Misallocation or Mismeasurement?

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

April 2017

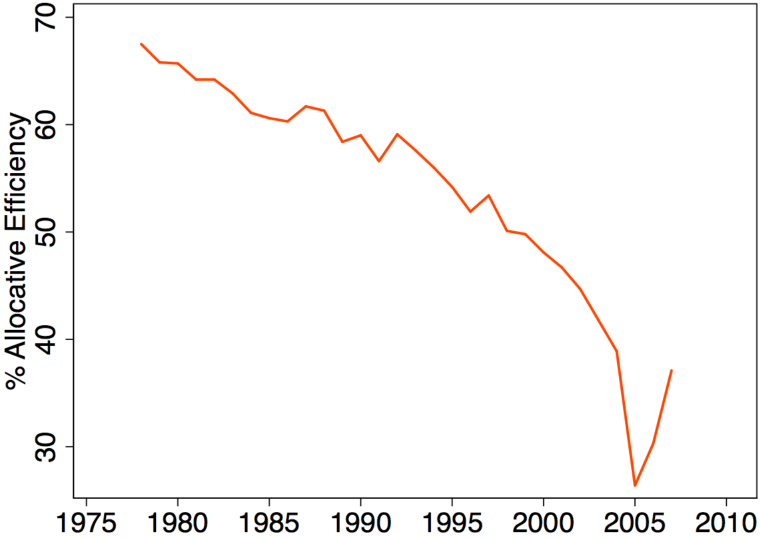
Motivation

- 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. manufacturing since the 1970s

- Major increase in TFPR dispersion (Kehrig, 2015)
  - Implies falling allocative efficiency
  - If true, lowered TFP growth by about 2.5 percent per year
  - Cumulates to 55 percent lower TFP by late 2000s
  - Given measured TFP growth was 1.7 percent per year, would imply residual TFP growth of 4.2 percent per year
- Real, or measurement error getting worse?

U.S. allocative efficiency



What we do

- Propose way to estimate 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
- Preview of results:
  - Reduces potential gains from reallocation in the U.S. by 70%
  - Eliminates the severe decline in U.S. allocative efficiency
  - Reduces potential gains from reallocation in India by 40%
  - Leaves U.S. at 35% higher allocative efficiency than India

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)

Measurement error in the Indian data

- Book values instead of market values for capital
- Unreported number of contract workers
- End-of-previous-year $\neq$ beginning-of-current-year stocks
- Jumps in reported age across years for many plants

Measurement error in the U.S. data

- Book values instead of market values for capital
- Census is frequently forced to impute data
  - SSA, IRS data on a subset of plants, variables
  - Sometimes impute based on other plants
  - See White, Reiter and Petrin (2016) for a critique
  - Nishida, Petrin, Rotemberg and White (2017, in progress)

Table of Contents

Simple model setup

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

Numerical example

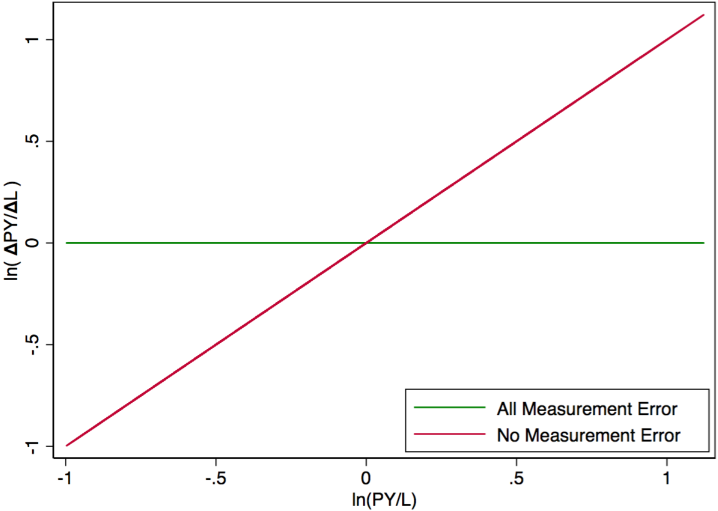
- τ_i — so the true distortion is fixed over time
- g_i — so additive measurement error is fixed over time
- A_{it} — so productivity is time-varying

	PY	L	$\frac{PY}{L}$	$\widehat{PY}$	$\widehat{\frac{PY}{L}}$	$\blacktriangle PY$	$\blacktriangle L$	$\frac{\blacktriangle PY}{\blacktriangle L}$
Firm 1	100	50	2	120	2.4	50	25	2
Firm 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
- Later we generalize in order to:
  - Allow measurement error in inputs
  - Allow shocks to measurement error and to distortions
  - Infer the signal from covariance b/w levels and first differences

Projection of first differences on levels



Full model vs. simple model

- Capital, labor, and intermediates
- Distortions hitting each input
- Multiple sectors
- Shocks to τ and shocks to measurement error
- Assumptions: additive measurement error orthogonal to τ, A

Table of Contents

Model (Setup)

- Closed economy, S sectors, N_s firms, L workers, K capital
- $Q = C + X, \quad X = \sum_s^S \sum_i^{N_s} X_{si}$
- $Q = \prod_{s=1}^S Q_s^{\theta_s}$
- $Q_s = \left(\sum_i^{N_s} Q_{si}^{1-\frac{1}{\epsilon}} \right)^{\frac{1}{1-\frac{1}{\epsilon}}}$
- $Q_{si} = A_{si} (K_{si}^{\alpha_s} L_{si}^{1-\alpha_s})^{\gamma_s} X_{si}^{1-\gamma_s}$
- max $R_{si} - (1 + \tau_{si}^L) w L_{si} - (1 + \tau_{si}^K) r K_{si} - (1 + \tau_{si}^X) X_{si}$
  - $R_{si} \equiv P_{si} Q_{si}$
  - Monopolistic competitor takes input prices as given

Model (Aggregate TFP)

- $TFP \equiv \dfrac{C}{L^{1-\tilde{\alpha}} K^{\tilde{\alpha}}}$
  - where $\tilde{\alpha} \equiv \dfrac{\sum_{s=1}^{S} \alpha_s \gamma_s \theta_s}{\sum_{s=1}^{S} \gamma_s \theta_s}$

- $TFP = \overline{T} \times \prod_{s=1}^{S} TFP_s^{\frac{\theta_s}{\sum_{s=1}^{S} \gamma_s \theta_s}}$
  - $\overline{T}$ = reflects sectoral distortions (set aside)
  - $TFP_s \equiv \dfrac{Q_s}{(K_s^{\alpha_s} L_s^{1-\alpha_s})^{\gamma_s} X_s^{1-\gamma_s}}$

Model (Sectoral TFP)

Suppressing s here and whenever possible:

$$TFP = \left[\sum_{i}^{N} A_i^{\epsilon-1} \left(\frac{\tau_i}{\tau} \right)^{1-\epsilon} \right]^{\frac{1}{\epsilon-1}}$$

- $\tau_i \equiv \left[\left(1+\tau_i^L\right)^{1-\alpha} \left(1+\tau_i^K\right)^{\alpha} \right]^{\gamma} \left(1+\tau_i^X\right)^{1-\gamma}$
- $\tau \equiv \left[\left(1+\tau^L\right)^{1-\alpha} \left(1+\tau^K\right)^{\alpha} \right]^{\gamma} \left(1+\tau^X\right)^{1-\gamma}$
- where $1+\tau^L \equiv \left[\sum_{i=1}^{N} \frac{R_i}{R} \frac{1}{1+\tau_i^L} \right]^{-1}$ and so on

Model (Sectoral TFP Decomposition)

$$TFP = AE \cdot PD \cdot \bar{A} \cdot N^{\frac{1}{\epsilon-1}}$$

- $AE \equiv$ Allocative Efficiency
- $PD \equiv$ Productivity Dispersion
- $\bar{A} \equiv$ Average productivity
- $N^{\frac{1}{\epsilon-1}} \equiv$ Variety

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)

Table of Contents

Indian *Annual Survey of Industries* (ASI)

- Survey of Indian manufacturing plants
  - Long panel 1985–2011
  - Used in Hsieh and Klenow (2009, 2014)
- Sampling frame
  - 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
- Variables used
  - Gross output (R_i), intermediate inputs (X_i), labor (L_i), wage bill (wL_i), and capital (K_i)

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

- U.S. Census Bureau data on manufacturing plants
  - Long panel, 1978–2007 analyzed so far
  - Used in Hsieh and Klenow (2009, 2014)
- Sampling frame
  - 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
- Variables used
  - Gross output (R_i), intermediate inputs (X_i), labor (L_i), wage bill (wL_i), and capital (K_i)

Measurement error in the Indian ASI

	Frequency	Magnitude
Age	12.4%	4 years
EOY & BOY capital stocks	25.7%	15.4%
EOY & BOY goods inventories	22.0%	24.8%
EOY & BOY materials inventories	22.3%	20.2%

There is measurement error in age if age in year t is not equal to 1 + age in year $t-1$. The magnitude of this measurement error is the median absolute deviation. There is measurement error in stocks and inventories if the deviation of the BOY value in year t from the EOY value in year $t-1$ is greater than 1%. The magnitude of this measurement error is the standard deviation of the absolute value of the percentage measurement error.

Data cleaning steps

		Indian ASI	U.S. LRD
Step	Cleaning	Remaining Obs	Remaining Obs
1	Starting sample of plant-years	1,159,641	1,767,000
2	Missing no key variables	924,547	1,589,000
3	Common sector concordance	899,793	1,523,000
4	Trimming extreme TFPR & TFPQ	844,875	1,428,000

- The last step trims 1% tails of MRP and TFPQ deviations from sector-year averages

Inferring allocative efficiency from the data

$$\widehat{AE} = \left[\sum_i^N \left(\frac{TFPQ_i}{TFPQ} \right)^{\epsilon-1} \left(\frac{TFPR_i}{TFPR} \right)^{1-\epsilon} \right]^{\frac{1}{\epsilon-1}}$$

- $\widehat{A}_i = TFPQ_i = \dfrac{\left(\widehat{R}_i\right)^{\frac{\epsilon}{\epsilon-1}}}{(\widehat{K}_i^{\alpha}\widehat{L}_i^{1-\alpha})^{\gamma}\widehat{X}_i^{1-\gamma}}$
- $TFPR_i = \dfrac{\widehat{R}_i}{(\widehat{K}_i^{\alpha}\widehat{L}_i^{1-\alpha})^{\gamma}\widehat{X}_i^{1-\gamma}}$

Inferring allocative efficiency from the data

$$\widehat{AE} = \left[\sum_i^N \left(\frac{TFPQ_i}{TFPQ}\right)^{\epsilon-1} \left(\frac{TFPR_i}{TFPR}\right)^{1-\epsilon}\right]^{\frac{1}{\epsilon-1}}$$

- $TFPQ = \left[\sum_i^N TFPQ_i^{\epsilon-1}\right]^{\frac{1}{\epsilon-1}}$
- $TFPR = \left(\frac{\epsilon}{\epsilon-1}\right)\left[\frac{MRPL}{(1-\alpha)\gamma}\right]^{(1-\alpha)\gamma}\left[\frac{MRPK}{\alpha\gamma}\right]^{\alpha\gamma}\left[\frac{MRPX}{1-\gamma}\right]^{1-\gamma}$
  - $MRPK = \left[\sum_i \frac{\widehat{R}_i}{\widehat{R}}\frac{1}{MRPK_i}\right]^{-1}$ and so on
  - $MRPK_i = \left(\frac{\epsilon-1}{\epsilon}\right)\alpha\gamma\frac{\widehat{R}_i}{\widehat{K}_i}$ and so on

Inferring aggregate AE from the data

Aggregating within-sector allocative efficiencies:

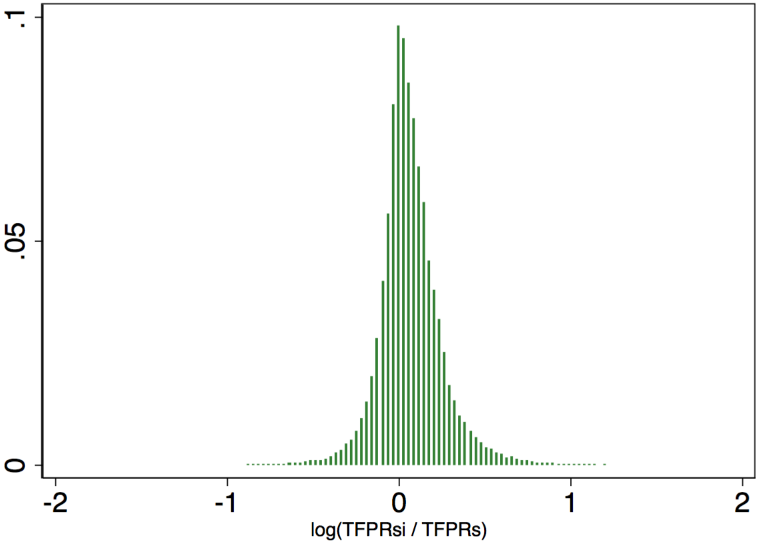
$$\widehat{AE}_t = \prod_{s=1}^{S} \widehat{AE}_{st}^{\frac{\theta_{st}}{\sum_{s=1}^{S} \gamma_s \theta_{st}}}$$

Parameterization:

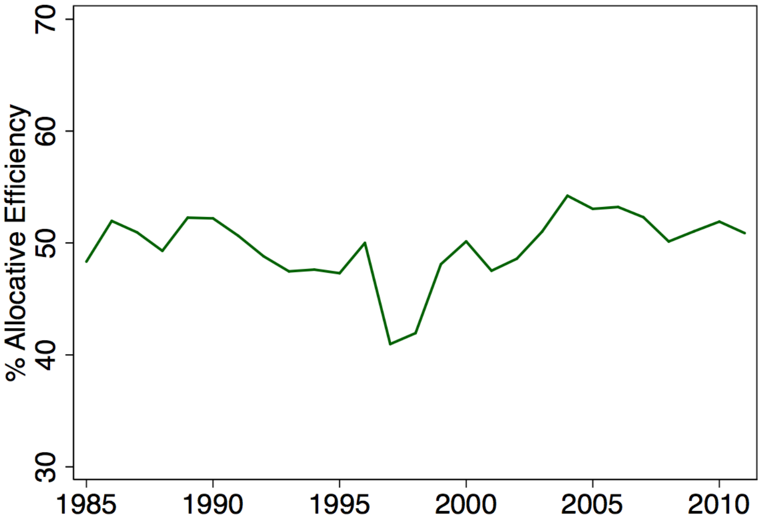
- $\epsilon = 4$ based on Redding and Weinstein (2016)
- α_s and γ_s inferred from sectoral cost-shares ($r = .2$)
- θ_{st} inferred from sectoral shares of aggregate output

Table of Contents

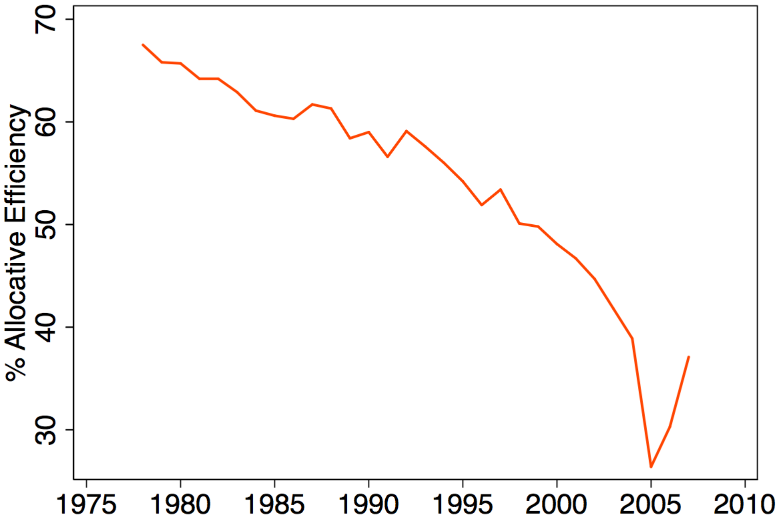
Density of $TFPR$ (standard deviation .178)



Indian allocative efficiency (49% on average)



U.S. allocative efficiency



India vs. U.S. allocative efficiency

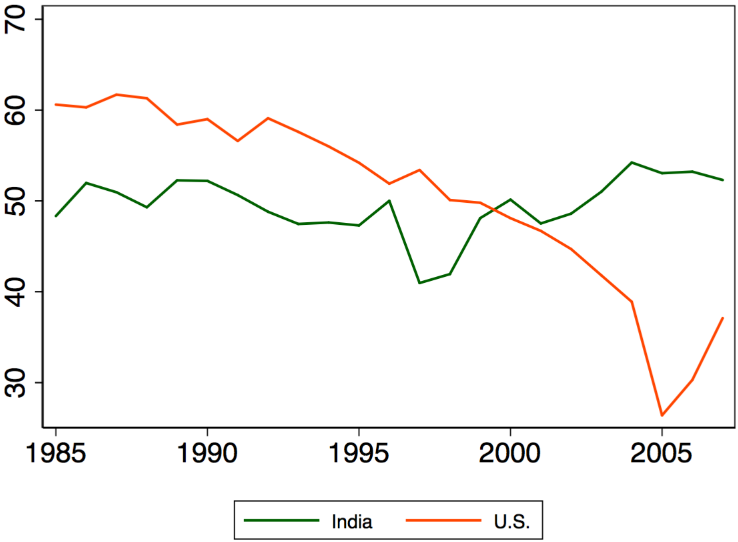


Table of Contents

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
- f_i and g_i = additive measurement errors
- ϕ_i and χ_i = multiplicative measurement errors

Measurement error and *TFPR*

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

- Marginal product = $\left(\frac{R_i}{\widehat{R}_i} \frac{\widehat{I}_i}{I_i}\right) \times$ Average product

Measurement error and $\Delta TFPR$

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

- Δ is the growth rate of a variable relative to the sector s mean.

If only *multiplicative* measurement error:

$$\Delta TFPR_i = \Delta\tau_i + \Delta\chi_i - \Delta\phi_i$$

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

- We focus on additive measurement error
  - Conservative, as multiplicative also overstates TFPR differences
- f_i and g_i are mean zero and orthogonal to $\ln \tau_i$ and $\ln A_i$
- τ_i and $\left(\widehat{R}_i / R_i \cdot I_i / \widehat{I}_i\right)$ are each lognormally distributed

Measurement error and revenue growth

To simplify exposition for awhile: no measurement error in inputs

$$\Delta \widehat{I}_i \;=\; \Delta I_i \;=\; (\epsilon - 1)\,\Delta A_i \;-\; \epsilon\,\Delta \tau_i$$

Revenue growth is:

$$\Delta \widehat{R}_i \;=\; \frac{R_i}{\widehat{R}_i}(\epsilon - 1)(\Delta A_i - \Delta \tau_i) + \frac{\blacktriangle g_i}{\widehat{R}_i}$$

Measurement error and the elasticity of revenue wrt inputs

Define $$\widehat{\beta} \equiv \frac{s_{\Delta\widehat{R},\Delta I}}{s^2_{\Delta I}}$$

Then

$$E\left\{\widehat{\beta} \mid \ln(TFPR)\right\} = E\left\{\frac{R}{\widehat{R}}\left(1+\widehat{\Omega}_\tau\right) \mid \ln(TFPR)\right\}$$

where $$\widehat{\Omega}_\tau \equiv \frac{s_{\Delta\tau,\Delta I}}{s^2_{\Delta I}}$$

Elasticity in a simplified case

If $\Delta\tau_i$ and ΔA_i are i.i.d., $\widehat{\Omega}_\tau$ does not depend on $\ln(TFPR)$. So:

$$E\left\{\widehat{\beta} \mid \ln(TFPR)\right\} = (1+\Omega_\tau)\, E\left\{\frac{R}{\widehat{R}} \mid \ln(TFPR)\right\}$$

$$= (1+\Omega_\tau)\, E\left\{1 - \frac{\widehat{R}-R}{\widehat{R}} \mid \ln(TFPR)\right\}$$

$$\approx (1+\Omega_\tau)\, E\left\{1 - \ln\left(\frac{\widehat{R}}{R}\right) \mid \ln(TFPR)\right\}$$

Elasticity in a simplified case

$$E\left\{\widehat{\beta} \mid \ln(TFPR)\right\} \approx (1+\Omega_\tau)\, E\left\{1 - \ln\left(\frac{\widehat{R}}{R}\right) \mid \ln(TFPR)\right\}$$

$$\approx (1+\Omega_\tau)\left[1 - \frac{\sigma^2_{\ln\frac{\widehat{R}}{R}}}{\sigma^2_{\ln TFPR}} \ln(TFPR)\right]$$

$$\equiv (1+\Omega_\tau)\left[1 - (1-\lambda)\cdot \ln(TFPR)\right]$$

where $\lambda \equiv \dfrac{\sigma^2_{\ln\tau}}{\sigma^2_{\ln TFPR}}$

Compression factor λ

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

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

$$\sigma^2_{\ln \tau} = \lambda \cdot \sigma^2_{\ln(TFPR)}$$

Thus λ is the fraction of TFPR dispersion due to "true" distortions (vs. additive measurement error)

Measurement error in both revenue and inputs

$$E\left\{\,\widehat{\beta}\mid \ln(TFPR)\,\right\} = (1+\Omega_\tau-\Omega_f)\,E\left\{\frac{R\widehat{I}}{\widehat{R}I}\mid \ln(TFPR)\right\}$$

$$\approx (1+\Omega_\tau-\Omega_f)\,[1-(1-\lambda)\cdot\ln(TFPR)]$$

$$\text{where}\quad \Omega_f \equiv \frac{\sigma^2_{\blacktriangle f/I}}{\sigma^2_{\Delta I+\blacktriangle f/I}}$$

Still get $\sigma^2_{\ln\tau} = \lambda\cdot\sigma^2_{\ln(TFPR)}$

$\Delta\tau$, ΔA, $\blacktriangle f/I$ are $i.i.d.$, so Ω_τ, Ω_f do not depend on $\ln(TFPR)$

Baseline specification

$$\Delta\widehat{R}_{it} = \Phi \cdot \ln(TFPR_{it}) + \Psi \cdot \Delta\widehat{I}_{it}$$

$$-(1-\lambda) \cdot \Psi \cdot \ln(TFPR_{it}) \cdot \Delta\widehat{I}_{it} + D_t + \xi_{it}$$

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

Baseline specification

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

- $\ln(TFPR_{it})$ is a Tornqvist of current and previous year
- Weight by Tornqvist gross output shares
- Winsorize 1% tails of $\Delta\widehat{R}_{it}$ and $\Delta\widehat{I}_{it}$

Segue on simulations

- Conduct simulations to validate methodology
- 40,000 plants over 50 years
- A_{it} and τ_{it} follow geometric random walks

$$ln(x_{it}) = ln(x_{it-1}) + \eta_{it}^{x} \text{ where } \eta_{it}^{x} \sim N(0, \sigma_{\eta,x}^{2})$$

- g_{it} (similarly f_{it}) random walk in levels

$$g_{it} = g_{it-1} + \eta_{it}^{g} \cdot R_{it} \text{ where } \eta_{it}^{g} \sim N(0, \sigma_{\eta,g}^{2})$$

- Compare $\lambda \equiv \frac{\sigma_{\ln \tau}^{2}}{\sigma_{\ln(TFPR)}^{2}}$ to $\widehat{\lambda}$ from our baseline specification
  - Use $\epsilon = 4$

Simulation results

Simulation	Parameters	$\lambda \equiv \frac{\sigma^2_{\ln \tau}}{\sigma^2_{\ln(TFPR)}}$	$\widehat{\lambda}$
1	$\sigma_{\eta,\tau}, \sigma_{\eta,A}, \sigma_{\eta,g} = .10$, $\sigma_{\eta,f} = 0$	0.487	**0.509**
2	$\sigma_{\eta,\tau}, \sigma_{\eta,A} = .10$, $\sigma_{\eta,g} = 0$, $\sigma_{\eta,f} = 0.10$	0.483	**0.512**
3	$\sigma_{\eta,\tau}, \sigma_{\eta,A}, \sigma_{\eta,g}, \sigma_{\eta,f} = .10$	0.322	**0.362**

Simulations with adjustment costs

Additional assumptions:

1. Inputs chosen one period ahead
2. Time-invariant distortions: $\tau_i \neq 0$, $\Delta\tau_i = 0$
   - $ln(\tau_i) \sim N(0, \sigma_\tau^2)$

Simulation	Parameters	$\frac{\sigma^2_{\ln\,\tau}}{\sigma^2_{\ln(TFPR)}}$	$\frac{\sigma^2_{\ln\,(ARP)}}{\sigma^2_{\ln(TFPR)}}$	$\widehat{\lambda}$
1	$\sigma_\tau = 0.013, \sigma_{\eta,\tau} = 0, \sigma_{\eta,A} = .10, \sigma_{\eta,g}, \sigma_{\eta,f} = 0$	0.755	**1.000**	**0.998**
2	$\sigma_\tau = 0, \sigma_{\eta,\tau} = 0, \sigma_{\eta,A}, \sigma_{\eta,g} = .10, \sigma_{\eta,f} = 0$	0.000	**0.043**	**0.061**
3	$\sigma_\tau = 0.013, \sigma_{\eta,\tau} = 0, \sigma_{\eta,A}, \sigma_{\eta,g} = .10, \sigma_{\eta,f} = 0$	0.100	**0.133**	**0.120**

Baseline estimates for 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

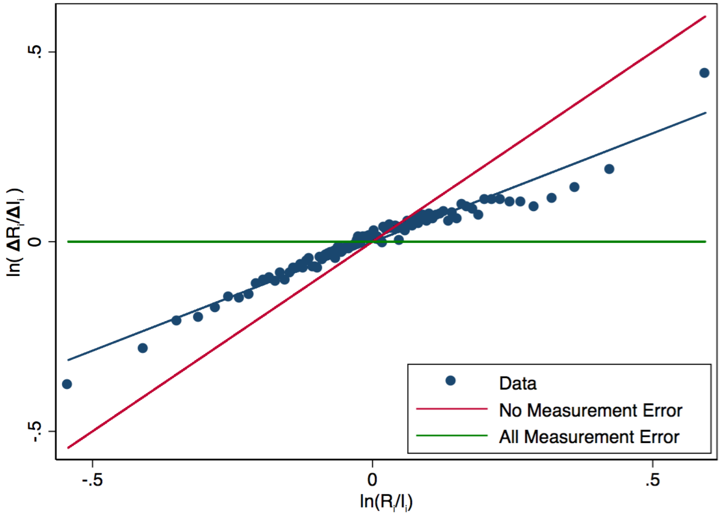
India

	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)



What if measurement error or τ mean revert?

$$E\left\{\widehat{\beta} \mid \ln(TFPR)\right\} = E\left\{\frac{\widehat{RI}}{\widehat{RI}}\left(1 + \widehat{\Omega}_\tau - \widehat{\Omega}_f\right) \mid \ln(TFPR)\right\}$$

$$= E\left\{\frac{\widehat{RI}}{\widehat{RI}} \mid \ln(TFPR)\right\} \bullet E\left\{1 + \widehat{\Omega}_\tau - \widehat{\Omega}_f \mid \ln(TFPR)\right\}$$

because the conditional covariance is zero.

- If measurement error or τ are mean-reverting, then Ω_τ and Ω_f increase at the extremes of $TFPR$
- Capture with square of $\ln(TFPR)$

Allowing for mean reversion

$$
\begin{aligned}
\Delta\widehat{R}_{it} = {} & \Phi \cdot \ln(TFPR_{it}) + \Psi \cdot \Delta\widehat{I}_{it} \\
& - (1-\lambda) \cdot \Psi \cdot \ln(TFPR_{it}) \cdot \Delta\widehat{I}_{it} \\
& + \Gamma \cdot \ln(TFPR_{it})^2 + \Lambda \cdot \ln(TFPR_{it})^2 \cdot \Delta\widehat{I}_{it} \\
& + \Upsilon \cdot \ln(TFPR_{it})^3 - (1-\lambda) \cdot \Lambda \cdot \ln(TFPR_{it})^3 \cdot \Delta\widehat{I}_{it} \\
& + D_t + \xi_{it}
\end{aligned}
$$

Indian estimates allowing for mean reversion

	All Years	1985–1993	1994–2001	2002–2011
Baseline $\widehat{\lambda}$	0.547 (0.035)	0.562 (0.050)	0.510 (0.080)	0.576 (0.027)
$\widehat{\lambda}$ with mean reversion	**0.520** **(0.041)**	**0.547** **(0.060)**	**0.465** **(0.090)**	**0.562** **(0.029)**

U.S. estimates allowing for mean reversion

	All Years	1978–1982	1983–1987	1988–1992	1993–1997	1998-2002	2003-2007
Baseline $\widehat{\lambda}$	.229 (.026)	0.358 (0.027)	0.336 (0.034)	0.326 (0.031)	0.326 (0.037)	0.192 (0.032)	0.095 (0.070)
$\widehat{\lambda}$ with mean reversion	**0.205** **(0.018)**	**0.371** **(0.029)**	**0.312** **(0.037)**	**0.318** **(0.033)**	**0.318** **(0.038)**	**0.129** **(0.041)**	**0.020** **(0.054)**

Measurement error in *relative* inputs

$1/3$ of potential gains reflect relative input distortions measured by differences in factor shares — but this, too, could be mismeasurement.

Measurement error need not be the same across inputs.

Let:

$$\widehat{K}_i = K_i + f_{K_i}$$

$$\widehat{L}_i = L_i + f_{L_i}$$

Differences in $\widehat{K}_i/\widehat{L}_i$ can reflect f_{K_i} versus f_{L_i}.

We ask if differences in $\widehat{K}_i/\widehat{L}_i$ decline when inputs increase.

Do the same for differences in intermediates versus value added.

Allowing for relative measurement error across inputs

$$\Delta\widehat{K}_i - \Delta\widehat{L}_i =$$

$$\frac{-\left[\left(\frac{\widehat{K}_i - K_i}{K_i}\right) - \left(\frac{\widehat{L}_i - L_i}{L_i}\right)\right]\Delta\widehat{V}_i + \left(\frac{K_i}{\widehat{K}_i}\cdot\frac{L_i}{\widehat{L}_i}\right)\cdot\Delta Z_i}{1 - \alpha\left(\frac{\widehat{K}_i - K_i}{K_i}\right) - (1-\alpha)\left(\frac{\widehat{L}_i - L_i}{L_i}\right)}$$

$$\text{where} \quad \Delta Z_i = \Delta\tau_{L_i} - \Delta\tau_{K_i} + \Delta f_{K_i} - \Delta f_{L_i}$$

We assume ΔZ_i is orthogonal to $\left(\frac{\widehat{K}_i - K_i}{K_i}\right) - \left(\frac{\widehat{L}_i - L_i}{L_i}\right)$.

Estimating relative measurement error across inputs

$$
\begin{aligned}
\Delta\left(\frac{\widehat{K}_{it}}{\widehat{L}_{it}}\right) &= \Phi \cdot \ln\left(\frac{\widehat{K}_{it}}{\widehat{L}_{it}}\right) + \Pi \cdot \Delta\widehat{V}_{it} \\
&\quad -(1-\lambda_{KL}) \cdot \ln\left(\frac{\widehat{K}_{it}}{\widehat{L}_{it}}\right) \cdot \Delta\widehat{V}_{it} + D_t + \xi_{it}
\end{aligned}
$$

- $\ln\left(\frac{\widehat{K}_{it}}{\widehat{L}_{it}}\right)$ is a Tornqvist of current and previous year
- $\Delta\widehat{V}_{it}$ is growth of capital and labor inputs
- We winsorize 1% tails of $\Delta\left(\frac{\widehat{K}_{it}}{\widehat{L}_{it}}\right)$ and $\Delta\widehat{V}_{it}$
- And proceed similarly to estimate λ_{VX}

Indian estimates with relative measurement error

	All Years	1985–1993	1994–2001	2002–2011
$\widehat{\lambda}$ with mean reversion	0.520 (0.041)	0.547 (0.060)	0.465 (0.090)	0.562 (0.029)
$\widehat{\lambda}_{KL}$	**0.927** **(0.022)**	**0.910** **(0.035)**	**0.888** **(0.039)**	**0.976** **(0.033)**
$\widehat{\lambda}_{VX}$	**0.912** **(0.011)**	**0.895** **(0.014)**	**0.902** **(0.019)**	**0.928** **(0.020)**

U.S. estimates with relative measurement error

	All Years	1978–1982	1983–1987	1988–1992	1993–1997	1998-2002	2003-2007
$\widehat{\lambda}$ with mean reversion	0.205 (0.018)	0.371 (0.029)	0.312 (0.037)	0.318 (0.033)	0.318 (0.038)	0.129 (0.041)	0.020 (0.054)
$\widehat{\lambda}_{KL}$	**0.797** **(0.009)**	**0.822** **(0.020)**	**0.777** **(0.016)**	**0.815** **(0.017)**	**0.780** **(0.026)**	**0.777** **(0.030)**	**0.831** **(0.026)**
$\widehat{\lambda}_{VX}$	**0.838** **(0.006)**	**0.884** **(0.010)**	**0.883** **(0.011)**	**0.840** **(0.011)**	**0.821** **(0.018)**	**0.839** **(0.014)**	**0.811** **(0.021)**

Table of Contents

Correcting $TFPR_i$

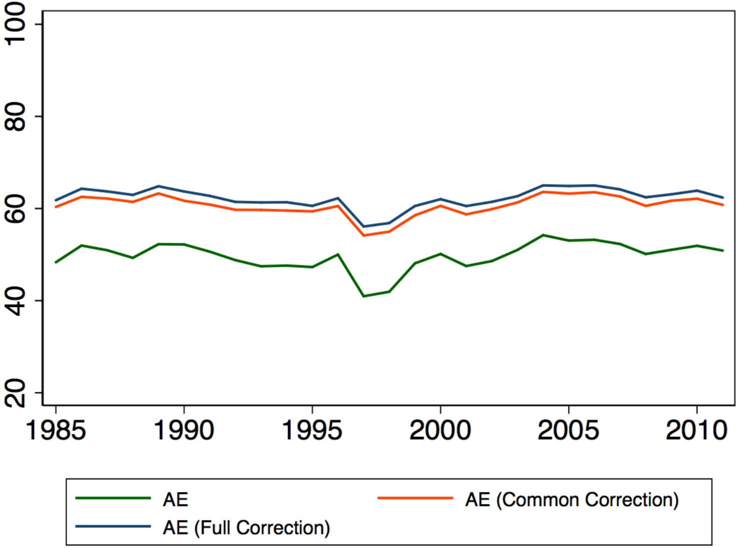
If all measurement error is common across inputs:

$$\widetilde{TFPR_i} \propto \exp\left(\widehat{\lambda} \cdot \ln(TFPR_i) + \epsilon_i\right)$$

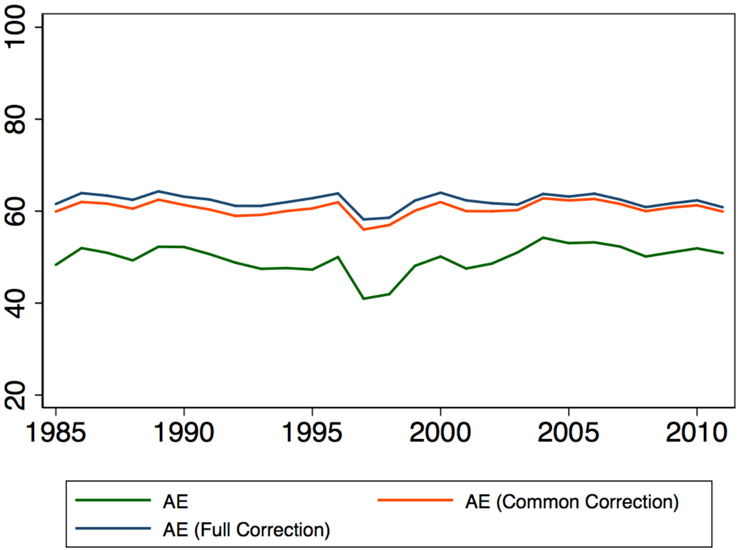
- $\epsilon_i \sim N\left(0,\ (\widehat{\lambda} - \widehat{\lambda}^2)\ \sigma^2_{\ln(TFPR)}\right)$

Similar correction in the presence of relative measurement error

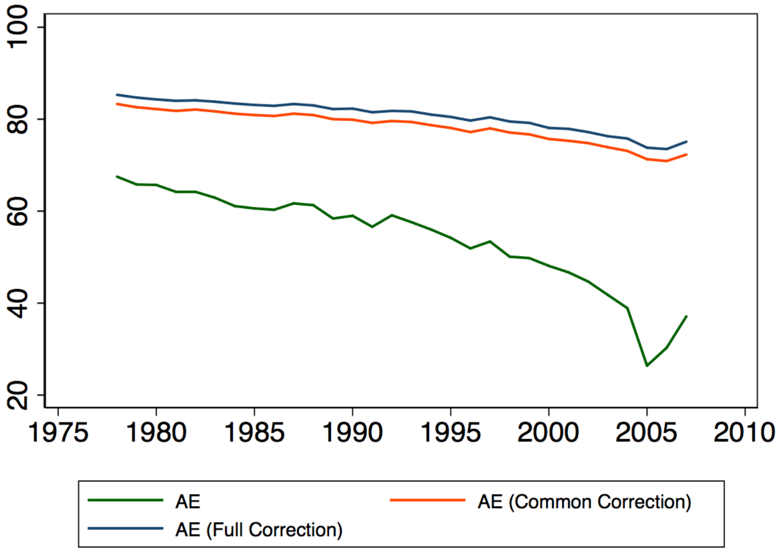
Allocative efficiency in India



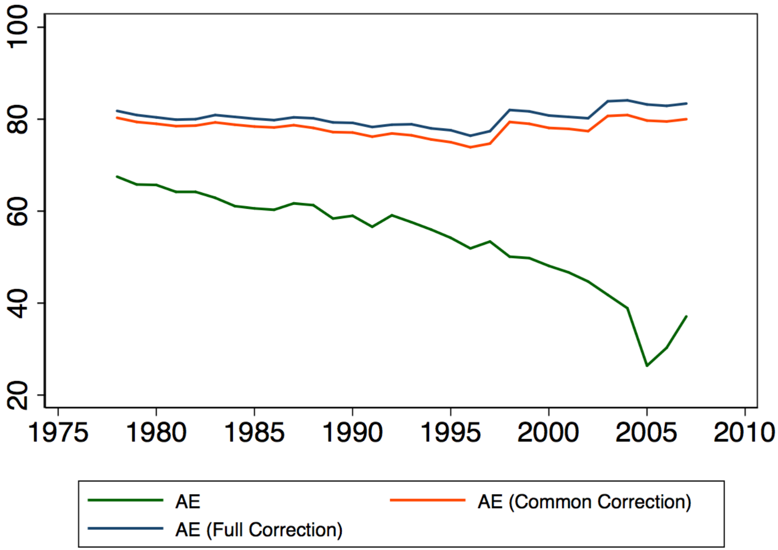
Allocative efficiency in India ($\hat{\lambda}$ in windows)



Allocative Efficiency in U.S.



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



Uncorrected vs. corrected gains from reallocation

	India 1985–2011		U.S. 1978–2007	
	Mean	S.D.	Mean	S.D.
Uncorrected Gains	102%	13.7%	95.6%	53.5%
Corrected Gains (Common & Relative)	61%	4.0%	24.4%	3.0%
Shrinkage	40%	71%	74%	94%

Allocative efficiency: U.S. relative to India

Conclusion

- Propose way to estimate true dispersion of marginal products
  - Projects first differences on levels for average products
  - Requires measurement error to be additive and uncorrelated with distortions, productivity
- Implemented on 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 $\frac{2}{3}$
- Implemented on 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.?
  - Rising survey non-response rate?
  - SIC to NAICS switch in 1998?
- 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
- Analyze cross-sector distortions
- Relate corrected wedges to size and policies