Misallocation or Mismeasurement?

Mark Bils, University of Rochester and NBER
Pete Klenow, Stanford University and NBER
Cian Ruane, Stanford University

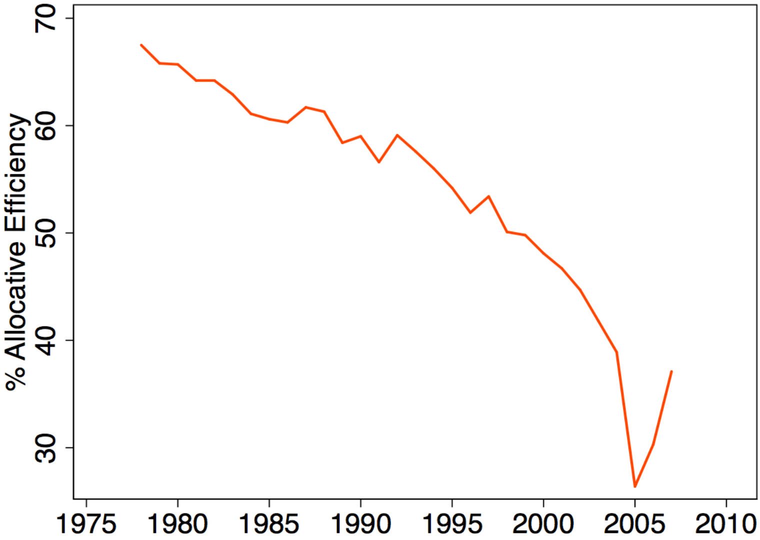
NBER Summer Institute CRIW and Macro-Productivity

July 2017

Motivation

- Large gaps in average revenue products (TFPR) across plants
  - Syverson (2011)
- Huge purported gains from reallocation of inputs
  - Banerjee & Duflo (2005)
  - Restuccia & Rogerson (2008)
  - Hsieh & Klenow (2009, 2014)
- But differences in measured average products need not reflect differences in true marginal products

U.S. allocative efficiency



What we do

- Propose a way to estimate dispersion in marginal products under:
  - Measurement error
  - Misspecification due to overhead costs
- Apply to:
  - manufacturing plants in the U.S. 1978–2007
  - manufacturing plants in India 1985–2011

Others skeptical of misallocation

- Adjustment costs
  - Asker, Collard-Wexler & De Loecker (2014)
  - Kehrig & Vincent (2016)
- Overhead costs
  - Bartelsman, Haltiwanger & Scarpetta (2013)
  - Haltiwanger, Kulick & Syverson (2016)
- Variable production elasticities
  - Song & Wu (2015)
- Imputation errors
  - White, Petrin & Reiter (2016), Rotemberg & White (2017)

Simple model for intuition

- $Y = \left(\sum_i Y_i^{1-\frac{1}{\epsilon}}\right)^{\frac{1}{1-\frac{1}{\epsilon}}}, P = \left(\sum_i P_i^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}}$
- $Y_i = A_i L_i$
- $\max\ (1 - \tau_i^Y) P_i Y_i - w L_i$
  - Monopolistic competitor takes w, Y, and P as given
- $\widehat{P_i Y_i} \equiv P_i Y_i + g_i$

Simple model TFPR

- P_i = markup $\times$ marginal cost
- $P_i = \left(\dfrac{\epsilon}{\epsilon - 1}\right) \times \left(\tau_i \cdot \dfrac{w}{A_i}\right)$, where $\tau_i \equiv \dfrac{1}{1 - \tau_i^Y}$
- $P_i Y_i \;\propto\; \tau_i \cdot L_i$
- $TFPR_i \;\equiv\; \dfrac{\widehat{P_i Y_i}}{L_i} \;\propto\; \tau_i \cdot \dfrac{P_i Y_i + g_i}{P_i Y_i}$

Numerical example

- τ_i — so the true distortion is fixed over time
- g_i — so additive measurement error is fixed over time
- A_{it} — so productivity is time-varying

	PY	L	$\frac{PY}{L}$	$\widehat{PY}$	$\widehat{\frac{PY}{L}}$	$\blacktriangle PY$	$\blacktriangle L$	$\frac{\blacktriangle PY}{\blacktriangle L}$
Plant 1	100	50	2	120	2.4	50	25	2
Plant 2	50	50	1	40	0.8	25	25	1

Lessons from the numerical example

- $\blacktriangle \widehat{P_{it} Y_{it}} / \blacktriangle L_{it} = \tau_i$ when constant measurement error, distortions

- Regressing $\ln\left(\blacktriangle \widehat{P_{it} Y_{it}} / \blacktriangle L_{it}\right)$ on $\ln(\text{TFPR})$ yields:
  - 1 if there is no measurement error in TFPR
  - 0 if all TFPR dispersion is due to measurement error
  - $\sim$ 2/3 in the numerical example above

- Want to generalize this way of estimating $\dfrac{\sigma^2_{\ln \tau}}{\sigma^2_{\ln(TFPR)}}$

Full Model (with no Measurement Error)

- Closed economy, S sectors, N_s firms, L workers, K capital
- $Q = \prod_{s=1}^{S} Q_s^{\theta_s}$
- $Q_s = \left(\sum_i^{N_s} Q_{si}^{1-\frac{1}{\epsilon}} \right)^{\frac{1}{1-\frac{1}{\epsilon}}}$
- $Q_{si} = A_{si} (K_{si}^{\alpha_s} L_{si}^{1-\alpha_s})^{\gamma_s} X_{si}^{1-\gamma_s}$
- $\max\ R_{si} - (1+\tau_{si}^L) w L_{si} - (1+\tau_{si}^K) r K_{si} - (1+\tau_{si}^X) X_{si}$

Model (Sectoral TFP Decomposition)

$$TFP = \underbrace{\left[\frac{1}{N}\sum_{i}^{N}\left(\frac{A_i}{\tilde{A}}\right)^{\epsilon-1}\left(\frac{\tau_i}{\tau}\right)^{1-\epsilon}\right]^{\frac{1}{\epsilon-1}}}_{AE=\text{Allocative Efficiency}} \times \underbrace{\left[\frac{1}{N}\sum_{i}^{N}\left(\frac{A_i}{\bar{A}}\right)^{\epsilon-1}\right]^{\frac{1}{\epsilon-1}}}_{PD=\text{Productivity Dispersion}}$$

$$\times \underbrace{N^{\frac{1}{\epsilon-1}}}_{\text{Variety}} \times \underbrace{\bar{A}}_{\text{Average Productivity}}$$

- $\tilde{A} = \left[\frac{1}{N}\sum_i^N (A_i)^{\epsilon-1}\right]^{\frac{1}{\epsilon-1}}$ (power mean)
- $\bar{A} = \prod_{i=1}^N A_i^{\frac{1}{N}}$ (geometric mean)

Indian *Annual Survey of Industries* (ASI)

- Survey of Indian manufacturing plants
  - Long panel 1985–2011
  - All plants $>$ 100 or 200 workers (45% of plant-years)
  - Probabilistic if $>$ 10 or 20 workers (55% of plant-years)
  - $\sim$ 43,000 plants per year

U.S. *Longitudinal Research Database* (LRD)

- U.S. Census Bureau data on manufacturing plants
  - Long panel, 1978–2007 analyzed so far
  - Annual Survey of Manufacturing (ASM) plants
  - $\sim$ 50k plants per year with at least one employee
  - Probabilistic sampling for $\sim$ 34k plants, certainty for other $\sim$ 16k

Inferring aggregate AE from the data

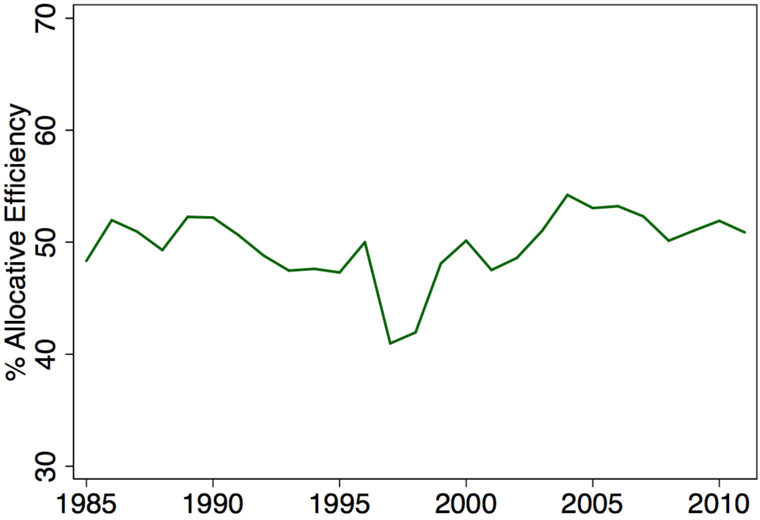
Aggregating within-sector allocative efficiencies:

$$\widehat{AE}_t = \prod_{s=1}^{S} \widehat{AE}_{st}^{\frac{\theta_{st}}{\sum_{s=1}^{S} \gamma_s \theta_{st}}}$$

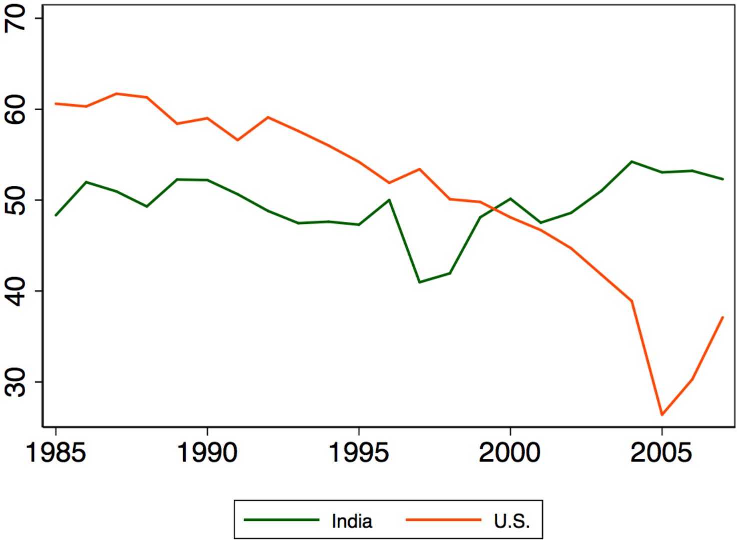
Parameterization:

- $\epsilon = 4$ based on Redding and Weinstein (2016)
- α_s and γ_s inferred from sectoral cost-shares ($r = .2$)
- θ_{st} inferred from sectoral shares of aggregate output

Indian allocative efficiency (49% on average)



India vs. U.S. allocative efficiency



Measurement error in revenue and inputs

$$\widehat{I}_i \equiv \phi_i \cdot I_i + f_i$$

$$\widehat{R}_i \equiv \chi_i \cdot R_i + g_i$$

- I_i and R_i = true inputs and revenues
- $\widehat{I}_i$ and $\widehat{R}_i$ = measured inputs and revenues

$$TFPR_i \equiv \frac{\widehat{R}_i}{\widehat{I}_i} \propto \left(\frac{\epsilon}{\epsilon - 1}\right) \tau_i \left(\frac{\widehat{R}_i}{R_i} \, \frac{I_i}{\widehat{I}_i}\right)$$

Measurement error and $\Delta TFPR$

$$\Delta TFPR_i = \Delta\tau_i + \Delta\left(\frac{\widehat{R}_i}{R_i}\right) - \Delta\left(\frac{\widehat{I}_i}{I_i}\right)$$

If only *additive* measurement error:

$$\Delta TFPR_i = \frac{\Delta\tau_i}{\widehat{R}_i/R_i} - \left(\frac{\widehat{R}_i - R_i}{\widehat{R}_i} - \frac{\widehat{I}_i - I_i}{\widehat{I}_i}\right)\Delta I_i$$

$$+ \frac{\blacktriangle g_i}{\widehat{R}_i} - \frac{\blacktriangle f_i}{\widehat{I}_i}$$

Previewing our baseline specification

We will regress revenue growth on input growth for a panel of plants:

$$\Delta \widehat{R}_i = a_0 + a_1 \Delta \widehat{I}_i + a_2 \ln(TFPR_i) + a_3 \ln(TFPR_i) \cdot \Delta \widehat{I}_i + e_i$$

Additive measurement error will show up as $a_3 < 0$

Can also include higher order terms in $\ln(TFPR_i)$

Measurement error: key assumptions/limitations

- We focus on additive measurement error
  - Conservative, as multiplicative also overstates TFPR differences
- Impose τ_i and $\left(\widehat{R}_i/R_i \cdot I_i/\widehat{I}_i\right)$ each lognormally distributed to estimate expected $\ln(\tau_i)$ given $\ln(TFPR_i)$
- To correct misallocation need covariance of $\ln(\tau_i)$ and $\ln\left(\widehat{R}_i/R_i \cdot I_i/\widehat{I}_i\right)$. Set to zero. Will not hold if f_i and g_i not orthogonal to $\ln \tau_i$, or not mean zero.

Compression factor λ

λ can be used to estimate the variance of true distortions τ_i:

$$E\left\{\ln \tau_i \mid \ln(TFPR_i)\right\} = \lambda \cdot \ln(TFPR_i)$$

$$\sigma^2_{\ln\tau} = \lambda \cdot \sigma^2_{\ln(TFPR)} - \sigma_{\ln\tau, \frac{\widehat{R}_i I_i}{R_i \widehat{I}_i}}$$

If $\sigma_{\ln\tau, \frac{\widehat{R}_i I_i}{R_i \widehat{I}_i}} = 0$, then $\lambda = \dfrac{\sigma^2_{\ln\tau}}{\sigma^2_{\ln(TFPR)}}$

Baseline specification

$$\begin{aligned}\Delta\widehat{R}_{it} &= \Phi \cdot \ln(TFPR_{it}) + \Psi \cdot \Delta\widehat{I}_{it} \\ &\quad -(1-\lambda)\cdot\Psi\cdot\ln(TFPR_{it})\cdot\Delta\widehat{I}_{it} + D_t + \xi_{it}\end{aligned}$$

- $\lambda = \dfrac{\sigma^2_{\ln\tau}}{\sigma^2_{\ln(TFPR)}}$
- $\Psi = 1 + \Omega_\tau - \Omega_f$
- D_t = sector-year fixed effects

Baseline estimates pooling all years

	India 1985–2011	U.S. 1978–2007
$\widehat{\Phi}$	0.052 (0.005)	0.053 (0.002)
$\widehat{\Psi}$	0.967 (0.005)	0.794 (0.004)
$\widehat{\lambda}$	**0.547** **(0.035)**	**0.229** **(0.026)**
Observations	277,239	1,141,000

The dependent variable is revenue growth. $\widehat{\Phi}$ is the coefficient on *TFPR*, $\widehat{\Psi}$ on composite input growth, and $1-\widehat{\lambda}$ on the product of the two. Standard errors are clustered.

Baseline estimates in windows

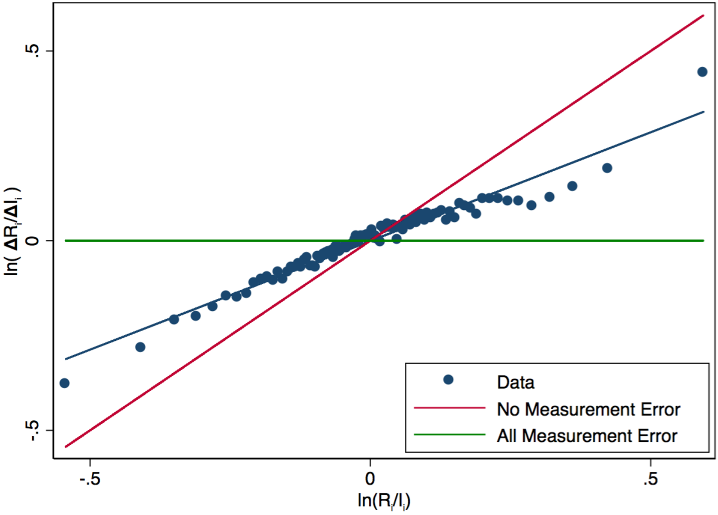
India

	1985–1993	1994–2001	2002–2011
$\widehat{\lambda}$	**0.562** **(0.050)**	**0.510** **(0.080)**	**0.576** **(0.027)**

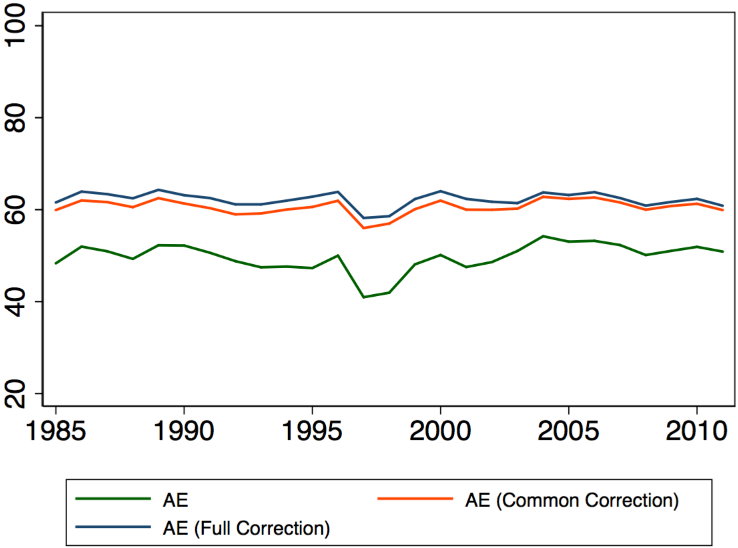
U.S.

	1978–1982	1983–1987	1988–1992	1993–1997	1998–2002	2003–2007
$\widehat{\lambda}$	**0.358** **(0.027)**	**0.336** **(0.034)**	**0.326** **(0.031)**	**0.326** **(0.037)**	**0.192** **(0.032)**	**0.095** **(0.070)**
var(TFPR)	.064	.076	.080	.090	.105	.135

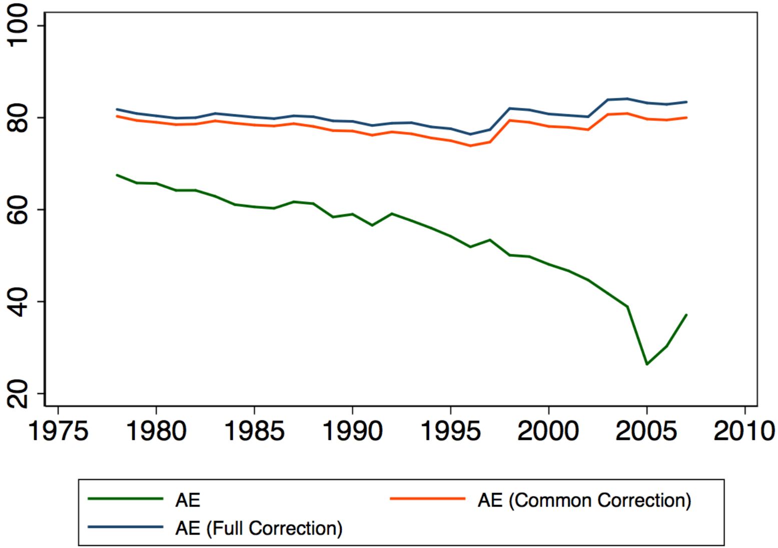
Indian first differences vs. levels ($\blacktriangle \hat{R}/\blacktriangle \hat{I}$ vs. R/I)



Adjusted allocative efficiency in India ($\hat{\lambda}$ in windows)



Adjusted allocative efficiency in U.S. ($\hat{\lambda}$ in windows)



Allocative efficiency: U.S. relative to India

Conclusion

- Propose way to estimate truer dispersion in marginal products
- For Indian ASI:
  - Marginal products are $\frac{1}{2}$ as dispersed as average products
  - Potential gains from reallocation reduced by $\frac{2}{5}$
  - Time-series volatility reduced by $\frac{2}{3}$
- For U.S. ASM:
  - Eliminates sharp downward trend in allocative efficiency
  - Leaves U.S. allocative efficiency higher than in India

Next steps

- Why did measurement error get worse in the U.S.?
- Should the Census have noticed it?
  - Variance of ln R and ln I rose 13.3% and 10.9% from 1978–2007
  - Correlation with each other fell from 0.993 to 0.979
  - 25–58 times higher variance for ln R and ln I than for ln TFPR