MIT 14.76/760: Firms, Markets, Trade and Growth
Sp 2026, Lectures 6-7: Misallocation (part I)

Dave Donaldson

Introduction

- Recall, a major goal of development is to improve some given aggregator

$$U = U(\mathbf{Q})$$

- Where each Q_i is produced by a firm using the technology

$$Q_i = F_i(\mathbf{x}_i, A_i)$$

- And we face a constraint on the aggregate (i.e. nationwide) inputs that can be used

$$\sum_i x_{im} \leqslant X_m$$

- Can imagine trying to raise U by either:
    - Improving the A_i: previous lectures
    - Making sure that the scarce X_m is allocated to its best uses: next phase of lectures

Defining Misallocation

- Denote the best allocation as $\{\mathbf{x}_i^*\}_i$ and the actual allocation as $\{\overline{\mathbf{x}}_i\}_i$
- Define

$$\overline{p}_i \frac{\partial F_i(\overline{\mathbf{x}}_i, A_i)}{\partial x_{im}} \equiv \overline{VMPX}_{im}$$

- Recall that *misallocation* is occurring at this allocation whenever

$$\exists i, m : \quad \overline{VMPX}_{im} \neq \text{constant}_m$$

- For various reasons, it is common to instead work with the following "wedge" between $\overline{VMPX}_{im}$ and prevailing price of the input, $\overline{w}_m$:

$$\overline{\mu}_{im} \equiv \frac{\overline{VMPX}_{im}}{\overline{w}_m}$$

Why might misallocation happen?

- Market failures (i.e. departures from First Welfare Theorem), such as:
  - Taxes, subsidies, regulations
  - Corruption, bribes, expropriation
  - Asymmetric information (e.g. credit constraints)
  - Incomplete contracts; missing markets
  - Market power (e.g. oligopoly, oligopsony in labor and materials markets)
  - Pure externalities (knowledge spillovers, pollution)
  - Irrational firms (don't choose profit-maximizing $\mathbf{x}_i$), e.g. due to agency problems

- Subsequent lectures will dig into some of these possibilities in detail

But how bad is misallocation?

- ...And do differences in misallocation appear large enough to account for cross-country differences in development?
- Hard to know! (Challenge is to measure $\overline{\mu}_{im}$ for all firms i and inputs m.)
- But this and next lecture will discuss attempts to do so. We will start with a stripped-down version of the hugely influential work of
  - Hsieh and Klenow (2009)
  - Restuccia and Rogerson (2008)
  - Banerjee and Duflo (2005)
- In particular, we'll start with the HK (2009) setup and restrict it to:
  - Only one sector (i.e. $\alpha_s = \alpha$ for all s))
  - Only one input into production, "labor" (i.e. set $\alpha = 0$)

Assumption about firms' technologies

- HK (2009) assume that each firm i has a CRTS production function:
  - Recall that "labor" (L_i) is only type of input (in our simplified version)
  - So, without loss, write production function as $Q_i = F_i(\mathbf{x}_i, A_i) = A_i L_i$
- Hence we have

$$\overline{\mu}_{im} \equiv \overline{\mu}_{i,L} = \frac{\overline{p}_i}{\overline{w}} \frac{\partial F_i(\overline{\mathbf{x}}_i, A_i)}{\partial L_i} = \frac{\overline{p}_i}{\overline{w}} A_i = \frac{\overline{R}_i}{\overline{w}\overline{L}_i}$$

where $R_i \equiv p_i Q_i$ is revenues

- The striking thing about this is that the (hard to measure) $\frac{\partial F_i(\overline{\mathbf{x}}_i, A_i)}{\partial L_i}$ has disappeared and is replaced by easier to measure objects (revenues, total labor expenses). How did we pull that off?!

How dispersed are these wedges?

- HK (2009) sought to answer this question using firm-level data from the formal manufacturing sectors (N = several '000) of China, India, and the US
- What HK actually do is slightly different from our presentation here
  - HK have two factors (K and L) with, effectively, a separate wedge on each (think: $\overline{\mu}_{i,L}$ and $\overline{\mu}_{i,K}$)
  - And HK assume not only that $F_i(\cdot)$ is CRTS but also Cobb-Douglas (i.e. $Q_i = A_i K_i^{\alpha} L_i^{1-\alpha}$)
  - So HK show that what matters is really a combined wedge, $TFPR_i \equiv (\overline{\mu}_{i,K})^{\alpha}(\overline{\mu}_{i,L})^{1-\alpha}$
- But HK later show that things would be very similar (quantitatively) if have only "labor" (i.e. $\alpha = 0$).
  - Remarkably, this implies that the costs of "firms using wrong K/L mix" is very small compared to costs of "some firms have too much/little of both K and L".
- So for our purposes, HK's $TFPR_i$ is very close to our $\overline{\mu}_{i,L}$

Distribution of TFPR (i.e. of $\approx \overline{\mu}_{i,L}$)

(Recentered to have mean =1 within each country. And note the log scale.)

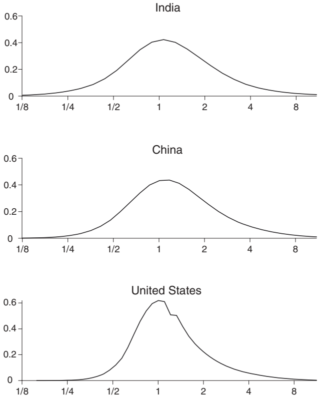


FIGURE II
Distribution of TFPR

What is this saying? Are you surprised?

The Cost of Misallocation

- We have seen evidence for striking amounts of $\overline{\mu}_{i,L}$ dispersion within these 3 countries, and considerably more dispersion in India and China than in the US
- How costly is this dispersion in $\overline{\mu}_{i,L}$ in terms of real GDP p.c.?
- HK (2009) sought to answer this too, but doing so requires stronger assumptions than what we've seen so far.
- Unsurprisingly, what we require is a sense of the demand side (i.e. preferences, $U(\cdot)$): how much does real GDP (i.e. aggregate consumer welfare) fall when the economy is producing too little of some product?

Assumption about preferences

- HK (2009) assume that utility is produced from firms' products, combined via constant elasticity of substitution (CES) aggregate

$$U(\mathbf{Q}) = \left[\sum_i Q_i^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

- We tend to think that $\sigma > 1$ is more realistic—why?

- Straightforward exercise in consumer optimization implies demand curve (for some endogenous constant a)

$$D_i = a p_i^{-\sigma}$$

The Cost of Misallocation

- Let aggregate TFP be defined by:

$$TFP(\overline{\boldsymbol{Q}}) \equiv \frac{U(\overline{\boldsymbol{Q}})}{L}$$

- Can then solve for

$$\left(\frac{TFP(\overline{\boldsymbol{Q}})}{TFP(\boldsymbol{Q}^*)}\right)^{\sigma-1} = \frac{\sum_i \left(\frac{\overline{\mu}_{i,L}}{T}\right)^{1-\sigma} A_i^{\sigma-1}}{\sum_i A_i^{\sigma-1}} \qquad (1)$$

where $T \equiv \sum_i \overline{\lambda}_i \overline{\mu}_{i,L}$ is the (weighted) average of $\overline{\mu}_{i,L}$, based on sales weights $\overline{\lambda}_i \equiv \frac{\overline{p}_i \overline{Q}_i}{\sum_k \overline{p}_k \overline{Q}_k}$

- Equation (1) has the property that $TFP(\overline{\boldsymbol{Q}}) < TFP(\boldsymbol{Q}^*)$ iff $\overline{\mu}_{i,L}$ is not equal to some common value for all firms – exactly as we expect (since only wedge dispersion can cause misallocation)

The Cost of Misallocation

- To quantify this formula we now require two things...
- First, an estimate of σ, the (own-price) elasticity of demand for any good in this economy.
  - HK (2009) use the value of $\sigma = 3$ for this.
- Second, a measure of A_i for each firm.
  - To achieve this, HK (2009) observe that by combining the production function ($Q_i = A_i L_i$) with the the demand curve ($D_i = a p_i^{-\sigma}$) we can obtain

$$A_i = \frac{a^{\frac{1}{1-\sigma}} (p_i Q_i)^{\frac{\sigma}{\sigma-1}}}{L_i}$$

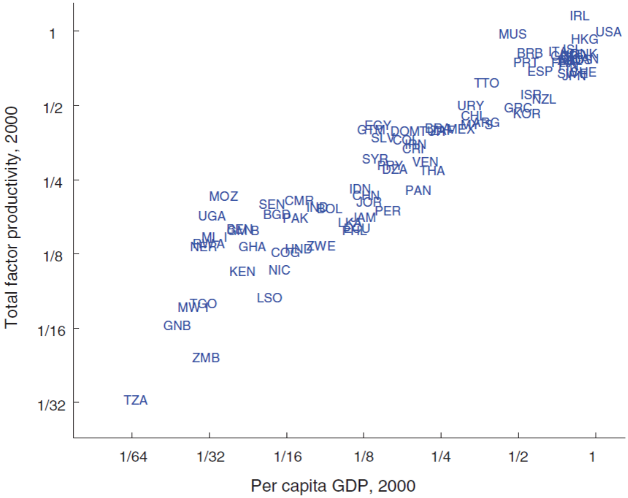
  - This can be estimated (up to the bit involving a, which cancels out of (1) anyway) from data on revenues (i.e. $p_i Q_i$) and number of workers (i.e. L_i) for each firm i

How costly are distortions? (“TFP gains" $\equiv \frac{TFP(\boldsymbol{Q}^*)}{TFP(\overline{\boldsymbol{Q}})} - 1$)

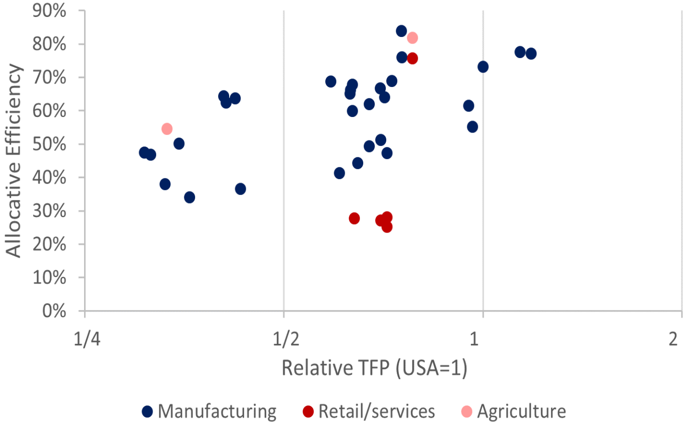
TABLE IV
TFP GAINS FROM EQUALIZING TFPR WITHIN INDUSTRIES

China	1998	2001	2005
%	115.1	95.8	86.6
India	1987	1991	1994
%	100.4	102.1	127.5
United States	1977	1987	1997
%	36.1	30.7	42.9

Can this account for the TFP differences across countries that you saw in Lecture 1?

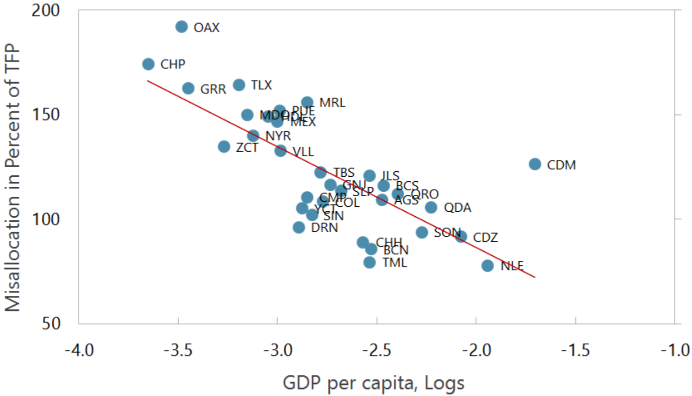


Can this account for TFP differences across countries?



(HK exercise repeated in 37 countries. "Allocative Efficiency" $\equiv \frac{TFP(\overline{\boldsymbol{Q}})}{TFP(\boldsymbol{Q}^*)}$. Source: Klenow's IGC/BREAD lecture, 2023.)

Can this account for TFP differences *within* a country?



(HK exercise repeated within each Mexican state, 2013. "Misallocation in Percent of TFP" $\equiv \frac{TFP(\boldsymbol{Q}^*)}{TFP(\overline{\boldsymbol{Q}})} - 1$. Source: Misch and Saborowski (2018).)

How Should Firm Sizes Change?

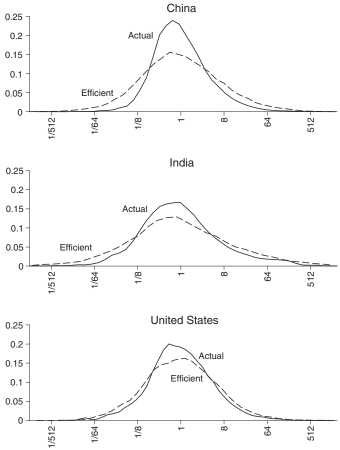


FIGURE III
Distribution of Plant Size

Most efficient firms held back, least efficient helped

Where do wedges come from?

- That is, why does $\overline{\mu}_{i,L} \neq \overline{\mu}_{j,L}$ for two firms $i \neq j$?
- A benefit of what we've done so far is that this question doesn't matter for either
  - Measuring $\overline{\mu}_{i,L}$ for each i
  - Measuring $\frac{TFP(\boldsymbol{Q}^*)}{TFP(\overline{\boldsymbol{Q}})}$
- But, still, we might hope that the inferred $\overline{\mu}_{i,L}$ are higher in places where we think that real-world policies cause distortions

Correlations with observed policies: Indian delicensing

TABLE XIII
REGRESSION OF SECTOR TFPR DISPERSION ON DELICENSING AND SIZE RESTRICTIONS IN INDIA

	(1)	(2)	(3)
Delicensed 1991	−0.298		−0.298
	(0.117)		(0.117)
Delicensed 1991 × post-1991	0.032		−0.056
	(0.036)		(0.040)
Size restriction		0.368	
		(0.173)	
Delicensed 1991 ×			0.415
post 1991 × size restriction			(0.120)

Correlations with observed policies: Indian small sector reservation

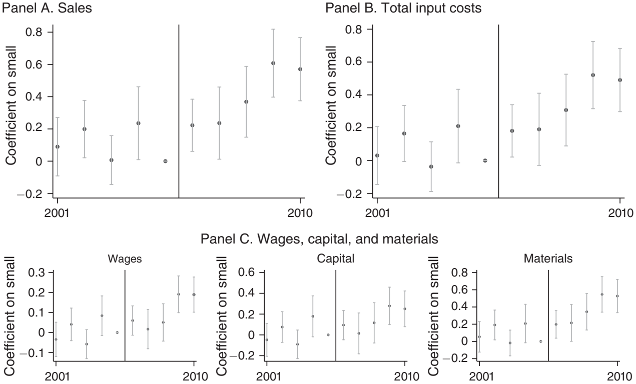
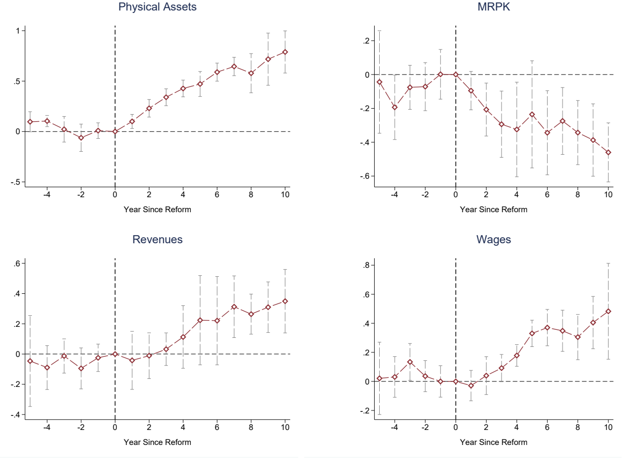


Figure 6. Event-Study Plot of Coefficients: Effect of Direct Eligibility

(Policy that subsidizes input use for qualifying small firms in certain sectors. Source: Rotemberg (2019). If measured wedge is R_i/C_i for total input costs C_i, panels A and B suggest that the policy didn't appear to cause the measured wedge to change.)

Correlations with observed policies: Indian FDI liberalization ("MRPK"=$\bar{r} \times \overline{\mu}_{i,K}$)



(Source: Bau and Matray (2023). This suggests the FDI liberalization caused $\bar{r}$ and/or $\overline{\mu}_{i,K}$ to fall.)

Aside: mapping to HK (2009)

- Our presentation here has been a stripped-down version of HK (2009). Recall, we set:
  - # sectors $s = 1$ (so compare to HK's expressions for any given s)
  - Labor is only factor (so compare to HK by setting their $\alpha_s = 1$)

- In addition, HK micro-found the wedge $\overline{\mu}_{i,L}$ by assuming that firm i behaves in following way:
  - It is monopolistically competitive in its output market: can set its own p_i but takes a (in demand curve $D_i = ap_i^{-\sigma}$) as given
  - And faces a "tax" on sales of size $\tau_{Y,i}$
  - But takes input price w as given (i.e. competitive behavior in labor market)
  - Together, this means that the firm's problem is:

$$\max_{L_i, p_i, Q_i} \; (1 - \tau_{Y,i}) p_i Q_i - wL_i \quad \text{s.t.} \quad Q_i = A_i L_i \quad \text{and} \quad Q_i = ap_i^{-\sigma}$$

  - Results in $VMPL_i/w \equiv \overline{\mu}_{i,L} = \frac{\sigma}{\sigma-1}(1 - \tau_{Y,i})^{-1}$

Potential criticisms of the HK (2009) method?

Potential criticisms of the HK (2009) method?

- Measurement error: worse in China and India? (See Rotemberg and White, 2018, and Bils, Cian and Klenow, 2020)
- Wrong production function: more wrong for China and India?
- Adjustment dynamics: in reality, we are never seeing firms in “steady-state”, as always getting hit by shocks. If such shocks are of higher variability in China/India then might mimic these results. (Asker, Collard-Wexler and de Loecker, 2014)
- Exercise takes TFPQ dispersion as given. Could that itself be distorted? (e.g. Hsieh and Klenow, 2014)
- Which policy intervention(s) could fix things?
- Whole exercise held aggregate L fixed. Rules out possibility that aggregate factor supply (e.g. capital or human capital) is also distorted.
- What about rent-seeking?
- If production features IRTS then markups/wedges are necessary for production to happen