Misallocation in the Market for Inputs: Enforcement and the Organization of Production

Johannes Boehm   Ezra Oberfield

Sciences Po   Princeton

21 October 2020

ERWIT

Misallocation in the Market for Inputs

- How important are distortions for productivity/income?

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

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

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
  - $\Rightarrow$ Might have wrong producers doing wrong tasks
  - But firms may overcome hold-up problems with some suppliers through informal means
  - $\Rightarrow$ Distortions may not show up as a wedge

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
  - $\Rightarrow$ Might have wrong producers doing wrong tasks
  - But firms may overcome hold-up problems with some suppliers through informal means
  - $\Rightarrow$ Distortions may not show up as a wedge
- Counterfactual: improving courts $\Rightarrow \nearrow$ TFP $\approx 4\%$

Literature

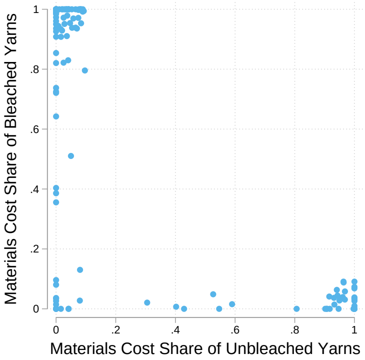
- Factor Misallocation: Restuccia & Rogerson (2008), Hsieh & Klenow (2009, 2014), Midrigan & Xu (2013), Hsieh Hurst Jones Klenow (2016), Garcia-Santana & Pijoan-Mas (2014)
- Firm heterogeneity and linkages in GE: Oberfield (2018), Eaton, Kortum, and Kramarz (2016), Lim (2016), Lu Mariscal Mejia (2016), Chaney (2015), Kikkawa, Mogstad, Dhyne, Tintelnot (2017), Acemoglu & Azar (2018), Kikkawa (2017)
  - Sourcing patterns: Costinot Vogel Wang (2012), Fally Hillberry (2017), Antras de Gortari (2017), Antras Fort Tintelnot (2017)
- Aggregation properties of production functions: Houthakker (1955), Jones (2005), Lagos (2006), Mangin (2015)
- Courts and economic performance: Johnson, McMillan, Woodruff (2002), Chemin (2012), Acemoglu and Johnson (2005), Nunn (2007), Levchenko (2007), Antras Acemoglu Helpman (2007) Laeven and Woodruff (2007), Ponticelli and Alencar (2016), Amirapu (2017)

Data

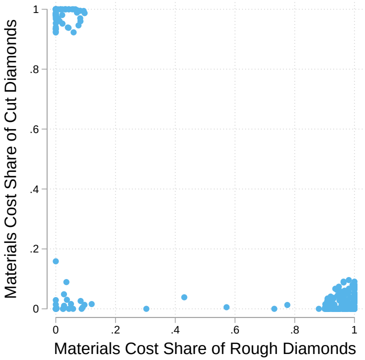
- Indian Annual Survey of Industries (ASI), 2001-2013
  - All manufacturing plants with > 100 employees, $1/5$ of plants between $20(10) - 100$
  - Drop plants without inputs, not operating, extreme materials share
  - $\sim 25,000$ plants per year

Data

- Indian Annual Survey of Industries (ASI), 2001-2013
  - All manufacturing plants with > 100 employees, $1/5$ of plants between $20(10) - 100$
  - Drop plants without inputs, not operating, extreme materials share
  - $\sim 25,000$ plants per year



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



(d) Input mixes for Polished Diamonds (92104)

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
- Homogeneous (H) vs. Relationship-specific (R) inputs
  - from Rauch classification: 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

Reduced-form regressions: Summary

- R-intensive industry + courts in your state slow $\Rightarrow$ lower materials share in total cost

$$\text{materialsInCost}_j = \underbrace{\beta}_{-0.0167^{**}} \text{courtSpeed}_s \times \text{R-intensity}_\omega + \alpha_s + \alpha_\omega + \varepsilon_j$$

Robustness:

- Instrumenting court speed by age of court ($\rightarrow$ determinants of court speed)
- With industry $\times$ state controls
- Results from time variation also consistent with this

Details

Reduced-form regressions: Summary

- R-intensive industry + courts in your state slow ⇒ lower materials share in total cost

$$\text{materialsInCost}_j = \underbrace{\beta}_{-0.0167**} \text{courtSpeed}_s \times \text{R-intensity}_\omega + \alpha_s + \alpha_\omega + \varepsilon_j$$

- Courts in your state slow ⇒ lower share of R vs. H inputs

$$X^R{}_j/(X^R{}_j + X^H{}_j) = \underbrace{\beta}_{-0.00546*} \text{courtSpeed}_s + \alpha_\omega + \varepsilon_j$$

Robustness:

- Instrumenting court speed by age of court (→ determinants of court speed)
- With industry × state controls
- Results from time variation also consistent with this

Details

Reduced-form regressions: Summary

- R-intensive industry + courts in your state slow ⇒ lower materials share in total cost

$$\text{materialsInCost}_j = \underbrace{\beta}_{-0.0167^{**}} \text{courtSpeed}_s \times \text{R-intensity}_\omega + \alpha_s + \alpha_\omega + \varepsilon_j$$

- Courts in your state slow ⇒ lower share of R vs. H inputs

$$X^R{}_j/(X^R{}_j + X^H{}_j) = \underbrace{\beta}_{-0.00546^{*}} \text{courtSpeed}_s + \alpha_\omega + \varepsilon_j$$

- R-intensive industry + courts in your state slow ⇒ plant has longer vertical span of production ("produce shirts from yarn instead of from cloth")

$$\text{verticalSpan}_j = \underbrace{\beta}_{0.0195^{+}} \text{courtSpeed}_s \times \text{R-intensity}_\omega + \alpha_s + \alpha_\omega + \varepsilon_j$$

Robustness:

- Instrumenting court speed by age of court (→ determinants of court speed)
- With industry × state controls
- Results from time variation also consistent with this

Details

Model

Goals

- 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

Why not Hsieh-Klenow?

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)

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

Firms minimize cost over all techniques (from all recipes)

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

- Number of 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, number of 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.

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

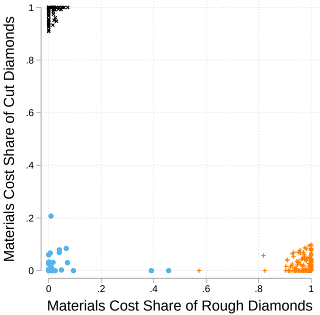
$$\left(\frac{C'_\omega}{C_\omega} \right)^{-\gamma} = \sum_{\rho \in \varrho(\omega)} R_{\omega\rho} \left[\left(\frac{t_x^{*\prime}}{t_x^*} \right)^{\alpha_R^\rho} \prod_{\hat{\omega} \in \Omega^\rho} \left(\frac{C'_{\hat{\omega}}}{C_{\hat{\omega}}} \right)^{\alpha_{\hat{\omega}}^\rho} \right]^{-\gamma}$$

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: 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

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

- Assume: Wedges drawn from inverse Pareto distribution $T_d(t_x) = t_x^{\tau_d}$

$$\bar{t}_x^d = 1 + \frac{1}{\zeta_R + \tau_x^d}$$

To back out τ_x^d, need ζ_R

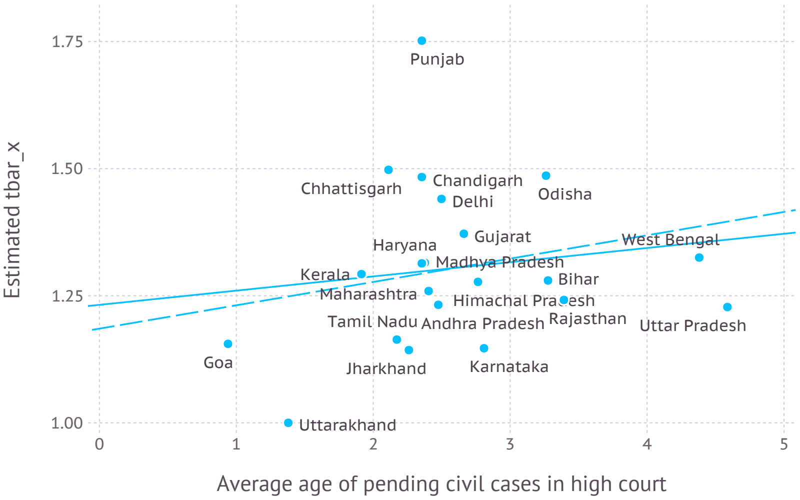
$$\log(X_{i\omega}^{DOM}/X_{i\omega}^{IMP}) = \zeta \log(1 + \text{tariff}_{i\omega t}) + \lambda_t + \lambda_{i\omega} + \eta_{i\omega t}$$

	Dependent variable: $\log(X_{\omega\hat{\omega}t}^{DOM}/X_{i\omega\hat{\omega}t}^{IMP})$		
	(1)	(2)	(3)
$\log(1 + \iota_{\hat{\omega}t})$	0.617	0.218	1.209*
	(0.44)	(0.77)	(0.52)
Industry $\times$ Input FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Level	5-digit	5-digit	5-digit
Sample	All inputs	R only	H only
R^2	0.601	0.580	0.623
Observations	23692	12002	11690

Robust errors in parentheses, clustered at the state $\times$ industry level. Sample

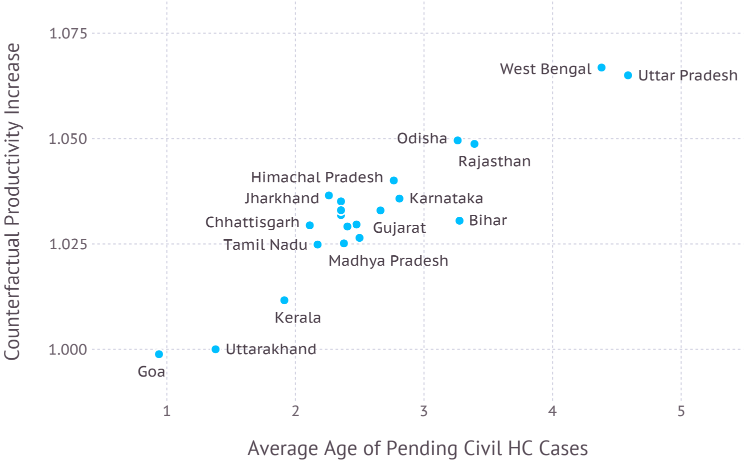
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

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).



1 year faster $\Rightarrow \approx 2\%$ higher income per capita

Halve wedges

Conclusion

- Huge amount of **heterogeneity** in intermediate input use, even within narrow industries
  - Some of it is due to **differences in organization/technology**
    - $\Rightarrow$ **Recipes**
  - Some of it is due to **differences in location**
    - $\Rightarrow$ Identify this as **wedges** (if asymmetric in intermediate inputs)
- Framework for studying and identifying stochastic frictions in an economy with input-output linkages
- Applied to the formal Indian manufacturing sector, suggests that courts are important

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

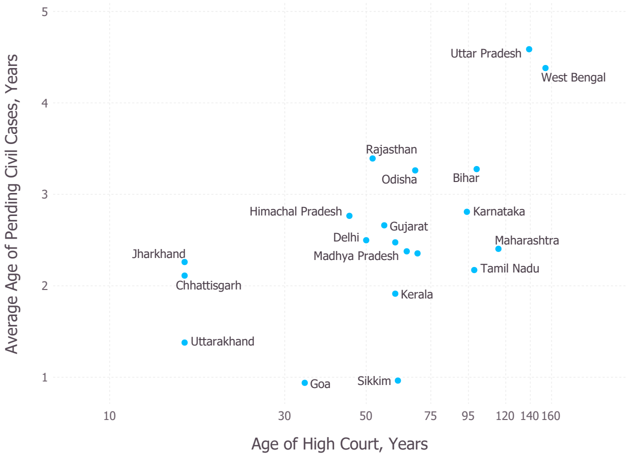
Back

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



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

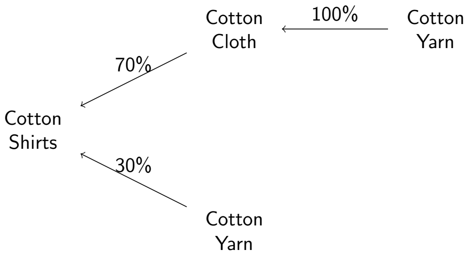
	Dependent variable: $X_j^R/(X_j^R + X_j^H)$					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg age of Civil HC cases	-0.00547*	-0.00621**	-0.00530*	-0.0144**	-0.0146**	-0.0167**
	(0.0022)	(0.0023)	(0.0024)	(0.0044)	(0.0044)	(0.0045)
Log district GDP/capita		-0.00389	-0.00384		-0.00912+	-0.00980+
		(0.0045)	(0.0046)		(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

Vertical Distance Between Goods

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: $0.3 \times 1 + 0.7 \times 1.0 \times 2 = 1.7$

Vertical Distance Between Goods – Examples

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

Input group	Average vertical distance
Fabrics Or Cloths	1.67
Yarns	2.78
Raw Cotton	3.55

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

ASIC code	Input description	Vertical distance
73105	Aluminium Casting	1.23
73104	Aluminium Alloys	1.46
73103	Aluminium	1.92
22301	Alumina (Aluminium Oxide)	2.92
31301	Caustic Soda (Sodium Hydroxide)	3.81
23107	Coal	3.85
22304	Bauxite, raw	3.93

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

($\Leftrightarrow$ inputs are further away from outputs)

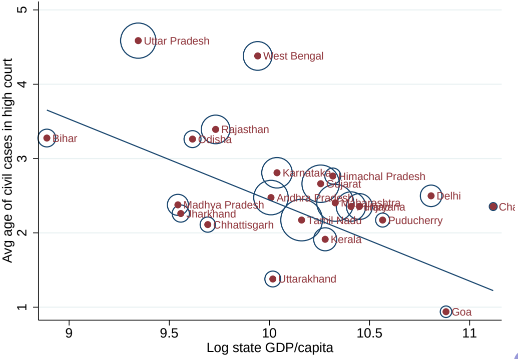
	Dependent variable: Avg Vertical Distance of Inputs from Output					
	(1)	(2)	(3)	(4)	(5)	(6)
Avg Age Of Civil Cases * Rel. Spec.	0.0195^{+}	0.0341^{*}	0.0320^{*}	0.0292	0.0414^{+}	0.0437^{*}
	(0.011)	(0.014)	(0.014)	(0.019)	(0.022)	(0.021)
LogGDPC * Rel. Spec.		0.0517^{+}	0.0309		0.0613^{+}	0.0471
		(0.029)	(0.034)		(0.037)	(0.040)
Rel. Spec. $\times$ 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 $\times$ industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

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

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** (0.0055)	-0.0155* (0.0066)		
Avg age of Civil HC cases (adj.)			0.0172** (0.0037)	
Adjusted Court Quality * Rel. Spec.			-0.0328** (0.0064)	-0.0282** (0.0064)
Log district GDP/capita	0.00694+ (0.0038)		0.00578 (0.0038)	
LogGDPC * Rel. Spec.		-0.00159 (0.012)		0.00390 (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 own HC's benches.)

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

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

Note: sample is smaller because some plants don't use relspec. or homog. inputs.

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
Observations	208527	196748	208527	196748

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

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

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

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
		(0.095)	(0.092)		(0.095)	(0.093)
Caste HHI * Rel. Spec.		-0.133	-0.202^{+}		-0.155	-0.213^{+}
		(0.13)	(0.12)		(0.13)	(0.12)
Corruption * Rel. Spec.		0.570*	0.463^{+}		0.476^{+}	0.442^{+}
		(0.26)	(0.25)		(0.26)	(0.25)
5-digit Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE			Yes			Yes
Estimator	OLS	OLS	OLS	IV	IV	IV
R^2	0.432	0.443	0.453	0.432	0.443	0.453
Observations	163344	154028	154021	163344	154028	154021

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

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

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
R^2	0.484	0.484	0.484	0.484
Observations	196748	196748	196748	196748

Standard errors in parentheses, clustered at the state × industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
"State × Rel. Spec. controls" are interactions of GDP/capita, trust, language herfindahl, caste herfindahl, and corruption with relationship-specificity.

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
5-digit Industry FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
R^2	0.453	0.453	0.453	0.453
Observations	154021	154021	154021	154021

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

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

"State × Rel. Spec. controls" are interactions of GDP/capita, trust, language herfindahl, caste herfindahl, and corruption with relationship-specificity.

Back

Summary Stats, Recipe Classification

Table: Statistics on products and recipes

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

Table: 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
Cost share of X_H	.33	.30	0	.998
Number of inputs with cost share $>$ 1%	4.4	4.6	1	37
Number of inputs with cost share $>$ 0.1%	6.4	12.6	1	205

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

Back

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
Estimator	OLS	OLS	IV	IV
R^2	0.480	0.484	0.480	0.484
Observations	208527	196748	208527	196748

Standard errors in parentheses, clustered at the state $\times$ industry level.

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

Size and Age

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

Back

Wedges and Plant Characteristics

Table: Wedges and Plant Characteristics

	Age	Size	Multiproduct	# Products
	(1)	(2)	(3)	(4)
Avg Age Of Civil Cases * Rel. Spec.	0.620^{+}	-0.0253	-0.0121	-0.0580
	(0.32)	(0.040)	(0.0076)	(0.037)
5-digit Industry FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
R^2	0.214	0.339	0.301	0.295
Observations	353392	359820	360316	360316

Back

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

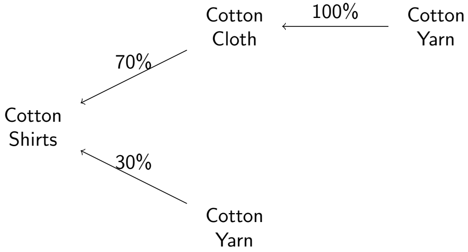
$$G\left(z_l \min\left\{l, \frac{\tilde{y}_l}{\psi_l}\right\}, z_x \min\left\{x, \frac{\tilde{y}_x}{\psi_x}\right\}\right) - \tilde{y}_l - \tilde{y}_x \geq y_b$$

- 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:

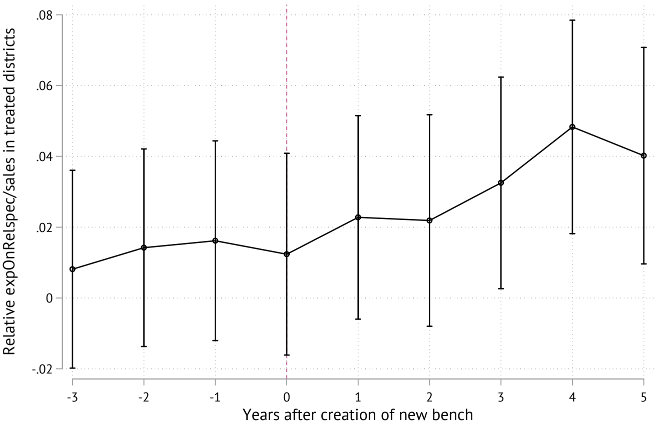
- Dharwad, Gulbarga (Karnataka, July 2008)
- Madurai (Tamil Nadu, July 2004)

	X^R/Sales	$s_R - s_H$	Materials/TotalCost	Vert. Distance
	(1)	(2)	(3)	(4)
(New Bench in District)$_d \times$ (Post)$_t$	0.0126**	0.00960	-0.00305	0.00678
	(0.0043)	(0.0076)	(0.0033)	(0.010)
(New Bench in District)$_d \times$ (Post)$_t \times$ (Rel.Spec)$_\omega$			0.0142	-0.0764*
			(0.010)	(0.031)
Plant $\times$ 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

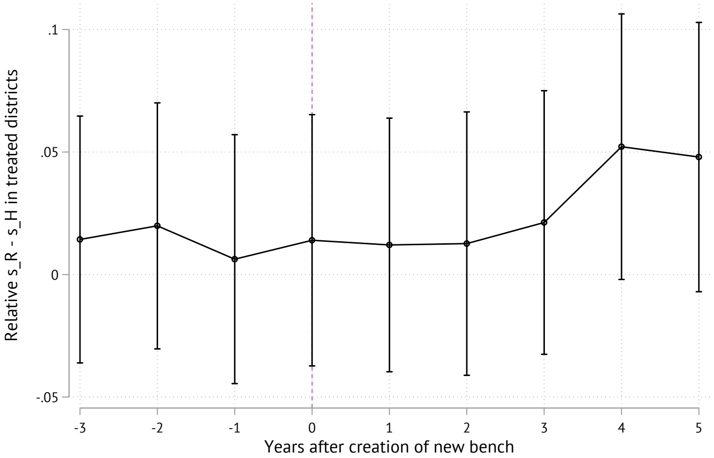
Figure: Expenditure on rel.spec. inputs in sales



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

Time variation: new benches

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



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

Back 1 Back 2 Back 3

A Hsieh-Klenow exercise

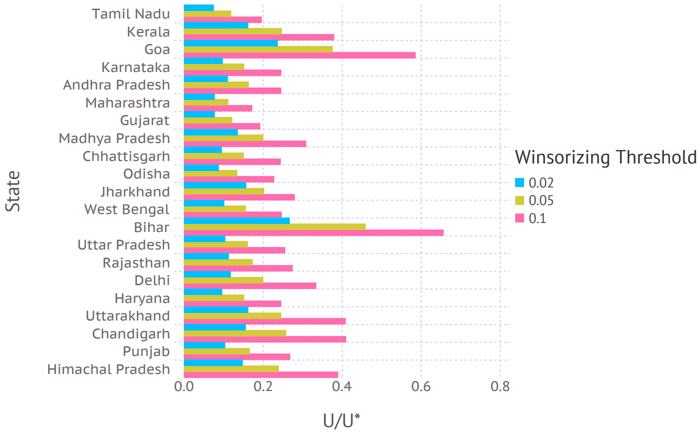
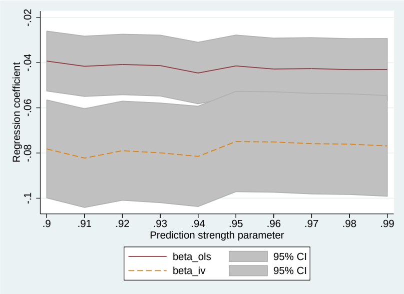


Figure: Hsieh-Klenow Exercise Results, By State

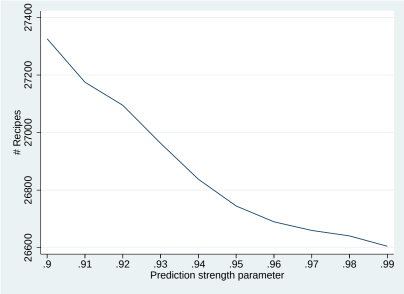
Back

Robustness: How Finely to Define Recipes

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



(a) Regression Coefficients



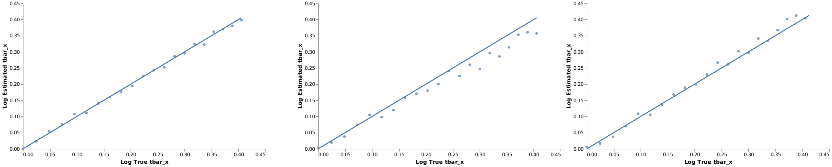
(b) Number of Recipes

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

Back

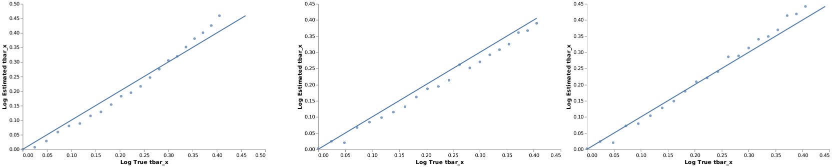
Large-Sample Monte Carlo Experiments

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: Number of observations not skewed across states

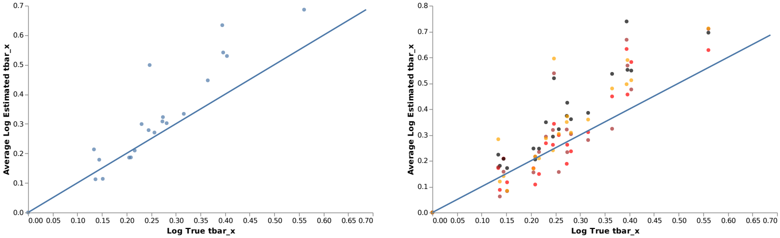


(a) One cluster per product (b) Two clusters per product (c) Four clusters per product

Back

Figure: Number of observations skewed across states

Small-sample Monte Carlo Experiments



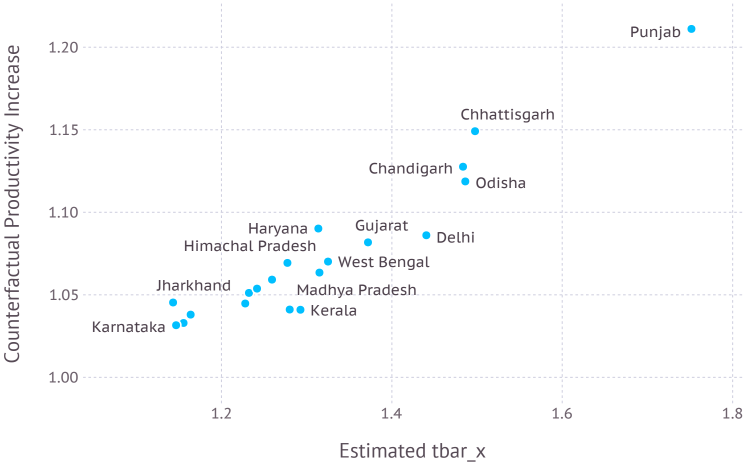
(a) Average of four MC simulations (b) All four MC simulation results

Figure: 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 across four runs, the right panel shows estimates from each individual run (coded in four different colors).

Back

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



Back

Why not Hsieh and Klenow (2009)?

Hsieh-Klenow takes variation in factor shares as evidence of distortions.

- Problematic in the case of intermediate inputs: there's just so much variation in the data (you get crazy numbers) HK
- Unlikely that all plants have Cobb-Douglas PF in intermediate inputs
- HK relies heavily on second moments: measurement error becomes problematic (Rotemberg-White, Bils et al.)

Challenge: departing from Hsieh-Klenow/Cobb-Douglas without having to rely on price/quantity data

Back