Value Chains and Productivity Transmission

Ezra Oberfield

March 27, 2026

STEG Virtual Course on Private Enterprises, Productivity, and Economic Growth

Input-Output Architecture and Growth

- How does the economic environment shape how the economy is organized?
  - How is economic activity organized across firms
  - How do firms organize themselves

- What determines aggregate productivity?
  - Traditional perspective: Aggregate of Individual Producers
  - Endogenous production network perspective: relationships are an essential part of aggregate productivity

- Organization depends on:
  - Search Frictions (Miyauchi, 2024)
  - Contracting frictions (Boehm, 2020; Boehm Oberfield, 2020, Startz)
  - Geographical/trade frictions (Chaney, 2014; Eaton, Kortum, Kramarz)
  - Reliability/fragility (Elliott Golub Leduc, 2022; Acemoglu, Tahbaz Salehi, 2025)
  - Quality mismatch (Demir Fieler Xu Yang, 2024)
  - Cultural Proximity (Fujiy Khanna Toma, 2025)
  - Market power/thickness
  - Financial frictions/trade credit

Production Networks and Development: Plan for today

1. How does the organization of economic activity within and across firms change with development:
    - Growth and the Fragmentation of Production (with Johannes Boehm)
2. Contracting frictions distort production and relationships
    - Misallocation in the Market for Inputs: Enforcement and the Organization of Production (with Johannes Boehm)

A Theory of Input-Output Architecture (2018)

- Fixed continuum of entrepreneurs $j \in [0,1]$, each produces differentiated variety
- To produce use labor and one input

$$y_b = \frac{1}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z_{bs} x_s^\alpha l^{1-\alpha} \qquad \implies c_b = \frac{1}{z_{bs}} c_s^\alpha w^{1-\alpha}$$

- Draw many potential suppliers: Poisson with mean M. For each, $z_{bs} \sim H(z)$
- Define $q_j = w/c_j$, so that $q_b = z_{bs} q_s^\alpha$. Let $F(q)$ be CDF of q in cross section
  - M is arrival rate of potential supplier
  - $M \int \left[1 - F((q/z)^{1/\alpha})\right] dH(z)$ is arrival rate of supplier such that $z_{bs} q_s^\alpha > q$

$$\underbrace{F(q)}_{\Pr(q_j \leq q)} = \underbrace{e^{-M \int \left[1-F((q/z)^{1/\alpha})\right] dH(z)}}_{\text{Prob no such arrivals}} \qquad F \text{ is fixed point}$$

A Theory of Input-Output Architecture (2018)

- Fixed continuum of entrepreneurs $j \in [0,1]$, each produces differentiated variety
- To produce use labor and one input

$$y_b = \frac{1}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z_{bs} x_s^\alpha l^{1-\alpha} \qquad \Longrightarrow \; c_b = \frac{1}{z_{bs}} c_s^\alpha w^{1-\alpha}$$

- Draw many potential suppliers: Poisson with mean M. For each, $z_{bs} \sim H(z)$
- Define $q_j = w/c_j$, so that $q_b = z_{bs} q_s^\alpha$. Let $F(q)$ be CDF of q in cross section
  - M is arrival rate of potential supplier
  - $M \int \left[1 - F((q/z)^{1/\alpha})\right] dH(z)$ is arrival rate of supplier such that $z_{bs} q_s^\alpha > q$

$$\underbrace{F(q)}_{\Pr(q_j \leq q)} = \underbrace{e^{-M \int \left[1-F((q/z)^{1/\alpha})\right] dH(z)}}_{\text{Prob no such arrivals}} \qquad F \text{ is fixed point}$$

- Functional form: Arrival of potential supplier with $z_{bs} > z$ is $M[1 - H(z)] = m z^{-\zeta}$.

$$F(q) = e^{-[\Gamma(1-\alpha)m]^{\frac{1}{1-\alpha}} q^{-\zeta}}$$

A Theory of Input-Output Architecture (2018)

- Model can be used when changes in economic environment:
  - change set of potential suppliers
  - change which supplier one chooses
  - changes nature of interactions with a given supplier

Other Classes of Production Network Models

- Chaney (2014): Continuum of suppliers, limited by search (random + friend of friend)
- Lim (2018): Continuum of suppliers, random fixed cost for each
- Global Value Chains: (Costinot Vogel Wang, 2013; Fally Hillberry, 2018; Antras de Gortari 2020) / Global Sourcing (Antras Fort Tintelnot, 2017)

Growth and the Fragmentation of Production (with Johannes Boehm)

Specialization and Productivity

- Smithian Growth: Specialization and Development
- This paper: specialization in value chain *among plants* and growth
- Empirical facts about organization and performance using manufacturing data from India
  - Measure specialization in value chains
  - macro correlations: vertical specialization $\Leftrightarrow$ income per capita
  - micro correlations: vertical specialization $\Leftrightarrow$ plant size
  - Causality? Mechanisms?
- Model: Capture mechanisms, implications for growth

Manufacturing Plants in India

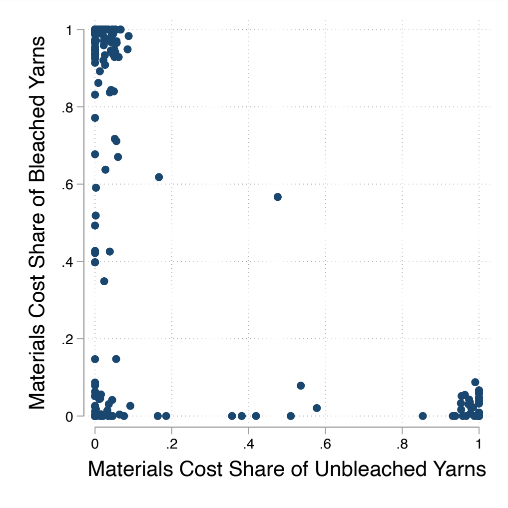
Data: **Indian Annual Survey of Industries**, 1989/90-2014/15 (with gaps)

- Plant-level panel survey of formal manufacturing plants
  - All plants that have 100+ employees
  - 1/5 of all plants between 20 (10 if using power) and 100 employees

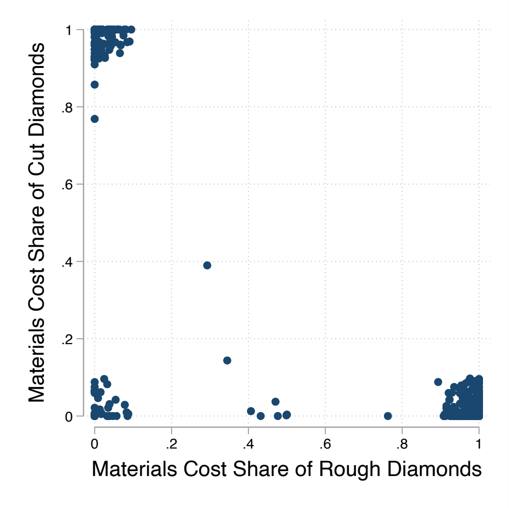
Most important part of the survey:

- Quantities, unit values & 5-digit product codes for all manufacturing output and intermediate inputs (domestic and imported)

Within narrow industries, firms use different inputs



(a) Input mixes for Bleached Cotton Cloth (63303)

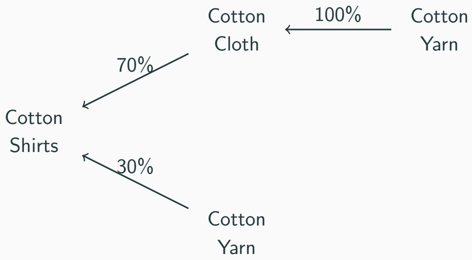


(b) Input mixes for Polished Diamonds (92104)

Measuring the vertical span of production (Boehm & Oberfield, 2020)

Fix an output industry.

1. Vertical Distance of Input in supply chain: is input close or far from output?
   - Similar to upstreamness of Alfaro et al. (2019)



- Cloth: Close (1 plant away)
- Yarn: Farther (1.7 plants away)

2. Vertical Span of Plant: Does shirt producer use inputs that are distant or close?
   - Purchase cloth? $\implies$ vertical span $= 1$
   - Purchase yarn? $\implies$ vertical span $= 1.7$
   - Use many inputs: weighted average distance of each input

Vertical Distance of inputs from output – Examples

Table 1: Vertical distance examples for 63428: *Cotton Shirts*

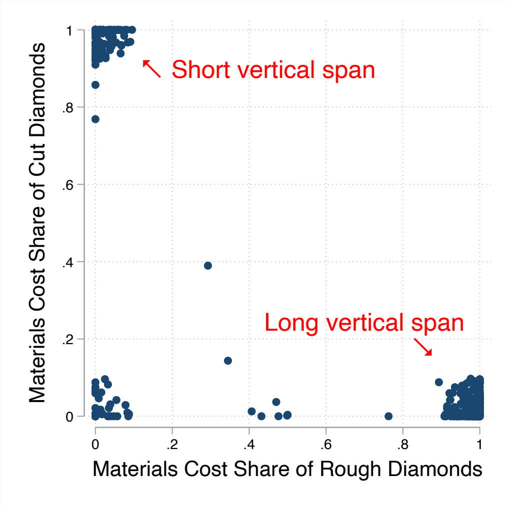
	Mean Vertical Distance
Fabrics/Cloths	1.66
Yarns	2.58
Ginned & pressed cotton	3.44
Raw cotton	4.09

Table 2: Vertical distance examples for 73107: *Aluminium Ingots*

	Vertical Distance
Anodes, copper	1.00
Aluminium scrap	1.19
Aluminium oxide	1.25
Bauxite, calcined	2.18
Caustic soda (sodium hydroxide)	2.39
Bauxite, raw	3.03
Coal	3.43

Long and short vertical span

Figure 1: Input mixes for Polished Diamonds (92104)



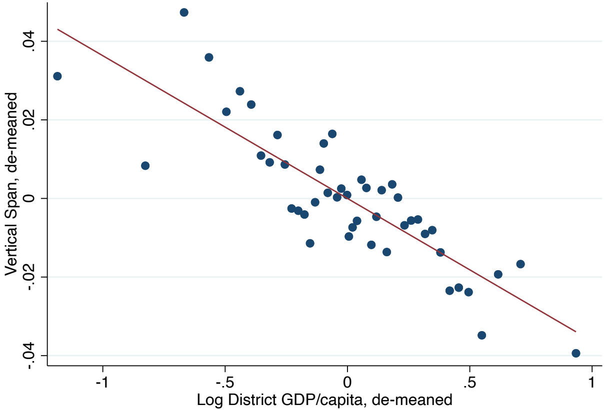
Motivating Facts about Vertical Specialization

Macro facts: Vertical specialization is positively correlated with development

Within industry × year:

Plants in **richer districts** are on average **more vertically specialized**

Holds in the **time dimension**:
In Indian states that grew faster, plants vert. specializing more ▸ show



Binscatter with n=50. Y and X variable de-meaned by industry x year. SP plants only.

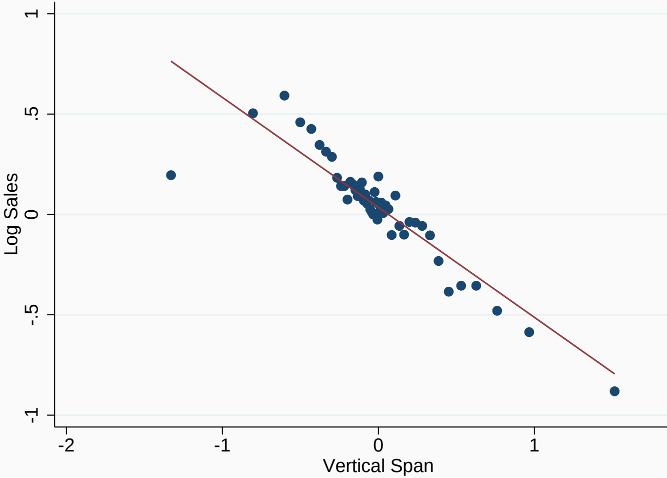
▸ Single-plant

Micro facts: Vertical specialization is positively correlated with plant size

Within industry × year:

Plants with **higher sales** tend to have **shorter vertical span**

Holds also in the **time dimension**: Plants that v. specialize more grow faster ▸ show



SP plants only. Regression includes industry x year FE's.

▸ Single-plant

Causality? Size and Span jointly determined

1. Indian Trade Liberalization:
   - **Until end of 80s:** India in near-autarky
     - Import licensing system
     - Very high tariffs. Large variation (up to 355%), average ∼80%. Was set in the 1950s.
   - **July 1991:** Balance of Payments crisis. Removal of import licensing system, starts cutting tariffs.
   - **1992-1997:** Tariffs come down to average of 35%, ending up fairly uniform. ▸ Tariffs
   - ⇒ **tariff change was determined in the 50's**
   - ⇒ **tariff changes are uncorrelated with 1992 industry characteristics** (Khandelwal and Topalova, 2010: "as exogenous to the state of the industries as a researcher might hope for").
   - See also Panagariya (2004), Sivadasan (2009), Khandelwal and Topalova (2010), Goldberg et al. (2010).

2. Hummels et al. (2014)-type instrument: shift in export demand:
   - Industries export to different destinations
   - Changes in destination demand for product (leaving out India)

▸ Tariff details

Tariff changes act as demand & supply shocks

	Dependent variable: Log Sales			
	(1)	(2)	(3)	(4)
$\log(1+\tau_{it}^{\text{output}})$	0.315**	0.385**		0.726**
	(0.10)	(0.11)		(0.16)
$\log(1+\bar{\tau}_{it}^{\text{input}})$		-0.192^{+}		-0.607**
		(0.12)		(0.20)
$\text{ExDemand}_{\omega t}$			0.0467**	0.0507**
			(0.0080)	(0.010)
Year FE	Yes	Yes	Yes	Yes
Plant × Industry FE	Yes	Yes	Yes	Yes
R^2	0.943	0.943	0.947	0.949
Observations	160254	160233	162421	124098

Standard errors in parentheses, clustered at the state × industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

⇒ Use Δ **import tariffs** in the output good or **Hummels instrument** as a **demand shifter** ▸ 90s

Demand $\nearrow \Rightarrow$ firms reduce vertical span

	Dependent variable: Vertical Span					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Sales	-0.0191**	-0.0197**	-0.512+	-0.253*	-0.164*	-0.268**
	(0.0020)	(0.0024)	(0.28)	(0.13)	(0.073)	(0.094)
$\log(1+\bar{\tau}^{\text{input}}_{j\omega t})$		-0.0211		0.0198		0.0552
		(0.050)		(0.065)		(0.10)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})\overline{\text{span}}_j$		-0.0905		-0.219*		-0.372**
		(0.056)		(0.099)		(0.13)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})(\text{distance}_{\omega i}-\overline{\text{span}}_j)$		-0.121		-0.337*		-0.498*
		(0.095)		(0.15)		(0.21)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant × Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV 1	IV 1	IV 2	IV 2
R^2	0.765	0.731	-1.049	-0.232	-0.0816	-0.227
First stage w. id. robust F			5.566	10.91	26.53	15.70
Observations	186628	145165	138204	137059	143428	110578

Standard errors in parentheses, clustered at the state-industry level. $^{+}$ $p<0.10$, * $p<0.05$, ** $p<0.01$

Instrument: Col. 3, 4: log output tariff: Col 5, 6: Export demand shifter

▸ Sel: SP/MP
▸ Changes ▸ 90s
▸ GIV ▸ RTS

Demand $\nearrow$ $\Rightarrow$ firms reduce the actual number of inputs

	Dependent variable: # Inputs					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Sales	0.0359**	0.0374**	-1.136*	-0.468+	-0.383+	-0.542*
	(0.0030)	(0.0035)	(0.54)	(0.26)	(0.20)	(0.25)
$\log(1+\bar{\tau}_{j\omega t}^{\text{input}})$		-0.441**		-0.366*		-0.207
		(0.16)		(0.14)		(0.19)
$\sum_i \alpha_i \log(1+\bar{\tau}_{it}^{\text{input}})\overline{\text{span}}_j$		0.125		-0.126		-0.337
		(0.10)		(0.14)		(0.21)
$\sum_i \alpha_i \log(1+\bar{\tau}_{it}^{\text{input}})(\text{distance}_{\omega i} - \overline{\text{span}}_j)$		0.162		-0.202		-0.722*
		(0.10)		(0.25)		(0.30)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant $\times$ Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV 1	IV 1	IV 2	IV 2
R^2	0.856	0.847	-4.788	-0.874	-0.574	-1.059
First stage w. id. robust F			5.566	10.91	26.53	15.70
Observations	186628	145165	138204	137059	143428	110578

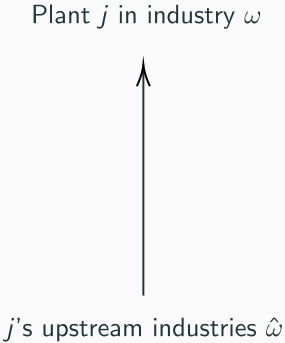
Changes within plant-products
Standard errors in parentheses, clustered at the state-industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

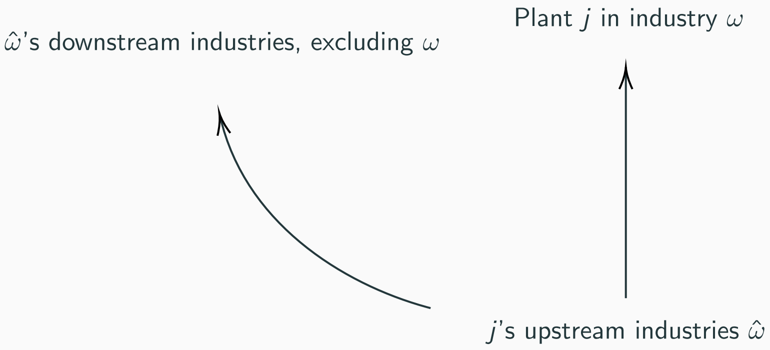
Young (1928): Economies of scale? Increasing Returns? Network Externalities?

Plant j in industry ω

Young (1928): Economies of scale? Increasing Returns? Network Externalities?

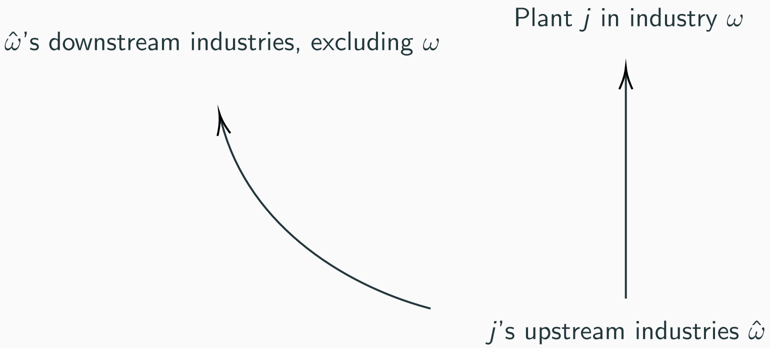


Young (1928): Economies of scale? Increasing Returns? Network Externalities?

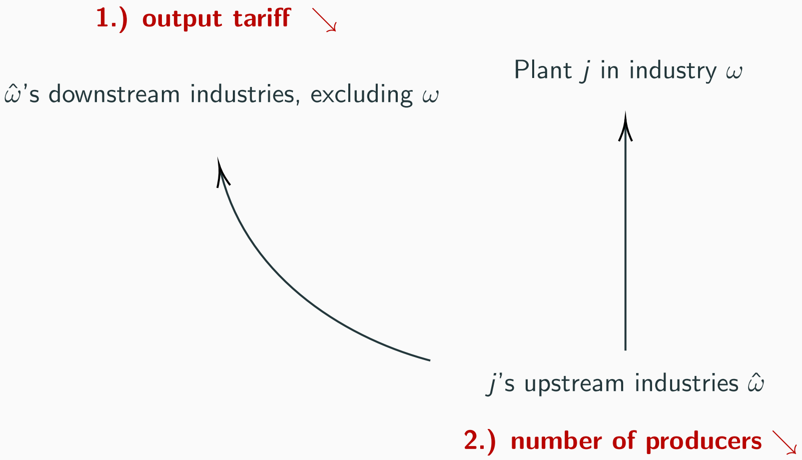


Young (1928): Economies of scale? Increasing Returns? Network Externalities?

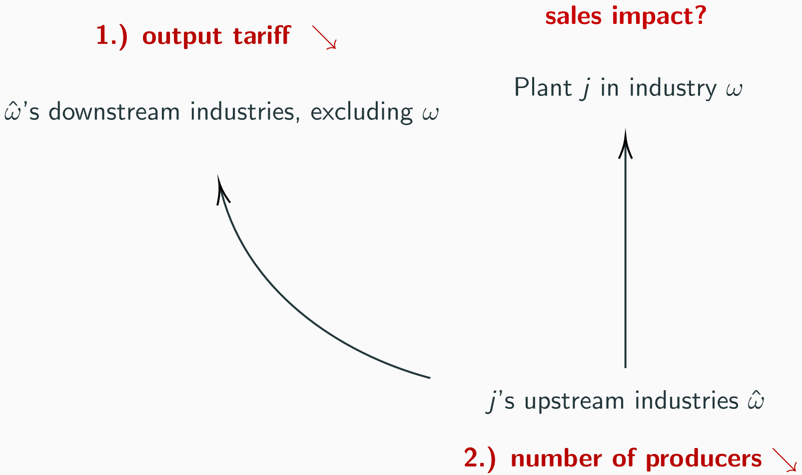
1.) output tariff $\searrow$



Young (1928): Economies of scale? Increasing Returns? Network Externalities?



Young (1928): Economies of scale? Increasing Returns? Network Externalities?



Upstream entry and sales

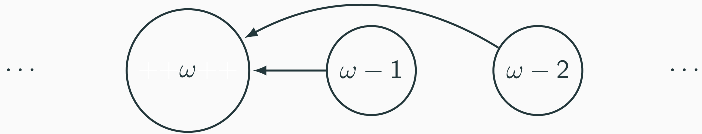
	Dependent variable: log Sales					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg. log #Producers in Upstream Ind.	0.0383**	0.0375**	0.0611**	0.0735**	0.254**	0.121**
	(0.0050)	(0.0060)	(0.017)	(0.018)	(0.036)	(0.018)
$\log(1+\bar{\tau}^{\text{input}}_{j\omega t})$		-0.0242		-0.0207		-0.0162
		(0.096)		(0.097)		(0.097)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})(\text{distance}_{\omega i} - \overline{\text{span}}_j)$		-0.331**		-0.334**		-0.337**
		(0.11)		(0.11)		(0.11)
Industry × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant × Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV 1	IV 1	IV 2	IV 2
R^2	0.952	0.954	0.000636	0.000292	-0.0301	-0.00391
First-stage w. id. F			615.6	517.1	88.48	437.5
Observations	199039	142006	199039	142006	199039	142006

Standard errors in parentheses, clustered at the industry-year level. $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

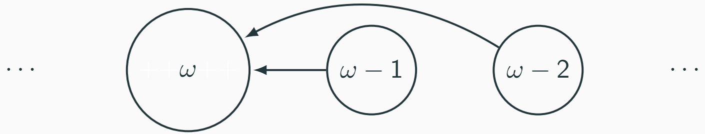
▸ 90s

▸ Sales

Simple Model



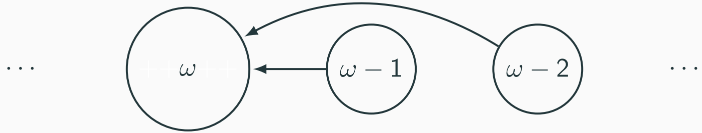
- Industries on a (large) circle, indexed by ω
- Entry stage: Firm j in ω born with q_j $\implies c_{j\omega} = \frac{1}{q_j} w^{1-\alpha} \tilde{c}^{\alpha}_{j,\omega-1}$
- Two ways of producing, $\tilde{c}_{j,\omega-1} = \min\{\tilde{c}^{o}_{j,\omega-1}, c^{i}_{j,\omega-1}\}$:
  1. Buying $\omega-1$ from a supplier ('shirts from cloth')
  2. Buying $\omega-2$ from a supplier ('shirts from cotton')
  - Exerts search effort $h_{j1}, h_{j2} \implies$ matching process $\implies$ potential suppliers $S_{j,\omega-1}$, $S_{j,\omega-2}$



- Industries on a (large) circle, indexed by ω
- Entry stage: Firm j in ω born with q_j $\implies$ $c_{j\omega} = \frac{1}{q_j} w^{1-\alpha} \tilde{c}^{\alpha}_{j,\omega-1}$
- Two ways of producing, $\tilde{c}_{j,\omega-1} = \min\{\tilde{c}^{o}_{j,\omega-1}, c^{i}_{j,\omega-1}\}$:
  1. **Buying $\omega - 1$ from a supplier**

     - # of suppliers of $\omega - 1$ with match-spec. prod. $> z \sim Poisson\big(h_{j1} m_{\omega-1} z^{-\zeta}\big)$
     - h_{j1}: search effort; $m_{\omega-1} \equiv M(J_{\omega-1})$: matching efficiency;

$$\tilde{c}^{o}_{j1} = \min_{s \in S_{j,\omega-1}} \frac{\text{price}_s}{\text{match-specific prod}_{js}} \sim EV(h_{j1} m_{\omega-1} \bar{c}^{-\zeta}_{\omega-1}, \zeta)$$



- Industries on a (large) circle, indexed by ω
- Entry stage: Firm j in ω born with q_j $\implies c_{j\omega} = \frac{1}{q_j} w^{1-\alpha} \tilde{c}^{\alpha}_{j,\omega-1}$
- Two ways of producing, $\tilde{c}_{j,\omega-1} = \min\{\tilde{c}^{o}_{j,\omega-1}, c^{i}_{j,\omega-1}\}$:

1. **Buying $\omega - 1$ from a supplier** $\tilde{c}^{o}_{j,\omega-1} \sim EV(h_1 m_{\omega-1} \bar{c}^{-\zeta}_{\omega-1}, \zeta)$

2. **Buying $\omega - 2$ from a supplier**

- # of suppliers of $\omega - 2$ with match-spec. prod. $> z \sim Poisson\left(h_{j2} m_{\omega-2} z^{-\zeta}\right)$
- task-specific prod. $b_{j,\omega-1}$, s.t. $b^{\zeta}_{j,\omega-1} \sim Stable(\alpha)$

$$\tilde{c}^{i}_{j,\omega-1} = \frac{1}{b_{j,\omega-1}} w^{1-\alpha} \tilde{c}^{\alpha}_{j2} \sim EV\left(h^{\alpha}_{j2} m^{\alpha}_{\omega-2} (w^{1-\alpha} \bar{c}^{\alpha}_{\omega-2})^{-\zeta}, \zeta\right)$$

Probability of Outsourcing

For firm j in ω that exerts search effort h_{j1}, h_{j2}:

- the effective cost of ω_1 is

$$\tilde{c}_{j,\omega-1} = \min\{\tilde{c}^o_{j,\omega-1}, \tilde{c}^i_{j,\omega-1}\} \sim EV\Big(h_{j1} m_{\omega-1} \bar{c}^{-\zeta}_{\omega-1} + h^{\alpha}_{j2} m^{\alpha}_{\omega-2} (\bar{c}^{\alpha}_{\omega-2} w^{1-\alpha})^{-\zeta}, \zeta\Big)$$

- and the probability of buying $\omega - 1$ is

$$\frac{h_{j1} m_{\omega-1} \bar{c}^{-\zeta}_{\omega-1}}{h_{j1} m_{\omega-1} \bar{c}^{-\zeta}_{\omega-1} + h^{\alpha}_{j2} m^{\alpha}_{\omega-2} (\bar{c}^{\alpha}_{\omega-2} w^{1-\alpha})^{-\zeta}}$$

Search problem: Nonhomotheticity

- Profits from sales to households, isoelastic demand, isoelastic search costs:

$$\max_{h_{j1},h_{j2}} \underbrace{A_\omega q_j^{\varepsilon-1}\left\{\left[h_{j1}m_{\omega-1}\bar{c}_{\omega-1}^{-\zeta}+h_{j2}^{\alpha}m_{\omega-2}^{\alpha}(\bar{c}_{\omega-2}^{\alpha}w^{1-\alpha})^{-\zeta}\right]^{-\frac{\alpha}{\zeta}}w^{1-\alpha}\right\}^{1-\varepsilon}}_{\mathbb{E}(\pi_j|q_j,h_{j1},h_{j2})}-\sum_{i=1,2}\frac{wk}{1+\gamma}h_{ji}^{1+\gamma}$$

Proposition

Under the optimal search effort, the probability of using ω_1 is

- *increasing in Hicks-neutral productivity q,*
- *increasing in the final consumer's demand for ω*

Size $\leftrightarrow$ **Span relationship** in the data

Non-homotheticity: Some intuition

- Two key ingredients
  - Make and Buy are substitutes
  - Make requires labor

- Expect to be large $\implies$ search effort $\uparrow$

$\implies$ Labor expensive relative to intermediates ($\omega - 1$ and $\omega - 2$)

$\implies$ Buy (to save on labor cost)

- Expect to be small $\implies$ search effort $\downarrow$

$\implies$ Intermediates ($\omega - 1$ and $\omega - 2$) expensive relative labor

$\implies$ Make (to save on intermediates)

- Reinforced by search direction
  - i.e., if expecting to buy, direct search toward suppliers of $\omega - 1$ ($h_1/h_2 \uparrow$)

Demand & entry

Consumption: Representative household has standard nested CES preferences

$$u = \left(\sum_{\omega} \delta_{\omega}^{\frac{1}{\eta}} u_{\omega}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}, \qquad u_{\omega} = \left(\int_{J_{\omega}} u_{\omega j}^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}}, \qquad \varepsilon > \eta > 1$$

Market Structure: Firms sell to firms further downstream, and to final consumers.

- Firms price at marginal cost when selling to other firms
- Firms are monopolistically competitive when selling to final consumers.

Entry: Representative entrepreneur chooses

$$\max_{\{J_{\omega}\}} \sum J_{\omega} \bar{\pi}_{\omega} - w \frac{k^E}{1 + 1/\chi} \left(\sum_{\omega} (\delta_{\omega}^E)^{1/\beta} J_{\omega}^{\frac{1+\beta}{\beta}} \right)^{\frac{\beta}{1+\beta} \frac{\chi+1}{\chi}}$$

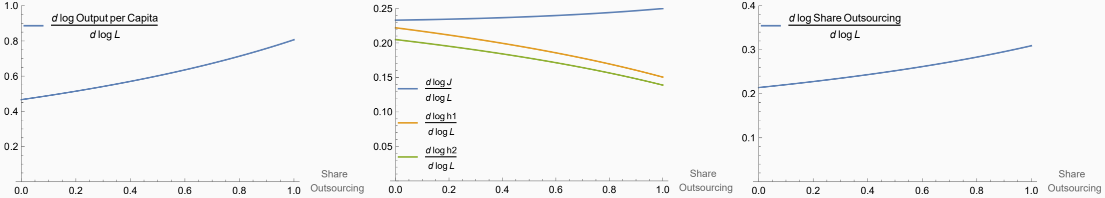
- Nests free entry ($\beta, \chi \to \infty$) and inelastic entry ($\beta = \chi = 0$) as special cases.
- Assume $\beta, \chi < \infty$: increase in demand is not fully absorbed by new entrants

Economy-wide increase in scale: $L \uparrow$

Consider symmetric industries, same q for all firms:

$$\frac{d \log \bar{c}/w}{d \log L} = -\frac{s}{1-s}\frac{1}{\zeta}\left\{\mu \frac{d \log J}{d \log L} + \frac{s_1}{s}\frac{d \log h_1}{d \log L} + \frac{s_2}{s}\frac{d \log h_2}{d \log L}\right\}$$

where $s = \underbrace{O\alpha}_{s_1} + \underbrace{(1-O)\alpha^2}_{s_2}$ is aggregate expenditure share on intermediate inputs



(a) Response of Income/capita **(b)** Response of Search Effort, Entry **(c)** Response of Share Outsourcing

Economy-wide increase in scale: $L \uparrow$

$$\frac{d \log \bar{c}/w}{d \log L} = -\frac{s}{1-s}\frac{1}{\zeta}\left\{\mu \frac{d \log J}{d \log L} + \frac{s_1}{s}\frac{d \log h_1}{d \log L} + \frac{s_2}{s}\frac{d \log h_2}{d \log L}\right\}$$

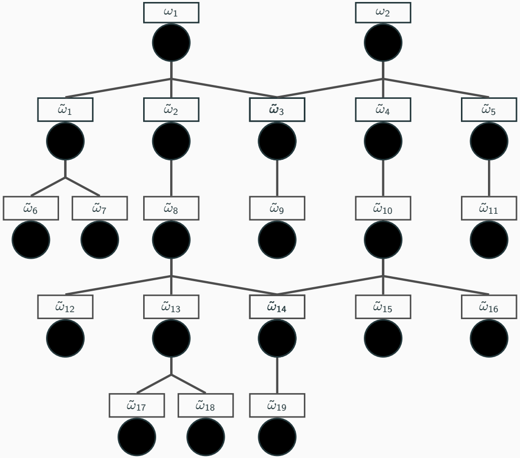
- $L \uparrow \implies$ Search effort, entry $\uparrow$
- Aggregate increasing returns to scale: $L \uparrow \implies u \uparrow$
- Like I/O multiplier: Response is larger if firms are more specialized
- Feedback: $L \uparrow \implies$ more specialization ($s \uparrow$)

  $\implies$ **Response gets stronger with development**

- Channel absent in canonical I/O models Acemoglu, Carvalho, Ozdaglar, Tahbaz-Salehi (2012), Jones (2011,2013), Baqaee-Farhi (2019,2020), Boehm-Oberfield (2020)
  - Reminiscent of Ciccone (2002)

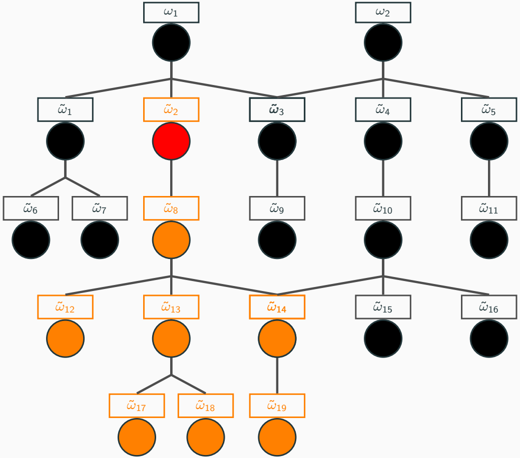
Full Model: Technology Menu

Industries arranged as a directed forest



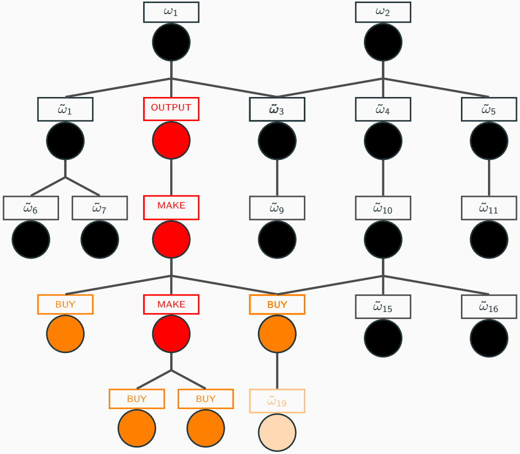
Full Model: Technology Menu

Each firm produces using **production modules** that make up a **production tree**:



Full Model: Technology Menu

The firm faces a make-or-buy decision for each non-leaf production module.



Misallocation in the Market for Inputs: Enforcement and the Organization of Production (2020, with Johannes Boehm)

Misallocation in the Market for Inputs

- How important are distortions for productivity/income?
- Our focus: Distortions in use of intermediate inputs
  - Role of courts & contract enforcement
- Manufacturing Plants in India
  - In states with worse enforcement... input bundles are systematically different
- Quantitative structural model:
  - Imperfect enforcement may distort technology & organization choice

  ⇒ Might have wrong producers doing wrong tasks
  - But firms may overcome hold-up problems with some suppliers through informal means

  ⇒ Distortions may not show up as a wedge

Data

- Court Quality: Average age of pending cases Correlation with GDP/capita
  - Calculated from microdata of pending high court cases
  - Best states: 1 year, worst states: 4.5 years
- Standardized vs. Relationship-specific (Rauch)
  - Standardized ≈ sold on an organized exchange, ref. price in trade pub.
  - Relationship-specific ≈ everything else
  - Standardized: 30.1% of input products, 50.0% of spending on intermediates
- We exclude energy, services (treat those as primary inputs)
- For reduced-form evidence, use single-product plants

Slower courts + Industry depends on Rel.spec. Inputs ⇒ Lower Materials Cost Share

	Dependent variable: Materials Expenditure in Total Cost					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg Age Of Civil Cases * Rel. Spec.	-0.0167**	-0.0155*	-0.0165*			
	(0.0046)	(0.0066)	(0.0069)			
LogGDPC * Rel. Spec.		-0.00159	-0.0130			
		(0.012)	(0.015)			
Rel. Spec. × State Controls			Yes			Yes
5-digit Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.480	0.482	0.484			
Observations	208527	199544	196748			

Standard errors in parentheses, clustered at the state × industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

IV: Main determinant of backlogs is court age (backlogs accumulate over time)

Slower courts + Industry depends on Rel.spec. Inputs ⇒ Lower Materials Cost Share

	Dependent variable: Materials Expenditure in Total Cost					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg Age Of Civil Cases * Rel. Spec.	-0.0167**	-0.0155*	-0.0165*	-0.0156+	-0.0206*	-0.0237*
	(0.0046)	(0.0066)	(0.0069)	(0.0085)	(0.0098)	(0.0094)
LogGDPC * Rel. Spec.		-0.00159	-0.0130		-0.00836	-0.0230
		(0.012)	(0.015)		(0.016)	(0.018)
Rel. Spec. × State Controls			Yes			Yes
5-digit Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.480	0.482	0.484	0.480	0.482	0.484
Observations	208527	199544	196748	208527	199544	196748

Standard errors in parentheses, clustered at the state × industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

- Moving from avg age of 1 year to 4 years: ⇒ M-share ↓ $4.7 - 6.2pp$ more in industries that rely on relationship goods than in industries that rely on standardized inputs

Measurement Error | State characteristics controls | Industry characteristics controls | Time Variation

Slow courts ⇒ tilt input mix towards homogeneous inputs

	Dependent variable: $X_j^R/(X_j^R + X_j^H)$					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg age of Civil HC cases	-0.00547*	-0.00621**	-0.00530*	-0.0144**	-0.0146**	-0.0167**
	(0.0022)	(0.0023)	(0.0024)	(0.0044)	(0.0044)	(0.0045)
Log district GDP/capita		-0.00389	-0.00384		-0.00912^{+}	-0.00980^{+}
		(0.0045)	(0.0046)		(0.0051)	(0.0051)
State Controls			Yes			Yes
5-digit Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.441	0.446	0.449	0.441	0.446	0.449
Observations	225590	204031	199339	225590	204031	199339

Standard errors in parentheses, clustered at the state × industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Full set of controls Time Variation

Courts slow + Industry depends on Rel.spec. Inputs ⇒ Plants have longer vertical span of production

(⇔ inputs are further away from outputs)

	Dependent variable: Avg Vertical Distance of Inputs from Output					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg Age Of Civil Cases * Rel. Spec.	0.0195^{+}	0.0341*	0.0320*	0.0292	0.0414^{+}	0.0437*
	(0.011)	(0.014)	(0.014)	(0.019)	(0.022)	(0.021)
LogGDPC * Rel. Spec.		0.0517^{+}	0.0309		0.0613^{+}	0.0471
		(0.029)	(0.034)		(0.037)	(0.040)
Rel. Spec. × State Controls			Yes			Yes
5-digit Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.443	0.451	0.453	0.443	0.451	0.453
Observations	163334	156191	154021	163334	156191	154021

Standard errors in parentheses, clustered at the state × industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Definition | State characteristics controls | Industry characteristics controls | Time Variation

Model: How Costly are Distortions?

- Weak contract enforcement like tax on certain inputs
  - Main identifying assumption: slow courts do not distort use of homog. inputs
- But many ways to avoid problem...
  - Informal enforcement, relatives
  - Long term relationship
  - Switch to different mode of production
  - ⇒ ...so distortion might not show up as a wedge
- Our approach: Model these choices
  - Multiple ways of producing using different suppliers
  - Distortions differ across suppliers
  - Use structure to back out distortions from observed input use
- Things we don't want to attribute to misallocation
  - Heterogeneity in production technology across plants
  - Heterogeneity across locations in
    - Preferences over goods
    - Prevalence of various industries
  - Measurement error

Model

- Many industries indexed by $\omega \in \Omega$
  - Differ by suitability for consumption vs. intermediate use
  - Rubber useful as input for tires, not textiles

- Mass of measure J_ω of firms (varieties) in industry ω

- Household has nested CES preferences

$$U = \left[\sum_{\omega} \upsilon_{\omega}^{\frac{1}{\eta}} U_{\omega}^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}} \qquad U_\omega = \left[\int_0^{J_\omega} u_{\omega j}^{\frac{\varepsilon_\omega - 1}{\varepsilon_\omega}} dj \right]^{\frac{\varepsilon_\omega}{\varepsilon_\omega - 1}}$$

Production

Firms can use different production functions ("recipes") to produce output ω:

Recipe $\rho \in \varrho(\omega)$: production function $G_{\omega\rho}(\cdot)$

- uses labor, set of intermediate inputs $\hat{\Omega}^{\rho} = \{\hat{\omega}_1, ..., \hat{\omega}_n\}$
- $G_{\omega\rho}(\cdot)$ is CRS, inputs are complements

Production

Firms can use different production functions ("recipes") to produce output ω:

Recipe $\rho \in \varrho(\omega)$: production function $G_{\omega\rho}(\cdot)$

- uses labor, set of intermediate inputs $\hat{\Omega}^{\rho} = \{\hat{\omega}_1, ..., \hat{\omega}_n\}$
- $G_{\omega\rho}(\cdot)$ is CRS, inputs are complements

Techniques: sets of productivity and supplier draws, specific to a recipe ρ. Each of them contains

- a set of potential suppliers $S_{\hat{\omega}}(\phi)$
- for each supplier:
    - an input-augmenting productivity draw: common component $b_{\hat{\omega}}(\phi)$, supplier-specific component z_s
    - a distortion t_x (see next slide)

$$y_b = G_{\omega\rho}\left(b_l l, b_{\hat{\omega}_1} z_{s_1} x_{\hat{\omega}_1}, ..., b_{\hat{\omega}_n} z_{s_n} x_{\hat{\omega}_n}\right)$$

Distortions

- If input $\hat{\omega}$ is relationship-specific: distortion $t_x \in [1, \infty)$, CDF $T(t_x)$
- If input $\hat{\omega}$ is homogeneous: no distortion

- **Weak Enforcement**:
  - Equivalent to tax (paid with labor) that is thrown in ocean Why?
  - One Microfoundation Details
    - Goods can be customized, but holdup problem
    - Court quality determines size of loss before contract is enforced
  - Interpretation: $t_x = \min \left\{ t_x^{formal}, t_x^{informal} \right\}$

- Labor wedge: t_l, common to all firms
  - Workers can steal, but stealing effort is wasteful

Functional Form Assumptions

- # suppliers for input $\hat{\omega}$ with match specific productivity $> z$ is Poisson with mean

$$z^{-\zeta_{\hat{\omega}}}, \qquad \zeta_{\hat{\omega}} \in \{\zeta_R, \zeta_H\}$$

- Among those of type ω, # techniques for recipe ρ with each productivity better than $\{b_l, b_{\hat{\omega}_1}, ..., b_{\hat{\omega}_n}\}$ is $\sim$ Poisson with mean

$$B_{\omega\rho} b_l^{-\beta_l^\rho} b_{\hat{\omega}_1}^{-\beta_{\hat{\omega}_1}^\rho} ... b_{\hat{\omega}_n}^{-\beta_{\hat{\omega}_n}^\rho}, \qquad \beta_l^\rho + \beta_{\hat{\omega}_1}^\rho + ... + \beta_{\hat{\omega}_n}^\rho = \gamma$$

- Define normalized tail exponents

$$\alpha_L^\rho \equiv \frac{\beta_l^\rho}{\gamma}, \qquad \alpha_{\hat{\omega}_i}^\rho \equiv \frac{\beta_{\hat{\omega}_i}^\rho}{\gamma} \qquad \Rightarrow \qquad \alpha_L^\rho + \sum_i \alpha_{\hat{\omega}_i}^\rho = 1$$

$$\alpha_R^\rho \equiv \sum_{\hat{\omega} \in \hat{\Omega}_R^\rho} \alpha_{\hat{\omega}}^\rho \qquad \alpha_H^\rho \equiv \sum_{\hat{\omega} \in \hat{\Omega}_H^\rho} \alpha_{\hat{\omega}}^\rho \qquad \Rightarrow \qquad \alpha_L^\rho + \alpha_H^\rho + \alpha_R^\rho = 1$$

Aggregation

Proposition: Among firms that produce ω, the fraction of firms with unit cost $\geq c$ is

$$e^{-(c/C_\omega)^\gamma}$$

where

$$\begin{aligned} C_\omega &= \left\{ \sum_{\rho \in \varrho(\omega)} \kappa_{\omega\rho} B_{\omega\rho} \left((t_x^*)^{\alpha_R^\rho} (t_l)^{\alpha_L^\rho} \prod_{\hat{\omega} \in \hat{\Omega}^\rho} C_{\hat{\omega}}^{\alpha_{\hat{\omega}}^\rho} \right)^{-\gamma} \right\}^{-1/\gamma} \\ t_x^* &= \left\{ \int t_x^{-\zeta_R} dT(t_x) \right\}^{-1/\zeta_R} \\ \kappa_{\omega\rho} &= \text{constant} \end{aligned}$$

Proposition: Among firms in ω using recipe ρ, average and aggregate exp. shares on:

$$\text{Labor: } \alpha_L^\rho + \left(1 - \frac{1}{\bar{t}_x}\right) \alpha_R^\rho, \quad \hat{\omega} \in \hat{\Omega}_\rho^R : \frac{\alpha_{\hat{\omega}}^\rho}{\bar{t}_x}, \quad \hat{\omega} \in \hat{\Omega}_\rho^H : \alpha_{\hat{\omega}}^\rho,$$

where $\bar{t}_x \equiv \left[\int t_x^{-1} d\tilde{T}(t_x) \right]^{-1}$

Counterfactual?

Question:

- Change wedge distribution from T to T', what is impact on agg. output?

From data, need two sets of shares

- HH_ω: share of the household's spending on good ω
- Among those of type ω, let $R_{\omega\rho}$ be the share of total revenue of those that use recipe ρ.

$$\frac{U'}{U} = \left(\sum_{\omega} HH_\omega \left(\frac{C'_\omega}{C_\omega} \right)^{\eta-1} \right)^{\frac{1}{\eta-1}}$$

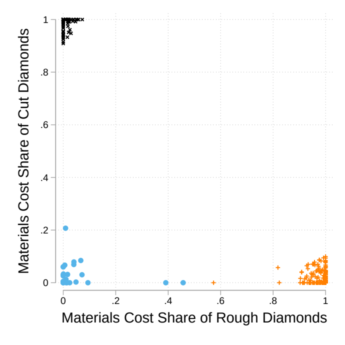
$$\left(\frac{C'_\omega}{C_\omega} \right)^{-\gamma} = \sum_{\rho \in \varrho(\omega)} R_{\omega\rho} \left[\left(\frac{t_x^{*\prime}}{t_x^*} \right)^{\alpha_R^\rho} \prod_{\hat{\omega} \in \Omega^\rho} \left(\frac{C'_{\hat{\omega}}}{C_{\hat{\omega}}} \right)^{\alpha_{\hat{\omega}}^\rho} \right]^{-\gamma}$$

Identification

- *Same* across states: Recipe technology
  - Production function (G_ρ)
  - Shape of technology draws ($\beta_l^\rho, \{\beta_{\hat{\omega}}^\rho\}$)
  - Shape of match-specific productivity draws, (ζ)
- *Different* across states:
  - Measure of producers of each type (J_ω)
  - Household tastes (υ_ω)
  - Comparative/absolute advantage: (recipe productivity, $B_{\omega\rho}$)
  - Distribution of wedges (T)
- Main identifying assump.: Slow courts do not distort use of homog. inputs
- Other Assumptions:
  - No trade across states
  - L is labor equipped with other primary inputs (capital, energy, services)

Identifying Recipes in the Data

Figure 2: Example: Polished Diamonds



- Cluster analysis to determine recipes within each product: Ward's method (requires # clusters) Ward's Method
- Prediction strength method (Tibshirani-Walther 2005) to find # clusters

- Robustness to different degree of "fineness" of recipes Fineness
- Monte Carlo simulations to get small-sample and large-sample properties of combined procedure MC small MC large

$\Rightarrow$ 26,776 recipes (avg. 5.9 recipes per product)

Moments for GMM

Proposition: Let s_{Rj}, s_{Hj}, s_{Lj} be firm j's revenue shares.

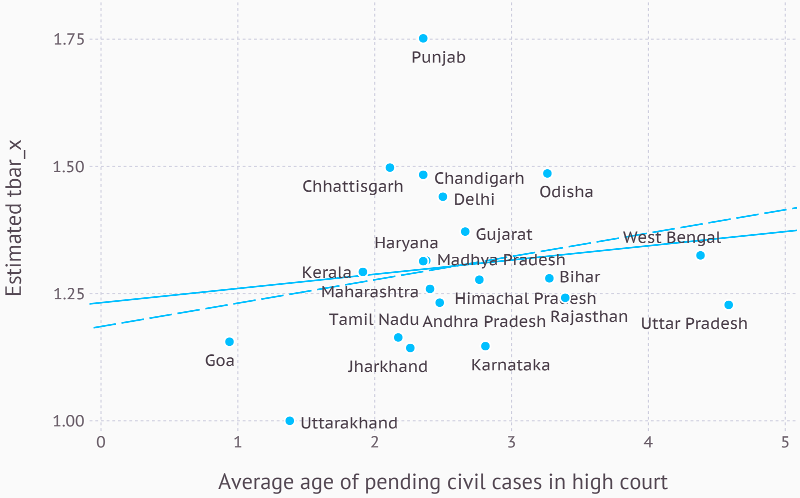
- The first moments of revenue shares among firms that use recipe ρ satisfy:

$$\mathbb{E}\left[\bar{t}_x^d \frac{s_{Rj}}{\alpha_R^\rho} - \frac{s_{Hj}}{\alpha_H^\rho}\right] = 0$$
$$\mathbb{E}\left[\frac{s_{Lj} + s_{Rj}}{\alpha_L^\rho + \alpha_R^\rho} - \frac{s_{Hj}}{\alpha_H^\rho}\right] = 0$$

$\Rightarrow$ Identification of wedges

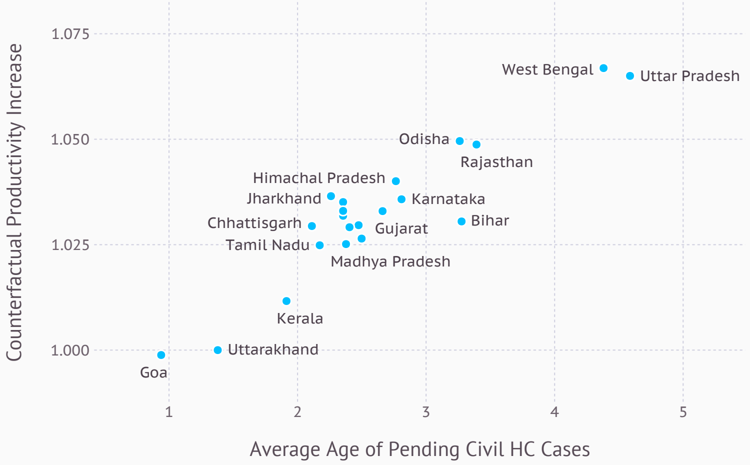
- from within-recipes variation instead of within-industries
- from first moments only

Intermediate input wedges are correlated with court quality



Gains From Improving Courts

Counterfactual sets court quality to 1. Impose $\gamma = 1$ (or first-order approx).



The Network Origins of Firm Dynamics: Contracting Frictions and Dynamism with Long-Term Relationships (with Johannes Boehm, Ruairidh South, and Mazhar Waseem)

- Another response to contracting frictions: engage in long term relationships
    - Static benefit: better performance
    - Dynamic cost: Might not switch to new good supplier
- Evidence from Pakistan value added tax data
    - Contracting frictions (slow courts) increase average relationship length...
    - ...but reduce dynamism
    - Hard for productive young firms to grow

Conclusion

- Huge amount of **heterogeneity** in intermediate input use, even within narrow industries
    - Differences in organizational form and types of inputs used
    - Systematic changes with development
    - Systematic changes with institutions
    - Systematic changes with...
- What are the most important policy levers?
- What are the implications for growth?
- Lots to learn...lots of rich data (and richness in the data) to learn from

Appendix

Increased vertical specialization is positively correlated with state growth

Within plant, over time:

	Dependent variable: Vertical Span		
	(1)	(2)	(3)
Log GDP/capita$_{st}$	-0.0716*	-0.0601*	-0.0551*
	(0.028)	(0.026)	(0.026)
Year FE	Yes	Yes	
Plant FE	Yes	Yes	
5-digit Industry FE		Yes	
5-digit Industry × Year FE			Yes
Plant × 5-digit Industry FE			Yes
R^2	0.592	0.656	0.808
Observations	270003	269399	163668

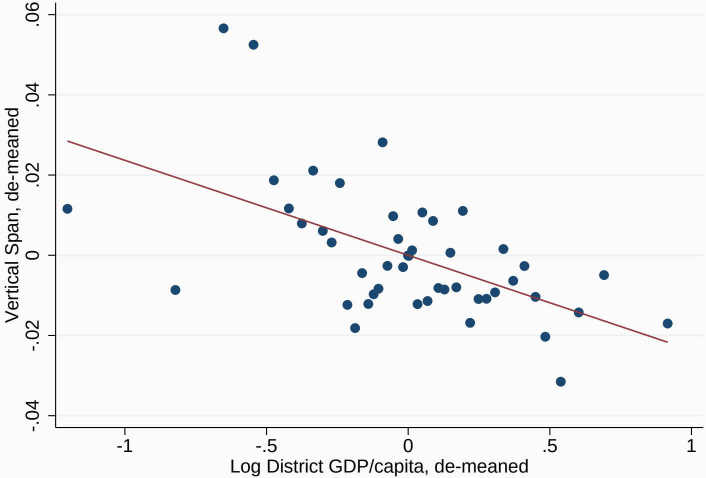
Standard errors in parentheses, clustered at the state × 5-dgt industry level. SP plants only.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Standard errors in parentheses, clustered at the state × 5-dgt industry level. SP plants only.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back Bin Scatter

Fact 1: In richer districts, plants are more specialized (Single Plants only)

Within industry $\times$ year:



Binscatter with n=50. Y and X variable de-meaned by industry x year. SP plants only.

Back

Fact 2: Vertical specialization positively correlated with growth (Single Plant Only)

Within plant, over time:

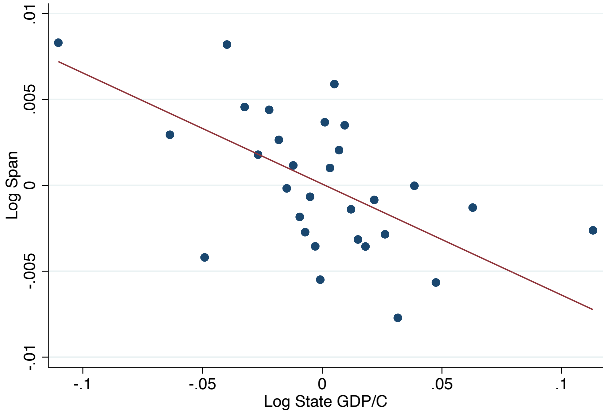
	Dependent variable: Vertical Span		
	(1)	(2)	(3)
Log GDP/capita$_{st}$	-0.0552	-0.0647	-0.0741^{+}
	(0.048)	(0.045)	(0.043)
Year FE	Yes	Yes	Yes
Plant FE	Yes	Yes	
5-digit Industry FE		Yes	
Plant × 5-digit Industry FE			Yes
R^2	0.644	0.720	0.780
Observations	95727	94754	61073

Standard errors in parentheses, clustered at the state × 5-dgt industry level. SP plants only.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

Macro Fact in Changes: Increased vertical specialization is positively correlated with state growth

Within plant×industry, year:



Binscatter with n=30. SP plants only. Variables de-meaned by plant-industry and year.

Sales growth is correlated with increased vertical specialization

	Dependent variable: Δ log Sales			
	(1)	(2)	(3)	(4)
Δ Vertical Span	-0.0655**	-0.0445**	-0.0284*	-0.0259*
	(0.0082)	(0.0087)	(0.013)	(0.011)
Year FE	Yes			
Product $\times$ Year FE		Yes	Yes	Yes
Plant FE			Yes	
Plant $\times$ Product FE				Yes
R^2	0.00819	0.149	0.432	0.431
Observations	120436	111244	83026	74707

Changes within plant-products

Standard errors in parentheses, clustered at the state-industry level.

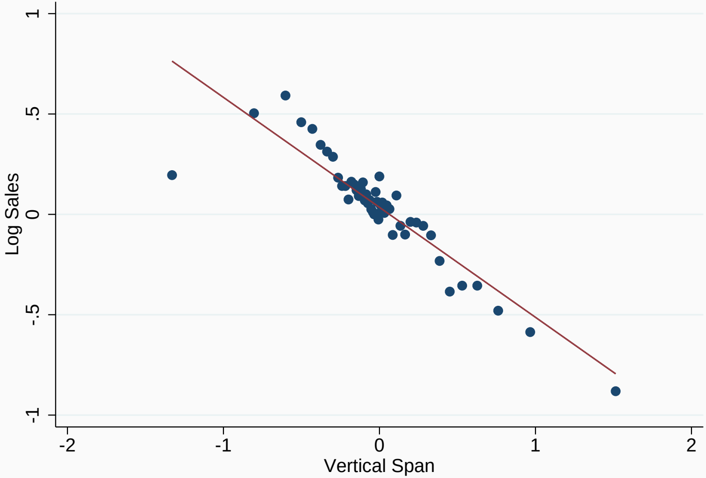
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

SP plants only.

Back

More specialized plants are larger (Single Plant only)

Plants with higher sales tend to have shorter vertical span (within industry $\times$ year)



SP plants only. Regression includes industry x year FE's.

Back

Fact 4: Sales growth is correlated with increased vertical specialization

	Dependent variable: Δ log Sales			
	(1)	(2)	(3)	(4)
Δ Vertical Span	-0.0655**	-0.0445**	-0.0284*	-0.0259*
	(0.0082)	(0.0087)	(0.013)	(0.011)
Year FE	Yes			
Product $\times$ Year FE		Yes	Yes	Yes
Plant FE			Yes	
Plant $\times$ Product FE				Yes
R^2	0.00819	0.149	0.432	0.431
Observations	120436	111244	83026	74707

Changes within plant-products

Standard errors in parentheses, clustered at the state-industry level.

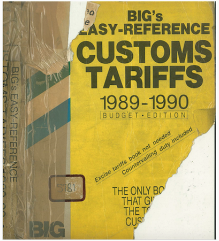
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

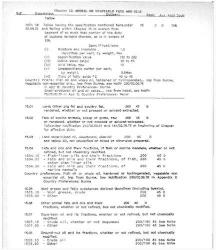
back

Details on tariff construction

We use UNCTAD tariffs, complemented by hand-digitized effective tariff rates for early years of the liberalization (1990-1996).

- Exclude agricultural tariffs (which changed in response to domestic shocks)
- Exclude mechanical & electrical machinery (HS headings 84, 85): long list of exceptions





Back

Plants with shorter span are larger: Details

	Dependent variable: Log Sales				
	(1)	(2)	(3)	(4)	(5)
Vertical Span	-0.719**	-0.670**	-0.431**	-0.432**	-0.193**
	(0.024)	(0.023)	(0.034)	(0.034)	(0.015)
Age				0.00616**	-0.00314**
				(0.0012)	(0.00068)
Log Employment					1.067**
					(0.015)
Year FE	Yes	Yes	Yes	Yes	Yes
5-digit Industry FE	Yes	Yes			
District FE		Yes			
Industry × District × Year FE			Yes	Yes	Yes
R^2	0.394	0.440	0.700	0.701	0.859
Observations	353659	295094	140610	136831	136608

Standard errors in parentheses, clustered at the 5-dgt industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

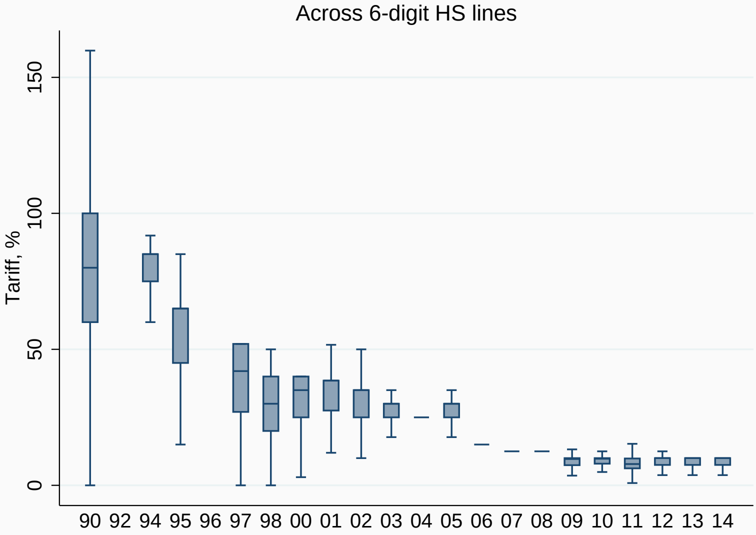
Back

Other cross-sectional covariates

	Dependent variable: Vertical Span					
	(1)	(2)	(3)	(4)	(5)	(6)
Materials Share of Cost	-0.250**			-0.119**		
	(0.018)			(0.015)		
Importer Dummy		-0.163**			-0.0143**	
		(0.0094)			(0.0055)	
Share of R-Inputs in Materials Cost			-0.260**			-0.181**
			(0.021)			(0.021)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes			
Plant x Industry FE				Yes	Yes	Yes
R^2	0.310	0.309	0.322	0.774	0.765	0.773
Observations	332356	353694	347548	173141	186641	181958

Standard errors in parentheses, clustered at the 5-dgt industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Import Tariffs, India, 1990-2014



Back

Changes since 1990: tariffs and sales

	Dep. var.: Δ_{1990}^{t} log Sales	
	(1)	(2)
$\Delta_{1990}^{t} \log(1+\tau_{\omega t}^{\text{output}})$	1.302^{+}	1.533^{+}
	(0.75)	(0.79)
$\Delta_{1990}^{t} \log(1+\bar{\tau}_{\omega t}^{\text{input}})$		-1.188
		(0.77)
Year FE	Yes	Yes
R^2	0.0852	0.0903
Observations	2376	2376

Standard errors in parentheses, clustered at the state $\times$ industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

Changes since 1990: vertical span and demand

	Dependent variable: Δ^t_{1990} Vertical Span		
	(1)	(2)	(3)
Δ^t_{1990} log Sales	-0.147$^+$	-0.166$^+$	-0.237$^+$
	(0.084)	(0.086)	(0.12)
$\Delta^t_{1990} \log(1+\bar{\tau}^{\text{input}}_{it})$		0.194	1.421$^+$
		(0.24)	(0.77)
$\Delta \sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})(\text{distance}_{\omega i} - \overline{\text{span}}_j)$			-0.747
			(0.75)
$\Delta \sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})\overline{\text{span}}_j$			-1.031$^+$
			(0.62)
Year FE	Yes	Yes	Yes
R^2	-0.194	-0.255	-0.498
Observations	2179	2179	2128

Standard errors in parentheses, clustered at the state $\times$ industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Δ^t_{1990} log sales is instrumented by the change in the log output tariff since 1990.

Back

Demand shocks affect vertical specialization

	Dependent variable: Δ Vertical Span					
	(1)	(2)	(3)	(4)	(5)	(6)
Δ log Sales	-0.0158**	-0.0160**	-0.0165**	-0.0301*	-0.0457+	-0.0652
	(0.0020)	(0.0020)	(0.0024)	(0.013)	(0.025)	(0.052)
$\Delta \log(1+\bar{\tau}^{\text{input}}_{j\omega t})$		-0.0173	0.00538		-0.0354	-0.0188
		(0.020)	(0.047)		(0.044)	(0.051)
$\Delta \sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})(\text{distance}_{\omega i} - \overline{\text{span}}_j)$			-0.00198			-0.0731
			(0.087)			(0.14)
$\Delta \sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})\overline{\text{span}}_j$			-0.00455			-0.0478
			(0.041)			(0.096)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.00207	0.00208	0.00229	0.000325	-0.00220	-0.00774
Observations	123666	123666	94795	90115	90115	89301

Standard errors in parentheses, clustered at the state-industry level. $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Columns (3), (4) instrument Δ log sales by the change in the log output tariff.

Back

Vertical span and demand, generated instruments à la Wooldridge (2002, Ch 6)

IV for log sales: fitted values of a Poisson regression of sales on exogenous variables.

	Dependent variable: Vertical Span		
	(1)	(2)	(3)
Log Sales	-0.0909**	-0.0841**	-0.0933**
	(0.021)	(0.020)	(0.021)
$\log(1+\bar{\tau}^{\text{input}}_{j\omega t})$		-0.0931*	-0.0265
		(0.039)	(0.057)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})\overline{\text{span}}_j$			-0.133*
			(0.064)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})(\text{distance}_{\omega i}-\overline{\text{span}}_j)$			-0.203+
			(0.11)
Year FE	Yes	Yes	Yes
Plant × Product FE	Yes	Yes	Yes
Weak-Id robust 1st stage F	221.3	230.4	215.6
Estimator	G-IV	G-IV	G-IV
R^2	-0.0212	-0.0169	-0.0218
Observations	138204	138204	13705[illegible]

Standard errors in parentheses, clustered at the state-industry level.

Back

Selection 1: SP and MP

	Dependent variable: Vertical Span					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Sales	-0.0116**	-0.0116**	9.814	-0.213	-0.110*	-0.103
	(0.0016)	(0.0020)	(196.5)	(0.38)	(0.054)	(0.066)
$\log(1+\bar{\tau}^{\text{input}}_{j\omega t})$		-0.0205		-0.0378		-0.0677
		(0.042)		(0.047)		(0.072)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})\overline{\text{span}}_j$		-0.0109		-0.0922		-0.169^{+}
		(0.044)		(0.15)		(0.092)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})(\text{distance}_{\omega i}-\overline{\text{span}}_j)$		0.138*		0.0155		-0.133
		(0.067)		(0.19)		(0.12)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant × Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV 1	IV 1	IV 2	IV 2
R^2	0.764	0.732	-280.4	-0.117	-0.0263	-0.0186
First stage w. id. robust F			0.00261	1.162	36.16	43.04
Observations	338887	245188	232632	231254	260217	186983

Standard errors in parentheses, clustered at the state-industry level. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

Selection 2: Always SP vs at some point before MP

	Dependent variable: Vertical Span					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Sales	-0.0184**	-0.0188**	-0.542+	-0.278+	-0.164*	-0.269**
	(0.0020)	(0.0024)	(0.31)	(0.15)	(0.074)	(0.095)
Log Sales × MP before	-0.00147**	-0.00187**	0.00784+	0.00404+	0.0000738	0.000932
	(0.00041)	(0.00047)	(0.0048)	(0.0024)	(0.00088)	(0.0013)
$\log(1+\bar{\tau}^{\text{input}}_{j\omega t})$		-0.0191		0.0254		0.0545
		(0.050)		(0.068)		(0.100)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})\overline{\text{span}}_j$		-0.0918+		-0.227*		-0.370**
		(0.055)		(0.10)		(0.13)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})(\text{distance}_{\omega i}-\overline{\text{span}}_j)$		-0.117		-0.361*		-0.499*
		(0.095)		(0.17)		(0.21)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant × Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV 1	IV 1	IV 2	IV 2
R^2	0.765	0.731	-1.174	-0.283	-0.0817	-0.227
First stage w. id. robust F			2.592	4.762	13.16	7.909
Observations	186628	145165	138204	137059	143428	110578

Back

Increasing or Decreasing RTS? Domestic Sales on Export Demand, 2010-14

	Dependent variable: Log Domestic Sales					
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{ExDemand}_{\omega t}$	0.0214*	-0.000381	0.00264	0.0208**	0.0160	0.0222
	(0.0092)	(0.017)	(0.018)	(0.0066)	(0.013)	(0.014)
$\log(1+\tau_{it}^{\text{output}})$		-0.765	-0.648		-0.609	-0.562
		(0.52)	(0.52)		(0.41)	(0.41)
$\log(1+\bar{\tau}_{it}^{\text{input}})$		-0.397	-0.614		-1.343^{+}	-1.771^{+}
		(1.19)	(1.55)		(0.74)	(0.95)
$\sum_i \alpha_i \log(1+\bar{\tau}_{it}^{\text{input}})\overline{\text{span}}_j$			0.433			0.400
			(0.83)			(0.44)
$\sum_i \alpha_i \log(1+\bar{\tau}_{it}^{\text{input}})(\text{distance}_{\omega i} - \overline{\text{span}}_j)$			0.968			-0.0418
			(1.59)			(0.86)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant × Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	SP	SP	SP	All	All	All
R^2	0.953	0.956	0.956	0.957	0.959	0.959
Observations	64375	40279	38335	109802	65959	63433

Changes within plant-products
Standard errors in parentheses, clustered at the state-industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

- Consistent with Alfaro-Urena, Manelici and Vasquez (2022) and Albornoz, Brambilla and Ornelas (2021)

Demand $\nearrow$ $\Rightarrow$ firms reduce the actual number of inputs

	Dependent variable: # Inputs					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Sales	0.0359**	0.0374**	-1.136*	-0.468+	-0.383+	-0.542*
	(0.0030)	(0.0035)	(0.54)	(0.26)	(0.20)	(0.25)
$\log(1+\bar{\tau}^{\text{input}}_{j\omega t})$		-0.441**		-0.366*		-0.207
		(0.16)		(0.14)		(0.19)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})\overline{\text{span}}_j$		0.125		-0.126		-0.337
		(0.10)		(0.14)		(0.21)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})(\text{distance}_{\omega i}-\overline{\text{span}}_j)$		0.162		-0.202		-0.722*
		(0.10)		(0.25)		(0.30)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant $\times$ Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV 1	IV 1	IV 2	IV 2
R^2	0.856	0.847	-4.788	-0.874	-0.574	-1.059
First stage w. id. robust F			5.566	10.91	26.53	15.70
Observations	186628	145165	138204	137059	143428	110578

Changes within plant-products

Standard errors in parentheses, clustered at the state-industry level.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

Firms are more likely to adopt inputs when faced with a cost decrease

	Dependent variable: Input Used Dummy $\mathbf{1}(X_{j\hat{\omega}t} > 0)$	
	(1)	(2)
$\log(1 + \tau_{it})$	-0.0506**	-0.0373**
	(0.0067)	(0.0071)
Year FE	Yes	Yes
Plant $\times$ Input FE	Yes	Yes
Plant $\times$ Product FE		Yes
R^2	0.337	0.361
Observations	2460831	2454899

Standard errors in parentheses.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Supply and demand shifters determine entry

	Dependent variable: log Producers $\lvert J\rvert_{d\omega t}$	
	(1)	(2)
$\log(1+\bar{\tau}_{it}^{\text{input}})$	-0.108**	-0.0496**
	(0.025)	(0.015)
$\log(1+\tau_{it}^{\text{output}})$	0.186**	0.251**
	(0.021)	(0.013)
Year FE	Yes	
State FE	Yes	
Industry FE	Yes	
State $\times$ Year FE		Yes
State $\times$ Industry FE		Yes
R^2	0.481	0.844
Observations	548180	537013

Standard errors in parentheses.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

The left-hand side is the log number of producers of a good ω at time t in state d.

Back

Firms reduce the effective number of inputs when demand $\nearrow$

	Dependent variable: Inverse Input HHI					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Sales	0.0101	0.00930	-1.804^{+}	-1.100^{+}	-1.705^{+}	-2.135^{+}
	(0.0072)	(0.0082)	(1.08)	(0.63)	(0.99)	(1.18)
$\log(1+\bar{\tau}^{\text{input}}_{j\omega t})$		-0.745^{+}		-0.574^{+}		-0.354
		(0.42)		(0.30)		(0.50)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})\overline{\text{span}}_j$		0.428		-0.0487		-1.362^{+}
		(0.34)		(0.29)		(0.78)
$\sum_i \alpha_i \log(1+\bar{\tau}^{\text{input}}_{it})(\text{distance}_{\omega i} - \overline{\text{span}}_j)$		0.630		-0.138		-1.906^{+}
		(0.39)		(0.55)		(1.16)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant $\times$ Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV 1	IV 1	IV 2	IV 2
R^2	0.807	0.799	-3.137	-1.132	-2.464	-3.560
First stage w. id. robust F			4.732	10.81	28.71	15.73
Observations	192795	145189	142265	137084	147328	110594

Changes within plant-products
Standard errors in parentheses, clustered at the state-industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

Sample of 1990 plants: upstream industry size and sales

	Dependent variable: log Sales					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg. log #Producers in Upstream Ind.	0.0655**	0.0560**	0.0551**	0.0201	0.119**	0.115**
	(0.013)	(0.018)	(0.018)	(0.043)	(0.044)	(0.044)
$\log(1+\bar{\tau}_{j\omega t}^{\text{input}})$			0.540*			0.519*
			(0.26)			(0.26)
Year FE	Yes			Yes		
Industry × Year FE		Yes	Yes		Yes	Yes
Plant × Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.916	0.943	0.943	0.00262	-0.000638	0.000690
Observations	13683	9768	9757	13683	9768	9757

Standard errors in parentheses, clustered at the industry-year level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Sample: all SP plants observed in 1990

(except (3) and (6) which further condition on $t \leq 2000$)

Back

Upstream industry size and sales

	Dependent variable: log Sales					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg. log Sales in Upstream Ind.	0.00367**	0.00251**	0.00251**	0.00642*	0.00930**	0.00936**
	(0.00034)	(0.00038)	(0.00038)	(0.0029)	(0.0026)	(0.0026)
$\log(1 + \bar{\tau}_{j\omega t}^{\text{input}})$			0.0241			0.0193
			(0.085)			(0.086)
Year FE	Yes			Yes		
Industry × Year FE		Yes	Yes		Yes	Yes
Plant × Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.942	0.952	0.952	0.000572	-0.00304	-0.00311
Observations	215805	199039	198727	215805	199039	198727

Standard errors in parentheses, clustered at the industry-year level.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

Fact 4: Sales growth is correlated with increased vertical specialization

	Dependent variable: Δ log Sales			
	(1)	(2)	(3)	(4)
Δ Vertical Span	-0.0693**	-0.0668**	-0.0577**	-0.0546**
	(0.0085)	(0.0085)	(0.011)	(0.011)
Δ R-Share in Materials	-0.0242*	-0.0247*	-0.0270+	-0.0346*
	(0.012)	(0.012)	(0.015)	(0.014)
Δ Vertical Span × Δ R-Share in Materials	-0.0359*	-0.0408*	-0.0549*	-0.0544*
	(0.016)	(0.016)	(0.025)	(0.023)
Constant	0.194**	0.194**	0.181**	0.171**
	(0.0046)	(0.0025)	(0.0015)	(0.00030)
Year FE	Yes	Yes	Yes	Yes
Product FE		Yes		
Plant FE			Yes	
Plant × Product FE				Yes
R^2	0.00825	0.0409	0.305	0.314
Observations	116199	115643	89440	80377

Unit Costs and Tariff changes

	Dependent variable: Δ^t_{1990} log Unit Cost	
	(1)	(2)
$\Delta^t_{1990}\log(1+\tau^{\text{output}}_{it})$	-0.789**	-0.949**
	(0.10)	(0.17)
$\Delta^t_{1990}\log(1+\bar{\tau}^{\text{input}}_{j\omega t})$		0.226
		(0.17)
Year FE	Yes	Yes
R^2	0.0566	0.0583
Observations	920	916

Standard errors in parentheses, clustered at the state $\times$ industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

Generalizations

Extreme value math extends to any finite "production tree"

- Any (finite) number of inputs in each stage
- Any (finite) depth of the tree

Conditional on search effort choices, the distributions of input unit costs are EV

Search choices depend on Hicks-neutral productivity and upstream cost distributions

$\Rightarrow$ solve search problem recursively starting with most upstream (leaf) nodes

Full Model:

- (Imperfectly) elastic entry into industries ω on a large "production tree"
- Positive profits from sales to households, marginal cost pricing to firms further downstream
- Firms born with Hicks-neutral q. Increasing returns to scale through input search.
- Potentially network economies through arrival rate of draws also depending on upstream sector characteristics.

▸ Conclusion

Discrete Choice Math

- Lowest cost way of acquiring good $\omega - 1$

$$\min\left\{\min_{s\in S_1}\frac{p_s}{z_s}\ ,\ \frac{1}{b_2}\min_{s\in S_2} w^{1-\alpha}\left(\frac{p_s}{z_s}\right)^{\alpha}\right\}$$

- Arrival of suppliers with $z_s > z$ is Poisson with arrival rate $\propto z^{-\zeta}$

$$\min_{s\in S_1}\frac{p_s}{z_s} \sim Weibull(scale_1, \zeta) \tag{1}$$

$$\min_{s\in S_2} w^{1-\alpha}\left(\frac{p_s}{z_s}\right)^{\alpha} \sim Weibull\left(scale_2, \frac{\zeta}{\alpha}\right)$$

(3)

Back

Discrete Choice Math

- Lowest cost way of acquiring good $\omega - 1$

$$\min \left\{ \min_{s \in S_1} \frac{p_s}{z_s} \,,\ \frac{1}{b_2} \min_{s \in S_2} w^{1-\alpha} \left(\frac{p_s}{z_s} \right)^{\alpha} \right\}$$

- Arrival of suppliers with $z_s > z$ is Poisson with arrival rate $\propto z^{-\zeta}$

$$\min_{s \in S_1} \frac{p_s}{z_s} \sim Weibull(scale_1, \zeta) \tag{1}$$

$$\min_{s \in S_2} w^{1-\alpha} \left(\frac{p_s}{z_s} \right)^{\alpha} \sim Weibull\left(scale_2, \frac{\zeta}{\alpha} \right) \tag{2}$$

$$\frac{1}{b_2} \min_{s \in S_2} w^{1-\alpha} \left(\frac{p_s}{z_s} \right)^{\alpha} \sim Weibull\left(scale_3, \zeta \right) \tag{3}$$

- Follows from:

$$Z \sim \text{standard exponential},\ Y \sim \alpha\text{-Stable} \qquad \Rightarrow \qquad (Z/Y)^{\alpha} \sim Z$$

Back

Nested CES Example

Imagine the production function was a Nested CES:

$$y_j = q\left\{(A_1 h_1 x_1)^{\frac{\eta-1}{\eta}} + \left[(A_0 l)^{\frac{\phi-1}{\phi}} + (A_2 h_2 x_2)^{\frac{\phi-1}{\phi}}\right]^{\frac{\phi}{\phi-1}\frac{\eta-1}{\eta}}\right\}^{\frac{\eta}{\eta-1}}$$

Proposition

If $\gamma \geq \eta - 2$ *and* $\gamma \geq \phi - 2$, *then*

$$\frac{d\ln h_1}{d\ln q} > \frac{d\ln h_2}{d\ln q} \qquad \textit{iff} \qquad \eta > \phi$$

Our setting is a special case with $\eta \to \infty$ and $\phi \to 1$.

Back

Where does the nonhomotheticity come from?

- Imagine a production function where search effort is factor-augmenting.

$$\max_{h_1,h_2} \delta g\left\{C\left(w, \frac{p_1}{h_1}, \frac{p_2}{h_2}\right)\right\} - \frac{h_1^{1+\gamma}}{1+\gamma} - \frac{h_2^{1+\gamma}}{1+\gamma}$$

- *Levels* of optimal search effort are determined by cost shares:

$$0 = -\delta g' \frac{p_i}{h_i^2} C_i\left(w, \frac{p_1}{h_1}, \frac{p_2}{h_2}\right) - h_i^{\gamma}$$

- Relative elasticity of h_1 vs h_2 is therefore determined by relative *elasticity* of cost shares ... and these are encoded in the Morishima elasticities of substitution σ_{21}, σ_{12}
- If γ sufficiently large, $d \log h_1 / d \log q > d \log h_2 / d \log q$ iff $\sigma_{21} > \sigma_{12}$.
- In particular that's satisfied when there is perfect substitutability between a nested and non-nested production function:

$$y_j = \begin{cases} q_j f(l_{j0}, x_{j1}) & \text{or} \\ q_j f(l_{j0}, g(l_{j1}, x_{j2})) \end{cases}$$

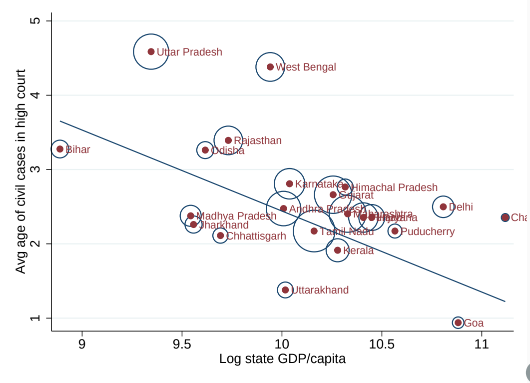
(assuming g is imperfectly substitutable...)

Back

Appendix

Slow Courts

- Contract disputes between buyers and sellers
- District courts can de-facto be bypassed, cases would be filed in high courts
- Court quality measure: average age of pending civil cases in high court



Back

Measurement: Quality of Closest Court, OLS

	Dependent variable: Materials Expenditure in Total Cost			
	(1)	(2)	(3)	(4)
Avg age of Civil HC cases	0.00991**			
	(0.0035)			
Avg Age Of Civil Cases * Rel. Spec.	-0.0151**	-0.0155*		
	(0.0055)	(0.0066)		
Avg age of Civil HC cases (adj.)			0.0172**	
			(0.0037)	
Adjusted Court Quality * Rel. Spec.			-0.0328**	-0.0282**
			(0.0064)	(0.0064)
Log district GDP/capita	0.00694+		0.00578	
	(0.0038)		(0.0038)	
LogGDPC * Rel. Spec.		-0.00159		0.00390
		(0.012)		(0.0093)
5-digit Industry FE	Yes	Yes	Yes	Yes
District FE		Yes		Yes
Estimator	OLS	OLS	OLS	OLS
R^2	0.461	0.482	0.461	0.482
Observations	201505	199544	201505	199544

Standard errors in parentheses, clustered at the state $\times$ industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

(Note: 'adjusted' court quality is the minimum avg. age in the state's own HC and a neighboring HC, if that neighboring HC has a bench that is closer than the closest of your

Measurement: Quality of Closest Court, IV

	Dependent variable: Materials Expenditure in Total Cost			
	(1)	(2)	(3)	(4)
Avg age of Civil HC cases	-0.00381			
	(0.0060)			
Avg Age Of Civil Cases * Rel. Spec.	-0.0283**	-0.0206*		
	(0.010)	(0.0098)		
Avg age of Civil HC cases (adj.)			-0.00972	
			(0.013)	
Adjusted Court Quality * Rel. Spec.			-0.0482*	-0.0373*
			(0.021)	(0.018)
Log district GDP/capita	-0.00535		-0.00616	
	(0.0039)		(0.0040)	
LogGDPC * Rel. Spec.		-0.00836		-0.000887
		(0.016)		(0.013)
5-digit Industry FE	Yes	Yes	Yes	Yes
District FE		Yes		Yes
Estimator	IV	IV	IV	IV
R^2	0.457	0.482	0.453	0.482
Observations	201505	199544	201505	199544

Standard errors in parentheses, clustered at the state $\times$ industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

Substituting with imports when courts are bad

	R-Imports in Total R		H-Imports in Total H	
	(1)	(2)	(3)	(4)
Avg age of Civil HC cases	0.0193**	0.00925**	0.0112**	0.00440**
	(0.0023)	(0.0018)	(0.0016)	(0.0013)
Log district GDP/capita		0.0224**		0.0180**
		(0.0027)		(0.0019)
Trust in other people (WVS)		0.110**		0.0564**
		(0.012)		(0.011)
Language Herfindahl		0.0162		-0.0292**
		(0.019)		(0.0093)
Caste Herfindahl		0.0584*		0.0171
		(0.028)		(0.013)
Corruption		0.0315		-0.0912**
		(0.028)		(0.022)
5-digit Industry FE	Yes	Yes	Yes	Yes
Estimator	IV	IV	IV	IV
R^2	0.227	0.251	0.180	0.197
Observations	168120	148165	168953	149623

Materials Share: state characteristics controls

	Dependent variable: Materials Expenditure in Total Cost			
	(1)	(2)	(3)	(4)
Avg Age Of Civil Cases * Rel. Spec.	-0.0167**	-0.0165*	-0.0156+	-0.0237*
	(0.0046)	(0.0069)	(0.0085)	(0.0094)
LogGDPC * Rel. Spec.		-0.0130		-0.0230
		(0.015)		(0.018)
Trust * Rel. Spec.		0.0295		0.0323
		(0.038)		(0.038)
Language HHI * Rel. Spec.		0.0601		0.0625
		(0.040)		(0.039)
Caste HHI * Rel. Spec.		0.126*		0.133*
		(0.053)		(0.053)
Corruption * Rel. Spec.		0.117		0.129
		(0.11)		(0.11)
5-digit Industry FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV	IV
R^2	0.480	0.484	0.480	0.484

Back

Composition of the Input Mix: full set of controls

	Dependent variable: $X_j^R/(X_j^R + X_j^H)$			
	(1)	(2)	(3)	(4)
Avg age of Civil HC cases	-0.00547*	-0.00530*	-0.0144**	-0.0167**
	(0.0022)	(0.0024)	(0.0044)	(0.0045)
Log district GDP/capita		-0.00384		-0.00980^{+}
		(0.0046)		(0.0051)
Trust		-0.00740		-0.00160
		(0.018)		(0.019)
Language HHI		-0.0553**		-0.0567**
		(0.021)		(0.022)
Caste HHI		-0.0428		-0.0525^{+}
		(0.028)		(0.029)
Corruption		-0.0676		-0.0844^{+}
		(0.044)		(0.045)
5-digit Industry FE	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV	IV
R^2	0.441	0.449	0.441	0.449
Observations	225590	199339	225590	199339

Back

Vertical Distance: state characteristics controls

	Dependent variable: Vertical Distance of Inputs from Output					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg age of Civil HC cases	0.00144	-0.0103		-0.00490	-0.00168	
	(0.0070)	(0.0076)		(0.011)	(0.011)	
Avg Age Of Civil Cases * Rel. Spec.	0.0230^{+}	0.0387**	0.0320*	0.0294	0.0459*	0.0437*
	(0.012)	(0.013)	(0.014)	(0.020)	(0.020)	(0.021)
Log district GDP/capita		-0.0350**			-0.0361**	
		(0.0072)			(0.0073)	
LogGDPC * Rel. Spec.		0.0328^{+}	0.0309		0.0625**	0.0471
		(0.017)	(0.034)		(0.020)	(0.040)
Trust		0.0401			0.0357	
		(0.055)			(0.056)	
Language Herfindahl		0.0559			0.0563	
		(0.054)			(0.054)	
Caste Herfindahl		0.0511			0.0541	
		(0.069)			(0.068)	
Corruption		-0.324*			-0.295^{+}	
		(0.16)			(0.16)	
Trust * Rel. Spec.		-0.160^{+}	-0.0941		-0.159^{+}	-0.0979
		(0.091)	(0.090)		(0.092)	(0.091)
Language HHI * Rel. Spec.		-0.120	-0.0885		-0.131	-0.0928

Back

Materials Share: industry characteristics controls

	Dependent variable: Materials Expenditure in Total Cost			
	(1)	(2)	(3)	(4)
Avg Age Of Civil Cases * Rel. Spec.	-0.0165*	-0.0137*	-0.0237*	-0.0162+
	(0.0069)	(0.0064)	(0.0094)	(0.0092)
Capital Intensity * Avg. age of cases		-0.103**		0.0139
		(0.037)		(0.064)
Industry Wage Premium * Avg. age of cases		-0.00139+		-0.00349*
		(0.00084)		(0.0015)
Industry Contract Worker Share * Avg. age of cases		-0.0105		0.0192
		(0.029)		(0.039)
Upstreamness * Avg. age of cases		0.00222		0.00657*
		(0.0015)		(0.0032)
Method	OLS	OLS	IV	IV
State × Rel. Spec. Controls	Yes	Yes	Yes	Yes
5-digit Industry FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes

Back

Vertical Distance: industry characteristics controls

	Dependent variable: Vertical Distance of Inputs from Output			
	(1)	(2)	(3)	(4)
Avg Age Of Civil Cases * Rel. Spec.	0.0320*	0.0261+	0.0437*	0.0253
	(0.014)	(0.014)	(0.021)	(0.022)
Capital Intensity * Avg. age of cases		-0.00400		0.213
		(0.073)		(0.15)
Industry Wage Premium * Avg. age of cases		0.00329		0.0106*
		(0.0021)		(0.0043)
Industry Contract Worker Share * Avg. age of cases		-0.0151		0.00351
		(0.025)		(0.048)
Upstreamness * Avg. age of cases		-0.00436		-0.00169
		(0.0036)		(0.0070)
Method	OLS	OLS	IV	IV
State × Rel. Spec. Controls	Yes	Yes	Yes	Yes

Back

Summary Stats, Recipe Classification

Table 3: Statistics on products and recipes

	Count
Products (5-digit ASIC)	4,530
Products with ≥ 3 plants	3,573
Products with ≥ 5 plants	3,034
Recipes	26,776
Recipes with ≥ 3 plants	6,280
Recipes with ≥ 5 plants	2,574
Avg. plants per recipe	8.2
SD plants per recipe	79.4

Table 4: Summary statistics on recipes

	Mean	Std. Dev.	Min	Max
Cost share of L	.40	.22	.0002	.999
Cost share of X_R	.27	.28	0	.999

Wedges and Enforcement

- Three ways weak enforcement might alter shares
  1. Wasted resources
  2. Quantity restrictions
  3. Higher effective input price

- Common feature: Wedge between shadow values of buyer and supplier

- Prediction of quantity restriction:
  - Larger wedges imply larger "markups"
  - But we do not see this

$$\frac{\text{revenue}}{\text{cost}} = \underbrace{\beta}_{<0} \text{Court Quality} \times \text{specificity} + \epsilon$$

Back

Sales/Cost Ratio

Table 5: Sales over Total Cost

	Dependent variable: Sales/Total Cost		
	(1)	(2)	(3)
Avg Age Of Civil Cases * Rel. Spec.	-0.0353**	-0.0347**	-0.0345**
	(0.0097)	(0.0094)	(0.0093)
Plant Age		0.000574**	0.000258+
		(0.00014)	(0.00014)
Log Employment			0.0314**
			(0.0016)
5-digit Industry FE	Yes	Yes	Yes
District FE	Yes	Yes	Yes
Estimator	OLS	OLS	OLS
R^2	0.114	0.110	0.115
Observations	208527	205109	204767

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Back

Sales/Cost Ratio, IV

Table 6: Sales over Total Cost

	Dependent variable: Sales/Total Cost		
	(1)	(2)	(3)
Avg Age Of Civil Cases * Rel. Spec.	-0.0494*	-0.0496*	-0.0508*
	(0.022)	(0.022)	(0.022)
Plant Age		0.000575**	0.000259^{+}
		(0.00014)	(0.00014)
Log Employment			0.0314**
			(0.0016)
5-digit Industry FE	Yes	Yes	Yes
District FE	Yes	Yes	Yes
Estimator	IV	IV	IV
R^2	0.114	0.110	0.115
Observations	208527	205109	204767

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Higher Price?

- Our baseline finding: distortion $\uparrow \quad \Rightarrow \quad$ materials share $\downarrow$
- If wedge acts like higher price, requires materials, primary inputs be substitutes
- Outside evidence: Close to Cobb Douglas, maybe complements
  - Oberfield-Raval (2018)
  - Atalay (2018)
- Can check with Indian Data
  - If cost of materials $\uparrow$, what happens to materials share?
    - If complements, $\uparrow$
    - If substitutes, $\downarrow$
  - What if suppliers rely more on rel. spec. inputs?

Elasticity of substitution at plant level

Dependence on R inputs of input industries as cost shifter

	Dependent variable: Materials Expenditure in Total Cost			
	(1)	(2)	(3)	(4)
Avg Age Of Civil Cases * Rel. Spec.	-0.0147^{+}	-0.0174^{+}	-0.0397**	-0.0421**
	(0.0080)	(0.0098)	(0.013)	(0.014)
LogGDPC * Rel. Spec.		-0.00849		-0.0178
		(0.013)		(0.017)
Avg Age Of Civ. Cases * Rel. Spec. of Upstream Sector	-0.00360	0.00265	0.0450*	0.0345^{+}
	(0.011)	(0.012)	(0.019)	(0.019)
Trust * Rel. Spec.		0.0250		0.0287
		(0.038)		(0.038)
Language HHI * Rel. Spec.		0.0346		0.0349
		(0.033)		(0.033)
Caste HHI * Rel. Spec.		0.109*		0.110*
		(0.050)		(0.050)
5-digit Industry FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes

Size and Age

Table 7: Plant Age and Size

	Dependent variable: Mat. Exp in Total Cost		
	(1)	(2)	(3)
Plant Age	-0.000733**		-0.000718**
	(0.000063)		(0.000061)
Log Employment		-0.00255**	-0.00171*
		(0.00085)	(0.00082)
5-digit Industry FE	Yes	Yes	Yes
District FE	Yes	Yes	Yes
Estimator			
R^2	0.488	0.487	0.489
Observations	211228	215688	210876

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Wedges and Enforcement

Market wage: w wage in excess of stealing

- If worker steals ψ^l units of output, needs to be paid $g^l(\psi^l)w$
- If supplier customizes incompletely by ψ^x, needs to be paid $g^x(\psi^x)\lambda_s$
- Contract specifies ψ^l, ψ^x. Workers choose ψ^l, supplier chooses ψ^x

Buyer minimizes cost:

$$\min g_l(\psi_l)wl + g_x(\psi_x)\lambda_s x$$

subject to

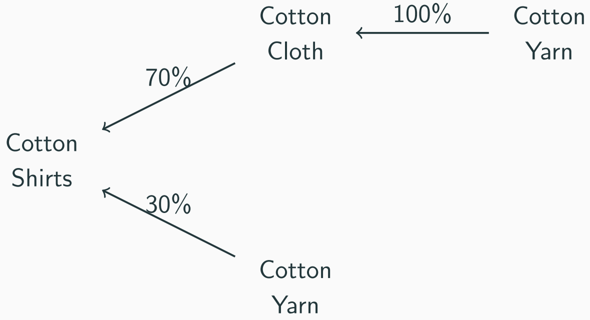
$$G\left(z_l \min\left\{l, \frac{\tilde{y}_l}{\psi_l}\right\}, z_x \min\left\{x, \frac{\tilde{y}_x}{\psi_x}\right\}\right) - \tilde{y}_l - \tilde{y}_x \geq y_b$$

- Weak enforcement: court only enforces claims in which damage is greater than a multiple $\tau - 1$ of transaction.
- Recover functional form if $g_l(\psi_l), g_x(\psi_x) \to 1$

Back

Vertical Distance

1. For a given product ω, construct the materials cost shares of industry ω on each input
2. Recursively construct the cost shares of the input industries (and inputs' inputs, etc...), excluding all products that are further downstream.
3. Vertical distance between ω and ω' is the average number of steps between ω and ω', weighted by the product of the cost shares.



$\Rightarrow$ Shirts $\leftarrow$ Cloth: 1; Shirts $\leftarrow$ Yarn: $0.3 \times 1 + 0.7 \times 1.0 \times 2 = 1.7$ Back

Identifying Recipes in the Data: Cluster Analysis

Use clustering algorithm to group plants that use similar input bundles.

Ward's method:

1. Start with the finest partition, i.e. the set of singletons $(\{j\})_{j \in J_\omega}$
2. In each step, merge two groups to minimize the sum of within-group distances from the mean:
$$\min_{\rho_n \geq \rho_{n-1}} \sum_{\rho \in \rho_n} \sum_{j \in \rho} \sum_{\omega} \left(m_{j\omega} - \overline{m}_{\rho\omega}\right)^2$$
   This creates a hierarchy of partitions.
3. Choose a partition (set of clusters) based on how many clusters you want.

Our implementation: cluster based on 3-digit and 5-digit input shares, pick # clusters based on # observations. Summary stats Back

Time variation: new benches

Two new high court benches during our sample period:

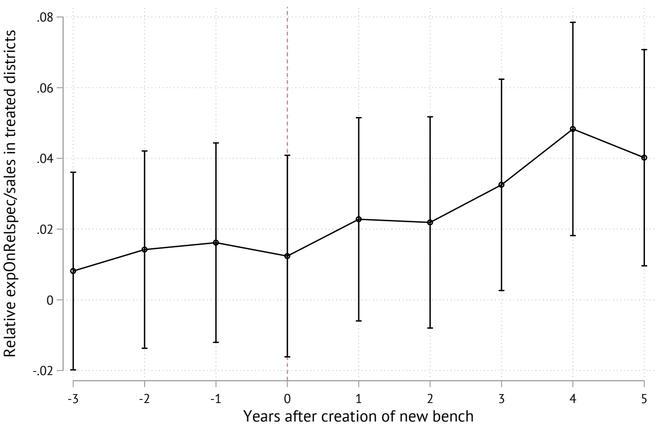
- Dharwad, Gulbarga (Karnataka, July 2008)
- Madurai (Tamil Nadu, July 2004)

	X^R/Sales	$s_R - s_H$	Materials/TotalCost	Vert. Distance
	(1)	(2)	(3)	(4)
$(\text{New Bench in District})_d \times (\text{Post})_t$	0.0126**	0.00960	-0.00305	0.00678
	(0.0043)	(0.0076)	(0.0033)	(0.010)
$(\text{New Bench in District})_d \times (\text{Post})_t \times (\text{Rel.Spec})_\omega$			0.0142	-0.0764*
			(0.010)	(0.031)
Plant × Product FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R^2	0.832	0.824	0.906	0.813
Observations	80427	74696	78462	77995

Back 1 Back 2 Back 3

Time variation: new benches

Figure 3: Expenditure on rel.spec. inputs in sales

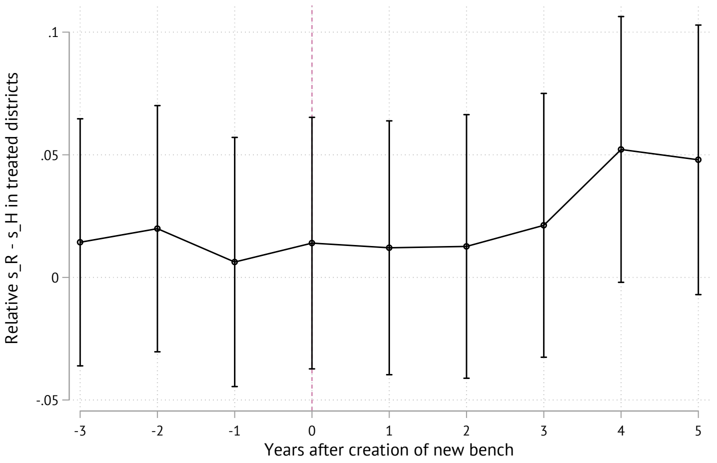


Treated districts vs. non-treated districts. Regression includes firm × product and year FE.

Back 1 Back 2 Back 3

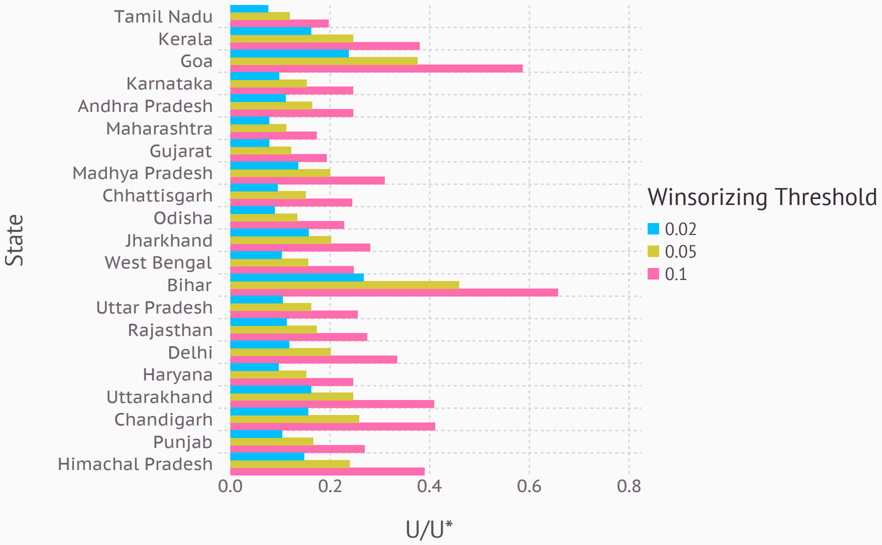
Time variation: new benches

Figure 4: $s^R - s^H$ on the LHS



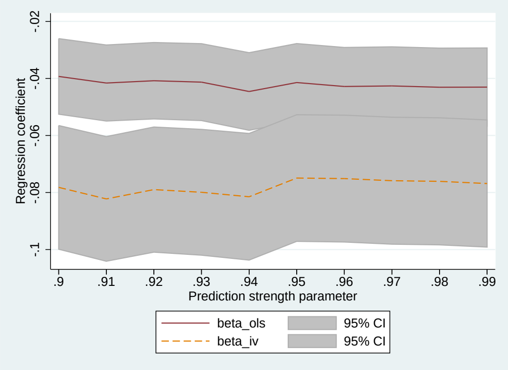
Treated districts vs. non-treated districts. Regression includes firm $\times$ product and year FE.

A Hsieh-Klenow exercise

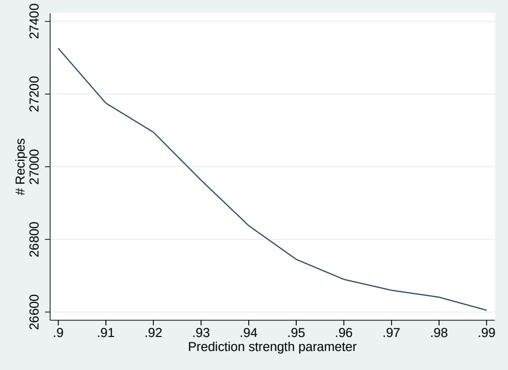


Robustness: How Finely to Define Recipes

Varying the hyperparameter for the Tibshirani-Walther cross-validation procedure generates similar number of recipes.



(a) Regression Coefficients

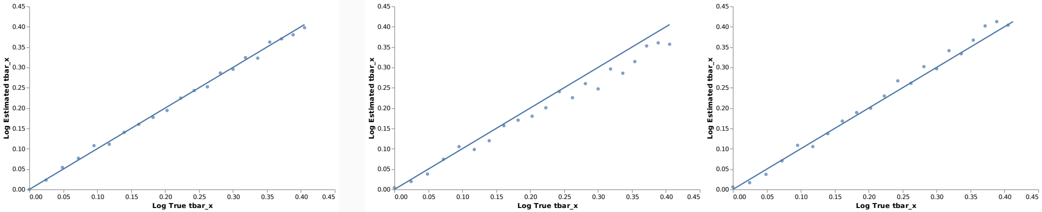


(b) Number of Recipes

Figure 6: Regression coefficients & number of recipes for different levels of recipe fineness

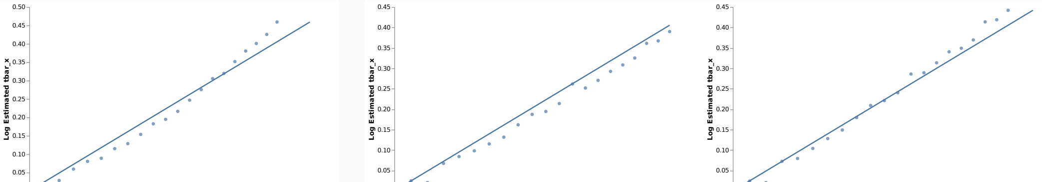
Large-Sample Monte Carlo Experiments

Simplest economy: two products, two recipes (varying R-intensity), 21 states with increasing $\bar{t}_x$

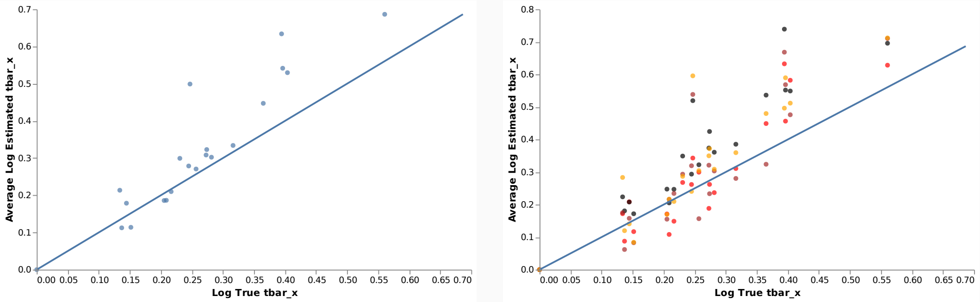


(a) One cluster per product **(b)** Two clusters per product **(c)** Four clusters per product

Figure 7: Number of observations not skewed across states



Small-sample Monte Carlo Experiments



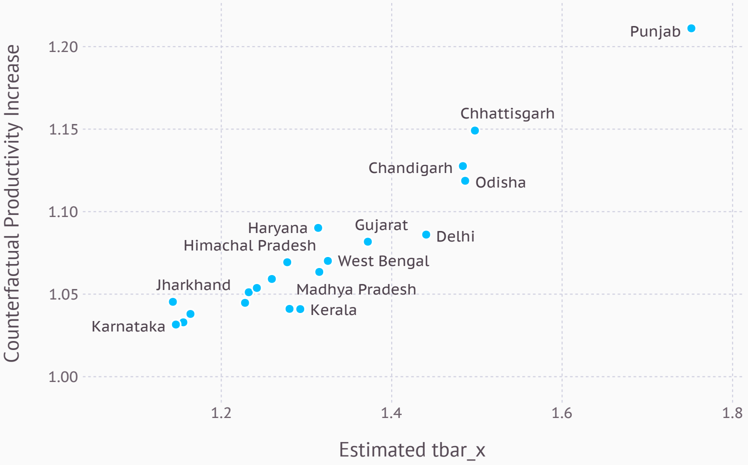
(a) Average of four MC simulations

(b) All four MC simulation results

Figure 9: MC results using actual number of observations and estimated $\bar{t}$

The figure shows actual (horizontal axis) vs. estimated (vertical axis) distortions from a simulated model economy, where the parameters are the point estimates from our benchmark estimation and the number of simulated plants is the same as in our actual dataset. The left panel shows average estimated distortions

Counterfactual: halve wedges $\bar{t}_x$



Back

Endogeneity: IV

- Since independence: # judges based on state population

⇒ backlogs have accumulated over time

- But: new states have been created, with new high courts and clean slate