Preliminary Draft

Labor Market Imperfections and Trade: Micro-Level Evidence from India *

Maria Bas†

Maxime Berrou ‡

Pamela Bombarda§

February 26, 2025

Abstract

This paper studies the effects of trade liberalization on informality in India over a long time period (1990-2010) relying on representative survey data on formal firms, informal firms, and workers. The identification strategy relies on a difference-in-differences methodology that exploits exogenous output and input tariff variation at the industry level. We present novel evidence on the impact of tariff changes on the share of employment of formal firms relative to informal ones within industry-state over time. Our findings show that foreign competition reduces the share of formal employment, increasing informality. Access to foreign inputs increases the share of formal employment. This effect is more pronounced in states with labor market regulations that favor workers, as firms hire more formal labor and can afford to bear labor costs. These results are confirmed by a micro-level analysis looking at the effects of tariff changes on the probability that a worker finds a formal job.

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

---

*We would like to thank Johannes Boehm, Mohammad Hoseini, and participants of CITP Nottingham 2024, CY University ThEMA, Paris 1 IELM, ETSG Athens 2024 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). This paper examines the role of trade reforms on employment in the context of a pervasive domestic distortion in developing countries: informality. 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, 90% of the workforce lacks social insurance, and 85% of manufacturing sector employment is informal (Mehrotra, 2019). Despite experiencing rapid economic growth since the structural reforms of the early 1990s, India remains a striking example of persistent informality.

The aim of this work is to examine the impact of trade liberalization on labor market reallocation in India over a two-decade period (1990–2010). Our empirical analysis leverages multiple data sources at both the firm and worker levels to evaluate the effects of trade liberalization on the allocation of firms and workers between the formal and informal sectors. Our analysis also takes into account the role played by differences across Indian states on labor market regulations. We investigate the complementarity between trade and labor market policies by studying if the effect of trade liberalization on the reallocation of firms and workers between the formal and informal sector depends on labor market laws. India offers a particularly compelling context for studying the impact of trade on informality, as it provides consistent and detailed data spanning decades for both formal and informal firms and workers. Furthermore, India's economy is characterized differences across states in fixed and variable labor regulation costs generating differences informality rates. As highlighted by Nataraj (2011), the median manufacturing firm in India is informal, employs just two workers, and operates with a capital base of $235. Additionally, the country underwent substantial trade liberalization episodes, beginning with reforms in 1991 that continued into the late 1990s, followed by a second wave of tariff reductions in the early 2000s.

The literature on international trade and informality highlights three channels through which trade reforms can affect the reallocation of firms and workers between the formal and informal sectors. One possible channel involves increased foreign competition, which reduces the domestic market shares of formal firms competing with foreign rivals, leading to the exit of less efficient formal firms and a rise in the share of informal firms (Goldberg and Pavcnik, 2003, Ben Yahmed and Bombarda, 2020). In Brazil, Dix-Carneiro and Kovak (2019) study the effects of trade liberalization in 1990s on the margins of labor market adjustment. They highlight that formally employed workers located in regions more exposed to foreign competition experienced a reduction in the probability of being employed in a formal job relative to those in less exposed regions. Still in Brazil, Ponczek and Ulyssea (2022) also show that regions more exposed to foreign competition observed higher informality and greater unemployment relatively to regions less exposed. The second channel is access to higher-quality foreign inputs, enabling formal firms to improve efficiency, lower marginal costs, and reduce prices, thereby increasing demand for their goods. This demand expansion drives growth in production and labor demand for formal firms. 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.

These different effects of trade reform on informality could vary with the degree of labor market regulations. Most evidence suggests that those regulations in India impede the development of the formal sector and its ability to effectively adapt to shocks (Hasan et al., 2007, Aghion et al., 2008, Panagariya, 2007). To account for informality, Ulyssea (2018) develops a model of trade and heterogeneous firms assuming that firms in the formal sector have to pay a fixed labor regulation cost and variable taxes to hire formal registered workers, while firms in the informal sector can sidestep fixed and variable labor regulation costs, contingent upon the probability of detection. Thereby labor market regulations affects the share of formal workers. We expect that in states that have low number of labor market regulations (lower fixed and variable labor costs), the effects of foreign competition on informality are stronger since it is easier for formal firms to fire workers. However, we expect that firms in states with high number of labor market regulations are bigger and more profitable and can afford the labor costs to register workers and so the share of formal firms is larger. In those states, we then expect that the effect of input tariff cuts on the reallocation of workers between informal to formal sector to be larger.

To test those predictions about the ways firms and workers may have been reallocated following trade liberalization, we exploit four datasets in order to obtain a complete view of the Indian economy over two decades, from 1990 to 2010: *(i)* worker-level data (Employment Unemployment Survey), *(ii)* formal manufacturing firm-level data (Annual Survey of Industries), *(iii)* informal manufacturing firm-level data (Unorganized Manufacturing Surveys) and *(iv)* economic census data covering the universe of economic units in India. The different channels of trade liberalization are captured by four-digit output tariffs, input tariffs and export tariffs faced by Indian firms from WITS. Combining the firm level data on formal and informal manufacturing firms, we create an aggregated industry-state panel and directly test these channels by examining how the number and share of informal and formal firms change in response to exogenous reductions in output tariffs (capturing the foreign competition channel) and input tariffs during India's trade liberalization episodes in the early 1990s and 2000s, controlling for reductions of export tariffs (market access channel). Our empirical analysis first focuses on firms' reaction to trade liberalization, and then confirms those results looking at workers' reaction. In both cases, our methodology relies on a difference-in-differences framework, which compares units of observation that were relatively more exposed to liberalization -in industries with greater tariff cuts- to units that were relatively less exposed -in industries with lower tariff cuts- , and controlling for confounding factors. Then, we investigate if the effects of trade liberalization on the reallocation of firms' and workers across formal and informal sectors is heterogeneous across locations depending on labor market regulations by splitting the sample between high- and low-labor market regulations states based on the dataset of Chakraborty et al. (2024).

Our findings suggest that output tariff reductions, through increased foreign competition, reduce the share of formal firms and the share of formal employment. Our results also show that industries and states facing higher input tariff cuts have raised the share of formal firms and the share of formal employment. Our estimates suggest that for the 45 percentage point reduction of output tariffs during the first period (1990-2000), the share of formal employment was reduced by 6 percent, while for the similar amount of reduction of average input tariffs (46 percentage points) in that period the share of formal employment increased by 21 percent. In the second period (2000-2010) the average reduction of output tariff was 20 percentage points which induced a reduction of the share of formal employment of 5 percent, while for a similar reduction of input tariffs (23 percentage points) the share of formal employment increased by 11,5 percent. Our results are confirmed at the individual level using worker-level data. Relying on individual data allows us to consider a richer set of variables to measure informality at the worker-level. This enables us to develop two different econometric models. In the first model, we consider industry level tariff and we compare individuals with similar observable characteristics working in industries differently exposed to tariffs cut. Then, we consider a local labor market approach using district tariffs where the assumption of perfect labor mobility is relaxed. We control for a rich set of individual characteristics (age, square of age, gender, marital status, education level, urban or rural location, etc.), time, location, and industry. The worker-level regressions confirm the results of the firm aggregated panel at the state-industry level. We find coefficients consistent with our previous results, which highlight the fact that there is reallocation of workers between the formal and informal sectors, even using alternative definitions of formality. Our estimates suggest that a 10 percentage point decrease of output tariffs increased the probability of informal work by roughly 3 percentage points, whereas input trade liberalization decreased it by 7 percentage point, relative to less impacted industries.

Our contribution to the literature is threefold. First, we provide one of the few studies on informality in India that establish a causal link between the main channels through which trade impacts informality, and tests those channels on both firm and worker level data. Our results suggest that international trade does impact firms' decision regarding their formality status, which in turn impacts worker's formality status. Our second main contribution is to look at the unequal effects of trade liberalization on informality depending on differences across states in labor market institutions. We are the first to provide evidence of the complementarity between trade and labor policies that affect the share of formal firms and employment in a context of a large developing country. Our final contribution is to address common measurement issues regarding informality. Since informality is a broad phenomenon that cannot fully be captured with a single outcome variable, we rely on the availability of different variables, both for workers and for firms. In our preferred specifications, we define formal workers as wage workers whose employer contributes to a retirement fund. Our results are also robust to alternative specifications in which formality is defined as having a long-term wage worker, and being a wage worker. Similarly, while we mainly define formal firms as firms registered under the Factories Act, we also consider other definitions (other types of registration, number of formal workers hired). To limit measurement issues further, specifically the fact that informal firms may fail to be accurately surveyed, we also use sampling weights provided in the surveys to aggregate the data and create a representative panel of the Indian economy at the district-level (for workers) and at the state-industry level (for firms).

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 7 further explores the role of labor market regulations , and section 8 develops a local labor market approach. Section 9 discusses the labor productivity gap and quantifies reallocations between sectors. Section 10 concludes.

2 Theoretical motivation

This section describes the main mechanisms through which globalization impacts formal and informal firms labor demand. The recent literature has developed general equilibrium trade models of heterogenous firms, intra-industry trade à la Melitz (2003) and formal and informal sector. Dix-Carneiro et al. (2024) 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). This model highlight three different channels through which trade reform affects firms' labor demand in the formal and informal sector. The first channel is increased foreign competition that reduces domestic market shares of formal firms competing with foreign ones and leads to the exit of least efficient formal firms, increasing the share of informal firms. In this setting, the decline in formal employment leads to both increase in unemployment and informal employment. This channel was already present in the models developed by Kovak (2013) and (Dix-Carneiro and Kovak, 2017, 2019) that showed that the presence of a large informal sector acted as a buffer to displaced workers due to foreign competition.

The second channel is access to foreign inputs that allows formal firms to source inputs from abroad that are of lower costs (Dix-Carneiro et al., 2024) or higher 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 lower costs of inputs can promote exports by increasing 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. 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. In this setting, firms with different productivity levels can produce in the formal or informal sector depending on their profitability. Firms producing in the formal sector have to pay a fixed labor regulation cost and variable taxes to hire formal registered workers. These formal firms can have access to foreign inputs that go through customs and must be registered. Firms in the informal sector can sidestep fixed and variable labor regulation costs, contingent upon the probability of detection, but face a limitation in accessing foreign inputs due to their unregistered status. Moreover, they assume that foreign inputs are complementary with skilled labor. Under these assumptions, the 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 increasing formal employment. In our framework, where foreign inputs are skilled-biased, the reallocation effect of input-trade liberalization will be more pronounced among skilled workers.

The last channel is related to the expansion of market access and new opportunities for exporting firms in the formal sector. Those firms that export expand their foreign demand, increase the demand of formal employment and 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 through which trade liberalization affects the reallocation of firms and workers between formal and informal sector varies with labor market laws. We highlight an unexplored mechanism through which the impact of trade reforms depends on labor policy. The theoretical framework developed by Ulyssea (2018) shows that labor policy affects the costs that formal firms faced when hiring and firing workers. Thereby, the effect of output tariff cuts that increases foreign competition and induce formal firms to fire workers that then reallocate to the informal sector should be stronger for formal firms that faced lower labor market regulations where it is easier to fire formal workers. On the other hand, since the share of formal workers is greater in labor markets with laws that protect more the rights of workers, the impact of input tariff cuts that decreases firms' marginal costs and increases more the demand faced by formal firms and thereby their labor demand should be higher for firms facing higher labor market regulations.

The main two contributions of this paper to the literature are first to provide causal empirical evidence on these predictions on how formal and informal firms adjust their labor demand when facing exogenous tariff cuts. The second contribution is to shed new light on the complementarity between trade and labor policy by investigating the heterogeneous effects of trade reform on the share of employment of formal firms at the industry-state level depending on differences across states on labor market regulations.

3 Indian Trade Reforms

3.1 Twenty years of trade reforms

In the decades following independence, the Indian economy was characterized by a highly regulated framework aimed at achieving self-sufficiency and minimizing dependence on external trade. International trade was tightly controlled, with little liberalization. The average tariff rates on imports were prohibitively high, averaging 87%, and non-tariff barriers covered approximately 90% of the value-added in manufacturing (Hasan et al., 2007). Import flows were further constrained by restrictive licensing requirements: only goods listed under the Open General License (OGL)—a positive list—could be imported, and importers were required to demonstrate actual user status, meaning the goods had to be used by the importer and not resold (Panagariya, 2004). Although the initial steps towards liberalization began in 1976 and continued through the 1980s, these reforms were limited and incremental in nature. They primarily addressed specific sectors or industries and failed to dismantle the overarching system of controls and restrictions. As noted by Panagariya (2004), "it was only during the second half of the 1990s that the government began to loosen its grip on investment and import licensing", signifying a shift towards broader and more substantial liberalization.

The 1991 economic reforms represented a paradigm shift for the Indian economy, transitioning it from a predominantly inward-looking model to one integrated with the global economy. These 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 to dismantle the long-standing constraints of the license raj, reduce bureaucratic hurdles, and liberalize trade to align India with global economic practices.

Figure A.1 plots the evolution of the Indian average output, input and export tariffs. In 1991, quantitative restrictions on imports were largely removed, and tariff levels were significantly reduced. The restrictive import licensing system was dismantled for most goods (Hasan et al., 2007). This shift made tariff barriers the primary mechanism governing Indian firms' participation in international trade. By 1992, the average output tariffs stood at approximately 65%. Over the course of the decade, sustained reductions were implemented as part of India's liberalization efforts, and by the late 1990s, following the conclusion of a Five-Year Plan, average tariff levels had stabilized at around 35%, marking a significant departure from the protectionist policies of the past. The industrial sector, manufacturing, and consumer goods industries were among the most liberalized, with substantial reductions in tariffs and the removal of import licensing. Additionally, sectors like automobiles, electronics, and capital goods benefited significantly from the reduction in barriers, fostering competition and enabling access to advanced technologies.

Tariff liberalization continued into the early 2000s, with a significant push around 2004, driven by the need to further integrate India into global trade and enhance its competitiveness following its commitments under the World Trade Organization (WTO) agreements. During this period, average applied tariff rates declined substantially, from 35% to 15%, reflecting a deliberate effort to simplify and rationalize the tariff structure. Peak tariff duties were also reduced, and the number of tariff slabs was streamlined, making the system more transparent and efficient (Batra, 2022). After 2005, however, output tariff rates stabilized around 15%, as the government adopted a more cautious approach, balancing liberalization with domestic industry protection. No significant tariff liberalization measures were undertaken between 2005 and 2010, and tariff rates remained relatively unchanged after that for the remainder of the decade.

Table A.1 provides evidence of the harmonization of the tariff schedule over the two waves of liberalization. The standard deviation of tariffs, as well as the maximum rate, decreased over time. However, this process took time. The protracted nature of tariff liberalization—spanning over two decades-means it unfolded in a non-uniform manner, targeting different categories of goods in distinct phases.in the early 1990s, tariff cuts specifically targeted capital goods and essential inputs for industry. Consumer goods continued to be regulated until roughly 1997-1998, and then were liberalized too. Figure A.2 plots average tariff reductions by two-digit industry. It shows that some industries, such as the "office, accounting and computing machinery" category, underwent immediate tariff cuts in 1991, but that other industries, like textiles, were more protected initially.

3.2 Data on Trade Barriers

The tariff data is sourced from the World Integrated Trade Solution (WITS). Our analysis focuses on tariffs from the European Union and the United States, which have been India's two largest trading partners for most of the studied period. Among the three types of tariffs provided by WITS, we use the effectively applied tariff rates (AHS) at the ISIC Rev. 3, which is equivalent to the 4-digit NIC-1998 (the Indian classification).

The output tariff at the product level (also referred to as the output tariff) is 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 computed as a weighted average of European and American tariffs. The weights are based on each partner's market share in total Indian exports for a given industry. These industry-specific weights are fixed at the beginning of the period to ensure that the effect of changes in export tariffs is not influenced by potentially endogenous shifts in partner countries' preferences or demand over time.

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.[1] Additionally, especially in the early 1990s, there were no concerns regarding data availability for these countries. Between 1990 and 2010, India significantly reduced its output tariff rates, but it did so uniformly rather than selectively targeting specific trade partners. India does not appear to have reduced its average tariff rates disproportionately relative to its other trading partners (see Figure A.3). Furthermore, India signed very few trade agreements during this time, and those it did sign were primarily with smaller economies and lacked the scope and depth of most free trade agreements seen elsewhere.[2] As noted by Batra (2022), "India's FTAs are limited to shallow integration provisions." India's relatively low engagement in preferential trade agreements also minimizes the risk, highlighted by Feodora (2023), of relying on Most Favored Nation (MFN) rates instead of actual preferential rates, which may not be available.

To build our measure of input tariff, we combine the above mentioned output tariff with Indian Input-Output table for the period 1998-1999. IO tables contain 115 industries in total, 51 of which are manufacturing. We follow the literature and construct the input tariff as the weighted average of the Indian output tariffs, τ_{kt}, yielding:

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

where the weights, $w_{kj}^{1998} = \frac{\text{input}_{kj}^{1998}}{\sum_k \text{input}_{kj}^{1998}}$, represent the cost shares of industry k in the production of a good in industry j, based on industry-level data in 1998. We check that the tariffs are not highly correlated with one another. [3]

[1] Data from the Observatory of Economic Complexity.

[2] 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.

[3] 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.

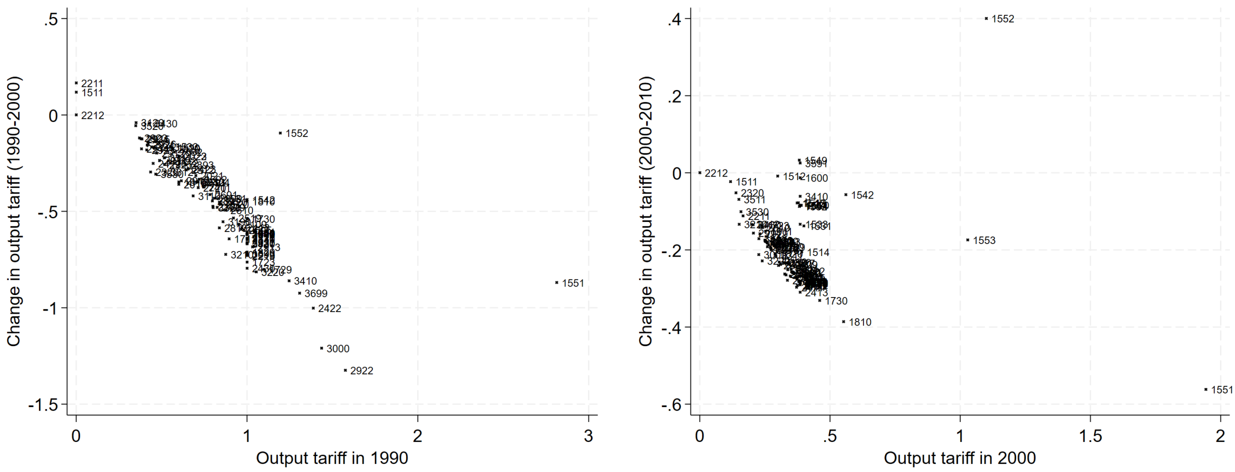
3.3 Exogeneity of Trade Policy

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—1991 and 2004. In 1991, the liberalization was driven by an acute economic crisis, with the balance of payments crisis necessitating intervention from the the IMF. The trade policy changes were part of the conditionality attached to the IMF's financial assistance, making the reforms externally imposed rather than domestically designed. By 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. However, the context suggests that 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. Additionally, the government's stated objective was to align India's applied tariff rates for non-agricultural products with those of ASEAN economies, as noted in the WTO's 2007 Trade Policy Review (WTO, 2007).

To validate these arguments, we analyze Indian output tariff data with respect to EU and US to determine whether a systematic relationship exists between initial 4-digit tariff levels and subsequent tariff changes over the two decades under study. The results are displayed in Figure 1. For the 1990s, panel (a) reveals a negative relationship: 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. For the 2000s, panel (b) shows a similar trend, though the relationship appears noisier. Higher tariffs continued to be associated with larger reductions; however, some industries retained relatively higher tariff levels by the end of the decade. A closer analysis highlights that many of the industries with minimal tariff cuts or increased protection belonged to the food products and beverages sector, consistent with India's historically protectionist stance toward agriculture and food-related industries. Overall, Figure 1 demonstrates 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 further shows that is also the case for input tariffs (see appendix A).

A related concern is that industries with the highest initial tariff levels, which experienced the greatest tariff reductions, might systematically differ in their characteristics compared to industries with lower initial tariffs. For example, larger or more formal industries could have been more heavily protected from foreign competition at the outset. Such a pattern could complicate

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.

the causal interpretation of the regression estimates, as trade policy might have been influenced by unobserved variables, such as the size of the informal sector or political considerations, thereby introducing potential reverse causality. To address this concern, follow Topalova and Khandelwal (2011) and examine whether initial industry characteristics are systematically related to tariff changes. For each decade, we regress initial industry characteristics on tariff changes, as presented in Table 1. Those characteristics are the relative size of the industry (industry employment total manufacturing employment), the share of employment in formal firms, the capital-labor ratio, the share of firms older than three years and output. In the first decade of liberalization, this is never the case of the main variables for output and input tariffs. In the second decade, this in only the case for input tariffs (industry size) and share of output tariff (share of formal firms).

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. Regression results of of 1990 industry characteristics on 1990-1999 tariff change (panel A), and of 2000 industry characteristics on 2000-2010 tariff change (panel B). Industry size$_j$ is industry j employment over total manufacturing employment. Formal employment share$_j$ is industry j employment in ASI-registered firms over industry j's ASI and UMES employment. Firms 3+ years share$_j$ is number of firms in industry j that are older than 3 years over total firms. All regressions are weighted by the square root of industry employment. Robust standard errors are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4 Indian Labor Market and Informality

Labor market institutions. Labor costs in India exhibit considerable variation across states, largely due to differences in labor market regulations. The Industrial Disputes Act (IDA) serves as a key legislative 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, leading to significant regulatory divergence (Hasan et al., 2007). Over time, some states have introduced amendments that facilitate hiring and firing practices (pro-employer states), while others have enacted provisions that enhance job security for workers (pro-worker states). These amendments have contributed to the emergence of distinct labor market regimes across Indian states (Chaurey, 2015).

A vast literature has intended to refine the classification proposed by Besley and Burgess (2004) which solely focused on amendments to the IDA, to take into account other dimensions, such as enforcement. We follow the recent literature and use the same classification as Chakraborty et al. (2024)[4]. Labor regulation status is defined at the state level and is time-invariant. States are classified in the following way: *(i)* Pro-employer states are those where laws tend to favor firms and it is easier to fire workers (Andhra Pradesh, Karnataka, Rajasthan, Tamil Nadu and Uttar Pradesh), *(ii)* pro-worker states where the labor regulation favor the rights and protection of workers (Gujarat, Maharastra, Orissa and West Bengal) and *(iii)* neutral states where the laws tend to be neutral (Assam, Bihar, Haryana, Jammu and Kashmir, Punjab, Kerala and Madhya Pradesh). 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 refered to as "high-LMR") and those with either neutral or pro-employer regulations ("low-LMR").

Data on firms and workers. We now present the datasets on firms and workers. Table A.2 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).[5] The ASI is a repeated cross-section[6] that provides comprehensive coverage of manufacturing establishments registered under the Industrial Disputes Act. Specifically, it includes establishments with 10 or more workers that use electricity, as well as establishments with at least 20 workers that do not use 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.[7]. Similar to the ASI, the UMES provides establishment-level information on key variables such as labor, wages, fixed assets, energy use, and sales. We follow previous studies by Hoseini and Briand (2020) and Chakraborty et al. (2021), and aggregate both ASI and UMES datasets to create a panel dataset at the state-industry level[8]. We restrict the analysis to years for which there is both ASI and UMES information: 1990, 1994, 2000, 2005 and 2010.

We then gather an additional data source that provides worker-level information for both 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

[4]It is based on Gupta et al. (2009).

[5]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.

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

[7]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)

[8]To proceed, we make use of the 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. More information is provided in Appendix A.

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 sample approximately 70,000 rural and 45,000 urban households per round. We use the information from these repeated cross-sectional surveys to construct a pseudo-panel at the individual level (Guillerm, 2017) [9]. 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.[10]

Finally, we also combine census from the 1990 and 1998 Economic Census of India to recover additional values and to build the employment shares for the local tariff used in section 8, and to build additional control variables[11].

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 workers and 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 comprehensive analysis difficult. Indian statistical authorities have designed their surveys to include data that facilitates determining the formality status of individuals and firms, thereby enabling the identification of informal activities.

To identify informal workers, empirical studies typically rely on variables indicating a lack of access to specific benefits or protections that workers are generally entitled to. 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. In our analysis, we define informality exploiting the information from both firm-level and individual-level data.

Using firm level data, we classify all firms in the ASI database as formal and those in the

[9]Regions in NSSEU are defined as groups of districts within a state that share similar agroclimatic conditions and socioeconomic characteristics. The country is divided into 77 such regions.

[10]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.

[11]Although economic censuses are available for the period 1990–2010 and provide information on all economic units in India, they provide limited coverage of key variables (Amirapu and Gechter, 2019). They do not allow to distinguish between the formal and the informal sector, making they unsuited for this analysis.

UMES database as informal[12]. Establishments included in the ASI are registered under the Factories Act (FA), which was enacted in 1948 [13]. 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[14]. This approach aligns with previous studies, such as Chakraborty et al. (2021) and Hoseini and Briand (2020), which treat ASI firms as formal when constructing ASI panels. 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. As a robustness check, we 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[15].

Using the NSSEU survey, we define a worker as informal exploiting two set of information. First, we consider access to specific social insurance benefits, particularly retirement benefits, which are consistently reported in the 1999–2000, 2004–2005, and 2009–2010 survey rounds. According to Indian labor regulations, all firms with 20 or more employees are required to contribute to an employee provident fund, providing a clear benchmark for identifying formal employment[16]. From the survey, we extract information regarding employers' contributions to a provident fund for employees' retirement. Specifically, 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 seems appropriate because all self-employed workers in the survey are individuals operating their own businesses. Second, to extend the analysis to also account for the earlier period of 1983–2000, we propose an alternative measure of informality based on the worker's type of wage employment. The survey asks wage workers if their wage work is "regular" (*i.e.*, not on the basis of periodic renewal of work contract). The correlation between the two measures is high and positive (0.67).

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

[12]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.

[13]It was replaced in 2020.

[14]Although ASI firms are formal, they may employ informal workers. This aspect of informality has been recently documented by the ILO Mehrotra, 2019.

[15]It is important to note that this question is not included in the UMES survey. Consequently, all UMES firms are still classified as informal, but under this new measure, some ASI firms will be considered informal.

[16]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.

Shleifer, 2014). We discuss those issues Table A.3 provides aggregated nationwide statistics ASI and UMES firms for 2005. It reveals that only 1% of firms and 19% of workers are employed in registered firms. Furthermore, Table A.3 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.4 for descriptive statistics of both ASI and UMES in 1990, 2000 and 2010). 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.5. 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.

Table A.6 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.[17]

Formality rates changed as India liberalized. Figure A.6 suggests that the liberalization of the economy may have led to an increase in the share of formal workers. Indeed, the share remained relatively stable until the early 1990s, and then increased steeply to reach approximately 16.5%. Overall, over the period of liberalization, the economy has gradually formalized, despite the fact that the vast majority of workers remain informal. Those results strengthen the relevance of the first part of our research question, which sets out to find empirical evidence of trade policy changes on informality.

Finally, since we are also interested in potential differential effects of investigate whether formality shares followed a different evolution depending on the type of labor market regulations. Figure A.7 shows that, on the aggregate, pro-worker states employ a greater share of formal workers than neutral and pro-employer states. Over the period, the gap between pro-worker and other is approximately 5 percentage points.

Table 3: Average firm-level characteristics and labor market regulations

	All		**Low-LMR**		**High-LMR**	
	Mean	SD	Mean	SD	Mean	SD
Labor	23.60	244.77	20.85	253.27	30.89	220.44
ln Output	3.94	6.08	3.78	5.95	4.32	6.35
ln Emoluments	0.09	0.24	0.09	0.21	0.10	0.28
Plus 3 years	0.89	0.31	0.89	0.31	0.90	0.30
Observations	198429		144188		54241	

Calculation based on micro-level ASI and UMES datasets for the year 2000. *All* contains the total sample for which labor market regulation is available, *Low-LMR* is a subsample of pro-employer and neutral states, and *High-LMR* is a subsample of pro-worker states.

Table 3 further shows that firms-level characteristics differ depending on the regulatory regime.

[17] 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 percent (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).

It takes into account both formal and informal firms. Firms in states where regulations are more stringent tend to be larger, produce more and pay more. In that case, those firms may be better able to adapt formalize when input tariffs decrease. The identification strategy we use to uncover such reallocations between sectors is described in the next section.

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. The analysis starts with firm-level data and is subsequently extended to consider individual-level regression models. Firm-level analysis provides a comprehensive overview of how trade reforms influence economic activity across formal and informal activities across state-sectors. Subsequently, we turn to individual-level regression models to validate and enrich our findings. The micro-level approach offers a more granular perspective, focusing on how trade reforms affect workers within the formal and informal manufacturing sectors.

5.1 Industry-state panel from firm level data

We investigate the relationship between trade liberalization and changes in firms' demand for formal labor using aggregated firm level data at the 4-digit industry level and state from the universe of informal (UMES dataset) and formal firms (ASI dataset) in the manufacturing sector in India for two different periods of waves of trade reform: *(i)* 1990-2000 and *(ii)* 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:

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

where FS_{jst} denotes the share of formal employees in 4-digit industry j and state s at time t computed as the share of employment of formal firms relative to total employment in both formal and informal firms.[18] Our estimations account for three key channels through which international trade can influence informality: market access, import competition, and access to foreign technology. 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 foreign 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 and controls for changes in the

[18] In some specification to define informality, we instead consider whether firms in UMES and ASI report contributing to employee provident funds.

relative price of foreign inputs used by firms. Easier access to foreign inputs is expected to promote formalization of the economy, leading to a negative coefficient for this coefficient. All our estimations also control for the export tariff, $\tau_{X,jt}$, to take into account foreign demand shocks. A reduction in export tariffs can boost labor demand in more productive formal-sector firms, encouraging a reallocation of workers from the informal to the formal sector. Industry-state fixed effects γ_{js} account for time-invariant characteristics specific to 4-digit industries within each state. These effects capture state-specific attributes of industries, which is particularly important in a large and diverse country like India, where geographic location contributes to significant heterogeneity. State-year fixed effects μ_{st} control for time-varying characteristics at the state level. This is a crucial adjustment, as most contemporaneous reforms in India were implemented at the state level, reflecting the country's federal structure. This fixed effect accounts for all policies likely to influence the size of the informal sector, such as the implementation of VAT in the early 2000s, which encouraged formalization, as well as labor market regulations, related to minimum wage laws and firing costs (see Hoseini and Briand, 2020 and Soundararajan, 2019 among others). A potential concern for identification is that industry-level, time-varying factors unrelated to trade policy might have influenced firms' decisions to hire formal versus informal labor. To address this, we include industry initial size trends $Trend_{jt}$, computed as the initial year total employment in the 4-digit industry interacted with each year. This enable us to compare industries that experienced differing tariff changes but exhibited similar labor force trends at the beginning of the period. Standard errors are clustered at the 4-digit industry level, which corresponds to the level of tariff variation. Our identification strategy relies on comparing industry-states exposed to varying levels of tariffs while ensuring they are otherwise similar in characteristics.

Finally, when we investigate the unequal effect of trade liberalization on the share of formal workers depending on labor laws varying across Indian states, we split our sample into states having high-labor market regulations and states with low-labor market regulations. Therefore, we estimate equation (2) in these two different subsample and compare the coefficients of output and input tariffs. This strategy allows us to control for other time-varying differences across states over time including state-year fixed effects.

5.2 Individual worker level survey

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 tariff cuts on the likelihood of workers becoming formal in response.[19] We estimate the following linear probability model, on a sample of formal and informal workers in manufacturing industries:

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

$$F_{ijdt} = \alpha + \beta_1 \tau_{O,jt} + \beta_2 \tau_{I,jt} + \beta_3 \tau_{X,jt} + \beta_4 X_{ijdt} + \delta i + \gamma_{dt} + \mu_j + Trend_{jt} + \varepsilon_{ijst} \tag{3}$$

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 (3) in two different sub-samples of workers located in states that have high- and low-labor market regulations.

6 The effects of trade liberalization on informality

6.1 Results using aggregate industry-state panel from firm level data

This section presents estimation results for equation (2) at the state-industry level for the two waves of trade liberalization. Table 4 presents the results. Columns (1) to (3) show the results for the first wave of trade reform between 1990-2000 using 1990, 1994 and 2000 decades, while results in columns (4) to (6) show the estimates for the second wave during the period 2000-2010 using 2000, 2005 and 2010 decades in the estimations.

Results presented in columns (1) and (4) of Table 4 are in line with the framework discussed in section 2. Output tariffs cuts increasing foreign competition has reduced the share of formal workers. The coefficient on input tariffs is negative and significant implying that industries facing larger input tariff cuts have increased the share of formal workers. This findings take into account the inclusion of an extensive set of fixed effects. First, we account for time-varying shocks at the state level by including state-year fixed effects. The inclusion of those fixed effects ensures that the results are not influenced by state-specific reforms, such as adjustments to minimum wage levels or modifications in the tax system, that could affect firms' decisions to hire formal labor. Additionally, we include state-industry fixed effects to address diversity across states and industries, capturing heterogeneity between industry-location pairs. Finally, we control for alternative reforms in broad categories of industries by introducing 2-digit industry-year fixed effects.

In section 3, we argued that the suddenness and unexpected nature of Indian trade reforms limited the risk that our results are picking up the effect other factors at the 4-digit industry-level and co-varying with tariffs over time. For instance, India and its main trade partners could have simultaneously liberalized their trade policies. Exporting Indian firms would in that case benefit from expanded market access abroad, increase their revenue and formalize, as in Vietnam (McCaig and Pavcnik, 2018). All regressions in table 4 include export tariffs as a control. However, as shown in Table A.1. there is almost no variation in the tariffs applied by the main trading partners to India during the period.

In columns (2) and (5), we take into account other reforms and unobservable shocks varying across 4-digit industries over time by including initial size industry trends.[20] The coefficients on both output and input tariffs are stable and robust when we take into account these shocks. Our estimates in column (2) suggest that for the 45 percentage point reduction of output tariffs during the first period (1990-2000), the share of formal employment was reduced by 6 percent, while for the similar amount of reduction of average input tariffs (46 percentage points) in that period the share of formal employment increased by 21 percent. In the second period (2000-2010) the average reduction of output tariff was 20 percentage points which induced a reduction of the share of formal employment of 5 percent, while for a similar reduction of input tariffs (23 percentage points) the share of formal employment increased by 11,5 percent. These findings suggest that input-trade liberalization more than compensate the negative effect of foreign competition on the share of formal employment and the net effect of trade reforms in both periods is a decrease in informality.

These results are robust to alternative specifications. In columns (3) and (6) we rely on an alternative definition of informality at the firm level relying on whether firms in both samples, ASI and UMES, declared that they contribute to the provident fund. Our estimates are robust and stable when using this alternative definition of informality. Next, we estimate the effects of tariff changes on the full sample period. Table B.1 shows that our results are robust when we do not split the sample by periods.

A related question is whether firms also reallocate between the formal and the informal sector due to trade liberalization. In order to investigate this we estimate equation 2 using the share of

[20]Our results are robust when we include a trend to vary at the state-4-digit industry level to account for differential dynamics prior to trade liberalization in different states-industries.

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. However this effects are smaller in magnitude compare to the effects of tariff cuts on the share of formal workers.

As an alternative explanation, we also consider the role of other policies, since trade reform was only one dimension of India's liberalization. By increasing incentives for firms to operate formally, contemporaneous reforms may have increased formality in some industries. This is a particularly strong concern in the 1990s, when the bulk of India's liberalization reforms was taking place. We use data from Aghion et al. (2008) and consider the effects of relaxing the restrictions on foreign direct investment and of dismantling licensing which regulated entry and production in some sectors in the early 1990s. Results presented in table B.4 show that our results for input tariffs are robust when we control for other industry level reforms that took place in the early 1990s.

Table 4: The effects of trade liberalization on the share of formal workers

	1990-2000			**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)
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 τ_{jt}	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 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. *Registered* is the share of workers employed in ASI firms over total labor in ASI and UMES firms. *Provident fund* is the share of workers employed in firms reporting to provide provident fund (PF) over total labor in ASI and UMES firms. Only manufacturing sectors. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, we investigate what drives the results presented in table 4. Focusing on the input channel, a potential source of heterogeneity is the capital intensity of industries. Industries that are initially more capital intensive may benefit relatively more from input trade liberalization since intermediate inputs and capital are complimentary in production. The results of B.3 confirms this intuition. The coefficient attached to input tariffs is larger in absolute value for industries that are initially more capital intensive.

6.2 Individual worker level 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[21]. Firms hire them temporarily, for instance, to avoid labor market regulations or to mitigate the impact of negative shocks (Chaurey, 2015). The aggregated nature of the analysis in section 6.1 does not allow us to explore that possibility[22]. 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 5: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[21] 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).

[22] 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).

In order to investigate the effects of trade liberalization on the likelihood of becoming a formal worker, we estimate equation (3) using the linear probability model (LPM) during the second wave of trade liberalization. Table 5 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 not impacted by trade liberalization and holding all other variables equal.

Appendix B show additional checks, where in Table B.5 detail the role of each individual characteristics, in Table B.6 we include additional trends, and in Table B.7 we check that the results are robust to differences in firm size category.

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 B.8. 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 B.8 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 B.9 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 B.10 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 B.11 and B.12. There does not seem to be a strong differential effect along those characteristics [23]. In short, workers are more likely to become formal through access to foreign inputs if they are male, live in urban areas, and have a medium level of education.

7 The role of labor market regulations

This section explores the role played by labor market policies on the relationship between trade liberalization and informality. More specifically, we investigate if the effects that we found previously on foreign competition and access to foreign inputs due to trade reforms are heterogeneous depending on the degree of labor market regulations that varies across Indian states[24].

In states where the degree of labor market regulations is low, the labor laws are favoring more firms relative to workers and it is easier for formal firms to fire workers. Therefore, we expect that the effects of output tariff reductions (foreign competition) on informality to be higher for states that have a low degree of labor market regulations (Low LMR). States where labor market regulations are more developed and workers have more rights (High LMR), have a higher proportion of formal firms that are bigger and more profitable and can afford the labor costs to register workers. Thereby, the share of formal firms is larger in those states and so the effects of input tariff cuts on the reallocation of workers between informal to formal sector should be greater.

In order to test those predictions, we split the sample into states that have labor market regulations that are favoring more workers and so firms have to pay higher fixed and variable labor costs (High LMR) and states that have labor market regulations favoring more firms where firms can more easily fire workers (High LMR) following the classification of Chakraborty et al. (2024).

Table 6 presents the results. Columns (1) and (2) present the results for the industry-state level estimations relying on the share of formal workers. We split the sample between high-LMR (column 1) and low-LMR states (column 2) and we compare the estimates on output and input tariffs. Our estimates suggest that the negative effect of foreign competition on the share of formal workers is stronger in low-LMR states, while the effect of input tariffs cuts on the share of formal employment is larger in high-LMR states. Columns (3) and (4) confirm this results using the micro-worker level data and the linear probability model. The effects of output tariffs

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

[24] Note that in the previous estimations we control for state-year or district-year fixed effects that take into account the differences across Indian states on labor market regulations.

cuts increase more the probability of becoming informal in states with low-LMR where firms can more easily fire workers, whereas the impact of input tariff cuts on the likelihood of becoming formal is greater in high-LMR where the share of bigger formal firms is greater and they can afford the higher labor market costs. These results suggest a complementarity between trade and labor policy. The gains from input-trade liberalization on formal employment go hand to hand with labor market laws that regulate the rights of workers.

Table 6: The role of labor market regulations

	Firms		**Workers**	
Samples	High LMR	Low LMR	High LMR	Low LMR
	(1)	(2)	(3)	(4)
$\tau_{O,jt}$	0.544	0.200**	0.131*	0.126***
	(0.333)	(0.083)	(0.074)	(0.046)
$\tau_{I,jt}$	-0.824***	-0.239*	-0.562***	-0.310***
	(0.158)	(0.122)	(0.113)	(0.064)
State × year FE	Yes	Yes	Yes	Yes
2-industry × year FE	Yes	Yes	No	No
3-industry FE	No	No	Yes	Yes
3-industry × state	Yes	Yes	No	No
Observations	637	1,332	13,189	24,839
R-squared	0.829	0.808	0.481	0.491

Notes: OLS estimation. The aggregated firm panel for the years 2000, 2005 and 2010 is used for columns (3) and (4), where the dependent variable is the share of ASI-employed labor over total labor in industry j, state s and time t. Worker survey for the years 1999-2000, 2004-2005 and 2009-2010 is used for columns (1) and (2), where the dependent variable is a binary variable for employer's contribution to benefits. *High LMR* is a subsample of states with pro-worker labor laws, and *low LMR* is a subsample of states with pro-employer and neutral laws. Only manufacturing sectors. All worker regressions include individual characteristics and pseudo-individual FE, and all firm regressions include 4-digit industry employment trends, capital per labor and emoluments per labor controls. All regressions control for export tariffs. Heteroskedasticity-robust standard errors clustered by 4-digit industries-state are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8 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.[25] The NSSEU dataset, spanning

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

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[26].

A growing body of recent research on the local labor market effects of trade reforms underscores that trade liberalization, driven by foreign 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 reports residing.[27] 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 employment shares for constructing local tariffs and other industry-level and local-level variables, we use economic censuses from 1990 and 1998.[28] 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,

[26]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 rele of labor market regulations, which was highlighted in section 7.

[27]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.

[28]Since NSSEU data cannot be aggregated at the district-sector level, we use Economic Census data.

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:

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

Similarly to equation (3), 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 7. As in the within-industry analysis, greater import competition decreases the probability of working formally, whereas input tariff reductions increase it.

Table 7: 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 × district FE	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes
Year × 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

9 Labor Reallocation and Aggregate Productivity

The presence of a substantial informal sector may contribute to significant resource misallocation at a macroeconomic level. Table 2 highlights these dynamics, showing that informal firms, while avoiding the costs of registration, face unique challenges that constrain their ability to grow. This will potentially affect aggregate output. In this section, we build on the macroeconomic development accounting literature to assess the potential impact of trade liberalization in India on aggregate labor productivity in the manufacturing sector, specifically through the reallocation of resources from informal to formal firms, in the spirit of McCaig and Pavcnik (2018).

We start by estimating the labor productivity gap between firms in the formal and informal sectors, using established methodologies widely adopted in development accounting research such as Caselli, 2005 and Gollin et al., 2014. Using a Cobb-Douglas production function formulation for formal and informal sectors, we have $Y = AK^{\alpha_s}L^{1-\alpha_s}$ where A is total factor productivity, K is capital, L is labor, $(1 - \alpha_s)$ is the output elasticity with respect to labor, and $s \in \{f, i\}$ denotes the formal and informal firm sectors respectively. Under the assumptions of homogeneous labor and perfectly competitive markets, wages equate to the marginal revenue product of labor (MRPL) yielding:

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

with $ARLP_s$ the average revenue product of labor in sector s. Thus, the gap in the marginal revenue product of labor between the two sectors is proportional to the observed gap in the average revenue product of labor across the same sectors.

$$\frac{w_f}{w_i} = \frac{\text{MRPL}_f}{\text{MRPL}_i} = \frac{\text{ARPL}_f}{\text{ARPL}_i} \tag{7}$$

Thus, two measures are used in the literature to compute productivity gaps: wages or revenue per worker. Since both ASI and UMES contain information on wages, revenue and labor, we are able to compute the labor productivity gap using both measures and to proceed in a similar way as McCaig and Pavcnik (2018) studying what the drivers of the gap are. For both sectors, we calculate the average revenue product of labor (ARPL), which is the aggregate revenue declared by firms divided by the number of workers. Next, we calculate the wage ratio by comparing total annual earnings per worker between the two datasets. Total annual earnings encompass wages/salaries as well as additional payments, such as bonuses and welfare expenses made by the employer, for wage workers.[29]

[29] The detailed methodology for these computations follows McCaig and Pavcnik (2018).

Table 8: Labor productivity gap between the formal and the informal sectors

	All		**Textile**		**Maharashtra**	
	Revenue based (1)	Wage based (2)	Revenue based (3)	Wage based (4)	Revenue based (5)	Wage based (6)
Unadjusted	17.2	6.71	15.4	7.51	8.43	2.37
Adjusted by human capital	13.1	5.10	12.2	5.97	5.01	1.41
+ measurement error in revenue and time worked	8.18		7.62		3.12	
+ differences in output-labor elasticity	5.45		5.08		2.08	
Share of workers reallocated to formal plants	0.096	0.096	0.130	0.130	0.094	0.094
Initial share of workers in the formal sector	0.177	0.177	0.124	0.124	0.279	0.279

Notes: The labor productivity gap reported in columns (1), (3), and (5) is based on the average revenue product of labor and subsequent adjustments. The average revenue product of labor is the ratio of aggregate revenue per worker in the formal sector to aggregate revenue per worker in the informal sector. The labor productivity gap reported in columns (2), (4), and (6) is based on the ratio of aggregate annual earnings per worker in the formal sector to aggregate annual earnings per worker in the informal sector, plus subsequent adjustments.

Table 8 reports the results for 2005.[30] Columns (1) and (2) show the labor productivity gap in manufacturing. The unadjusted gaps is for the revenue-based and wage-based measures are 17.2 and 6.71. Those gaps are relatively large compared to Vietnam where McCaig and Pavcnik (2018) find a value of 9.0, but those values are consistent with estimates of large productivity gaps in India (Nataraj, 2011). In columns (3) to (6), we run a similar analysis for different data samples to test whether those results are driven by specific industries or states. Columns (3)-(4) report the results for a textile and apparel, and columns (5)-(6) for the largest state economy, Maharashtra. The smaller gap in Maharashtra may be driven by the fact that it is a relatively dynamic state economy. Overall, the unadjusted gaps are large.

Then, closely following McCaig and Pavcnik (2018), we explore the effect of human capital differences between sectors as well as the role of potential measurement errors. First, since formal workers may be more educated and therefore more productive, we control for human capital differences between the formal and the informal sectors following the methodology developed by Gollin et al. (2014).[31] As expected, this adjustment reduces the gap between both sectors. We further control for potential measurement errors, focusing only on the revenue-based approach. Informal firms may under-report their true revenue by as much as 30%(de Mel et al., 2009). Furthermore, they may overstate hours actually worked (Fafchamps et al., 2014).[32] As expected,

[30]Future work will extend this analysis to years 1990, 2000 and 2010, to consider potential changes of the labor productivity gap over time.

[31]Information about average education level in each sector is unavailable in ASI and UMES, so we estimate those values from the labor force survey.

[32]We do not correct for differences in time worked because the related information provided by ASI and UMES

these corrections further reduce the labor productivity gap in all samples. Finally, we relax the assumption of identical output elasticity of labor from which equation 7 is derived. If the informal sector has a higher output labor elasticity, the labor productivity gap may be overestimated. That additional adjustment[33] reduces the manufacturing labor productivity gap in manufacturing to around 5.45. Overall, the large productivity gap in India suggests a potential for substantial efficiency gains if labor reallocation from the informal to the formal sector occurs.

Therefore, the next step is to estimate the share of workers reallocated from sectors following trade liberalization. $\hat{\beta}_1$ and $\hat{\beta}_2$ are the coefficients estimated in Table 5, which are multiplied with their respective tariff change over the period 1999-2010[34]We are in the process of extending this analysis to the 1990s. We then multiply this with the relative initial size of the industry j, and sum over all industries.

$$\text{Share reallocated} = \sum_j s_j \left(\hat{\beta}_1 \times \Delta\tau_{O,j} - \hat{\beta}_2 \times \Delta\tau_{I,j} \right) \tag{8}$$

Equation 8 is similar, but not identical, to the one used by McCaig and Pavcnik (2018). That is due to the difference in setting between India and Vietnam. In India, export tariffs did not vary much over the period (see Figure A.1), but output and input tariffs did. Therefore, we include the effective rate of protection in our analysis. Results appear in Table 8. The estimated share of workers reallocated is approximately 9.6% of manufacturing workers. As a comparison, McCaig and Pavcnik (2018) find for Vietnam that 5% of hours were reallocated to the formal sector. These calculations suggest large worker transitions to the formal sector in the aftermath of liberalization.

is not easily comparable. Thus, the inclusion of that correction is particularly important in our setting.

[33]We assume an elasticity ratio between the informal and formal sectors to be 1.5, as McCaig and Pavcnik (2018) who take values from Restrepo-Echavarria (2014).

[34].

10 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 is based on a difference-in-differences methodology that exploits exogenous variations of output and input tariffs on the share of employment of formal firms over total employment (of formal and informal firms) over two decades of trade reforms. While foreign competition reduces the share of formal employment increasing informality, the positive effect of access to foreign inputs on the share of formal workers more than compensate the negative effect of import competition. Moreover, our results suggest that this negative impact of import competition is concentrated on states where labor market regulations are weak and firms can fire workers more easily, while the positive effect of foreign inputs on the share of formal employment is larger in states where employment protection is higher. Our findings have clear policy implications: when liberalizing trade in developing economies, the effect of access to new varieties of foreign inputs should not be underestimated as well as the complementarity with labor policies.

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.

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.

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.

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.

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.

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.

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.

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.

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.

FEODORA, T. (2023): "Missing Tariffs," Working paper.

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

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.

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

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.

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.

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.

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.

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

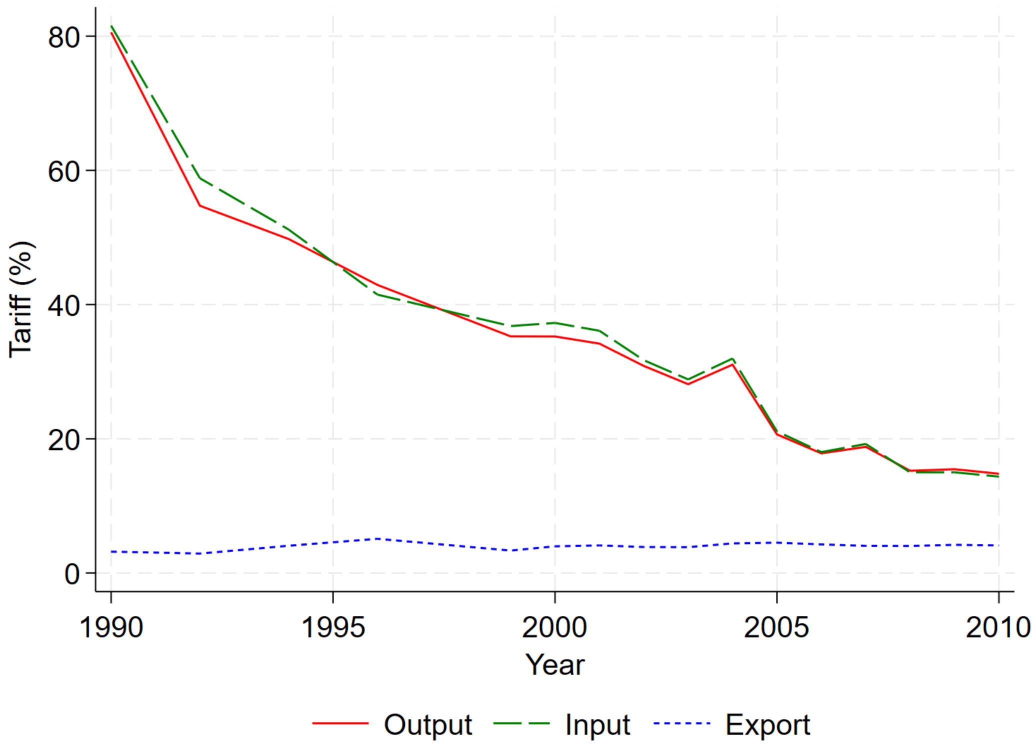
Appendices

A Descriptive Statistics

This section provides additional details about the datasets used for tariffs, firms and workers.

A.1 Tariffs

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



Source: WITS.

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 tariff for select two-digit industries, 1999-2010

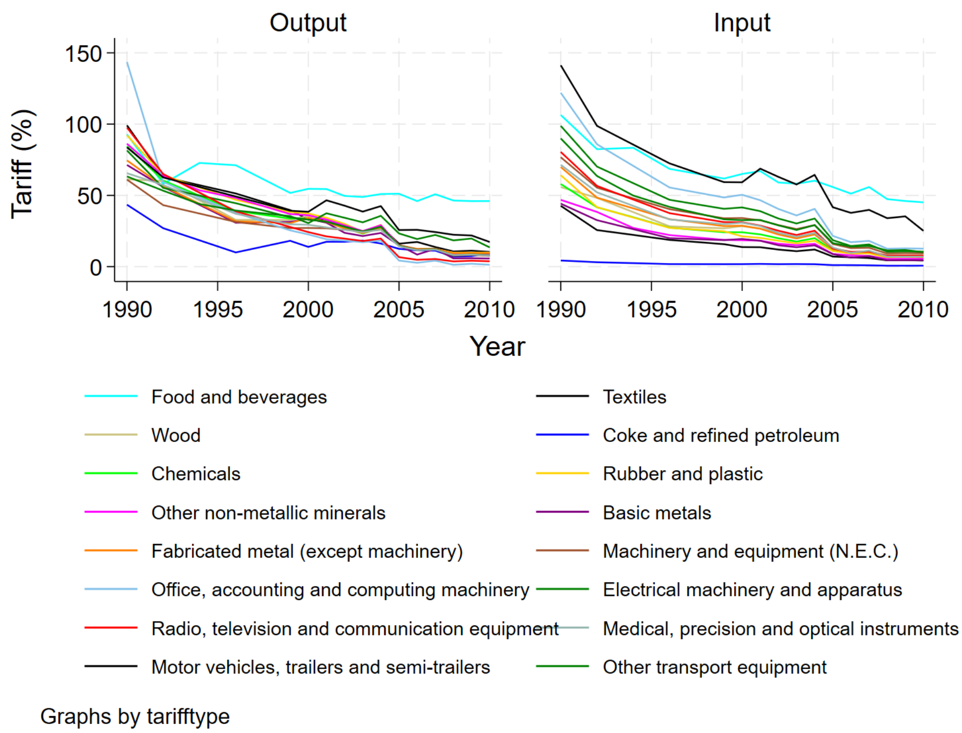


Figure A.3: Average Indian input tariff for the main trade partners of India, 1999-2010

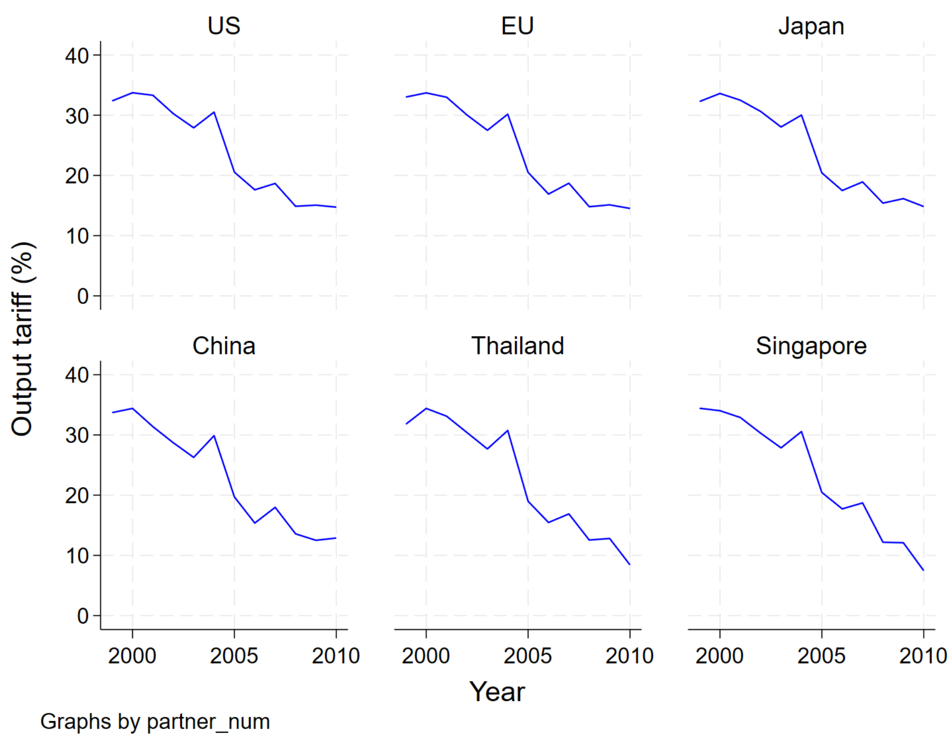
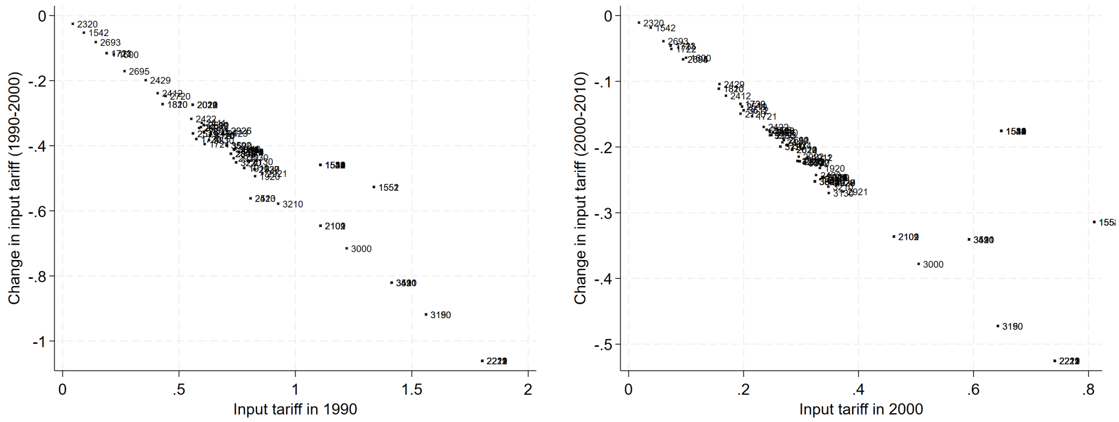


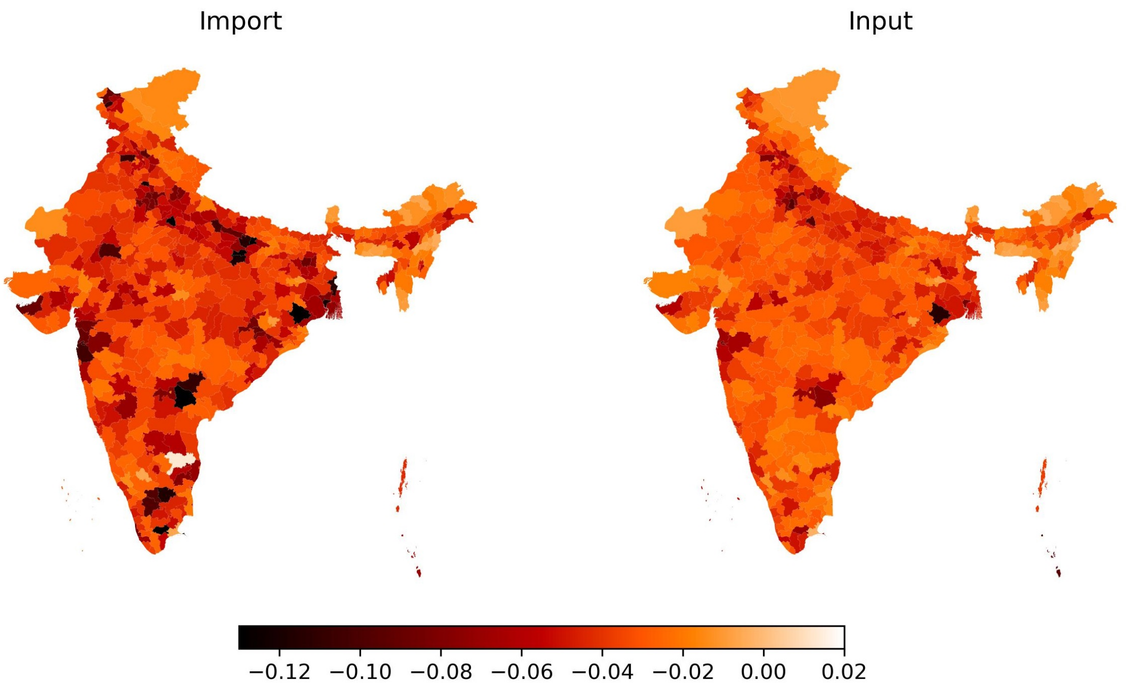
Figure A.4: Input tariff changes and initial tariff level, by decade



(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 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.[35]

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.[36] 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.

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

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

Table A.2: 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.3: Relative size of ASI and UMES

	ASI	UMES	$\frac{ASI}{ASI+UMES}$
Firms	117.63	17172.69	0.01
Labor	8515.59	36781.56	0.19
Capital	495410.95	110641.65	0.82
Output	1636564.26	891986.07	0.65
Emoluments	78950.49	49853.69	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.4: 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.5: 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.3 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).[37] [38]

[37]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.

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

Table A.6: 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.

Figure A.6: Average input tariff and formality share, 1983-2010

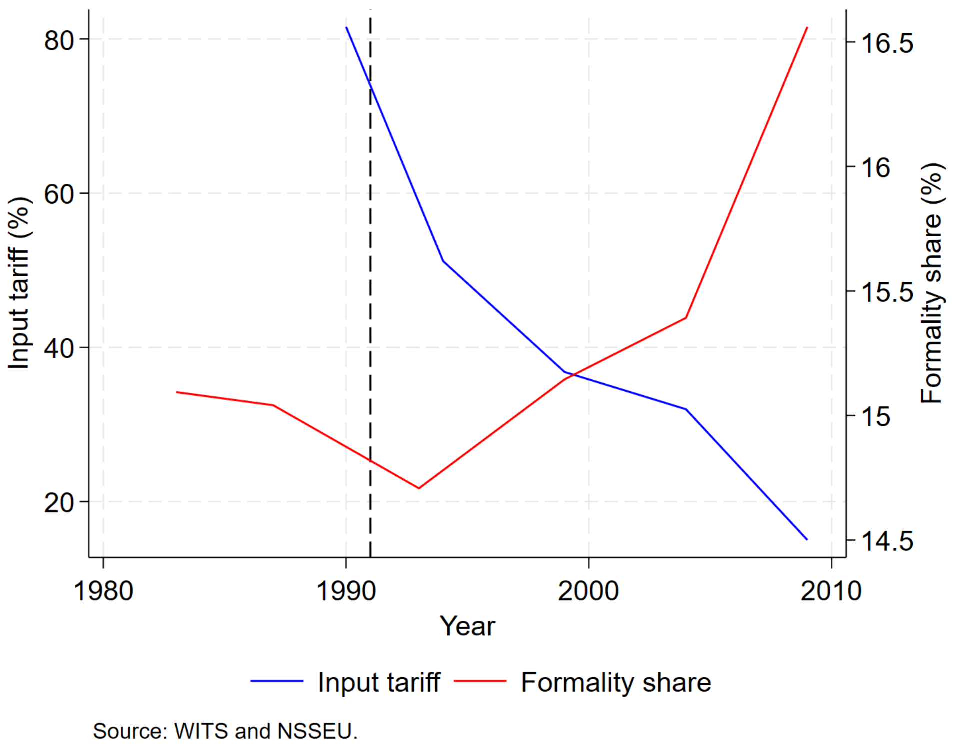
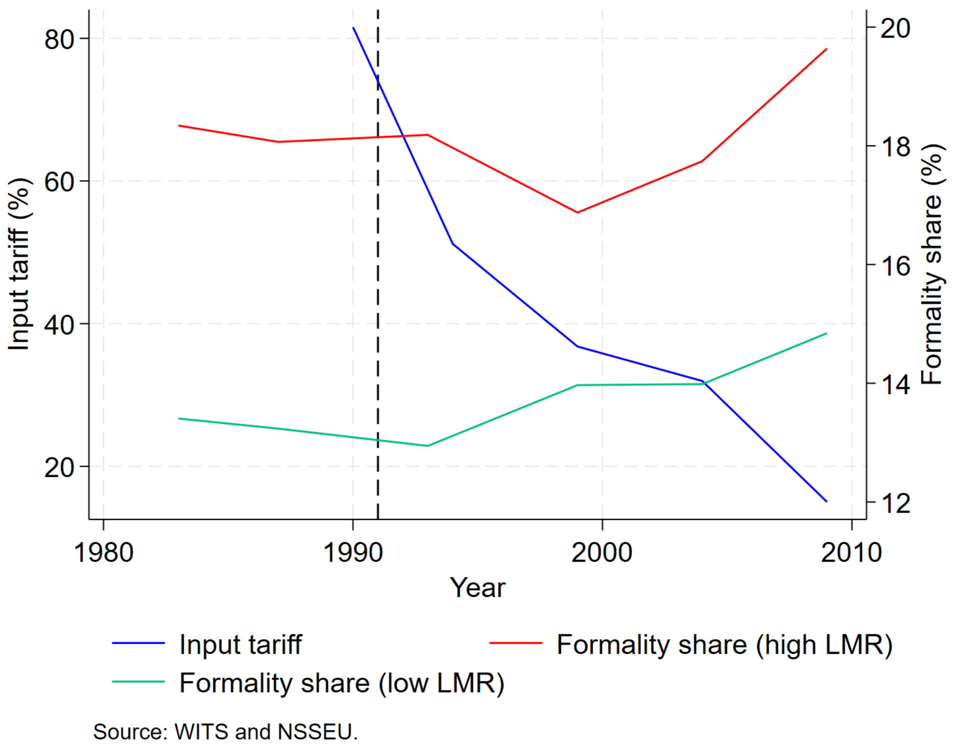


Figure A.7: Average input tariff and formality share by labor market regime, 1983-2010



B Additional Robustness Checks

B.1 Industry-state-level

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

Within-industry analysis			
	(1)	(2)	(3)
VARIABLES	Registered	Registered	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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 $\times$ year FE	Yes	Yes	Yes
2-industry $\times$ year FE	Yes	Yes	Yes
State $\times$ 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Industry-state-level regressions: heterogeneity by capital intensity

	1990-2000		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 capital intensity$_j$	-0.429***	-0.401***	-0.425***	-0.421***
	(0.045)	(0.048)	(0.066)	(0.060)
$\tau_{I,jt}$ × High capital intensity$_j$	-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 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. *High capital intensity* and *Low capital intensity* are binary variables. *Trends*$_{jt}$ is four-digit industry labor interacted with year. Only manufacturing sectors. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: 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 reform$_jt$ is the share of 6-digit HS code within an industry that were automatically open for FDI after 1991. Delicensed$_jt$ 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.2 Worker-level

Table B.5: 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 $\times$ 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 B.6: 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)
4d ind firm share 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 $\times$ 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 B.7: 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. $1-5\text{workers}_i$, $6-9\text{workers}_i$ and $20+\text{workers}_i$ 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.3 Worker-level regressions: heterogeneity

Table B.8: 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 B.9: 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 B.10: 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 B.11: 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)
Observations	42,881	42,881	42,510	42,510	42,493
R-squared	0.523	0.523	0.522	0.522	0.540
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

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 B.12: 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)
$\tau_{I,jt} \times$ Age 15-24$_i$			-0.266***
			(0.064)
$\tau_{I,jt} \times$ Age 25-34$_i$			-0.205***
			(0.070)
$\tau_{I,jt} \times$ Age 35-44$_i$			-0.231***
			(0.079)
$\tau_{I,jt} \times$ Age 45-54$_i$			-0.323***
			(0.093)
$\tau_{I,jt} \times$ Age 55+$_i$			-0.239**
			(0.096)
Age^2	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
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	0.542	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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$