Misallocation in the Market for Inputs:

Enforcement and the Organization of Production

Johannes Boehm
Sciences Po

Ezra Oberfield
Princeton

LSE Workshop on Networks in Macro & Finance

June 2017

Misallocation in the Market for Inputs

- How important are distortions for income differences?
- Our focus: Distortions in use of intermediate inputs
  - Role of courts & contract enforcement
- Margins
  - Which intermediate inputs to use?
  - How much to do in-house?
- Distortions
  - Might have wrong producers doing wrong tasks
  - Accumulate in supply chains

Manufacturing Plants in India

- New facts
  - Enormous variation in materials shares
    - but more variation in industries that use rel.-spec. inputs
  - In states with worse enforcement...input bundles systematically different
    - Industries using homogeneous inputs: higher materials share
    - Relative to those, industries using rel.-spec. inputs: lower materials shares
    - Within input bundles: shift toward homogeneous inputs
- Impact on aggregate productivity? $\Rightarrow$ Structural model
  - Key ingredients:
    - Firms can choose between different modes of production
    - Organization of production is endogenous
  - Key Challenge: Separate misallocation from heterogeneity
  - Preliminary results: Back out wedges on use of rel.-spec. inputs, labor
    - Correlated with court congestion
    - Reducing congestion in worst state to that of best state $\Rightarrow$ TFP $\uparrow\approx 6\%$.
    - Wedges are several times larger

Literature

- Factor Misallocation Literature: Restuccia & Rogerson (2008), Hsieh and Klenow (2009, 2014), Midrigan and Xu (2013), Hsieh Hurst Jones Klenow (2016)

- Multi-sector models with linkages: Jones (2011a,b), Bartelme and Gorodnichenko (2016), Boehm (2016), Ciccone and Caprettini (2016), Liu (2016), Bigio and Lao (2016), Caliendo, Parro, Tsyvinski (2017), Tang and Krishna (2017)

- Firm heterogeneity and linkages in GE: Oberfield (2016), Eaton, Kortum, and Kramarz (2016), Lim (2016), Lu Mariscal Mejia (2016), Chaney (2015), Kikkawa, Mogstad, Dhyne, Tintelnot (2017)

- Aggregation properties of production functions: Houthakker (1955), Jones (2005), Lagos (2006), Mangin (2015)

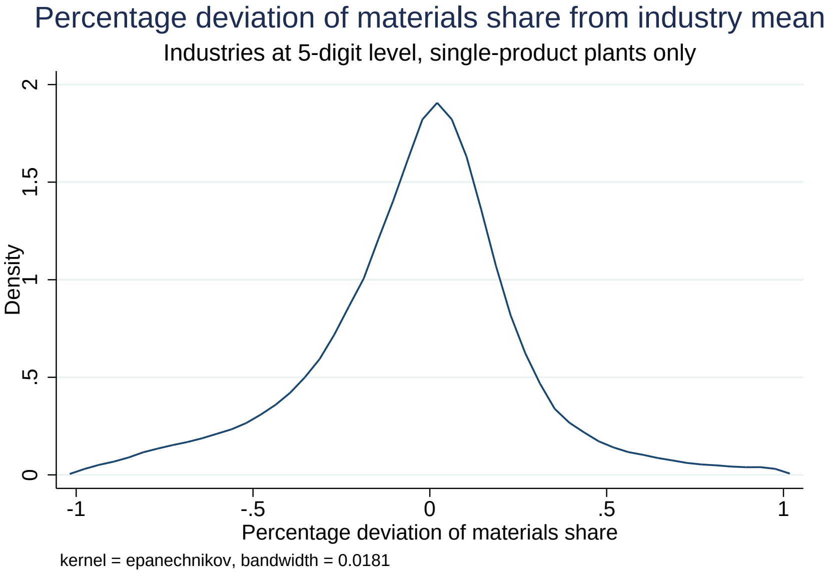
- Courts and economic performance: Chemin (2012), Acemoglu and Johnson (2005), Nunn (2007), Levchenko (2007), Antras Acemoglu Helpman (2007) Laeven and Woodruff (2007), Ponticelli and Alencar (2016)

Reduced Form Evidence

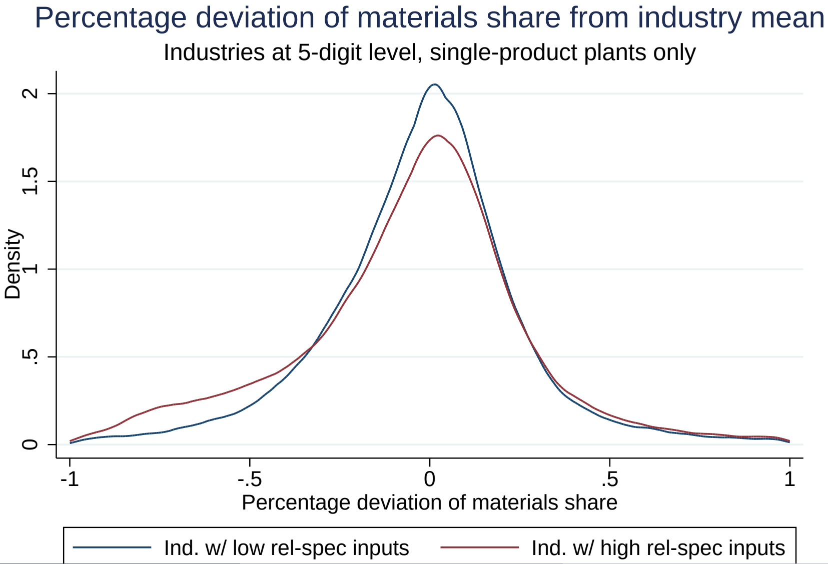
Data

- Indian Annual Survey of Industries (ASI), 2001-2010
  - All manufacturing plants with more than 100 employees, $1/5$ of plants between 20-100
  - Drop plants without inputs, not operating, extreme materials share
  - $\sim 25,000$ plants per year
- Standardized vs. Relationship-specific (Rauch)
  - Standardized $\approx$ sold on an organized exchange, ref. price in trade pub.
  - Relationship-specific $\approx$ 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

Large Variation in Materials Shares (within industries)

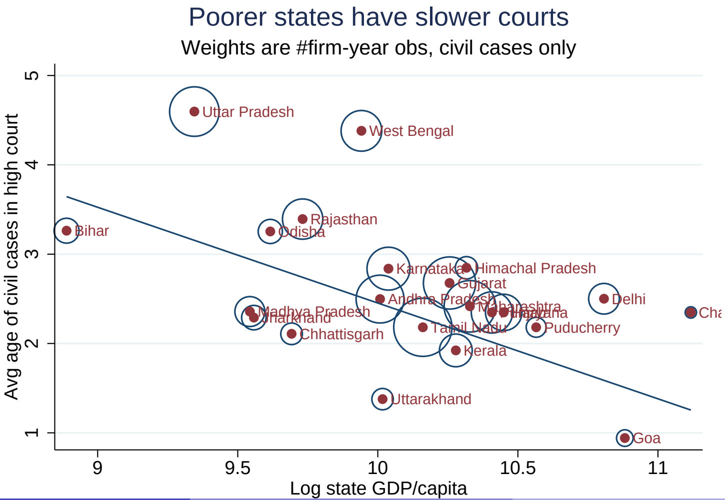


Different depending on industry's reliance on relationship-specific inputs

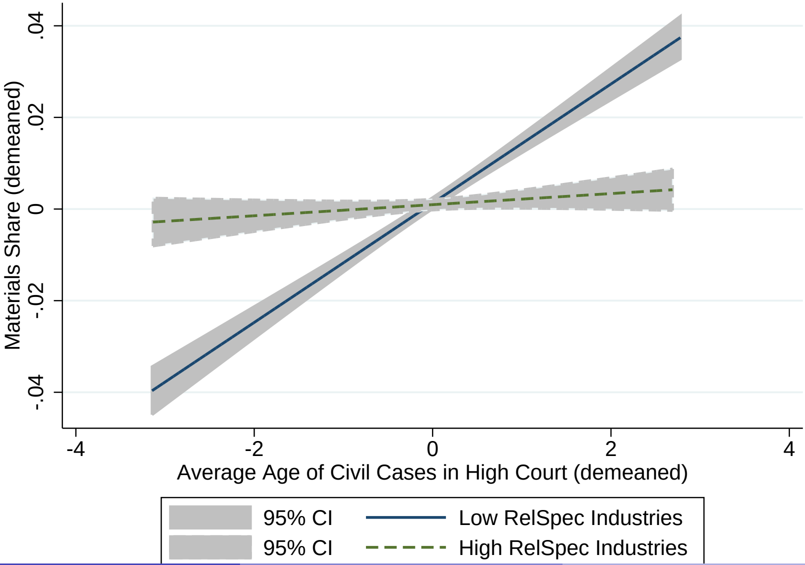


Slow Courts

- 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



Mat Share higher in states with more congested courts – but relatively lower in relationship-specific industries



Within Industry Regression

	(1) MatShare	(2) MatShare	(3) MatShare	(4) MatShare
Avg age of Civil HC cases	0.00715***	0.00904***	0.0135***	0.0147***
	(0.000592)	(0.000679)	(0.00131)	(0.00138)
Log district GDP/capita		0.00605***		0.00612***
		(0.00129)		(0.00129)
log Pop Density 2001		-0.00213***	-0.00109*	-0.00219***
		(0.000516)	(0.000475)	(0.000517)
AvgAgeOfCivCases * Rel. Spec.			-0.0128***	-0.0121***
			(0.00248)	(0.00257)
5-digit product FE	yes	yes	yes	yes
Observations	198127	183688	191004	183688
R^2	0.431	0.441	0.437	0.441

Standard errors in parentheses, clustered at state level
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

- Large asymmetry between industries that rely heavily on relationship-specific inputs vs industries that rely on standardized inputs

Within Industry, State Regression

	(1) MatShare	(2) MatShare
AvgAgeOfCivCases * Rel. Spec.	-0.0120***	-0.0105**
	(0.00256)	(0.00341)
Log GDP/capita * Rel. Spec.		-0.000602
		(0.00714)
5-digit product FE	yes	yes
State FE	yes	yes
Observations	209188	200663
R^2	0.470	0.476

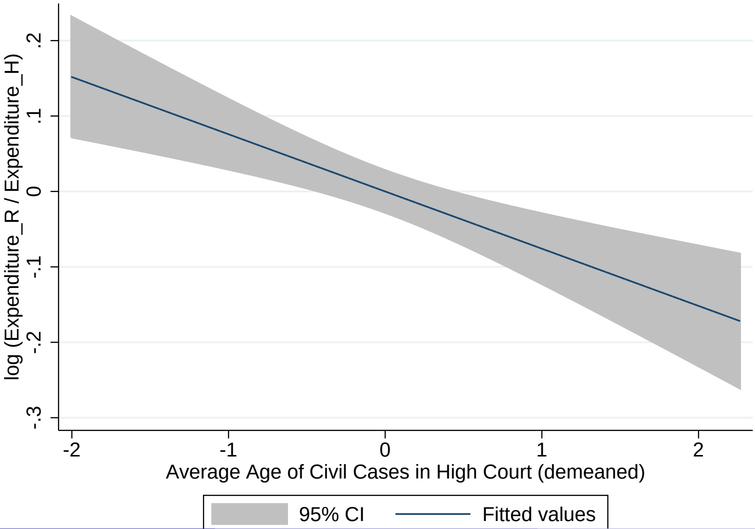
Standard errors in parentheses, clustered at state level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

- Moving from avg age of 1 year to 4 years: $\Rightarrow$ M-share $\downarrow 3.6pp$ more in industries that rely on relationship goods than in industries that rely on standardized inputs

In states with slow courts, input baskets are tilted towards homogeneous inputs

Within-industry relationship:

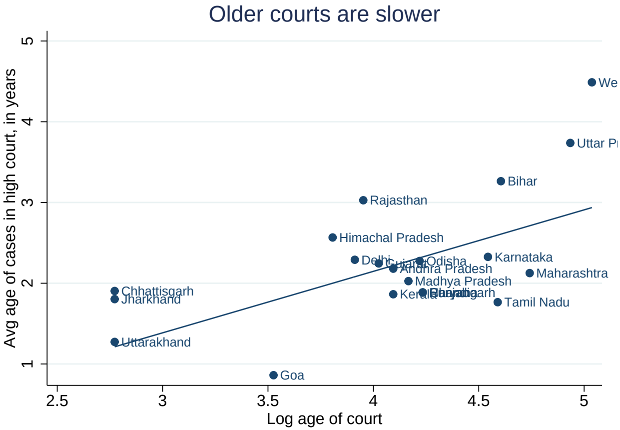


Endogeneity: IV

- Since independence: # judges based on state population

⇒ backlogs have been accumulating over time

- But: new states have been created, and therefore new high courts
- These courts start with a clean slate



IV makes coefficient larger

	(1) logshareRH	(2) logshareRH	(3) logshareRH
Avg age of civil HC cases (instr.)	-0.0544**	-0.0438*	-0.0580*
	(0.0205)	(0.0209)	(0.0292)
log pop density		-0.0220*	-0.0113
		(0.0101)	(0.0149)
log(gdpc)			-0.0806
			(0.0503)
Recipe FE	yes	yes	yes
Observations	24387	24387	22924
R^2	0.695	0.695	0.700

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Back

Model: How Costly are Distortions?

Goals

- Goal: Natural distribution of expenditure shares on different types of inputs
- Main identifying assumption: slow courts do not distort use of homog. inputs
  - Slow courts shift distribution
  - First moment matters! (contrast to Hsieh-Klenow)
- Things we don't want to attribute to misallocation
  - Heterogeneity in production technology across plants
  - Selection into method of production
  - Heterogeneity across locations in
    - Preferences over goods
    - Prevalence of various industries

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

$$U = \left[\sum_{\omega} \beta_{\omega}^{\frac{1}{\eta}} C_{\omega}^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}} \qquad C_{\omega} = \left[\int_0^{J_\omega} c_j^{\frac{\varepsilon_\omega - 1}{\varepsilon_\omega}} dj \right]^{\frac{\varepsilon_\omega}{\varepsilon_\omega - 1}}$$

Production

- **Technology**: Firms draw many ways of producing, uses most cost-effective
  - Recipe $\rho \in \varrho(\omega)$: broad class, uses inputs from particular industries, $\hat{\omega}_1^\rho, ..., \hat{\omega}_n^\rho$
  - A technique is production function using
    - particular suppliers $s_1, ..., s_n$
    - Match-specific input-augmenting productivities $z_l, z_{x1}, ...z_{xn}$

$$y_b = G_\rho\left(z_l l, z_{x1} x_{s_1}, ..., z_{xn} x_{s_n}\right), \qquad G \text{ is CRS, inputs are complements}$$

Production

- **Technology**: Firms draw many ways of producing, uses most cost-effective
  - Recipe $\rho \in \varrho(\omega)$: broad class, uses inputs from particular industries, $\hat{\omega}_1^\rho, ..., \hat{\omega}_n^\rho$
  - A technique is production function using
    - particular suppliers $s_1, ..., s_n$
    - Match-specific input-augmenting productivities $z_l, z_{x1}, ...z_{xn}$

$$y_b = G_\rho\bigg(z_l l, z_{x1} x_{s_1}, ..., z_{xn} x_{s_n}\bigg), \qquad G \text{ is CRS, inputs are complements}$$

- Techniques arrive randomly: Among those of type ω,
  - $\#$ techniques for recipe ρ with each productivity better than $\{z_l, z_{x1}, ..., z_{xn}\}$ is $\sim$ Poisson with mean

$$m_{\omega\rho} z_l^{-\zeta_l^\rho} z_{x1}^{-\zeta_{x1}^\rho} ... z_{xn}^{-\zeta_{xn}^\rho}$$

  - with $\zeta_l^\rho + \zeta_{x1}^\rho + ... + \zeta_{xn}^\rho = \gamma_\omega$

- Define normalized tail exponents

$$\alpha_l^\rho \equiv \frac{\zeta_l^\rho}{\gamma_\omega}, \qquad \alpha_{xi}^\rho \equiv \frac{\zeta_{xi}^\rho}{\gamma_\omega} \qquad \Rightarrow \qquad \alpha_l^\rho + \sum_i \alpha_{xi}^\rho = 1$$

Contract Enforcement

- **Weak Enforcement**: For each technique two types of wedges

$$t_l, t_{x1}, ..., t_{xn} \sim T_\rho \left(t_l, t_{x1}, ..., t_{xn}\right)$$

  - Equivalent to tax (paid with output) that is thrown in ocean Why?
  - One Microfoundation Details
    - Goods can be customized, but holdup problem
    - Workers can steal, but stealing effort is wasteful
    - Court quality determines size of loss before contract is enforced

- Depends on sourcing industry
  - i Homogeneous: $t_{xi} = 1$
  - i Relationship-specific: $t_{xi} \in [0, 1]$

Aggregation

Proposition: Let $q_j = \frac{w}{MC_j}$, $F_\omega(q)$ be CDF among firms in industry ω. Then

$$F_\omega(q) = e^{-(q/Q_\omega)^{-\gamma_\omega}}$$

where

$$\begin{aligned} Q_\omega &= \left\{ \sum_{\rho \in \varrho(\omega)} m_{\omega\rho} \kappa_{\omega\rho} \left(t^*_{\omega\rho} \prod_i Q_{\hat{\omega}_i^\rho}^{\alpha_{xi}^\rho} \right)^{\gamma_\omega} \right\}^{1/\gamma_\omega} \\ t^*_{\omega\rho} &= \left\{ \int \left(t_l^{\alpha_l^\rho} t_{x1}^{\alpha_{x1}^\rho} \ldots t_{xn}^{\alpha_{xn}^\rho} \right)^{\gamma_\omega} T\left(dt_l, dt_{x1}, \ldots, dt_{xn}\right) \right\}^{1/\gamma_\omega} \\ \kappa_{\omega\rho} &= \text{constant} \end{aligned}$$

Proposition: Among firms in ω using recipe ρ, share of total exp. on:

Labor: $\dfrac{\alpha_l^\rho \bar{t}_l^\rho}{\alpha_l^\rho \bar{t}_l^\rho + \sum_i \alpha_{xi}^\rho \bar{t}_{xi}^\rho}$, input i : $\dfrac{\alpha_{xi}^\rho \bar{t}_{xi}^\rho}{\alpha_l^\rho \bar{t}_l^\rho + \sum_i \alpha_{xi}^\rho \bar{t}_{xi}^\rho}$

where $\bar{t}_{xi}^\rho \equiv \int t_{xi} \tilde{T}(dt)$, $\bar{t}_l^\rho \equiv \int t_l \tilde{T}(dt)$, summarize distortions

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.

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

$$\frac{Q'_\omega}{Q_\omega} = \left\{ \sum_{\rho \in \varrho(\omega)} R_{\omega\rho} \left[\frac{t^{*'}_{\omega\rho}}{t^*_{\omega\rho}} \prod_i \left(\frac{Q'_{\hat{\omega}_i^\rho}}{Q_{\hat{\omega}_i^\rho}} \right)^{\alpha^\rho_{xi}} \right]^{\gamma_\omega} \right\}^{1/\gamma_\omega}$$

Identification

- Same across states: Recipe technology
  - Production function (G_ρ)
  - Shape of technology draws ($\zeta_{\rho l}, \{\zeta_{\rho x i}\}$)

- Different across states
  - Measure of producers of each type (J_ω)
  - Prevalence of different recipes ($m_{\omega\rho}$)
  - Household Preferences (β_ω)
  - Distribution of wedges for each recipe (T_ρ)

- Main identifying assump.: Slow courts do not distort use of homog. inputs

- Other Assumptions
  - Plants in state d draw t_x, t_l from $T_{\rho d}(t_x, t_l)$
    - t_x applies to all relationship-specific inputs
    - No wedge for homogenous inputs
  - No trade across states
  - L is labor equipped with other primary inputs (capital, energy, services)

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

Identifying Recipes in the Data

Cluster analysis uncovers different ways to produce a product.

Example: cloth, bleached, cotton (code 63303)

	input value, %	Description	# firm-years
Recipe 1	95	yarn bleached, cotton	54
	2	grey cloth (bleached / unbleached)	
	2	chemical & allied substances & products, n.e.c	
	1	colour, chemicals	
Recipe 2	35	grey cloth (bleached / unbleached)	39
	13	yarn, finished / processed - cotton (knitted)	
	6	fabrics, cotton	
	5	colour, chemicals	
	5	yarn dyed, synthetic	
	35	(others)	
Recipe 3	98	yarn unbleached, cotton	22
	1	cotton raw - others (pressed)	
	1	colour, chemicals	
Recipe 4	90	yarn, grey-cotton	18
	6	dye stuff	
	2	cotton woven	
	1	maize atta/flour/maida/sooji	
	1	benders (starch)	

Moments for GMM

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

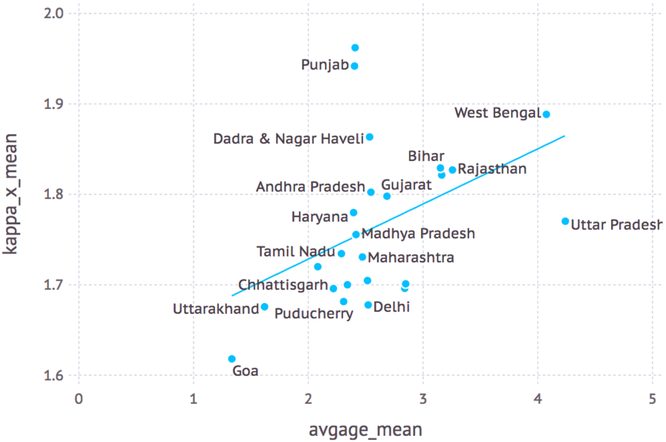
- The first moments of revenue shares among firms that use recipe ρ satisfy:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\bar{t}_x^\rho}\frac{s_{Rj}}{\alpha_R^\rho} - \frac{s_{Hj}}{\alpha_H^\rho}\right] &= 0 \\
\mathbb{E}\left[\frac{1}{\bar{t}_l^\rho}\frac{s_{Lj}}{\alpha_L^\rho} - \frac{s_{Hj}}{\alpha_H^\rho}\right] &= 0
\end{aligned}$$

- If, in addition, G_ρ is CES, T_ρ is Pareto,the second moments of revenue shares satisfy:

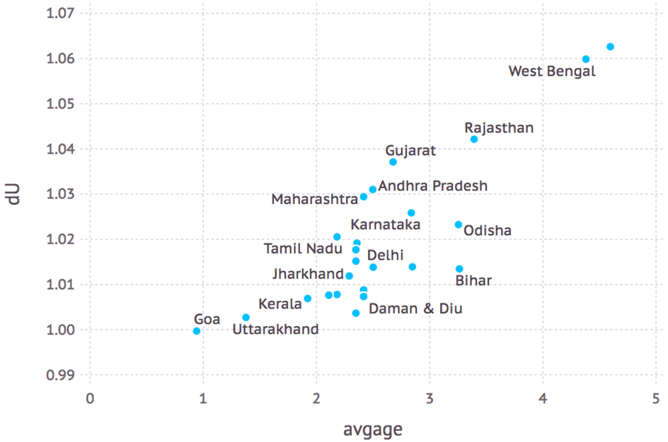
$$\begin{aligned}
\mathbb{E}\left[\left(\frac{2}{\bar{t}_x^\rho} - 1\right)\frac{s_{Rj}^2}{\alpha_R^\rho\left(\alpha_R^\rho + \frac{1-\sigma_\rho}{\gamma_\omega}\right)} - \frac{s_{Hj}^2}{\alpha_H^\rho\left(\alpha_H^\rho + \frac{1-\sigma_\rho}{\gamma_\omega}\right)}\right] &= 0 \\
\mathbb{E}\left[\left(\frac{2}{\bar{t}_l^\rho} - 1\right)\frac{s_{Lj}^2}{\alpha_L^\rho\left(\alpha_L^\rho + \frac{1-\sigma_\rho}{\gamma_\omega}\right)} - \frac{s_{Hj}^2}{\alpha_H^\rho\left(\alpha_H^\rho + \frac{1-\sigma_\rho}{\gamma_\omega}\right)}\right] &= 0
\end{aligned}$$

Intermediate input wedges are correlated with court quality



Gains From Improving Courts

Counterfactual sets court quality to 1.



Formal definition of shocks

Simple model:
Joint CDF of shocks:

$$Z\left(z_l, z_x\right) = \left(z_l/\underline{z}_l\right)^{-\zeta_l}\left(z_x/\underline{z}_x\right)^{-\zeta_x}$$

Define

$$m = M\underline{z}_l^{\zeta_l}\underline{z}_x^{\zeta_x}$$

Holding m fixed, we then look at the limiting economy in which $\underline{z}_l$, $\underline{z}_x \to 0$.

Back

Full model:
Joint CDF of shocks:

$$Z\left(z_l, z_{x1}, ... z_{xn}\right) = \left(z_l/\underline{z}_l\right)^{-\zeta_l^\rho}\left(z_{x1}/\underline{z}_{x1}\right)^{-\zeta_{x1}^\rho} ... \left(z_{xn}/\underline{z}_{xn}\right)^{-\zeta_{xn}^\rho}$$

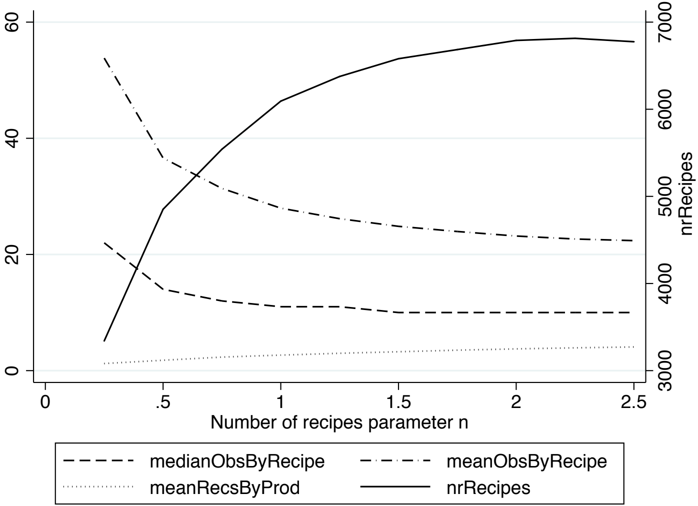
Define

$$m_\omega^\rho = M_\omega^\rho\underline{z}_l^{\zeta_l^\rho}\underline{z}_{x1}^{\zeta_{x1}^\rho} ... \underline{z}_{xn}^{\zeta_{xn}^\rho}$$

Holding m_ω^ρ fixed, we then look at the limiting economy in which $\underline{z}_l$, $\{\underline{z}_{xn}\} \to 0$

Back

Cluster statistics based on number of potential clusters per industry



Back

Wedges and Enforcement

- Two ways weak enforcement might alter shares
  1. Wasted resources
  2. Quantity restrictions

- 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

Auxiliary regressions

	(1) MatShare	(2) MatShare	(3) Sales/Cost
Age	-0.000685***		
	(0.0000410)		
log(employment)		-0.0116***	
		(0.000394)	
AvgAgeHC * Rel. Spec.			-0.0449***
			(0.0116)
5-dgt Industy FE	yes	yes	yes
State FE			yes
Observations	162083	166110	164031
R^2	0.449	0.449	0.112

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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^l. 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

The Cross-Sectional Distribution

- Let F be the CDF of efficiency in the economy (endogenous)

The Cross-Sectional Distribution

- Let F be the CDF of efficiency in the economy (endogenous)
- LLN: $F(q) = \Pr(q_j \leq q)$, depends on
  - How many techniques an entrepreneur discovers
  - Efficiency each technique delivers

Back

The Cross-Sectional Distribution

- Let F be the CDF of efficiency in the economy (endogenous)
- LLN: $F(q) = \Pr(q_j \leq q)$, depends on
  - How many techniques an entrepreneur discovers

    $$\#\text{ techniques } \sim Poisson(M)$$

  - Efficiency each technique delivers

Back

The Cross-Sectional Distribution

- Let F be the CDF of efficiency in the economy (endogenous)
- LLN: $F(q) = \Pr(q_j \leq q)$, depends on
  - How many techniques an entrepreneur discovers

$$\#\ \text{techniques} \sim Poisson\,(M)$$

  - Efficiency each technique delivers $\mathcal{C}\left(\tau_l/z_l, \tau_x/z_x q_s\right)^{-1}$
    - ⋆ Productivity of each technique: $z \sim Z(\cdot)$
    - ⋆ Efficiency of each supplier: $q_s \sim F(\cdot)$
    - ⋆ Wedges: $\tau \sim T(\cdot)$

Back

Intermediate input wedges are correlated with court quality

	(1)	(2)	(3)
	logshareRH	logshareRH	logshareRH
Avg age of civil HC cases	-0.0228***	-0.0192***	-0.0391***
	(0.000458)	(0.000435)	(0.000581)
log pop density		-0.0265***	-0.0163***
		(0.000385)	(0.000447)
log(gdpc)			-0.0592***
			(0.00120)
Recipe FE	yes	yes	yes
Observations	38430	38430	36168
R^2	0.061	0.164	0.230

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$