are for all states, in column 5 the sample is restricted to above-median unionization states, and column 6 to below-median unionization states. High unionization$_s$ takes value 1 if the share of unionized workers in state s in 1993 is above median, and 0 otherwise. High historical unionization$_s$ takes value 1 if the historical unionization value taken from Aghion et al. (2008) is above median, and 0 otherwise. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Enforcement: To evaluate the heterogeneous effects of trade across states, in this part we use an alternative measure of labor market institutions, enforcement. It is measured by the distance to the closest regional labor office in charge with enforcing regulations.

Table C.3: Trade liberalization and enforcement

Within-industry analysis

VARIABLES	(1) Formal	(2) Formal	(3) Formal	(4) Formal	(5) Formal
$\tau_{O,jt}$	0.115**	0.146***	0.113**	0.137**	0.106*
	(0.047)	(0.049)	(0.046)	(0.054)	(0.054)
$\tau_{I,jt}$	-0.326***	-0.339***	-0.321***	-0.306***	-0.306***
	(0.067)	(0.074)	(0.070)	(0.071)	(0.054)
$\tau_{O,jt} \times$ Dist. LO$_d$	0.105*		0.110*	0.171**	0.150*
	(0.058)		(0.065)	(0.082)	(0.076)
$\tau_{I,jt} \times$ Dist. LO$_d$		0.007	-0.011	0.069	0.083
		(0.048)	(0.051)	(0.065)	(0.061)
$\tau_{O,jt} \times$ Dist. capital$_d$				-0.086	-0.062
				(0.066)	(0.061)
$\tau_{I,jt} \times$ Dist. capital$_d$				-0.098	-0.166**
				(0.064)	(0.070)
District $\times$ quarter-year FE	Yes	Yes	Yes	Yes	Yes
3- industry FE	Yes	Yes	Yes	Yes	No
3- industry FE $\times$ state FE	No	No	No	No	Yes
Individual char.	Yes	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes	Yes
Export tariff	Yes	Yes	Yes	Yes	Yes
Observations	34,490	34,490	34,490	34,490	34,148
R-squared	0.538	0.538	0.538	0.538	0.575

LPM estimation using the NSS-EU round 1999, 2004 and 2009. Tariff in industry j at time t. Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Only manufacturing industries. Enforcement is measured as the distance between a worker's district and the nearest "Chief Labor Commissioner Office,". *Dist.* LO_d takes value 1 if the district is above the median distance from the state's labour office and zero otherwise. *Dist. capital*$_d$ takes value 1 if the district is above the median distance from the state's capital and zero otherwise. Individual characteristics include age, square of age, years of education, household size, religion and urban location. Pseudo-individual FE include gender, state and year of birth and literacy. Heteroskedasticity-robust standard errors clustered by industry are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Employment protection legislation: we investigate the effect of an increase in the stringency of regulations made to ensure employers provide fair working conditions to employees. The reform in Andhra Pradesh in 2003 increased significantly the penalties for firms avoiding regulations by hiring contract labor.

Table C.4: Trade Liberalization and the Reform in Andhra Pradesh

Dependent variables	Share of formal workers in industry j and state s			
	Registered			**Provident Fund**
	(1)	(2)	(3)	(4)
$\tau_{O,jt}$	0.283	0.241**	0.332*	0.316*
	(0.181)	(0.106)	(0.177)	(0.164)
$\tau_{I,jt}$	-0.651***	-0.644***	-0.635***	-0.611***
	(0.138)	(0.135)	(0.138)	(0.130)
$\tau_{O,jt} \times$ Andhra Pradesh$_s \times$ Post$_t^{2003}$	-0.121		-0.217	-0.194
	(0.222)		(0.226)	(0.217)
$\tau_{I,jt} \times$ Andhra Pradesh$_s \times$ Post$_t^{2003}$		0.323	0.404*	0.402**
		(0.206)	(0.214)	(0.201)
State $\times$ year FE	Yes	Yes	Yes	Yes
2-industry $\times$ year FE	Yes	Yes	Yes	Yes
3-industry $\times$ state FE	Yes	Yes	Yes	Yes
Industry trends	Yes	Yes	Yes	Yes
Export tariff	Yes	Yes	Yes	Yes
Observations	2,302	2,302	2,302	2,302
R-squared	0.798	0.799	0.800	0.811

Notes: OLS estimation by decades using 1990, 1994, 2000, 2005 and 2010. Andhra Pradesh$_s$ takes value 1 if state s is Andhra Pradesh and 0 otherwise. Post$_t^{2003}$ takes value 1 if the year is after 2003 and 0 otherwise. Heteroskedasticity-robust standard errors clustered by 4-digit state-industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D Aggregate productivity

Table D.1: Productivity Gap and Formal-Informal Reallocation by tariff change

	All		**Large Δ tariff**		**Small Δ tariff**	
	Revenue based	Wage based	Revenue based	Wage based	Revenue based	Wage based
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 1990-2000						
Unadjusted	34.1	17.2	26.2	10.2	41.7	32.3
Adjusted by human capital	27.7	13.9	21.9	8.58	33.4	25.8
+ measurement error in revenue and time worked	17.2		13.6		20.8	
+ differences in output-labor elasticity	11.5		9.09		13.8	
Share of workers reallocated in the 1990s	0.146	0.146	0.211	0.211	0.111	0.111
Initial share of workers in the formal sector	0.209	0.209	0.253	0.253	0.160	0.160
Panel B: 2000-2010						
Unadjusted	16.0	6.68	10.8	4.53	24.4	10.4
Adjusted by human capital	12.9	5.42	9.06	3.79	19.5	8.33
+ measurement error in revenue and time worked	8.09		5.64		12.1	
+ differences in output-labor elasticity	5.39		3.76		8.12	
Share of workers reallocated in the 2000s	0.050	0.050	0.065	0.065	0.039	0.039
Initial share of workers in the formal sector	0.178	0.178	0.180	0.180	0.161	0.161

Notes: Panel A is based on 1994 data, and panel B on 2005 data. Formal sector information comes from ASI, informal sector information from UMES. The sample comprises all manufacturing industries in columns (1) and (2), industries facing above-median tariff reductions in columns (3) and (4), and industries facing below-median tariff reductions in columns (5) and (6). The productivity gap reported in columns (1), (3), and (5) is based on the average revenue product of labor, plus subsequent adjustments. The average revenue product of labor is the ratio of total revenue divided by total employment within each sector. The labor productivity gap reported in columns (2), (4), and (6) is based on the ratio of total annual earnings divided by total employment in each sector, plus subsequent adjustments. Human capital information comes from the NSSEU survey for round 1999-2000.

Table D.2: Productivity Gap and Formal-Informal Reallocation by labor market regulation

	All		**Pro-worker**		**Pro-employer**	
	Revenue based	Wage based	Revenue based	Wage based	Revenue based	Wage based
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 1990-2000						
Unadjusted	34.1	17.2	30.1	17.4	37.5	18.3
Adjusted by human capital	27.7	13.9	24.1	13.9	31.0	15.1
+ Measurement error in revenue and time worked	17.2		15.0		19.3	
+ Differences in output-labor elasticity	11.5		10.0		12.8	
Share of workers reallocated in the 1990s	0.146	0.146	0.145	0.145	0.148	0.148
Initial share of formal workers	0.209	0.209	0.215	0.215	0.190	0.190
Panel B: 2000-2010						
Unadjusted	16.0	6.68	15.2	5.33	18.3	8.37
Adjusted by human capital	12.9	5.42	12.1	4.26	15.1	6.91
+ Measurement error in revenue and time worked	8.09		7.59		9.42	
+ Differences in output-labor elasticity	5.39		5.06		6.28	
Share of workers reallocated in the 2000s	0.050	0.050	0.049	0.049	0.050	0.050
Initial share of formal workers	0.178	0.178	0.168	0.168	0.182	0.182

Notes: Panel A is based on 1994 data, and panel B on 2005 data. Formal sector information comes from ASI, informal sector information from UMES. The sample comprises all manufacturing industries in columns (1) and (2), industries located in pro-worker states in (3) and (4), industries located in pro-employer states in (5) and (6). The productivity gap reported in columns (1), (3), and (5) is based on the average revenue product of labor, plus subsequent adjustments. The average revenue product of labor is the ratio of total revenue divided by total employment within each sector. The labor productivity gap reported in columns (2), (4), and (6) is based on the ratio of total annual earnings divided by total employment in each sector, plus subsequent adjustments. Human capital information comes from the NSSEU survey for round 1999-2000.

A Supplementary Online Appendix using NSSEU

In this section, we extend our analysis to labor informality to corroborate our industry-state level results. Drawing on data from the NSSEU labor force survey, we classify workers as formal if their employer contributes to a retirement fund, and as informal if such contributions are absent. As NSSEU data do not track individuals over time, we employ methods to estimate individual fixed effects using independent repeated cross-sectional data (Guillerm, 2017). All estimations include fixed effects based on observable individual characteristics—specifically, year of birth, gender, municipality of residence, migration status (internal and international), and years of education. Our sample focuses on individuals aged between 15 and 65 who report to be employed in manufacturing industries, for which we have tariff information. We consider individuals whose primary activity is reported as either self-employment or wage employment. Self-employment are defined as those individuals working as helpers in a family business or as own-account workers. Wage employment encompasses workers directly hired by firms, regardless of the firm's accounting status or the duration of employment (both long-term and short-term). Since we focus on reallocations within the active labor force, our final sample excludes unemployed individuals, those currently undergoing training, and individuals whose primary activity is categorized as domestic duties.[47]

To estimate the impact of trade liberalization on the likelihood of formal employment, we employ a linear probability model. As in the firm-level analysis, our identification strategy relies on a difference-in-differences framework. This approach compares individuals in industries more exposed to liberalization—those experiencing larger tariff cuts—to individuals in less exposed industries, while controlling for a range of confounding factors. Our results are consistent with the firm-level findings. They show that the average 20-percentage-point reduction in output tariffs between 1990 and 2010 increased the likelihood of informality by approximately 3 percentage points, highlighting the impact of intensified import competition. In contrast, input-side liberalization—measured through reductions in input tariffs—decreased the probability of informality by over 7 percentage points, compared to individuals in industries less affected by trade liberalization. We further examine the heterogeneity of these effects and find that the impact of improved access to foreign inputs on formalization is mainly concentrated among male workers, urban residents, and individuals with medium to high levels of education.

A.1 Identification Strategy

This section presents the identification strategy using micro-level worker data from the NSSEU for the second period 2000–2010 to examine if the industry-level analysis is validated at the micro-level. We focus on the second period under analysis where this dataset is available for the same definition of informality. The identification strategy at the micro-level investigates the effects of

[47]Unemployed people are those who reported not having worked but were actively seeking employment and/or were available to work within the past 365 days.

tariff cuts on the likelihood of workers becoming formal in response.[48] We estimate the following linear probability model, on a sample of formal and informal workers in manufacturing industries:

$$F_{ijdt} = \alpha + \beta_1 \tau_{O,jt} + \beta_2 \tau_{I,jt} + \beta_3 \tau_{X,jt} + \beta_4 X_{ijdt} + \delta i + \gamma_{dt} + \mu_j + Trend_{jt} + \varepsilon_{ijst} \tag{1}$$

where F_{ijdt} is a binary variable taking value 1 if the worker i reports working for an employer who contributes to their provident fund (for retirement) in 4-digit industry j in district d, and 0 otherwise. As above, we include three trade channels that may reallocate workers between formal and informal work: output tariff $\tau_{O,jt}$, input tariff $\tau_{I,jt}$ and we control for export tariff $\tau_{X,jt}$. X_{ijdt} are a set of individual characteristics such as age squared, years of education, religion, and marital status. Since NSSEU data are a repeated cross-section, we need to rely on pseudo-panel methods to estimate individual fixed effects models when only independent repeated cross-sectional data are available (Guillerm, 2017). Thus, in all our estimations we include pseudo individual fixed effects, δ_i. These pseudo individual fixed effects are based on individual characteristics such as year of birth, gender, district, and years of education. To control for time-invariant industry characteristics that can be correlated with tariffs, we include a 3-digit industry fixed effect, μ_j. We also incorporate 4-digit industry-specific trends to capture the differential effects of trade liberalization policies implemented in the previous decade on industries during the 2000s. To account for time-varying shocks affecting local conditions, such as urbanization or local economic growth, we include district-time fixed effects γ_{dt}. This control is particularly important because international trade and formalization often coincide with urbanization and economic growth (McCaig and Pavcnik, 2018).

Our difference-in-differences framework relies on comparing workers employed in industries that were impacted differently by tariff changes, while being otherwise similar in characteristics. Next, we also study if the effects of tariff changes on the likelihood of becoming formal depends on labor market regulations across states by estimating equation (1) in two different sub-samples of workers located in states that have high- and low-labor market regulations.

A.2 Results

One limitation of using aggregated data is that we may not be measuring well the intensive margin of informality, *i.e.*, the possibility that formal firms hire informal workers. For instance, import competition may drive firms to reduce costs by hiring informal labor (Chakraborty et al., 2021). More specifically, in India, firms exploit the intensive margin of informality through hiring contract workers to whom a number of regulations do not apply.[49] Firms hire them temporarily, for instance, to avoid labor market regulations or to mitigate the impact of negative shocks (Chau-

[48] In a robustness check, we also use a longer period, 1990-2010, using a different definition of informality.

[49] While contract workers are in principle entitled to social security, they are typically worse off than regular workers in terms of working conditions (Srivastava, 2016).

rey, 2015). The aggregated nature of the analysis in section ?? does not allow us to explore that possibility.[50] This section seeks to overcome those limitations by relying on individual worker level data. Shifting to the individual level data offers two key advantages relative to the aggregate analysis based on the share of formal employment at the industry-state level. First, the use of micro-level data allows us to incorporate additional controls, helping to address potential endogeneity concerns. Second, it provides a more detailed understanding of the effects of trade on worker-level outcomes within manufacturing industries.

Table A.1: The effects of trade liberalization on the probability of becoming a formal worker

Dependent variable	Indicator variable equal to on if worker i's employer contributes to provident fund			
	(1)	(2)	(3)	(4)
Output τ_{jt}	0.122		0.128*	0.158***
	(0.082)		(0.071)	(0.051)
Input τ_{jt}		-0.380***	-0.383***	-0.369***
		(0.110)	(0.105)	(0.103)
District × year FE	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes
3-industry FE	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes
Export τ_{jt}	No	No	No	Yes
Observations	61,058	61,058	61,058	61,058
R-squared	0.435	0.438	0.439	0.440

Notes: Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Tariff in industry j at time t. Only manufacturing sectors. Individual characteristics include age, square of age, years of education, marital status, household size and urban location. Pseudo-individual FE include gender, year of birth, religion, literacy and district. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In order to investigate the effects of trade liberalization on the likelihood of becoming a formal worker, we estimate equation (1) using the linear probability model (LPM) during the second wave of trade liberalization. Table A.1 presents the results. These estimates validate those from the aggregated panel of firm outcomes. Workers operating in industries which were impacted by larger output tariff cuts were more likely to become informal, but those in industries that had more significant input trade liberalization become more informal (column 4). The 20 percentage point average reduction of output tariffs during this period suggest that import competition increased the probability of becoming informal by roughly 3 percentage points, whereas input-trade liberalization decreased this probability by more than 7 percentage points, compared to industries

[50]While ASI does provide employment data on contract workers at the firm level, that number may be underestimated since contract workers hired through unlicensed contractors do not appear in the data Srivastava (2016).

not impacted by trade liberalization and holding all other variables equal.

A.3 Worker-level regressions: heterogeneity

We now investigate potential heterogeneous responses to trade liberalization across location and worker types. First, we check whether urban workers benefit more from input trade liberalization. Results are presented in table A.2. In column (2), we include a dummy variable Urban_i. Its coefficient is positive and significant, indicating that workers in urban areas are more likely to be formal than those in rural areas. In column (3), we control for the possibility that some industries may be more formal because their operations take place in urban areas. In column (4), we interact the input tariff with the urban dummy to capture the differential effect of input tariff reductions in urban areas. The coefficient is negative and significant, indicating greater reallocations of workers due to access to foreign inputs in urban areas. Since all regressions in table A.2 include both pseudo individual FE and worker characteristics, those results do not capture systematic differences between urban and rural workers that would make urban workers more complementary to foreign inputs (*e.g.*, skill differences). Rather, those results are likely driven by challenges specific to rural areas (for instance, inability to source inputs due to low-quality infrastructure or low enforcement of regulations).

We next turn to the role of gender. Table A.3 shows that women are less likely to be formal than men. Those findings are similar to those in most developing economies in which women have worse labor market outcomes than men. In addition, women do not formalize as much when there is greater access to foreign inputs.

Third, we consider the role of skills, which are associated to greater formality rates. Column (2) of A.4 shows that an additional year of education increases the probability of working formally, and column (3) further shows that the results are not driven by different initial education levels between industries. The role of skill on reallocations between sectors following input liberalization is explored in column (4). The direction of the effect depends on the degree of complementarity between the imported input and skill. Results show that more skilled workers benefit from greater reallocations. Those findings are in line with those of Bas and Bombarda (2023) for Mexico, another developing economy which benefited from imported inputs from developed partners.

We also consider heterogeneity by caste and age. Results are presented in Tables A.5 and A.6. There does not seem to be a strong differential effect along those characteristics.[51] In short, workers are more likely to be formally employed through access to foreign inputs if they are male, live in urban areas, and have a medium level of education.

[51] As expected, workers in a low caste are more likely to be informal, but this effect disappears when we control for individual characteristics.

Table A.2: Heterogeneous effects of trade liberalization at the worker-level: urban location

Dependent variable	Indicator variable equal to one if worker i is formally employed			
	(1)	(2)	(3)	(4)
$\tau_{O,jt}$	0.133***	0.135***	0.152***	0.149***
	(0.048)	(0.048)	(0.049)	(0.050)
$\tau_{I,jt}$	-0.304***	-0.314***	-0.352***	-0.313***
	(0.078)	(0.080)	(0.084)	(0.091)
Urban$_i$		0.025***	0.023***	0.036***
		(0.007)	(0.007)	(0.011)
Urban share$_j^{1999}$			0.103	0.102
			(0.068)	(0.068)
$\tau_{I,jt}$ × Urban$_i$				-0.064*
				(0.036)
District × quarter-year FE	Yes	Yes	Yes	Yes
3- industry FE	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes
Observations	42,864	42,864	42,493	42,493
R-squared	0.540	0.541	0.540	0.540

Notes: Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Tariff in industry j at time t. Only manufacturing sectors. Individual characteristics include age, square of age, years of education, marital status, household size and urban location. Pseudo-individual FE include gender, year of birth, religion, literacy and district. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Heterogeneous effects of trade liberalization at the worker-level: gender

Dependent variable	Indicator variable equal to one if worker i is formally employed		
	(1)	(2)	(3)
$\tau_{O,jt}$	0.135***	0.139***	0.135***
	(0.048)	(0.049)	(0.050)
$\tau_{I,jt}$	-0.314***	-0.307***	-0.320***
	(0.080)	(0.078)	(0.079)
Woman share$_j^{1999}$		-0.060	-0.068
		(0.067)	(0.068)
$\tau_{I,jt} \times$ Woman$_i$			0.134**
			(0.062)
District $\times$ quarter-year FE	Yes	Yes	Yes
3- industry FE	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes
Observations	42,864	42,493	42,493
R-squared	0.541	0.540	0.540

Notes: Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Tariff in industry j at time t. Only manufacturing sectors. Individual characteristics include age, square of age, years of education, marital status, household size and urban location. Pseudo-individual FE include gender, year of birth, religion, literacy and district. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Heterogeneous effects of trade liberalization at the worker-level: skill

Dependent variable	Indicator variable equal to one if worker i is formally employed			
	(1)	(2)	(3)	(4)
$\tau_{O,jt}$	0.150***	0.135***	0.127***	0.127***
	(0.054)	(0.048)	(0.042)	(0.042)
$\tau_{I,jt}$	-0.346***	-0.314***	-0.293***	
	(0.088)	(0.080)	(0.051)	
Years of education$_i$		0.014***	0.013***	0.014***
		(0.001)	(0.001)	(0.002)
Average years of education$_j^{1999}$			0.029***	0.030***
			(0.006)	(0.006)
$\tau_{I,jt}$ × Low-skill$_i$				-0.241***
				(0.061)
$\tau_{I,jt}$ × Mid-skill$_i$				-0.294***
				(0.057)
$\tau_{I,jt}$ × High-skill$_i$				-0.311***
				(0.054)
District × quarter-year FE	Yes	Yes	Yes	Yes
3- industry FE	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes
Observations	42,881	42,864	42,493	42,493
R-squared	0.527	0.541	0.543	0.544

Notes: Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Low-skill$_i$, Mid-skill$_i$ and High-skill$_i$ are binary variable respectively taking value 1 if the worker has less than primary school education, has completed primary school, middle school or high school, or has a degree beyond high school. Tariff in industry j at time t. Only manufacturing sectors. Individual characteristics include age, square of age, years of education, marital status, household size and urban location. Pseudo-individual FE include gender, year of birth, religion, literacy and district. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Heterogeneous effects of trade liberalization at the worker-level: caste

Dependent variable	Indicator variable equal to one if worker i is formally employed				
	(1)	(2)	(3)	(4)	(5)
$\tau_{O,jt}$	0.141***	0.141***	0.139***	0.140***	0.133***
	(0.053)	(0.053)	(0.053)	(0.052)	(0.048)
$\tau_{I,jt}$	-0.327***	-0.329***	-0.334***	-0.331***	-0.316***
	(0.087)	(0.087)	(0.088)	(0.088)	(0.081)
Scheduled caste$_i$		-0.016**	-0.016**	-0.012	0.003
		(0.008)	(0.007)	(0.009)	(0.008)
Scheduled caste share$_j^{1999}$			-0.025	-0.025	-0.005
			(0.098)	(0.098)	(0.087)
$\tau_{I,jt}$ × Scheduled caste$_i$				-0.022	0.005
				(0.043)	(0.041)
District × quarter-year FE	Yes	Yes	Yes	Yes	Yes
3- industry FE	Yes	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes	Yes
Observations	42,881	42,881	42,510	42,510	42,493
R-squared	0.523	0.523	0.522	0.522	0.540

Notes: Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Caste$_i$ equals to 1 if the individual belongs to a scheduled caste. Scheduled caste share$_j^{1999}$ is the share of employment in industry j in 1999 belonging to a scheduled caste. Tariff in industry j at time t. Only manufacturing sectors. Individual characteristics include age, square of age, years of education, marital status, household size and urban location. Pseudo-individual FE include gender, year of birth, religion, literacy and district. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Heterogeneous effects of trade liberalization at the worker-level: age

Dependent variable	Indicator variable equal to one if worker i is formally employed		
	(1)	(2)	(3)
Developed output τ_{jt}	0.135***	0.102*	0.101*
	(0.048)	(0.054)	(0.054)
$\tau_{I,jt}$	-0.314***	-0.245***	
	(0.080)	(0.069)	
Average age_j^{1999}		0.017***	0.017***
		(0.004)	(0.004)
$\tau_{I,jt}$ × Age 15-24$_i$			-0.266***
			(0.064)
$\tau_{I,jt}$ × Age 25-34$_i$			-0.205***
			(0.070)
$\tau_{I,jt}$ × Age 35-44$_i$			-0.231***
			(0.079)
$\tau_{I,jt}$ × Age 45-54$_i$			-0.323***
			(0.093)
$\tau_{I,jt}$ × Age 55+$_i$			-0.239**
			(0.096)
Age^2	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
District × quarter-year FE	Yes	Yes	Yes
3- industry FE	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes
Observations	42,864	42,493	42,493
R-squared	0.541	0.542	0.543

Notes: Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Tariff in industry j at time t. Only manufacturing sectors. Individual characteristics include age, square of age, years of education, marital status, household size and urban location. Pseudo-individual FE include gender, year of birth, religion, literacy and district. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.4 Robustness

Appendix B show additional checks, where in Table A.7 detail the role of each individual characteristics, in Table A.8 we include additional trends, and in Table A.9 we check that the results are robust to differences in firm size category. We also consider two alternative definitions of labor informality, defining formality regular workers and wage workers. Table A.10 presents the results. The coefficient attached to the output tariff is never significant, unlike in the baseline analysis. This suggest that workers in industries with greater import competition are not less likely to be regular workers or wage workers, while the baseline results suggest they are less likely to receive benefits. Those results are not contradictory in a setting where enforcement is imperfect: formal firms facing tougher competition may simply decide to stop paying formality costs attached to employing benefit-receiving workers. Workers may tolerate this loss of benefits despite their status as regular or wage workers if they have lower formal employment opportunities, which is likely to be the case in an industry facing import competition. On the other hand, the coefficient attached to the input tariff is negative and significant, in line with the baseline results.

Table A.7: The role of personal characteristics

VARIABLES	Formal (1)	Formal (2)	Formal (3)	Formal (4)	Formal (5)	Formal (6)	Formal (7)	Formal (8)
Output τ_{jt}	0.132**	0.132**	0.132**	0.139**	0.139**	0.128**	0.131**	0.159***
	(0.058)	(0.057)	(0.057)	(0.056)	(0.056)	(0.053)	(0.053)	(0.050)
Input τ_{jt}	-0.394***	-0.391***	-0.393***	-0.383***	-0.382***	-0.359***	-0.365***	-0.371***
	(0.129)	(0.129)	(0.128)	(0.127)	(0.127)	(0.113)	(0.114)	(0.102)
Age	0.002***	0.010***	0.010***	0.010***	0.008***	0.006***	0.006***	0.012***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Age^2		-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Household size			-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.007***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Woman				-0.048***	-0.048***	-0.021**	-0.021**	
				(0.014)	(0.014)	(0.010)	(0.010)	
Married					0.015***	0.018***	0.019***	0.014**
					(0.005)	(0.006)	(0.006)	(0.006)
Education						0.012***	0.012***	0.012***
						(0.002)	(0.002)	(0.002)
Urban							0.017**	0.017**
							(0.008)	(0.007)
District × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3-industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	63,173	63,173	63,173	63,173	63,173	63,150	63,150	61,058
R-squared	0.344	0.346	0.348	0.351	0.351	0.376	0.376	0.439

Notes: LPM with worker level data NSSEU between 1999 and 2010. Tariffs in industry j at time t. Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Only manufacturing sectors. Pseudo-individual FE include gender, district and 5-year cohorts. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Controlling for industry-level trends

VARIABLES	Formal (1)	Formal (2)	Formal (3)	Formal (4)	Formal (5)
Output τ_{jt}	0.158***	0.134**	0.120**	0.129**	0.103*
	(0.051)	(0.056)	(0.052)	(0.051)	(0.055)
Input τ_{jt}	-0.369***	-0.452***	-0.412***	-0.475***	-0.443***
	(0.103)	(0.089)	(0.084)	(0.101)	(0.069)
4 digit industry trend		-0.003**	-0.008***	-0.005***	-0.008***
		(0.001)	(0.002)	(0.002)	(0.002)
4d ind employment share trend			0.006***	0.005**	0.006***
			(0.002)	(0.002)	(0.002)
4d ind formal employment share trend				0.000***	
				(0.000)	
4d ind urban employment share trend					-0.010**
					(0.004)
District × year FE	Yes	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes	Yes
3d-industry FE	Yes	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes	Yes
Observations	61,058	55,619	55,619	55,619	55,619
R-squared	0.440	0.451	0.452	0.453	0.452

LPM estimation between 1999 and 2010. Tariffs in industry j at time t. Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Only manufacturing sectors. Individual characteristics include age, square of age, years of education, household size, religion and urban location. Pseudo-individual FE include gender, district and 5-year cohorts. Industry trends are variables constructed from the 1998 economic census, interacted with year. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: The role of firm size

Dependent variable	Indicator variable equal to one if worker i is formally employed				
	(1)	(2)	(3)	(4)	(5)
$\tau_{O,jt}$	0.032		0.031	0.065*	0.065**
	(0.046)		(0.044)	(0.034)	(0.032)
$\tau_{I,jt}$		-0.144**	-0.144**	-0.135**	-0.140***
		(0.058)	(0.057)	(0.056)	(0.052)
Firm size$_i$	0.019***	0.018***	0.018***	0.018***	
	(0.001)	(0.001)	(0.001)	(0.001)	
1 to 5 workers$_i$					-0.090***
					(0.010)
6 to 9 workers$_i$					-0.076***
					(0.010)
More than 20 workers$_i$					0.259***
					(0.019)
District × quarter-year FE	Yes	Yes	Yes	Yes	Yes
3- industry FE	Yes	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes	Yes
Observations	38,296	38,296	38,296	38,296	38,296
R-squared	0.612	0.613	0.613	0.615	0.618

Notes: Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Tariff in industry j at time t. Firm size$_i$ is the employment of the workers's employing firm. $1-5$workers$_i$, $6-9$workers$_i$ and $20+$ workers$_i$ equal to 1 if the worker's employing firm has 1 to 5, 6 to 9 and more than 20 workers. Individual characteristics include age, square of age, years of education, marital status, household size and urban location. Pseudo-individual FE include gender, year of birth, religion, literacy and district. Only manufacturing sectors. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Alternative definitions of informality

	(1) Regular worker	(2) Wage worker
$\tau_{O,jt}$	0.026	0.067
	(0.031)	(0.044)
$\tau_{I,jt}$	-0.176***	-0.315***
	(0.052)	(0.084)
Region × quarter-year FE	Yes	Yes
3- industry FE	Yes	Yes
Individual char.	Yes	Yes
Pseudo FE	Yes	Yes
Export τ_{jt}	Yes	Yes
Observations	85,288	85,288
R-squared	0.514	0.486

LPM estimation between 1987 and 2010. Tariffs in industry j at time t. *Regular worker* is a binary variable taking value 1 if the worker is a regular wage worker (*i.e.* in long-term employment with the firm), and 0 otherwise. *Wage worker* is a binary variable taking value 1 if the worker is a wage worker, and zero otherwise. Pseudo-individual FE include gender, region and 5-year cohorts. Regions are used instead of districts due to the lack of district information in the 1993-1994 wave of the survey. Individual characteristics include age, square of age, years of education, household size, religion and urban location. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.5 The structure of industry employment.

In this section, we investigate whether trade liberalization impacted the structure of industry employment through intersectoral reallocations. The previous analysis suggests that trade reallocates economic activity within industry. Data limitations do not allow us to follow workers over time, so we cannot formally know whether the same informal worker became formal or vice-versa. Our findings could be driven by the fact that industries facing greater tariff cuts became larger or smaller relative to relatively less impacted industries. Other empirical works have found that the effects of trade liberalization depend on the type of episode considered. Overall, there has been limited evidence of such reallocations in developing countries implementing unilateral trade reforms (Attanasio et al., 2004 for Colombia; Wacziarg and Wallack, 2004 for Mexico; Topalova, 2010 for India).

We start by carrying out a decomposition analysis similar to McCaig and Pavcnik (2018). The goal of such a decomposition is to show whether the change formality shares of industries are driven by compositional changes in industry size, or by changes in formality share within industry (keeping industry size fixed). Table A.11 suggests that the overall change in formality shares was driven by changes in formality within industry.[52] There is an overall formalization of the economy during that period. Focusing only on manufacturing industries, which were more directly impacted by tariff changes, the within-industry formalization is more than three times (0.036) the size of between-industry formalization (0.011).

Table A.11: Decomposition of the change in formality shares between and within 4-digit industries (1999-2010)

Sample	**Within**	**Between**	**Total**
All industries	0,012	-0,001	0,011
Manufacturing	0,036	0,011	0,048

Notes: This table decomposes the evolution of the share of formal workers between 1999 and 2010 between within- and between-industry reallocations, following the method used by McCaig and Pavcnik (2018).

To ensure that we are not merely capturing changes in the composition of employment towards more formal industries, we specify a model where the dependent variable is the share of employment in each industry. We then regress this share on our set of tariffs, following the approach of McCaig and Pavcnik (2018) and Topalova (2010), among others. This analysis is repeated for two samples: workers (based on the NSSEU survey), and firms (ASI and UMES surveys). Results are presented in Table A.12. The lack of statistically significant effect of tariff level on industry size is consistent with previous findings in India and other developing economies, and provides additional evidence that change in formality status of workers appears to have happened within

[52]Table A.11 focuses on the period 1999–2010, which allows us to use a benefit-based definition of worker formality.

industry.

Table A.12: Effects of trade liberalization on the relative size of manufacturing industries

	Share of employment in industry j over total manufacturing employment		
	1990s	**2000s**	**All**
	(1)	(2)	(3)
Panel A: workers (NSSEU)			
$\tau_{O,jt}$	0.000	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
$\tau_{I,jt}$	-0.002	-0.003	-0.002
	(0.006)	(0.006)	(0.006)
Observations	243	341	476
R-squared	0.835	0.813	0.802
Panel B: firms (ASI and UMES)			
$\tau_{O,jt}$	0.008	-0.018	0.004
	(0.015)	(0.020)	(0.012)
$\tau_{I,jt}$	-0.014	0.005	-0.025
	(0.019)	(0.014)	(0.024)
Observations	286	299	486
R-squared	0.917	0.928	0.876
3-industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes

Notes: Dependent variable is the share of employment in industry j (measured as the sum of all individuals working in industry j at time t) over total employment in the sample at time t. Panel A is based on the NSSEU worker survey, and Panel B is based on ASI-UMES firm survey aggregated at the industry level. Manufacturing industries only. All regressions include export tariff controls, year and 3-digit industry FE. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$