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

Maria Bas[1] Maxime Berrou[2,3] Pamela Bombarda[2]

[1]University of Paris 1, Centre d'Economie de la Sorbonne (CES)

[2]CY Cergy Paris Université

[3]ESSEC Business School

August 26, 2025

Table of Contents

Introduction

Informality in developing economies

- Large size of the informal sector, for firms and workers
- Coexistence of the two sectors within industries and locations (≠ dual view of informality)

⇒ Reallocations between the two sectors are possible.

Can trade policy play a role? How do labor market institutions shape its impact? We investigate the channels through which tariff liberalization in India reallocated economic activity across sectors, over the period 1990-2010.

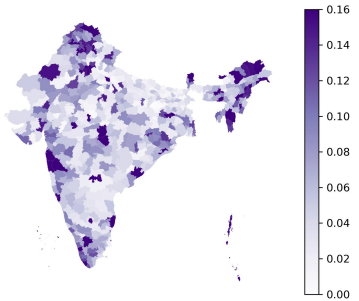
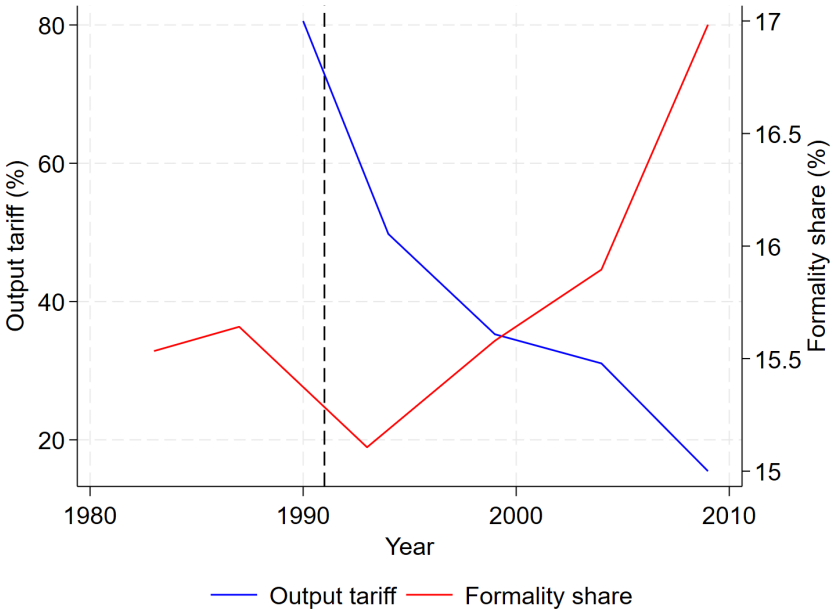


Figure: Share of formal workers in Indian districts (2004, NSSEU survey)

Formality shares by industry

Introduction

Figure: Evolution of average output tariff and formality, 1990-2010



Source: WITS and NSSEU.

Preview of results

We find evidence that trade policy liberalization **reallocates workers between the formal and the informal sectors**.

Those changes are consistent with theories of heterogeneous firms predicting trade-induced changes in formal firms' labor demand.

Unilateral trade liberalization implies two opposite effects, driven by output competition and access to foreign inputs.

Those results are **economically relevant:**

- We find that the formal-informal labor productivity gap is large in India
- Quantification exercise as in McCaig and Pavcnik (2018) suggests overall formalization of the economy throughout the period 1990-2010 (10% of workers formalised in the 1990s, 5% in the 2000s)

Table of Contents

Literature

1. Empirical evidence on trade and informality:

1. *Import competition* (Goldberg and Pavcnik (2003) for Brazil and Colombia; Dix-Carneiro and Kovak (2019) for Brazil; Chatterjee et al. (2024) for India)
2. *Access to foreign inputs* (Bas and Bombarda (2023) for Mexico)
3. *Expansion of market access* (McCaig and Pavcnik (2018) for Vietnam)

⇒ We disentangle between those channels in the same setting.

2. Trade, domestic institutions and firms' labor demand

- Input trade liberalization permits technology adoption Amiti and Konings (2007); Halpern et al. (2015)) and may change job composition (Bas and Paunov (2021)), increase their productivity (Topalova and Khandelwal (2011))
- Frictions in the domestic market for factors of production: sourcing decisions of manufacturing intermediates are distorted (Boehm and Oberfield (2020)), and stringent regulations/weak enforcement impact labor demand (Hasan et al. (2007) for India; Ponczek and Ulyssea (2022) for Brazil)
- Unions may protect workers facing trade-induced displacement (Cristea and Lopresti (2024) for U.S.)

⇒ Study the role of unions and imperfect enforcement in shaping firms' response to trade policy change.

3. Measurement of informality

- Data constraints (Goldberg and Pavcnik (2003))
- In low-income countries, nationally representative labor force data, covering workers in all types of employers, are more commonly available than firm-level data that capture the entire firm distribution (McCaig and Pavcnik (2018))

⇒ We build an aggregated panel of formal and informal firms over 1990-2010, and detailed worker-level surveys, to study the effects of tariff reductions on informality.

Conceptual framework

Informality is the part of the economy that **evades taxes and labor regulations**.

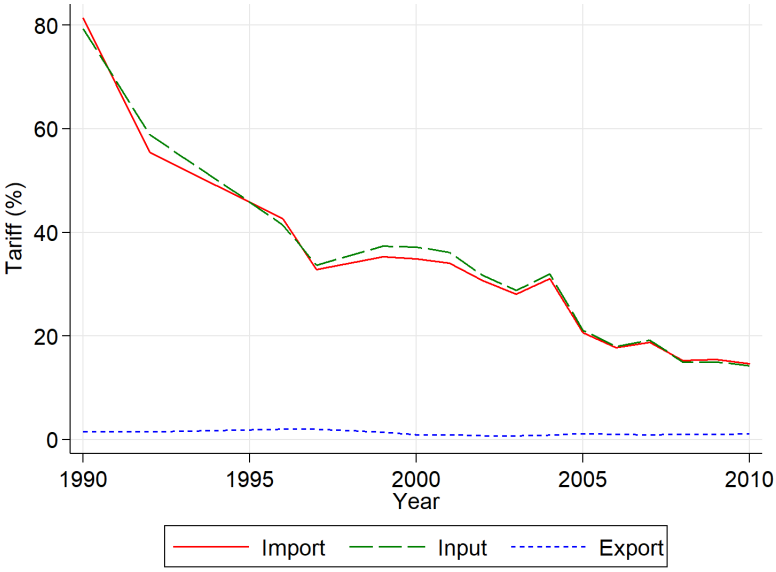
- Can be rationalized through a heterogeneous firms framework as in Ulyssea (2018), Dix-Carneiro et al. (2024), Bas and Bombarda (2023)
- Firms compare the relative cost of operating formally vs informally (*i.e.*, regulatory costs vs impeded growth)
- A productivity threshold below which firms are informal ⇒ size-dependent distortion

International trade reallocates economic activity between formality and informality as firms adjust their demand for formal labor:

1. Competition with foreign firms on the domestic market ↑ informality
   - Less incentive to operate at a low cost in autarky ⇒ output liberalization ↑ competition ⇒ some formal firms decide to operate informally
2. Access to foreign inputs ↓ informality
   - Low-quality domestic inputs and imperfect institutions impede formal firms' growth in autarky ⇒ input liberalization gives access to cheaper inputs for formal firms ⇒ formal firms' productivity and revenue ↑ increases ⇒ relative formality costs ↓
3. Market access ↓ informality
   - Formal firms only sell domestically ⇒ export liberalization gives access to a greater demand ⇒ some formal domestic firms ↑ their revenue and hire more formal workers

Table of Contents

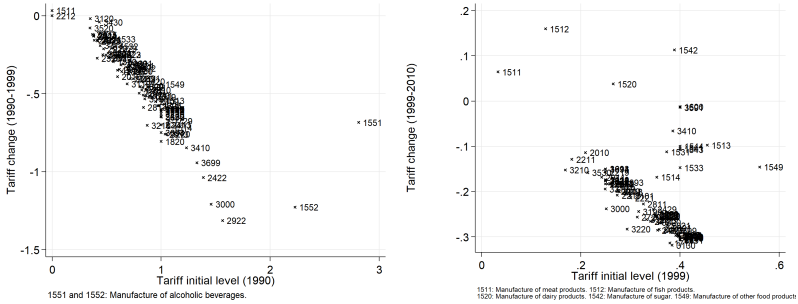
Large tariff liberalization in India starting in 1991



Source: WITS.

Figure: Average Indian industry tariff level, 1990-2010

Exogeneneity of tariff cuts



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

Figure: Industry output tariff change and initial tariff level, by decade

Informality in India

Some facts about the informal sector in India:

- **It is covers a large part of the economy.** 90% of workers and 99% of manufacturing firms are informal.
- **It is underperforming** relative the formal sector.
  - Informal *workers* tend to be younger, less educated, and earn less.
  - Informal *firms* are smaller and less productive

VARIABLES	(1) ln Labor	(2) ln Earnings per worker	(3) ln Capital per worker	(4) ln Sales per worker
Registered^{ASI}	2.455***	4.570***	0.411**	3.521***
	(0.115)	(0.276)	(0.195)	(0.314)
Observations	825,340	699,314	749,312	589,816
R-squared	0.689	0.560	0.365	0.543
4d-industry × year FE	Yes	Yes	Yes	Yes
state × year FE	Yes	Yes	Yes	Yes

Calculation based on the ASI and UMES data for 1990, 1994, 2000, 2005 and 2010. Only manufacturing industries. *Belongs to ASI* is a binary variable taking value 1 if the firm comes from the ASI dataset, and 0 if it comes from the UMES dataset. *ln Labor* is the log of all employees working for the firm. *ln Earnings per worker* is the log of total earnings paid by the firm over the last year, divided by the total number of employees. *ln Capital per worker* is the log of total fixed asset value divided by the total number of employees. *ln Sales per worker* is the log of total sales by the firm over the last year divided by the total number of employees.

Table: Formality premium

Table of Contents

Data

Firms and workers

Firm data, including both formal and informal firms, for 1990, 1994, 2000, 2005, 2010.

- *Formal firms:* Annual Survey of Industries (ASI)
- *Informal firms:* Unorganized Manufacturing Enterprises Survey (UMES)

$\Rightarrow$ Using the provided sampling weights, we build an **aggregated panel representative at the industry-state level**, following Hoseini and Briand (2020) and Chakraborty et al. (2021)

Worker data from the Employment Unemployment Survey from the National Sample Survey Organization (NSSEU), for 1999-2000, 2004-2005 and 2009-2010

Economic census data, meant to cover the universe of economic units in India, for 1990 and 1998

Data

Trade policy

Tariff data, at the 4-digit level, comes from the World Integrated Trade Solution (WITS)

All tariffs

- **Output tariff** is the simple average of the main partners of India at the beginning of each decade (Europe and US)
- **Export tariff** is a weighted average of those main partners, where the weight is the market share a country has over total Indian exports for an industry
- **Input tariff** for industry j are built using the output tariff in industry k, and 1998-1999 Indian IO tables (115x115 industries, with 50 manufacturing industries):

$$\tau_{jt}^{INP} = \sum_{k} \frac{\text{Cost}_{kj}^{1998}}{\sum_{k} \text{Cost}_{kj}^{1998}} \tau_{kt}^{OUT}$$

Measuring informality

Difficult to measure informality:

- Typically not observed by national statistics
- A broad/multi-dimensional phenomenon

$\Rightarrow$ Overcome those challenges exploiting survey design and variables

We identify variables in the data that can be interpreted as **making formality costly for the firm/employer**

1. **Formal firms** are identified by their *registration status* $\Rightarrow$ under the Factories Act of 1948 (working conditions and health requirements)
2. **Formal workers** are identified by their *access to benefits* $\Rightarrow$ whether the employer contributes to the worker's provident fund (for retirement)

$\Rightarrow$ We check that the way we measure informality accurately captures well-known facts about the informal sector

Measuring informality

Weighted results

Table: Description of Worker Level Data

	All Industries			Manufacturing		
	All	Informal	Formal	All	Informal	Formal
Age_i	35.85	35.21	40.15	34.15	33.36	38.13
	(12.35)	(12.55)	(9.85)	(12.06)	(12.16)	(10.70)
$Woman_i$	0.24	0.25	0.18	0.22	0.23	0.16
	(0.43)	(0.43)	(0.38)	(0.41)	(0.42)	(0.37)
Not $literate_i$	0.35	0.39	0.07	0.25	0.28	0.10
	(0.48)	(0.49)	(0.26)	(0.43)	(0.45)	(0.30)
High $School_i$	0.18	0.15	0.36	0.21	0.18	0.34
	(0.38)	(0.36)	(0.48)	(0.41)	(0.39)	(0.47)
$Urban_i$	0.35	0.31	0.68	0.61	0.58	0.76
	(0.48)	(0.46)	(0.47)	(0.49)	(0.49)	(0.42)
$Formal_i$	0.13	0.00	1.00	0.17	0.00	1.00
	(0.33)	(0.00)	(0.00)	(0.37)	(0.00)	(0.00)
Wage $Worker_i$	0.47	0.40	0.95	0.51	0.43	0.90
	(0.50)	(0.49)	(0.21)	(0.50)	(0.50)	(0.30)
Union $member_i$	0.14	0.06	0.69	0.16	0.08	0.59
	(0.35)	(0.24)	(0.46)	(0.37)	(0.26)	(0.49)
1-5 Workers$_i$	0.60	0.73	0.14	0.57	0.64	0.14
	(0.49)	(0.45)	(0.35)	(0.50)	(0.48)	(0.35)
6-9 Workers$_i$	0.08	0.08	0.08	0.09	0.11	0.03
	(0.27)	(0.27)	(0.28)	(0.29)	(0.31)	(0.17)
10-19 Workers$_i$	0.07	0.05	0.13	0.07	0.08	0.06
	(0.25)	(0.22)	(0.34)	(0.26)	(0.26)	(0.23)
20+ Workers$_i$	0.25	0.14	0.65	0.26	0.17	0.77
	(0.43)	(0.35)	(0.48)	(0.44)	(0.38)	(0.42)
Observations	193,470	168,636	24,834	22,792	19,010	3,782

Calculations are based on the NSSEU data for 1999-2000. 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.

Measuring informality

Table: Formality share of manufacturing industries

Industry (2d)	Formality share	Employment share
Manuf. of food products and beverages	0.107	0.115
Manuf. of tobacco products	0.069	0.084
Manuf. of textiles	0.103	0.181
Manuf. of wearing apparel/fur	0.051	0.140
Tanning and dressing of leather/luggage	0.105	0.027
Manuf. of wood and wood products	0.006	0.103
Manuf. of paper and paper products	0.256	0.012
Publishing, printing and reproduction of recorded media	0.223	0.019
Manuf. of coke and refined petroleum	0.675	0.002
Manuf. of chemicals and chemicals products	0.379	0.043
Manuf. of rubber and plastics products	0.246	0.016
Manuf. of other non-metallic mineral products	0.054	0.095
Manuf. of basic metals	0.435	0.021
Manuf. of fabricated metal products, except machinery	0.098	0.057
Manuf. of machinery and equipment N.E.C.	0.360	0.028
Manuf. of office, accounting and computing machinery	0.610	0.001
Manuf. of electrical machinery and apparatus	0.193	0.015
Manuf. of radio, television and communication equipment	0.368	0.003
Manuf. of medical, precision and optical instruments	0.228	0.003
Manuf. of motor vehicles, trailers and semi-trailers	0.508	0.014
Manuf. other transport equipment	0.538	0.008

Calculation based on the NSSEU data for 2004-2005 (round 61). Only manufacturing.

Table of Contents

Identification strategy

Source of identification. We exploit tariff change between industries and over time, and compare units of observations that relatively more exposed to tariff cuts to observations that were relatively less exposed.

A key concern is **tariff endogeneity**, since we compare industries that may be different in characteristics ⇒ we mitigate those concerns:

- Tariff cuts were related to initial tariff levels, for each decade Industry tariff Local tariff
- India did not select specific trade partners to liberalize its trade with Go
- No statistical correlation between tariff cuts and industry characteristics, as in Topalova and Khandelwal (2011) Go
- We also rely on the fact that correlation between tariffs is not too high to disentangle between the trade channels Go.

Three econometric models

Three approaches built on a **difference-in-differences methodology**

1. **Test our hypothesis** ⇒ Aggregated firm-level panel, 1990-2010
   - Are trade reallocations between formal and informal sectors driven by changes in firms' labor demand for formal/informal labor?

2. **Turn to workers' actual conditions** ⇒ Worker survey (within industry), 1999-2010
   - Do those changes in labor demand actually impact workers' access to benefits/protection?
   - Additional controls, mitigates endogeneity concerns

3. **Relax some assumptions** ⇒ Workers survey (local labor markets), 1999-2010
   - Allow for worker mobility between sectors within a district
   - Consider non-manufacturing workers

Aggregated firm panel analysis

Econometric model

We regress the share of formal labor on tariffs in a difference-in-differences framework:

$$\text{FormalShare}_{jst} = \alpha + \beta_1 \tau_{O,jt} + \beta_2 \tau_{I,jt} + \beta_3 \tau_{X,jt} + \gamma_{js} + \lambda_{jt} + \mu_{st} + \textit{Trend}_{jt} + \varepsilon_{jst} \quad (1)$$

In industry j, at time t and state s:

- FormalShare_{jst} is a measure of formality at the industry-state level
- $\tau_{O,jt}$ is the industry output tariff
- $\tau_{X,jt}$ is the industry export tariff
- $\tau_{I,jt}$ is the industry input tariff
- γ_{js}, μ_{st} and λ_{jt} are industry-state, and state-year and industry-year FE
- $\textit{Trend}_{jt}$ are industry-level trends (size of employment in 1990 or 2000)

Aggregated firm panel analysis

Share of formal labor

Table: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

▸ Share of formal firms
▸ Dynamic effect

Worker-level analysis: within industry

Econometric model

We run the following linear probability model in which we regress the probability for a formal worker to be formally employed on industry tariff levels:

$$F_{ijdt} = \alpha + \beta_1 \tau_{O,jt} + \beta_2 \tau_{I,jt} + \beta_3 \tau_{X,jt} + \beta_4 X_{it} + \delta_i + \gamma_{dt} + \mu_j + Trend_{jt} + \varepsilon_{ijdt} \qquad (2)$$

In industry j, at time t, state s and district d:

- F_{ijdt} is a measure of worker-level formality.
- $\tau_{O,jt}$ is the industry output tariff
- $\tau_{I,jt}$ is the industry input tariff
- $\tau_{X,jt}$ is the industry export tariff
- X_{it}: individual characteristics (age, age^2, urban, married, years of education, religion)
- δ_i: pseudo-individual fixed effect (district, gender, year of birth and literacy)
- γ_{dt}, μ_j are district-time and industry fixed effects
- $Trend_{jt}$ are industry-level trends

Worker-level analysis: within industry (1999-2010)

Results (1999-2010)

Table: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Worker-level analysis: within industry (1999-2010)

Robustness

More evidence on within-industry reallocations ▸ Industry size ▸ Decomposition

Role of each individual characteristic ▸ Go

Robustness of the main results:

- We include 4-digit industry trends constructed from the 1998 census ▸ Go
- We consider alternative definitions of informality (regular wage workers, wage workers) ▸ Go

We also investigate heterogeneous effects based on individual characteristics ▸ Urban ▸ Gender ▸ Skill ▸ Caste ▸ Age

Worker-level analysis: local labor markets (1999-2010)

Motivations

Local labor market approach:

- Relax the assumption of worker immobility between industries
- Consider the effects of trade on non-manufacturing sectors

For all three types of tariff τ_{jt}, the **district tariff** at time t in district d using census year 1998 is defined as:

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

◂ District characteristics ◂ Migration patterns

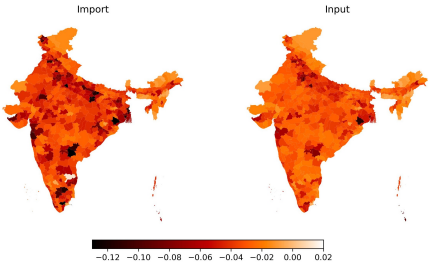


Figure: Indian local tariff change, 1999-2010

Worker-level analysis: local labor markets (1999-2010)

Econometric model

We run the following linear probability model in which we regress the probability for a worker to be formally employed on local tariff levels:

$$F_{ijdt} = \alpha + \beta_1 \tau_{O,dt} + \beta_2 \tau_{I,dt} + \beta_3 \tau_{X,dt} + \beta_4 X_{it} + \delta_i + \gamma_{dj} + \mu_t + \varepsilon_{ijdt} \quad (3)$$

In industry j, at time t, state s and district d:

- $\tau_{O,dt}$ is the district output tariff
- $\tau_{I,dt}$ is the district input tariff
- $\tau_{X,dt}$ is the district export tariff
- X_i: individual characteristics (age, age^2, urban, marital status, education level, religion)
- α_i: pseudo-individual fixed effect (district, gender, year of birth and literacy)
- γ_{dj} and μ_t are district-industry and quarter-year fixed effects

Worker-level analysis: local labor markets (1999-2010)

Results

VARIABLES	(1) Formal	(2) Formal	(3) Formal
Output τ_{dt}	0.009	0.070	0.231*
	(0.070)	(0.079)	(0.129)
Input τ_{dt}			-0.466*
			(0.257)
Observations	534,280	534,280	534,280
R-squared	0.698	0.698	0.698
Pseudo individual FE	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes
District $\times$ 4d-industry FE	Yes	Yes	Yes
Quarter-year FE	Yes	Yes	Yes
Export τ_{dt}	No	Yes	Yes

LPM estimation between 1999 and 2010. Tariff in district d at time t. Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. All industries. Individual characteristics include age, square of age, years of education, household size, religion and urban location. Pseudo-individual FE include gender, state and 5-year cohorts. Heteroskedasticity-robust standard errors clustered by district are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The role of labor market institutions

Motivations

Do different labor market institutions induce a differential effect of trade policy on informality?

⇒ Yes, if they modify the relative cost of formality.

We study two types of institutions:

1. Making informality *more* costly ⇒ **Unions**
   - In the spirit of Cristea and Lopresti (2024)
2. Making informality *less* costly ⇒ **Enforcement of regulations**
   - Proxied by distance to institutions, based on Ponczek and Ulyssea (2022)

The role of labor market institutions

Unions

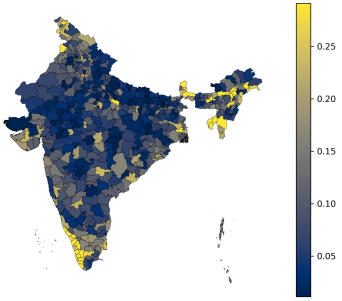
Formal workers are much more likely to be unionized than informal workers

- Trade unions in India exist in the organized sector (in large enough firms)
- Difficult for informal workers to take part in unions to secure better work conditions

There is **heterogeneity in local unionization shares**.

- Well distributed nationwide (≠ by state)
- Large variation in local rates (from less than 5% to more than 20%)

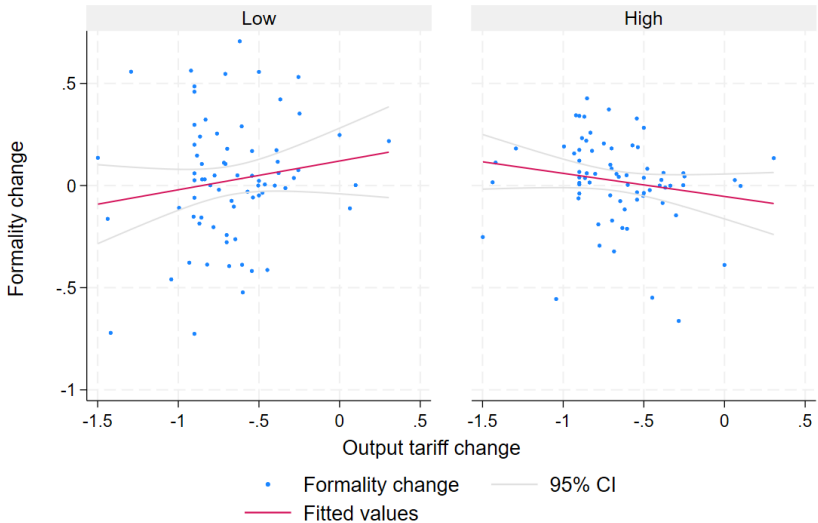
Figure: Unionization shares by district in 2009



The role of labor market institutions

Unions

Figure: Change in output tariff and formality, by unionisation rates



The role of labor market institutions

Unions

Within-industry analysis

VARIABLES	(1) Formal	(2) Formal	(3) Formal	(4) Formal
$\tau_{O,jt} \times$ Low unionisation$_{ru}$	0.190***			
	(0.059)			
$\tau_{O,jt} \times$ High unionisation$_{ru}$	0.099*			
	(0.050)			
$\tau_{I,jt} \times$ Low unionisation$_{ru}$	-0.335***			
	(0.082)			
$\tau_{I,jt} \times$ High unionisation$_{ru}$	-0.302***			
	(0.074)			
$\tau_{O,jt}$		0.176***	0.129***	0.190***
		(0.052)	(0.047)	(0.059)
$\tau_{I,jt}$		-0.312***	-0.311***	-0.335***
		(0.075)	(0.082)	(0.082)
$\tau_{O,jt} \times$ High unionisation$_{ru}$		-0.071*		-0.092*
		(0.041)		(0.049)
$\tau_{I,jt} \times$ High unionisation$_{ru}$			-0.015	0.033
			(0.035)	(0.043)
District $\times$ 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	40,011	40,011	40,011	40,011
R-squared	0.551	0.550	0.550	0.551

LPM estimation using the NSS-EU round 1999, 2004 and 2009. Tariff in industry j at time t. Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. Only manufacturing industries. *High unionization*$_{ru}$ is a dummy taking value 1 if the region-urban or region-rural area is above median unionization rate, and zero otherwise. *lOW unionization*$_{ru}$ is a dummy taking value 1 if the region-urban or region-rural area is below median unionization rate, and zero otherwise. Individual characteristics include age, square of age, years of education, household size, religion and urban location. Pseudo-individual FE include gender, state and year of birth and literacy. Heteroskedasticity-robust standard errors clustered by industry are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Aggregated firm panel results

The role of labor market institutions

Enforcement

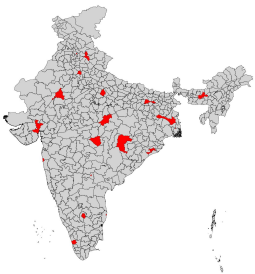
Use **distance from district to state-level institutions** (Ponczek and Ulyssea, 2022).

$$\text{Distance}_{i,j} = \text{Internal distance}_i + \text{Internal distance}_j + \text{Distance between centroids}_{i,j}$$

Chief labour commissioner offices:

- An organization set up to enforce labour laws and prevent industrial disputes
- Established centrally in 1945, and regional branches opened over time
- The branches are not located in every state (but in most large states)

Figure: Location of the regional labour offices



The role of labor market institutions

Enforcement

Within-industry analysis

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

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

Table of Contents

Implications for aggregate output

Macroeconomic development accounting literature to assess the effects of trade liberalization on aggregate output (McCaig and Pavcnik, 2018)

1. **Labour productivity gap.**
   - Labour productivity gap between sectors $s \in \{f, i\}$ is given by:

$$\frac{w_f}{w_i} = \frac{\text{MRPL}_f}{\text{MRPL}_i} = \frac{\text{ARPL}_f}{\text{ARPL}_i} \tag{4}$$

   - Two approaches using firm-level data (wage-based and revenue-based)
   - Adjusting for education differences, hours worked and different output elasticity of labour

2. **Share of workers reallocated between sectors.**
   - Given by the formula:

$$\sum_j \left(\hat{\beta} \times \Delta\tau_j\right) \times \mathit{IndustryShare}_j \tag{5}$$

Quantification in the spirit of McCaig and Pavcnik (2018)

Table: Labor productivity gap and reallocations 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)
Panel A: 1990s trade liberalization episode						
Unadjusted	36.6	17.2	30.8	10.7	23.6	9.18
Adjusted by human capital	29.7	13.9	25.8	8.97	15.8	6.16
+ measurement error in revenue and time worked	18.5		16.0		9.88	
+ differences in output-labor elasticity	12.3		10.7		6.58	
Share of workers reallocated to formal plants	0.099	0.099	0.133	0.133	0.099	0.099
Initial share of workers in the formal sector	0.209	0.209	0.175	0.175	0.353	0.353
Panel B: 2000s trade liberalization episode						
Unadjusted	17.2	6.68	15.4	7.51	8.29	2.33
Adjusted by human capital	14.0	5.42	12.8	6.28	5.56	1.56
+ measurement error in revenue and time worked	8.72		8.03		3.46	
+ differences in output-labor elasticity	5.81		5.35		2.31	
Share of workers reallocated to formal plants	0.047	0.047	0.060	0.060	0.051	0.051
Initial share of workers in the formal sector	0.178	0.178	0.124	0.124	0.283	0.283

Notes: Panel A is based on 1995 data, and panel B on 2005 data. 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. Human capital information comes from NSS-EU survey for the year 2000.

◂ High vs low tariff cuts ◂ High vs low unionization ◂ Pro-worker vs other states

Table of Contents

Conclusion

Our main contribution is exploiting an aggregated panel of firms and worker-level data covering 20 years of Indian trade reforms.

Our results **highlight the ambiguous effect of trade on informality:**

- Important to consider both output and input channels

Domestic labor market institutions, such as local union presence and imperfect enforcement, mitigate the effects of output competition, but do not appear to have effect through the input liberalisation channel.

We estimate a large labour productivity gap, and an overall formalisation of the economy.

Roughly 22 million manufacturing workers in India in the 1998 Economic Census $\Rightarrow$ potential large-scale movements of workers between sectors over the liberalisation period

Table of Contents

References I

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.

Amiti, M. and J. Konings (2007): "Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia," *American Economic Review*, 97(5), 1611–1638.

Bas, M. and P. Bombarda (2023): "Foreign Technology and Informal Employment: Evidence from Mexico," *Working paper*.

Bas, M. and C. Paunov (2021): "Input quality and skills are complementary and increase output quality: Causal evidence from Ecuador's trade liberalization," *Journal of Development Economics*.

Boehm, J. and E. Oberfield (2020): "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," *Quarterly Journal of Economics*, 135, 2007–2058.

Chakraborty, P., R. Singh, and V. Soundararajan (2021): "Import Competition, Formalization, and the Role of Contract Labor," *Working Paper*.

Chatterjee, S., K. Krishna, K. Padmakumar, and Y. Zhao (2024): "No Country for Dying Firms: Evidence from India," *Working paper*.

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

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

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

References II

Halpern, L., M. Koren, and A. Szeidl (2015): "Imported Inputs and Productivity," *American Economic Review*, 105, 3660–3703.

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*.

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

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).

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

Table of Contents

Formality shares in industries

Back

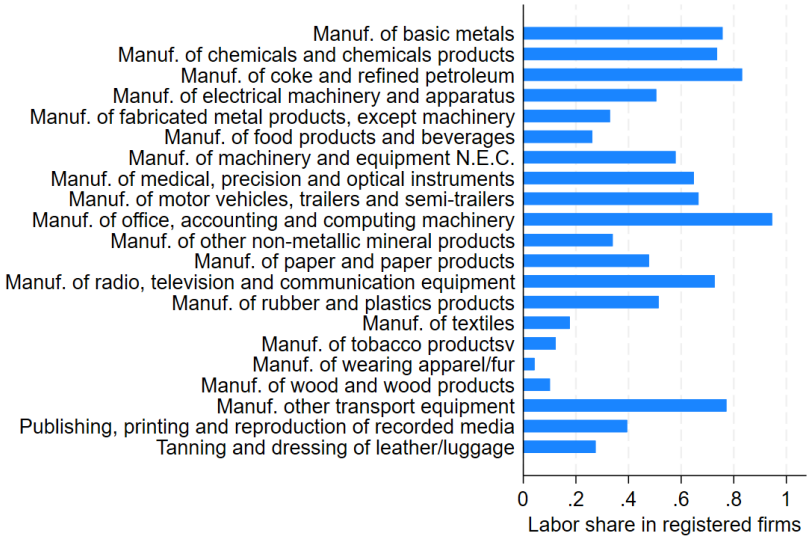


Figure: Share of formal workers in manufacturing industries in 2000

Measuring informality

	Industry-state firm panel mean/sd
Total labor	10240.65
	(27789.77)
Formal labor share (registered)	0.58
	(0.34)
Formal labor share (benefits)	0.52
	(0.34)
Total firms	2603.54
	(10414.62)
Formal firms share (registered)	0.25
	(0.29)
Formal firms share (benefits)	0.20
	(0.24)
Observations	1578

Calculation based on the ASI-UMES aggregated panel for 2005. Registration status is with respect to the Factories Act, 1948. Benefits are contribution to employees' provident fund.

Tariff change and initial industry characteristics

Back

Table: 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tariff evolution

Back

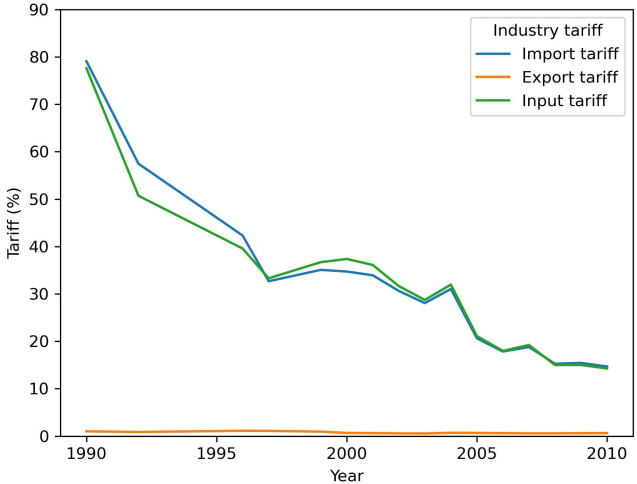


Figure: Indian tariffs evolution (1990-2010)

Tariff change across countries

Back

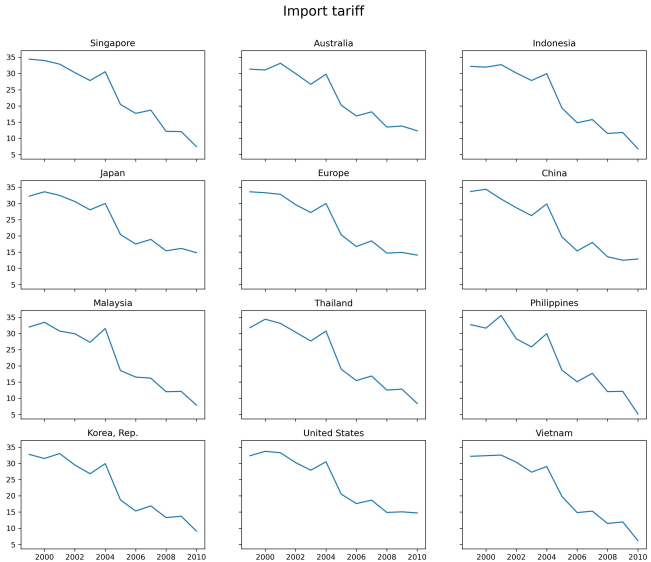
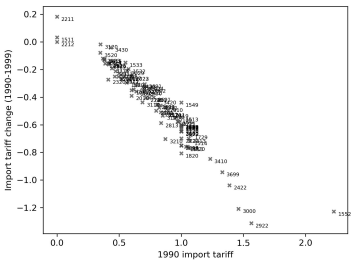


Figure: Country-specific variation in Indian output tariffs (1990-2010)

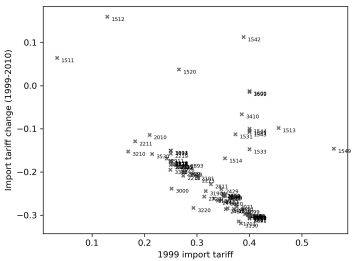
Exogeneity of tariffs

Industry tariffs

Back



1990-1999



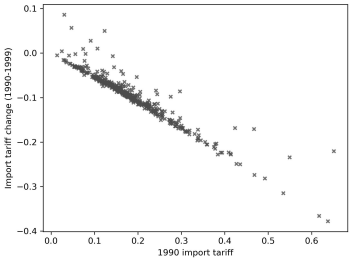
1999-2010

Figure: Industry import tariff change and initial tariff level, by decade

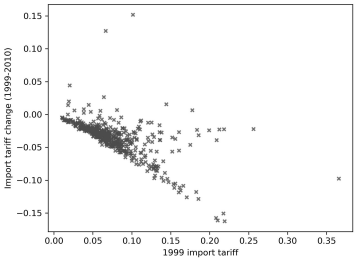
Exogeneity of tariffs

District tariffs

Back



1990-1999



1999-2010

Figure: District import tariff change and initial tariff level, by decade

Correlation between tariffs

Industry tariffs Back

	Output τ_{jt}	Input τ_{jt}	Export τ_{jt}
Output τ_{jt}	1.00		
Input τ_{jt}	0.60	1.00	
Export τ_{jt}	0.05	-0.06	1.00

Tariff in industry j at time t between 1990 and 2010.

Industry-state level results

Back

Share of formal firms

Within-industry analysis

	1990-2000			2000-2010		
VARIABLES	Registered	Registered	Provident Fund	Registered	Registered	Provident fund
Output τ_{jt}	-0.004	-0.017	-0.013	0.097***	0.107***	0.104***
	(0.015)	(0.017)	(0.013)	(0.017)	(0.021)	(0.014)
Input τ_{jt}	0.009	0.018	0.018	-0.116***	-0.077**	-0.084***
	(0.041)	(0.024)	(0.018)	(0.035)	(0.035)	(0.024)
Trend: 1990 industry firms		-0.000***	-0.000***			
		(0.000)	(0.000)			
Trend: 2000 industry firms					-0.000***	-0.000***
					(0.000)	(0.000)
Observations	2,775	2,775	2,775	3,247	3,247	3,247
R-squared	0.683	0.729	0.731	0.631	0.662	0.660
Year × state FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × 2d-ind FE	Yes	Yes	Yes	No	No	No
State × 3d-ind FE	Yes	Yes	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes	Yes	Yes

Panel for 1990, 1994, 2000, 2005 and 2010. Tariffs in industry j at time t. *ASI Labor Share* is the share of workers employed in ASI firms over total labor in ASI and UMES firms. *PF Labor Share* is the share of workers employed in firms reporting to provide provident fund (PF) over total labor in ASI and UMES firms. *Registered Labor Share* is the share of workers employed in firms reporting being registered under any regulation over total labor in ASI and UMES firms. Only manufacturing sectors. Heteroskedasticity-robust standard errors clustered by 2-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Aggregated firm panel analysis

Back

Dynamic effect

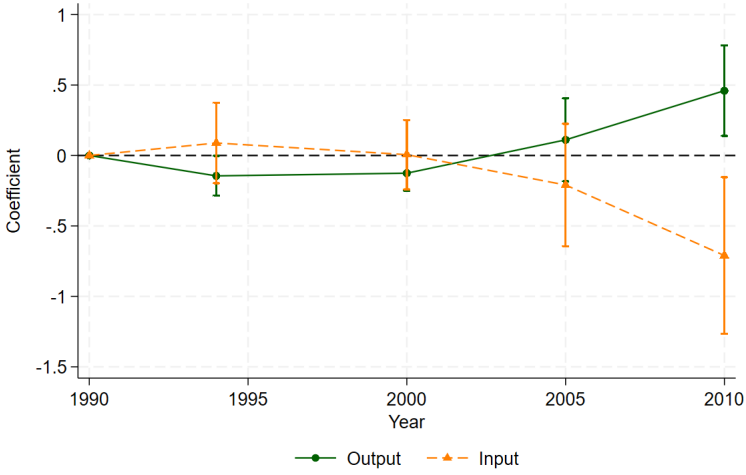


Figure: Dynamic effect of trade policy on the share of formal labor

Worker regressions

Industry size and tariffs Back

	Share of employment in industry j over total manufacturing employment			
	1987-1999		**1999-2010**	
	(1)	(2)	(3)	(4)
$\tau_{O,jt}$	0.000	0.001	-0.001	-0.007
	(0.002)	(0.002)	(0.002)	(0.006)
$\tau_{I,jt}$	-0.002	-0.002	-0.003	-0.000
	(0.006)	(0.007)	(0.006)	(0.008)
$\tau_{X,jt}$	-0.046	-0.065	0.001	-0.032***
	(0.040)	(0.066)	(0.046)	(0.008)
3d-industry FE	Yes	No	Yes	No
4d-industry FE	No	Yes	No	Yes
Round FE	Yes	Yes	Yes	Yes
Observations	243	214	341	341
R-squared	0.835	0.920	0.813	0.947

LPM estimation between for the period 1998-1999 and 1999-2010. Tariffs in industry j at time t. Dependent variable is the share of the workers reporting receiving retirement benefits from employer in each industry j. Only manufacturing sectors. Heteroskedasticity-robust standard errors clustered by 2-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Worker regressions

Industry size and tariffs

Back

Table: Decomposition of the change in formality shares between and within 4-digit industries (1999-2009)

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

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

Worker regressions

Personal characteristics

Back

VARIABLES	(1) Formal	(2) Formal	(3) Formal	(4) Formal	(5) Formal	(6) Formal	(7) Formal	(8) Formal
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)
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
Year × district FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3-digit sector 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

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. Pseudo-individual FE include gender, district and 5-year cohorts. Heteroskedasticity-robust standard errors clustered by 2-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Worker regressions

Including industry-level trends

Back

VARIABLES	(1) Formal	(2) Formal	(3) Formal	(4) Formal	(5) Formal
Import 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)
Observations	61,058	55,619	55,619	55,619	55,619
R-squared	0.440	0.451	0.452	0.453	0.452
Year × district FE	Yes	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes	Yes
3-digit sector FE	Yes	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes	Yes

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 2-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Worker regressions

Alternative definitions of informality

Back

VARIABLES	(1) Formal	(2) Formal (regular worker)	(3) Formal (wage worker)
Output τ_{jt}	0.120**	0.141***	0.192***
	(0.052)	(0.038)	(0.039)
Input τ_{jt}	-0.412***	-0.418***	-0.484***
	(0.084)	(0.079)	(0.094)
Observations	55,619	55,619	55,619
R-squared	0.452	0.500	0.509
Year × district FE	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes
3-digit sector FE	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes

LPM estimation between 1999 and 2010. Tariffs in industry *j* at time *t*. *Formal* equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. *Regular worker* refers to wage workers who are paid on a repeated, long-term basis. *Wage worker* are wage workers, as opposed to self-employed workers. Only manufacturing sectors. Pseudo-individual FE include gender, district and 5-year cohorts. Individual characteristics include age, square of age, years of education, household size, religion and urban location. Heteroskedasticity-robust standard errors clustered by 4-digit industries are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Worker regressions

Alternative definitions of informality

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

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

Worker regressions

Heterogeneity: urban or rural

Back

Table: 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)
$\tau_{O,jt}$	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)
$\text{Urban share}_j^{1999}$			0.103	0.102
			(0.068)	(0.068)
$\tau_{I,jt} \times \text{Urban}_i$				-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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Worker regressions

Heterogeneity: gender

Back

Table: 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)
$\tau_{O,jt}$	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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Worker regressions

Heterogeneity: skill

Back

Table: 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)
$\tau_{O,jt}$	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)
$\tau_{I,jt} \times$ Low-skill$_i$				-0.241***
				(0.061)
$\tau_{I,jt} \times$ Mid-skill$_i$				-0.294***
				(0.057)
$\tau_{I,jt} \times$ High-skill$_i$				-0.311***
				(0.054)
District $\times$ 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. 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Worker regressions

Heterogeneity: caste

Back

Table: 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)	(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)
$\tau_{I,jt} \times$ Scheduled caste$_i$				-0.022	0.005
				(0.043)	(0.041)
District × quarter-year FE	Yes	Yes	Yes	Yes	Yes
3- industry FE	Yes	Yes	Yes	Yes	Yes
Individual char.	Yes	Yes	Yes	Yes	Yes
Pseudo FE	Yes	Yes	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes	Yes
Observations	42,881	42,881	42,510	42,510	42,493
R-squared	0.523	0.523	0.522	0.522	0.540

Notes: Dependent variable equals to 1 if the worker reports receiving retirement benefits from employer, and to 0 otherwise. 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Worker regressions

Heterogeneity: age

Back

Table: 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)
$\tau_{O,jt}$	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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The role of labor market institutions

Back

Unions

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

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

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

The role of labor market institutions

State-level classification of regulations

Table: State-level classification of regulations, state-industry panel

Within-industry analysis

Samples	Firms			Workers		
	All	High LMR	Low LMR	All	High LMR	Low LMR
	(1)	(2)	(3)	(4)	(5)	(6)
$\tau_{O,jt}$	0.222***	0.527	0.220***	0.160***	0.150*	0.166***
	(0.083)	(0.360)	(0.083)	(0.049)	(0.079)	(0.049)
$\tau_{I,jt}$	-0.404***	-0.944***	-0.360***	-0.319***	-0.473***	-0.248***
	(0.127)	(0.164)	(0.135)	(0.077)	(0.105)	(0.067)
$\tau_{O,jt} \times$ High LMR$_s$	0.284			-0.011		
	(0.175)			(0.093)		
$\tau_{I,jt} \times$ High LMR$_s$	-0.385**			-0.079		
	(0.191)			(0.067)		
State $\times$ year FE	Yes	Yes	Yes	No	No	No
2-industry $\times$ year FE	Yes	Yes	Yes	No	No	No
3-industry $\times$ state FE	Yes	Yes	Yes	No	No	No
Industry trends	Yes	Yes	Yes	No	No	No
District $\times$ quarter-year FE	No	No	No	Yes	Yes	Yes
3- industry FE	No	No	No	Yes	Yes	Yes
Individual char.	No	No	No	Yes	Yes	Yes
Pseudo FE	No	No	No	Yes	Yes	Yes
Export τ_{jt}	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,972	637	1,329	36,379	12,879	23,500
R-squared	0.777	0.802	0.771	0.550	0.534	0.571

Notes: OLS estimation. The aggregated firm panel for the years 2000, 2005 and 2010 is used for columns (1) and (2) and (3), 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 (4), (5) and (6), 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. All regressions control for export tariffs. Heteroskedasticity-robust standard errors clustered by 4-digit industries-state are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Workers and informality in India (weighted table)

Back

	All mean/sd	**Tariff** mean/sd	**Wage worker** mean/sd	**Formal** mean/sd
Age	35.43	33.15	31.72	36.11
	(11.97)	(11.37)	(10.70)	(10.39)
Not literate	0.36	0.26	0.21	0.09
	(0.48)	(0.44)	(0.41)	(0.29)
High school	0.09	0.11	0.12	0.15
	(0.29)	(0.31)	(0.33)	(0.36)
Urban	0.24	0.46	0.56	0.68
	(0.43)	(0.50)	(0.50)	(0.47)
Formal	0.07	0.11	0.22	1.00
	(0.26)	(0.31)	(0.41)	(0.00)
Formal (regular worker)	0.16	0.31	0.64	0.95
	(0.37)	(0.46)	(0.48)	(0.21)
Formal (wage worker)	0.47	0.49	1.00	1.00
	(0.50)	(0.50)	(0.00)	(0.00)
1-5 workers	0.64	0.57	0.24	0.03
	(0.48)	(0.50)	(0.42)	(0.18)
6-9 workers	0.08	0.09	0.12	0.03
	(0.27)	(0.28)	(0.32)	(0.18)
10-19 workers	0.06	0.08	0.13	0.06
	(0.24)	(0.27)	(0.34)	(0.24)
20+ workers	0.13	0.19	0.37	0.74
	(0.34)	(0.39)	(0.48)	(0.44)
Observations	217377	23937	10130	2334

Calculation based on the NSSEU data for 2004-2005 (round 61). All summary statistics are weighted at the round level using sampling weights. *All* contains the total sample, *Tariff* is a subsample for which tariff information is available, *Wage worker* is a subsample for whichi tariff information is available and the worker is a wage worker, *Formal* is a subsample for which tariff information is available and the worker is formal (*i.e*, receives retirement benefits).

Initial district characteristics

Back

Table: Descriptive statistics of local areas

	Rural			Urban		
	1987/88	1999/00	2009/10	1987/88	1999/00	2009/10
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: labor market characteristics						
Share formal workers		0.06	0.05		0.24	0.21
		(0.08)	(0.05)		(0.09)	(0.08)
Share regular workers	0.10	0.09	0.09	0.40	0.38	0.39
	(0.09)	(0.09)	(0.08)	(0.10)	(0.11)	(0.12)
Share wage workers	0.39	0.42	0.44	0.56	0.56	0.56
	(0.16)	(0.17)	(0.17)	(0.09)	(0.09)	(0.11)
Panel B: liberalization policies						
Scaled output tariff	0.21	0.10	0.05	0.21	0.10	0.05
	(0.10)	(0.05)	(0.03)	(0.08)	(0.03)	(0.02)
Unscaled output tariff	0.74	0.37	0.17	0.74	0.36	0.16
	(0.13)	(0.06)	(0.06)	(0.10)	(0.04)	(0.05)
Scaled input tariff	0.15	0.08	0.04	0.15	0.08	0.04
	(0.07)	(0.03)	(0.02)	(0.06)	(0.03)	(0.01)
Unscaled input tariff	0.55	0.29	0.14	0.54	0.29	0.13
	(0.14)	(0.07)	(0.05)	(0.12)	(0.07)	(0.05)
Panel C: demographic characteristics						
Share not literate	0.58	0.46	0.33	0.24	0.18	0.13
	(0.18)	(0.18)	(0.16)	(0.11)	(0.09)	(0.08)
Share scheduled tribe/caste	0.19	0.20	0.22	0.13	0.15	0.16
	(0.12)	(0.13)	(0.15)	(0.07)	(0.08)	(0.08)
Share farming	0.73	0.72	0.64	0.11	0.09	0.08
	(0.16)	(0.17)	(0.18)	(0.08)	(0.06)	(0.06)
Share mining	0.01	0.01	0.01	0.01	0.01	0.01
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.01)
Share manufacturing	0.06	0.07	0.07	0.18	0.17	0.17
	(0.05)	(0.06)	(0.06)	(0.09)	(0.08)	(0.09)
Share services	0.07	0.07	0.07	0.28	0.28	0.26
	(0.07)	(0.08)	(0.05)	(0.10)	(0.10)	(0.08)
Share transport	0.01	0.02	0.03	0.07	0.07	0.08
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Share trade	0.04	0.05	0.06	0.18	0.23	0.22
	(0.02)	(0.03)	(0.04)	(0.05)	(0.05)	(0.08)
Share construction	0.04	0.04	0.11	0.06	0.08	0.11
	(0.07)	(0.05)	(0.09)	(0.03)	(0.04)	(0.04)
Log per capita consumption	6.23	6.24	6.37	6.58	6.67	6.75
	(0.23)	(0.23)	(0.25)	(0.20)	(0.17)	(0.22)
Observations	393	393	393	73	73	73

This table provides summary statistics of the variables included in the analysis at the local level. Due to the nature of the NSS data, it is not possible to aggregate urban and rural areas together at the district level until after the 1999/00 round. Results are therefore presented separately for rural areas (columns 1, 2 and 3) and urban areas (columns 4, 5 and 6). Data on benefits received by workers was not available in the 1987/88 round.

Migration in India

Back

Table: Migration patterns of local areas

	Rural		Urban	
	1987/88	1999/00	1987/88	1999/00
	(1)	(2)	(3)	(4)
Migrated	0.26	0.24	0.37	0.35
	(0.16)	(0.16)	(0.17)	(0.17)
Migrated in the past 10 years	0.95	0.95	0.89	0.90
	(0.05)	(0.04)	(0.06)	(0.06)
Migrated in the past 10 years in another district	0.79	0.82	0.78	0.80
	(0.14)	(0.13)	(0.11)	(0.10)
Migrated in the past 10 years in another district to seek employment	0.01	0.01	0.07	0.07
	(0.02)	(0.02)	(0.06)	(0.06)
Observations	392	393	73	73

This table provides summary statistics of the variables included in the analysis at the local level. Due to the nature of the NSS data, it is not possible to aggregate urban and rural areas together at the district level until after the 1999/00 round. Results are therefore presented separately for rural areas (columns 1 and 2) and urban areas (columns 3 and 4).

Quantification: high vs low tariff cuts

Back

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

	All		Large Δ tariff		Small Δ tariff	
	Revenue based (1)	Wage based (2)	Revenue based (3)	Wage based (4)	Revenue based (5)	Wage based (6)
Panel A: 1990s trade liberalization episode						
Unadjusted	36.6	17.2	27.7	10.2	43.5	32.3
Adjusted by human capital	29.7	13.9	23.2	8.57	34.8	25.9
+ measurement error in revenue and time worked	18.5		14.4		21.7	
+ differences in output-labor elasticity	12.3		9.64		14.4	
Share of workers reallocated to formal plants	0.099	0.099	0.144	0.144	0.075	0.075
Initial share of workers in the formal sector	0.209	0.209	0.252	0.252	0.160	0.160
Panel B: 2000s trade liberalization episode						
Unadjusted	17.2	6.68	11.9	4.53	25.6	10.4
Adjusted by human capital	14.0	5.42	10.0	3.79	20.5	8.33
+ measurement error in revenue and time worked	8.72		6.23		12.7	
+ differences in output-labor elasticity	5.81		4.15		8.52	
Share of workers reallocated to formal plants	0.047	0.047	0.060	0.060	0.036	0.036
Initial share of workers in the formal sector	0.178	0.178	0.180	0.180	0.161	0.161

Notes: Panel A is based on 1995 data, and panel B on 2005 data. 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. Human capital information comes from NSS-EU survey for the year 2000.

Quantification: high vs low unionization

Back

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

	All		High unionization		Low unionization	
	Revenue based (1)	Wage based (2)	Revenue based (3)	Wage based (4)	Revenue based (5)	Wage based (6)
Panel A: 1990s trade liberalization episode						
Unadjusted	36.6	17.2	24.2	11.5	55.0	26.2
Adjusted by human capital	29.7	13.9	19.9	9.48	44.4	21.2
+ measurement error in revenue and time worked	18.5		12.4		27.6	
+ differences in output-labor elasticity	12.3		8.30		18.4	
Share of workers reallocated to formal plants	0.099	0.099	0.102	0.102	0.097	0.097
Initial share of workers in the formal sector	0.209	0.209	0.265	0.265	0.167	0.167
Panel B: 2000s trade liberalization episode						
Unadjusted	17.2	6.68	14.2	4.79	20.2	9.78
Adjusted by human capital	14.0	5.42	11.7	3.94	16.3	7.90
+ measurement error in revenue and time worked	8.72		7.31		10.1	
+ differences in output-labor elasticity	5.81		4.87		6.79	
Share of workers reallocated to formal plants	0.047	0.047	0.046	0.046	0.047	0.047
Initial share of workers in the formal sector	0.178	0.178	0.202	0.202	0.156	0.156

Notes: Panel A is based on 1995 data, and panel B on 2005 data. 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. Human capital information comes from NSS-EU survey for the year 2000.

Quantification: high-LMR vs low-LMR

Back

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

	All		High unionization		Low unionization	
	Revenue based (1)	Wage based (2)	Revenue based (3)	Wage based (4)	Revenue based (5)	Wage based (6)
Panel A: 1990s trade liberalization episode						
Unadjusted	36.6	17.2	32.1	17.4	39.1	18.3
Adjusted by human capital	29.7	13.9	25.6	13.9	32.3	15.1
+ measurement error in revenue and time worked	18.5		15.9		20.1	
+ differences in output-labor elasticity	12.3		10.6		13.4	
Share of workers reallocated to formal plants	0.099	0.099	0.099	0.099	0.101	0.101
Initial share of workers in the formal sector	0.209	0.209	0.215	0.215	0.190	0.190
Panel B: 2000s trade liberalization episode						
Unadjusted	17.2	6.68	16.2	5.33	19.6	8.37
Adjusted by human capital	14.0	5.42	13.0	4.26	16.1	6.91
+ measurement error in revenue and time worked	8.72		8.10		10.0	
+ differences in output-labor elasticity	5.81		5.40		6.72	
Share of workers reallocated to formal plants	0.047	0.047	0.045	0.045	0.046	0.046
Initial share of workers in the formal sector	0.178	0.178	0.168	0.168	0.182	0.182

Notes: Panel A is based on 1995 data, and panel B on 2005 data. 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. Human capital information comes from NSS-EU survey for the year 2000.