Misallocation

Abhijit Banerjee

DSE 2016

Aggregative Growth

- The premise of neo-classical growth theory is that it is possible to do a reasonable job of explaining the broad patterns of economic change across countries, by looking at it through the lens of an aggregate production function.
- The aggregate production: $F(A, \overline{K}_P, \overline{K}_H, L)$, where $\overline{K}_P$ and $\overline{K}_H$ are the total amounts of physical and human capital invested, L is the total labor endowment of the economy and A is a technology parameter.
- The aggregate production function is not meant to be something that physically exists. It is rather a convenient construct.

An aggregation result

- Growth theorists, like everyone else, have in mind a world where production functions are associated with people: Everyone has the option of starting a firm, and when they do, they have access to an individual production function

$$Y = F(K_P, K_H, L, \theta), \tag{1}$$

where K_P and K_H are the amounts of physical and human capital invested in the firm, L is the amount of labor. θ is a productivity parameter.
- Assume that F is increasing in all its inputs. To make life simpler, assume that there is only one final good in tis economy and physical capital is made from it.
- Also assume that the population of the economy is described by a distribution function $G_t(W, \theta)$, the joint distribution of W and θ, where W is the wealth of a particular individual and θ is his productivity parameter. Let $\widetilde{G}(\theta)$ be the corresponding partial distribution on θ.

An aggregation result

- The lives of people, as often in economic models, is rather dreary: In each period, each person, given his wealth, his θ and the prices of the inputs, decides whether to set up a firm, and if so how to invest in physical and human capital.
- At the end of the period, once he gets returns from the investment and possibly other incomes, he consumes and the period ends. The consumption decision is based on maxmizing a utility function

$$\sum_{t=0}^{\infty} \delta^t U(C_t, \theta), 0 < \delta < 1. \tag{2}$$

An aggregation result

- The key assumption behind the construction of the aggregate production function is that all factor markets are perfect, in the sense that individuals can buy or sell as much they want at a given price.
- With perfect factor markets (and no risk) the market must allocate the available supply of inputs to maximize total output. We can therefore define $F(\overline{K}_P, \overline{K}_H, \overline{L}, \tilde{G}(\theta),)$ to be

$$\max_{\{K_P(\theta), K_H(\theta), L(\theta)\}} \left\{ \int_\theta F(K_P(\theta), K_H(\theta), L(\theta), \theta) d\widetilde{G}(\theta) \right\}$$

$$\begin{aligned} \text{subject to} \int_\theta K_P(\theta) d\theta &= \overline{K}_P, \\ \int_\theta K_H(\theta) d\theta &= \overline{K}_H, \\ \textit{and} \int_\theta L(\theta) d\theta &= \overline{L}. \end{aligned}$$

Interpretation of the result

- The distribution of wealth does not enter anywhere in this calculation. This reflects the fact that with perfect factor markets, there is no necessary link between what someone owns annd what gets used in the firm that he owns.
- The fact that we have written the production as $F(\overline{K}_P, \overline{K}_H, \overline{L})$, rather than a function of $\widetilde{G}(\theta)$, assumes that the distribution of productivties does not vary across countries.

Convexity?

No reason to expect a close relation between the "shape" of the individual production function and the shape of the aggregate function.

- Convexification by aggregation. Shapley-Folkman-Starr Theorem.
- Indeed in this environment where there are a continuum of firms, the (weak) concavity of the aggregate production function is guaranteed as long as the average product of the inputs in the individual production functions is bounded.
  - Proof?
  - Special case: it would be concave if the individual production functions were S-shaped in the sense of being convex to start out and then becoming concave.

Convergence

1. Since poor countries are capital scarce and the aggregate production function is concave the return on capital stock should be high.
2. $\frac{U'(c_t)}{U'(c_{t+1})}$ $=\delta r = \delta F'(K)$ where $F(K)$ is the aggregate production function.
3. Marginal utility must fall and hence consumption must grow where $\delta F'(K) > 1$.
4. Consumption must fall where $\delta F'(K) < 1$.
5. If F is concave, capital scarce economies will see faster consumption growth.
6. Convergence

Evidence on convergence

- Poorer countries do not grow faster. According to Mankiw Romer and Weil (MRW, 1992), there is no correlation between the growth rate and the initial level of Gross Domestic Product
- They converge after we control for differences in savings rates and differences in the rate of human capital accumulation **(conditional convergence)**

MRW

- An implication of there being conditional convergence is that A does not matter: can we test this directly. That is, can we explain the differences the level of GDP based on differences in capital, human capital levels?
- MRW assume a production function

$$Y = K^{\alpha} H^{\beta} (AL)^{1-\alpha-\beta}, \alpha = 0.3, \beta = 0.28$$

- They rewrite the production function

$$\begin{aligned} Y^{1-\alpha-\beta} &= (K/Y)^{\alpha} (H/Y)^{\beta} (AL)^{1-\alpha-\beta} \\ Y/L &= A (K/Y)^{\frac{\alpha}{1-\alpha-\beta}} (H/Y)^{\frac{\beta}{1-\alpha-\beta}} \end{aligned}$$

- Write this as

$$Y/L = A.X$$

MRW

- To compute K we use steady state: set

$$\frac{K}{Y} = \frac{I_K/Y}{\delta + n + g}$$

  where n is the growth rate of the labor force (L), g is the world average growth rate of output/worker ($g = 0.02$) and $\delta = 0.03$ is the depreciation rate. g captures autonomous productivity growth and $n + g$ is total steady state growth rate of Y.

- How do we compute H/Y?
  MRW use the ratio of secondary school enrollment in the total working age population to measure I_H/Y and then use the steady state condition

$$\frac{H}{Y} = \frac{I_H/Y}{\delta + n + g}$$

MRW

- How do we measure success?
- Since $\log(Y/L) = \log A + \log X$,

$$var[\log(Y/L)] = var[\log A] + var[\log X] + 2cov[\log A, \log X]$$

  The question is what to do with the covariance term?
  MRW use the measure

$$\frac{cov[\log Y/L, \log X]}{var[\log(Y/L)]} = \frac{var[\log X] + cov[\log A, \log X]}{var[\log(Y/L)]}$$

- MRW report great success for their model: 78% of all variance explained by this measure using PWT data from 1960-1985.

Klenow-Rodriguez-Clare

- Redo MRW.
- First fix the fact that the amount of time spent studying is nowhere counted in GDP. Therefore just exclude the human capital production sector. The fraction explained by X goes down to 76% (MRW1)
- Next update the data-set and restrict to countries that have enough education data. Goes back to 78% (MRW2)
- use a weighted average of primary, secondary and tertiary enrollment rates weighted by the corresponding population share to measure I_H/Y . Since the original measure of I_H/Y varies considerably more than this one (primary school enrollment varies less). The fraction goes to 40% (MRW3)
- Also using time spent in school as K_H assumes that all other resources scale up proportionally with time. However producing education is less physical capital intensive than other production. Adjusting for that makes the fraction go to 33% (MRW4)

A different approach to fitting the cross-country data: Lucas

- In a famous paper, Lucas (1990) starts from an aggregate production function $Y = AL^{1-\alpha}K^{\alpha}$, with $\alpha = 0.4$
- He assumes that both India and the US have the same production function but Indian labor has less human capital.
- He assumes that Indian labor is only 1/5th as productive as US labor, but this was based on an estimate from the 1960s. We update this estimate and assume that the difference is $3\frac{1}{3} : 1$.
- This implies that $A_u = (10/3)^{0.6} A_I \simeq 2A_I$.
- From the production function it follows that output per worker $y = Ak^{\alpha}$, where k is investment per worker in equipment.

Lucas

- Assuming that firms can borrow as much as they want at the rate r, profit maximization requires that $\alpha A k^{\alpha-1} = r$, from which it follows that the ratio of productvity in US and India, say,

$$\frac{y_U}{y_I} = \left(\frac{r_I}{r_U}\right)^{\frac{\alpha}{1-\alpha}} \left(\frac{A_U}{A_I}\right)^{\frac{1}{1-\alpha}}$$

- In 1990, $\frac{y_U}{y_I} = 11$. It follows that t $r_I \approx 5r_U$.
- The average stock market real return in the US is about 9%. This would imply a 45% rate for India.
- Lucas saw this as an obvious reason to reject the assumption that the TFP levels in the two countries are the same on the grounds that if the rates were indeed that different, capital would flow from US to India.

Technological theories

- Aghion-Howitt, Romer, Grossman Helpman argue that technology varies across countries
  - Schumpeterian approach
- Because some countries innovate
- While others have to import technologies at monopoly prices
- Moreover the technology does not transfer automatically
- And regulations may hinder technology transfer

Is technology the whole story?

- To explain the output per worker gap we need a TFP gap of a 1:2 ?
- US TFP growth rates seems to be of the order of 1-1.5% a year.
  - Even at 1.5% TFP takes about 45 years to go up by 200%.
  - Therefore in 2000, Indians would have been using machines discarded by the US in the 1950s.
- Why do not Indian firms move much closer to the US productivity levels by using machines that are (say) 20 years out of date and are presumably available at competitive prices.
- The Mckinsey Global Institute's Report on
  - Main sources of inefficiency in a range of industries in India (1999)
  - In many industries the better firms were using more or less the global best practice technologies, wherever they were economically viable.

Is technology the whole story?

- However most firms do not make use of these technologies.
  - And, according to the same Mckinsey report it is not because these technologies are not economically viable for them: The report on the apparel industry tells us that in the apparel industry

"Although machines such as the spreading machine provide major benefits to the production process and are viable even at current labour costs, they are extremeley rare in domestic (i.e. non-exporting) factories"(MGI, 2001)

- Despite this, technological backwardness is not one of the main sources of inefficiency higlighted in their report on the apparel industry.
- They emphasize that the scale of production is frequently too small,
  - the median producer is a tailors rather a firm that mass produces clothes. TFP is low not because the technology is wrong but because the firms are too small.

Is technology the whole story?

- The report says that there is some technological backwardness dairy processing industry and the telecommunications industry, but in both cases it is argued that all firms should find it profitable to upgrade along these dimensions.
- In these two cases there is however also a reference to the gains from what the report calls non-viable automation.
  - However the total productivity gain promised by what MGI calls non-viable of innovations. Both in the dairy processing industry and in the telecom case, this number is 15% or less, and in the automotive industry it is no larger.
- On balance the Schumepeterian approach is perhaps more important in highlighting what it takes for rich countries to keep growing (in particular the inter-temporal role of high skilled labor supply) than it is in explaining why the poorest countries don't grow faster

Back to Lucas: The prima facie case for sub-optimality

- There are indeed many investment opportunities in India (and in the developing world more generally) that yield 45% more.
  - Banerjee and Duflo (2004) estimated returns of over 90% for the medium sized firms in India that benefitted from a credit expansion
  - Banerjee-Duflo-Glennerster (2007) report average interest rates of 4% per month in urban slums of India.
  - According to Aleem (1989) the mean interest rate in his study area of Pakistan was 78.5%.
  - Udry and Anagol (2006) estimate returns in agriculture in Ghana: 50-250% annual.
  - De Mel, Mckenzie and Woodruff (2006) do a randomized evaluation in Sri Lanka of giving \$100-\$200 to tiny firms and find returns of 5-7% per month.
  - Cull, Mckenzie and Woodruff (2007) take 207 retail firms in Leon in the Guanajuato state in Mexico. Gave them a capital shock-\$140 in cash or equipment 1/3 of baseline capital stock
    - 20-35% returns per month

But Lucas was right on average

- On the other hand the ICOR (Incremental Capital Output Ratio) for the country as a whole measures the increase in output that came froma one unit increase in total capital stock in the past.
  - The inverse of the ICOR therefore gives an upper bound for the average marginal product for the economy.
  - For the late 90s the IMF estimates that the ICOR is over 4.5 for India and 3.7 for Uganda. The implied upper bound on the average marginal product is 22% for India and 27% in Uganda, which is not so much higher than the 9% usually assumed for the US.
- Suggests that these very high marginal rates co-exist with other much lower rates. In other words, there are mutiple marginal products and no single aggregate production function.

Returns on capital and aggregation

- The direct evidence on returns confirms this:
  - Aleem (1990) estimates the mean interest rate to be 78.5% and the standard deviation of interest rates to be 38.1%
  - Finance Corporations in India say that their interest rate on loans of less than a year vary between 48% per year and 5% per day. $(1.05)^{365} = 5.4212 \times 10^{7}\%$
  - For private bankers in India, the interest rate charged varies between 9% and 120% per annum (Timberg and Aiyer (1984))

In sum

- While there are some types of investment in human capital in developing countries that yield very high rates of return (especially in health), the average investment does not appear to be ticularly lucrative.
- Moreover, at least in education, the average rate of return seems even lower than it is in the case of physical.
- Also there is very little evidence of diminishing returns with respect to educational investment.
- Lucas was both right and wrong:

  capital flows do not eliminate large interest rate differences

  but we do need to explain why the **average** marginal product is so low in India even though the true "marginal product" is clearly very high.

Inefficiency and productivity

- Three different productivity effects:
  1. There may be across-the-board inefficiency, because everyone could have chosen the wrong technology or the wrong product mix.
  2. Capital may be misallocated across firms: there may be differences in productivity across firms because of differences in scale,
  3. Some entrepreneurs are more skilled than others, and the distribution of capital across these firms may be sub-optimal, in the sense that the most productive firms are too small.
- The discussion of technological backwardness above suggested while there is enormous variation across firms, there are usually some firms that have made all the right choices, while others have not, even though they would also benefit from doing so. For this reason we focus on the misallocation of capital across firms (i.e. explanations 2 and 3)

Why resources are misallocated?

Many theories of why resources are misallocated/misused.

1. Government/institutional failures: poor enforcement of contracts, poor property rights, protecting monopolies, favoritism, discrimination in favor of small/big firms.
2. Inadequate provision of infrastructure, public goods.
3. There is not enough redistribution of property towards those who will best make the best use of it.
4. Credit constraints
5. Poor insurance markets
6. Learning/information
7. Social norms/reputation based externalities
8. The role of the family
9. The role of behavioral issues

Quantifying the effects of misallocation

Hsieh and Klenow

- They start with a model where there is one final good produced using S intermediates

$$Y = \Pi_S Y_s^{\theta_s}$$

with

$$\sum_S \theta_s = 1$$

- Each intermediate is made with M_s further intermediates

$$Y^s = \left[\sum_{M_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

which are made from labor and capital

$$Y_{si} = A_{si} K_{si}^{\alpha_s} L^{1-\alpha_s}$$

Hsieh and Klenow

- The interpretation is that there are S industries and M_s differentiated producers within each industry.
- They need differentiated products because they assume that each firm faces a horizontal supply curve for each input.
- Introduces some diminishing returns acting through the price of the intermediate
- Two distortions
- Markup on the cost of capital τ_{Ksi}
  Tax on the price of the good τ_{Psi}

Hsieh and Klenow

- Profits are

$$\pi_{si} = (1 - \tau_{Psi}) P_{si} Y_{si} - wL_{si} - R(1 + \tau_{Ksi}) K_{si}$$

- From above the demand for intermediates is given by maximizing

$$P^s \left[\sum_{M_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} - \sum_{M_s} P_{si} Y_{si}$$

which tells us that

$$\frac{\sigma}{\sigma - 1} P^s (Y^s)^{\frac{1}{\sigma}} Y_{si}^{-\frac{1}{\sigma}} = P_{si}$$

which tells us that

$$P_{si} Y_{si} = \frac{\sigma}{\sigma - 1} P^s (Y^s)^{\frac{1}{\sigma}} Y_{si}^{\frac{\sigma-1}{\sigma}}$$

Hsieh and Klenow

- Therefore the firm maximizes

$$J_s Y_{si}^{\frac{\sigma-1}{\sigma}} - AC_{si} Y_{si}$$

where AC_{si} is the unit cost of a unit of Y_{si}

- Clearly optimality requires that

$$\begin{aligned} AC_{si} &= \frac{\sigma-1}{\sigma} J_s Y_{si}^{-\frac{1}{\sigma}} \\ &\textit{or} \\ TC_{si} &= \frac{\sigma-1}{\sigma} P_{si} Y_{si} \end{aligned}$$

Hsieh and Klenow

- From profit maximization and the Cobb-Douglas property,

$$
\begin{aligned}
MRPL_i &= \frac{w}{1-\tau_{Ysi}} = (1-\alpha)TC_{si} = \frac{\sigma-1}{\sigma}(1-\alpha)\frac{P_{si}Y_{si}}{L_{si}} \\
MRPK_i &= R\frac{1+\tau_{Ksi}}{1-\tau_{Ysi}} = \frac{\sigma-1}{\sigma}\alpha\frac{P_{si}Y_{si}}{K_{si}}
\end{aligned}
$$

where c_{sL} and c_{sK} are industry level constants

- Let us define

$$
\begin{aligned}
TFPQ_{si} &= A_{si} = \frac{Y_{si}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha}} \\
TFPR_{si} &= P_{si}A_{si} = \frac{P_{si}Y_{si}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha}} \\
&= \frac{P_{si}Y_{si}}{\left[\frac{\sigma-1}{\sigma}\alpha_s\frac{P_{si}Y_{si}(1-\tau_{Ysi})}{R(1+\tau_{Ksi})}\right]^{\alpha_s}\left[\frac{\sigma-1}{\sigma}(1-\alpha_s)\frac{P_{si}Y_{si}(1-\tau_{Ysi})}{w}\right]^{1-\alpha_s}} \\
&= C_s\frac{(1+\tau_{Ksi})^{\alpha_s}}{1-\tau_{Ysi}}
\end{aligned}
$$

where C_s is an industry level constant.

Hsieh and Klenow

- Notice when the τs are all zero $TFPR$ must be the same for all firms even though $TFPQ$ obviously differs. This is because the more productive firm faces a lower price exactly in proportion to its higher poroductivity.
- They define

$$\begin{aligned} \overline{\tau}_{Ys} &= \sum_{M_s} \tau_{Ys} \left[\frac{P_{si} Y_{si}}{P_s Y_s} \right] \\ \overline{\tau}_{Ks} &= \sum_{M_s} \tau_{Ks} \left[\frac{K_{si}}{K_s} \right] \end{aligned}$$

  where Y_s is sectoral output, K_s is sectoral capital stock and P_S is sectoral price.
- Then they can define

$$\overline{TFPR}_s = C_s \frac{(1+\overline{\tau}_{Ks})^{\alpha}}{1-\overline{\tau}_{Ys}}$$

  The extent of within industry misallocation is measured by the distribution of $\frac{TFPR_{si}}{\overline{TFPR}_s}$

Hsieh and Klenow

- Notice that this measure does not depend on what assumption we make about σ, R, w
- How do we get τ_{Ksi} and τ_{Ysi}? From

$$\begin{aligned}
\frac{w}{1-\tau_{Ysi}} &= \frac{\sigma-1}{\sigma}(1-\alpha)\frac{P_{si}Y_{si}}{L_{si}} \\
R\frac{1+\tau_{Ksi}}{1-\tau_{Ysi}} &= \frac{\sigma-1}{\sigma}\alpha\frac{P_{si}Y_{si}}{K_{si}}
\end{aligned}$$

we get that

$$\begin{aligned}
1-\tau_{Ysi} &= \frac{\sigma}{\sigma-1}\frac{wL_{si}}{(1-\alpha_s)P_{si}Y_{si}} \\
& \textit{and} \\
1+\tau_{Ksi} &= \frac{\alpha_s}{1-\alpha_s}\frac{wL_{si}}{RY_{Ksi}}
\end{aligned}$$

Hsieh and Klenow

- To generate these numbers they assume
  α comes from the US labor share for that industry
  σ comes from US IO literature. They pick $\sigma = 3$, which is at the bottom end of the distribution of what peeople find because the less substitutable the firms, the more cost from misallocation.
  Firm level data to generate wL_{si} and $P_{si}Y_{si}$
  R is set at 0.1(5% real interest + 5% depreciation). Note that the level of R does not affect the distribution of $\frac{TFPR_{si}}{\overline{TFPR_s}}$.

Hsieh and Klenow

- Uses 4-digit classification of manufacturing industries:
  ASI for India: a census of all firms with more than 100 workers and a 1/3 sample of all those between 20 and 100. Period: 1987-88 to1994-95
  Annual Surveys of Industrial Production for China. The Chinese data leaves out all firms with less than $600000 in revenue but is census of the rest. Data for 1998-2005
  US Census of Manufactures: 1977, 1987, 1997
- India has more misallocation than China which has more misallocation than the US (Figure 2, Table 3)
- How costly is this misallocation?

TABLE II
DISPERSION OF TFPR

China	1998	2001	2005
S.D.	0.74	0.68	0.63
75 – 25	0.97	0.88	0.82
90 – 10	1.87	1.71	1.59
India	**1987**	**1991**	**1994**
S.D.	0.69	0.67	0.67
75 – 25	0.79	0.81	0.81
90 – 10	1.73	1.64	1.60
United States	**1977**	**1987**	**1997**
S.D.	0.45	0.41	0.49
75 – 25	0.46	0.41	0.53
90 – 10	1.04	1.01	1.19

Notes. For plant i in industry s, $\text{TFPR}_{si} \equiv \frac{P_{si} Y_{si}}{K_{si}^{\alpha_s} (w_{si} L_{si})^{1-\alpha_s}}$. Statistics are for deviations of log(TFPR) from industry means. S.D. = standard deviation, 75 − 25 is the difference between the 75th and 25th percentiles, and 90 − 10 the 90th vs. 10th percentiles. Industries are weighted by their value-added shares. Number of plants is the same as in Table I.

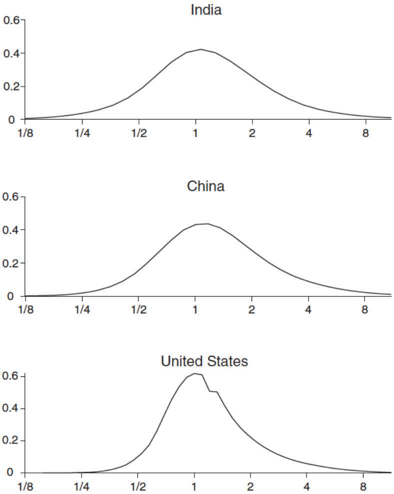

FIGURE II
Distribution of TFPR

Hsieh and Klenow

Little algebra leads to the following expression for GDP

$$Y = \Pi_S(TFP_s . K_s^{\alpha_s} L_s^{1-\alpha_s})$$

where

$$TFP_s = (\frac{1}{M_s}\sum_{M_s}\{A_{si} . \frac{\overline{TFPR_s}}{TFPR_{si}}\}^{\sigma-1})^{\frac{1}{\sigma-1}}$$

- Two things to be noted.
  1. If $\sigma > 2$, more variance of A_{si} increases TFP.
  2. Positive correlation between A_{si} and $\frac{TFPR_{si}}{\overline{TFPR_s}}$ is bad bews

Hsieh and Klenow

- How do we get A_{si}?
- Should be easy but we do not have Y_{si}.
- Use the fact that

$$P_{si}Y_{si} = \frac{\sigma}{\sigma-1}P^s(Y^s)^{\frac{1}{\sigma}}Y_{si}^{\frac{\sigma-1}{\sigma}}$$

to write

$$[\frac{\sigma}{\sigma-1}P^s(Y^s)^{\frac{1}{\sigma}}]^{\frac{1-\sigma}{\sigma}}(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}} = Y_{si}$$

- Hence

$$A_{si} = w^{1-\alpha_s}\kappa_s\frac{(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}}}{K_{si}^{\alpha_s}(w_{si}L_{si}^{1-\alpha_s})}$$

where w_{si} is the firm wage level, which may be higher or lower than the average wage because of human capital differences/longer hours.

Hsieh and Klenow

- We can always ignore the level of $w^{1-\alpha_s}\kappa_s$ for any intra-industry reallocation exercise: It is a constant that does not affect the variance or covariances of $\ln A_{si}$.
- $\ln A_{si}$ is highly positively correlated with $\ln TFPR_{si}$ and much more so in India and China than in the US (Table 4)
- Table 5 shows in China this is partly driven by state firms which low TFPR and low TFPQ.
- Full equalization of TFPR within sectors will increase TFP in India by 125%, in China by 90% and in US by 40%.
- Moving to the US joint distribution of TFPR and TFPQ would increase TFP in India by 45-50% and that in China by 30-45%

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

Notes. Entries are $100(Y_{\text{efficient}}/Y - 1)$ where $Y/Y_{\text{efficient}} = \prod_{s=1}^{S}\left[\sum_{i=1}^{M_s}\left(\frac{A_{si}}{A_s}\frac{\overline{\text{TFPR}}_s}{\text{TFPR}_{si}}\right)^{\sigma-1}\right]^{\theta_s/(\sigma-1)}$ and $\text{TFPR}_{si} \equiv \frac{P_{si}Y_{si}}{K_{si}^{\alpha_s}(w_{si}L_{si})^{1-\alpha_s}}$.

TABLE VI
TFP Gains from Equalizing TFPR Relative to 1997 U.S. Gains

China	1998	2001	2005
%	50.5	37.0	30.5
India	1987	1991	1994
%	40.2	41.4	59.2

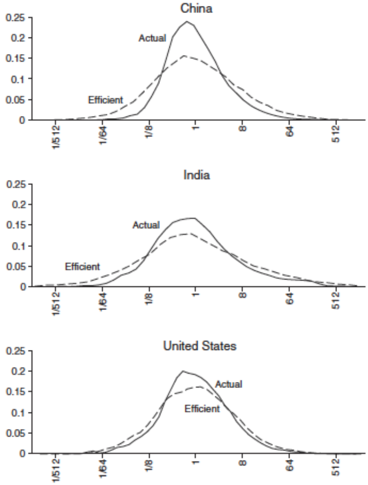

FIGURE III
Distribution of Plant Size

Hsieh and Klenow

- Is this a reflection of transition: strategic under-investment by new firms?
- In India the TFPR is clearly lowest for the youngest firms–the oldest firms are the most distorted. No relation in China, in the US the smallest firms have the highest TFPR ratio
- Is it because of learning by doing: is it that the low TFPR firms are actually investing in learning that they look different from the rest (this would explain why the youngest firms have low TFPR)
  TFPQ growth rates are higher for high TFPR firms in India and China, unlike in the US
- What is going on?

Hsieh and Klenow

- In India the smallest firms have the lowest TFPR ratio (Figure 5). In China similar pattern (though weaker). No pattern in the US:
- In both India and China a lot of small firms need to shrink (Table 8)
- Small firms have low productivity and therefore not much potential for significant underinvestment. Large firms are more productive and this where there is underinvestment.

Hsieh and Klenow

- Ownership effects as expected (table VII); Lower TFP firms exit (table VIII)
- Inputs and revenue are more or less correlated the same way in all 3 countries
- IV of current inputs by lagged valures reduces gains from equalizing TFPR levels from 87% to 72% in china, 127% to 108% in India and from 43% to 26% in the US
- Hsieh and Klenow take a baseline value of $\sigma = 3$, which implies that the share of profits is a 1/3. This is high by any standards—the number we saw before is 0.2, and the span of control models use 0.15 to 0.2 as the exponent on the entrepreneurial ability.

Does Misallocation persist: Transition dynamics

Banerjee Moll

- Individual production functions are assumed to be identical and a function of capital alone ($F(K)$) but otherwise quite general. In particular, we do not assume that they are concave.
- CRRA preferences : $U(C_t) = \frac{C_t^{1-\theta}}{1-\theta}$
- People are forward-looking and at each point of time they choose consumption and savings to maximize life-time utility.
- However the maximum amount they can borrow is linear and increasing in their wealth and decreasing in the current interest rate (leverage ratio $\lambda(r_t), \lambda'(r_t) < 0$).
- The budget constraint

$$W_{t+1} = \max_{K_t}\{F(K_t) - r_t K_t + r_t(W_t - C_t), s.t. \lambda(r_t)(W_t - C_t) \geq K_t$$

BM

- First order condition for savings

$$\frac{U'(C_t)}{U'(C_{t+1})} = \left[\frac{C_{t+1}}{C_t}\right]^{\phi} = \delta r_t$$

for net lenders and borrowers who are not credit constrained

- First order condition for savings

$$\frac{U'(C_{t+1})}{U'(C_t)} = \left[\frac{C_{t+1}}{C_t}\right]^{\phi} > \delta r_t$$

for borrowers who are credit constrained.

- Credit comes from other members of the same economy and the interest rate clears the credit market.
- We do not assume that everyone starts with the same wealth, but rather that at each point of time there is a distribution of wealth which evolves over time.

BM: Constant/diminishing returns

- With constant returns in production, we get the first best: inequality remains unchanged over time, production and investment is always efficient.
- With diminishing returns it is the poor who are credit constrained.
- This is because if a rich person is willing to borrow b dollars more at interest rate r, a poor person would also be willing, because of diminishing returns.
- Hence inequality falls over time and in the long run no one is credit constrained, although we do not necessarily get full wealth convergence. The long run interest rate converges to its first best level, and hence investment is efficient.

BM: increasing returns

- With increasing returns, it is the rich who are credit constrained.
- Hence inequality increases over time; we converge to a Gini coefficient of 1.
- Wealth becomes more and more concentrated with only the richest borrowing and investing. Because there is increasing returns this is also the first best outcome.

BM S-shaped production functions

- S-shaped production functions:

$$\begin{aligned} F(K_t) &= F_1(K_t), K_t < \overline{K}, F_1' > 0, F_1'' < 0 \\ &= AF_1(K_t - \overline{K}), K_t \geq \overline{K}, A > 1 \\ \textit{where } F_1(\overline{K}) &= AF_1(\overline{K} - K_1) \end{aligned}$$

- Consider the maximization problem of a decision-maker in this world who has wealth level W_t and maximizes $\sum_{t=0}^{\infty} \delta^t U(C_t)$ subject to the constraint

$$W_{t+1} = \max_{K_t} \{F(K_t) - rK_t + r(W_t - C_t), s.t. \lambda(r_t)(W_t - C_t) \geq K_t$$

- Also assume that $\delta r_t \leq 1$. Otherwise there is no steady state.
- The steady state interest rate must therefore be no more than $\frac{1}{\delta}$.

BM: S-shaped production functions

- Assume $\lambda(\frac{1}{\delta}) > 1$. That is at the steady interest rate there is demand for credit: otherwise the interest rate is indeterminate
- Then it must be the case that there are some lenders in equilibrium. Hence it must be that $\delta r = 1$ at the steady state (otherwise there would be no lenders left).
- Now in any steady state it must be that the borrowers are also not accumulating capital. For them it must be that $F'(K) = \frac{1}{\delta} = r$.
- No one is credit constrained.

BM: S-Shaped production functions

- There are two possible steady state levels of investment.
  - The solution to the steady state condition assuming that $F_1(K_t)$ is the only production function

    $$1 = \delta F_1'(K^{1*})$$

    For this to be a steady state a necessary condition is that $K^{1*} < \overline{K}$.
  - The solution to the steady state condition assuming that $AF_1(K - \overline{K})$ is the only production function

    $$1 = \delta A F_1'(K^{2*} - \overline{K})$$

    For this to be a steady state a necessary condition is that $K^{2*} > \overline{K}$.

BM: S-shaped production functions

These conditions are however not sufficient. To see consider the case where the lower steady state level of wealth is W^1. And let

$$\lambda(\frac{1}{\delta})W^1 > \overline{K}.$$

- Then the person at W^1 faces two choices:
  1. one is to save extra and try to reach $\overline{K}$ (in one or more periods).
  2. The other is to ignore the lure of increased productivity for higher current consumption, in which case the steady state is a true steady state.
  - Which one he chooses depends on;
  - The level of $\overline{K}$.; How far $\overline{K}$ is from ; size of $\lambda(\frac{1}{\delta})$;level of δ

BM S-shaped production functions

- Likewise there are addition condition for investing at K^{2*} to be a steady state outcome
- Suppose there is steady state where λ proportion of the population is investing at K^{1*} and the rest are investing at K^{2*}
- For this to happen people have to be at wealth W^1 and W^2 such that

$$\begin{aligned} F(K^{1*}) + \frac{1}{\delta}(W^1 - C^1 - K^{1*}) &= W^1 \\ F(K^{2*}) + \frac{1}{\delta}(W^2 - C^2 - K^{2*}) &= W^2 \\ \lambda(W^1 - C^1 - K^{1*}) + (1-\lambda)(W^2 - C^2 - K^{2*}) &= 0 \end{aligned}$$

  where the last equation comes from credit market clearing and the first two come from no wealth growth.
- Notice that these are 3 equations in 5 unknowns, $\lambda, W^1, W^2, C^1, C^2$.

BM S-shaped production functions

- A continuum of steady states with varying levels of per capita output.
- The poor stay poor (Dasgupta-Ray (1986), Galor-Zeira (1993)), Buera (2008)
- Per capita determined by initial conditions
- Unlike in the version of this model without credit constraints, where also there are many steady state wealth levels, but per capita output is constant
- Extensive margin misallocation (unlike in the decreasing or increasing returns economy.
- But no credit constraints in steady state and no intensive margin misallocation
- Fast convergence to this steady state

What explains persistent misallocation?

- Non-neoclassical savings functions?
- Preference for a quiet life?
- People do not know how productive they are?