Online Appendix

Misallocation or Mismeasurement?

Mark Bils
University of Rochester and NBER

Peter J. Klenow
Stanford University and NBER

Cian Ruane
International Monetary Fund

April 30, 2021

1 Model

1.1 Solving the firm's problem

Solving the representative firm's problem and normalizing the price index of the final good $P = 1$, we obtain the demand for sectoral output:

$$Q_s = \frac{1}{P_s}\theta_s Q \; .$$

We can also obtain the demand curve facing firm i in sector s:

$$P_{si} = \theta_s Q Q_s^{\frac{1-\epsilon}{\epsilon}} Q_{si}^{-\frac{1}{\epsilon}} \; .$$

With this we can solve the heterogeneous firms' problem. We obtain the standard result that prices are a constant markup over marginal cost:

$$P_{si} = \left(\frac{\epsilon}{\epsilon - 1}\right)\frac{1}{\gamma_s^{\gamma_s}}\left[\left(\frac{r}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}\right]^{\gamma_s}\left[\frac{1}{1-\gamma_s}\right]^{1-\gamma_s}\frac{1}{\tau_{si}A_{si}} \; .$$

1.2 Aggregating to the sector level

Aggregating to the sector level, we can express sectoral gross output as a function of sectoral inputs and sectoral productivity A_s:

$$Q_s = A_s (K_s^{\alpha_s} L_s^{1-\alpha_s})^{\gamma_s} X_s^{1-\gamma_s} \; ,$$

where

$$A_s = \left[\sum_{i=1}^{N_{st}} A_{si}^{\epsilon-1}\left(\frac{\tau_s}{\tau_{si}}\right)^{1-\epsilon}\right]^{\frac{1}{\epsilon-1}} \; .$$

The *average sectoral distortion* on labor is defined as follows:

$$1+\tau_s^L \equiv \left[\sum_{i=1}^{N_s} \frac{R_{si}}{R_s} \frac{1}{1+\tau_{si}^L}\right]^{-1} = \frac{\left[\sum_{i=1}^{N_s} \left[\frac{A_{si}}{\tau_{si}}\right]^{\epsilon-1}\right]}{\left[\sum_{i=1}^{N_s} \left[\frac{A_{si}}{\tau_{si}}\right]^{\epsilon-1} \frac{1}{1+\tau_{si}^L}\right]},$$

and similarly for τ_s^K and τ_s^X.

1.3 Aggregate consumption

Aggregate value added in this model ($C = Q - X$) can be expressed as follows:

$$C = \left(\frac{\epsilon}{\epsilon-1}\right)^{\frac{\sum_s^S (1-\gamma_s)\theta_s}{\sum_s^S \gamma_s\theta_s}} \times \prod_s^S \theta_s^{\frac{\theta_s}{\sum_s^S \gamma_s\theta_s}} \times \prod_s^S \left[\left(\alpha_s^{\alpha_s}(1-\alpha_s)^{1-\alpha_s}\right)^{\gamma_s} \gamma_s^{\gamma_s}(1-\gamma_s)^{1-\gamma_s}\right]^{\frac{\theta_s}{\sum_s^S \gamma_s\theta_s}} \times$$
$$\prod_s^S \left[\frac{\tau}{\tau_s}\right]^{\frac{\theta_s}{\sum_s^S \gamma_s\theta_s}} \times \left[1 - \left(\frac{\epsilon-1}{\epsilon}\right) \frac{1}{(1+\tau^X)} \sum_s^S \theta_s(1-\gamma_s)\left(\frac{1+\tau^X}{1+\tau_s^X}\right)\right] \times \left(\frac{1}{1+\tau^X}\right)^{\frac{\sum_s^S (1-\gamma_s)\theta_s}{\sum_s^S \gamma_s\theta_s}} \times$$
$$\left(\frac{L}{\sum_s^S \left[\theta_s\gamma_s(1-\alpha_s)\left(\frac{1+\tau^L}{1+\tau_s^L}\right)\right]}\right)^{\frac{\sum_s^S (1-\alpha_s)\gamma_s\theta_s}{\sum_s^S \gamma_s\theta_s}} \times \left(\frac{K}{\sum_s^S \left[\theta_s\gamma_s\alpha_s\left(\frac{1+\tau^K}{1+\tau_s^K}\right)\right]}\right)^{\frac{\sum_s^S \alpha_s\gamma_s\theta_s}{\sum_s^S \gamma_s\theta_s}} \times$$
$$\prod_s^S A_s^{\frac{\theta_s}{\sum_s^S \gamma_s\theta_s}},$$

where

$$\tau \equiv \left[(1+\tau^L)^{1-\alpha_s}(1+\tau^K)^{\alpha_s}\right]^{\gamma_s}(1+\tau^X)^{1-\gamma_s},$$

$$\tau^L \equiv \frac{1}{\sum_i^S \frac{R_s}{Q}\frac{1}{1+\tau_s^L}},$$

and similarly for τ^K and τ^X.

It is worth noting that the exponents on the sectoral productivity term $\left(\prod_{s=1}^{S} A_s^{\frac{\theta_s}{\sum_{s=1}^{S}\gamma_s\theta_s}}\right)$ sum to > 1. This is due to the amplification effect of intermediate inputs. A 1% increase in the productivity of each sector leads to a *greater* than 1% increase in aggregate consumption.

2 Data

2.1 Indian Annual Survey of Industries (ASI)

The ASI is a dataset put together by India's Ministry of Statistics and Programme Implementation (MOSPI). As of 2019, it can be freely downloaded here: http://microdata.gov.in/nada43/index.php/home. The reference period for each survey is the accounting year, which in India begins on the 1st of April and ends on the 31st of March the following year. Throughout the paper we reference the surveys by the earlier of the two years covered. Details of how the sampling methodology for the ASI changes over time are shown in Table A1.

2.11 Measurement of main variables

Gross Output: We construct gross output as the sum of the gross value of products sold, the change in finished good and semi-finished good inventories, and all other sources of revenue. The gross value of products sold includes distribution expenses, as well as taxes and subsidies. Other sources of revenue include the value of electricity sold, the value of own construction, the net value of resales and as well as receipts from industrial and non-industrial services rendered (e.g. contract or commission work).

Labor: We construct labor as the average number of personnel in the plant over the year. Personnel include wage or salary workers, supervisory/managerial

Table A1: Sampling Methodology for Indian ASI

Period	Census Sector	Sample Sector
1985–1986	12 less industrially developed states, 50 or more workers with power, 100 or more workers without power, industries with fewer than 50 plants in all of India	Stratified within state $\times$ 3-digit industry (NIC-70), 50% samples of remaining non-Census plants in alternate years
1987–1996	12 less industrially developed states, 100 or more workers, all joint returns, all plants within state $\times$ 4-digit industry if $<$ 4 plants, all plants within state $\times$ 3-digit industry if $<$ 20 plants	Stratified within state $\times$ 3-digit industry (NIC-87), minimum sample of 20 plants within strata, otherwise 1/3 of plants sampled
1997	12 less industrially developed states, plants with $>$ 200 workers, 'significant units' with $<$ 200 workers but contributed highly to value of output between 1993–1995, public sector undertakings	Stratified within state $\times$ 3-digit industry (NIC-87), minimum of 4 plants sampled per stratum
1998	Complete enumeration states, plants with $>$ 200 workers, all joint returns	Stratified within state $\times$ 4-digit industry (NIC-98), minimum of 8 plants per stratum
1999–2003	Complete enumeration states, plants with $\geq$ 100 or more workers, all joint returns	Stratified within state $\times$ 4-digit industry (NIC-98), minimum of 8 plants per stratum
2004–2006	6 less industrially developed states, 100 or more workers, all joint returns, all plants within state $\times$ 4-digit industry with $<$ 4 units	Stratified within state $\times$ 4-digit industry, 20% sampling, minimum of 4 plants
2007	5 less industrially developed states, 100 or more workers, all joint returns, all plants within state $\times$ 4-digit industry with $<$ 6 units	Stratified within state $\times$ 4-digit industry, minimum 6 plants, 12% sampling fraction: exceptions
2008–2013	6 less industrially developed states, 100 or more employees, all joint returns, all plants within state $\times$ 4-digit industry with $<$ 4 units	Stratified within district $\times$ 4-digit industry, minimum 4 plants, 20% sampling fraction

Notes: information regarding sampling methodology for each ASI wave is available in the metadata files here: http://microdata.gov.in/nada43/index.php/catalog/ASI/about.

staff, administrative/custodial employees and all unpaid workers (including family members).

Labor Cost: We construct labor costs as total payments to labor over the course of the year. These payments include wages and salaries, bonuses, contributions to old-age pension funds (and other funds), and all welfare expenses.

Capital: This is constructed as the average of the opening and closing book value of fixed assets (net of depreciation). These include all types of assets deployed for production and transportation, as well as living or recreational facilities (hospitals, schools, etc.) for factory personnel. It excludes intangible assets.

Intermediates: We construct intermediates as the sum of the value of materials consumed, fuels consumed and other intermediate expenses. Other intermediate expenses include repair and maintenance costs (plant/machinery, building, etc.), costs of contract and commission work, operating expenses (freight and transportation charges, taxes paid), non-operating expenses (communication, accounting, advertising), and insurance charges.

Industry: The official Indian industry classification is the National Industry Classification (NIC). The classification was revised in 1987, 1998, 2004 and 2008. We construct concordances between the various NIC revisions to construct our harmonized classification with 50 manufacturing industries. In terms of its level of disaggregation, our industry classification is close to 3-digit NAICS.

Plant identifiers: Official plant identifiers, enabling longitudinal linking across survey waves, are only available in the ASI surveys from 1998 onwards. Prior to 1998, we use plant identifiers available in an older version of the publicly available ASI. These were first used in Allcott et al. (2016). There are panel breaks between the following year pairs: 1986 to 1987, 1988 to 1989, and 2007 to 2008. Our panel therefore consists of 4 sub-periods within which we have plant identifiers: 1985–1986, 1987–1988, 1989–2007, and 2008–2011. In order to verify the quality of these identifiers prior to 1998, we examine whether the plant's reported year of incorporation is reported consistently across survey waves. In Figure A1 we show the share of panel plants whose reported year of incorporation changes from year to year (i.e. the firm's reported age does not change by one year between survey years). We don't see any evidence that this misreporting is higher prior to 1998 than afterwards.

Figure A1: Consistency of Age Reporting

Sources: Indian ASI. The figure shows the share of panel plants whose reported year of incorporation changes between year t and year t+1.

2.2 U.S. Annual Survey of Manufactures (ASM)

The ASM and CMF are restricted access databases of U.S. manufacturing plants put together by the U.S. Census Bureau. Both the ASM and CMF are mail-back surveys. We only use data for plants that also appear in the ASM sample. The ASM sample is redrawn in years ending in 4 and 9. We use sampling weights in all our analyses to make our results representative of the universe of U.S. manufacturing plants with at least one employee.

2.21 Measurement of main variables

Gross Output: We construct gross output as the sum of the value of shipments, the net value of resales, the change in finished good and semi-finished good inventories, and all other sources of revenue. Other sources of revenue include such payments as for contract work, installations, or repair work.

Labor: We construct labor as the average total number of employees.

Labor Cost: We construct labor costs as total payments to labor over the course of the year. These payments include wages and salaries, bonuses and other benefits. Other benefits include all fringe benefits paid by the firm, including federal insurance contributions, unemployment taxes, employee pension and welfare plans, and like.

Capital: The ASM does not contain yearly measures of the book value of capital. We therefore use the real market value of capital (measured in 1997 $), calculated using the perpetual inventory method. See Foster, Grim and Haltiwanger (2016) for more details.

Intermediates: We construct intermediates as the sum of the value of materials purchased, electricity and fuels purchased, and the change in material inventories. We do not include expenditures on services (such as marketing, advertising, etc.) because they are not available in every year of the ASM.

Industry: We use the harmonized sectoral classification from Fort and Klimek (2016) at the NAICS 3-digit level. We thereby have a balanced sectoral panel of 86 sectors. The Fort-Klimek (FK) sectors deal with the large reclassification of manufacturing plants into the service sector during the SIC to NAICS transition. It is available at the 6-digit NACIS level, but we use the 3-digit level to come closest to the 50 harmonized sectors we have for India.

Plant identifiers: We use plant identifiers available from the LBD. Once a plant is drawn in the ASM sample (in years ending in 4 and 9), it will be surveyed every year until the next resampling.

2.3 Data cleaning

We follow the same steps in both India and the U.S. when cleaning the samples. We first drop plants with missing or negative values of any of the main variables required to construct TFPR. A benefit of using gross output TFPR rather than value added TFPR is that value-added is frequently negative in plant-level surveys. We drop plants which don't belong to a consistently defined manufacturing industry. We then trim the 1% tails of $\frac{\text{TFPR}_{si}}{\text{TFPR}_s}$ and $\frac{\text{TFPQ}_{si}}{\text{TFPQ}_s}$ in each year, pooling all industries. After trimming the sample, we recalculate the sectoral factor shares α_s and γ_s, and then recalculate TFPR and TFPQ.

3 Simulations

In this appendix we first report our data moments and simulated moments for our baseline simulations. These are shown in Table A2. Our simulated moments are always extremely close to data moments. We next explore how our methodology behaves in the presence of measurement error in revenues, multiplicative measurement error, and adjustment costs. In Table A3 we show the parameter values underlying these simulations. In Table A4 we show the model moments, inferred share of distortions in TFPR dispersion and true share of TFPR dispersion for each simulation. We treat the calibration to the U.S. economy in 2006–2013 as the country and time window for illustrating the results.

The second rows of the Table A3 and Table A4 show our results when we calibrate our model to the U.S. economy in 2006–2013 assuming that all measurement error is in revenues rather than inputs: we set $\sigma_f = 0$ and calibrate σ_τ, σ_A, σ_a, and σ_g. (See Table A3 for resulting parameters.) Looking at Table A4, with

Table A2: Data Moments versus Simulated Moments

	σ^2_{TFPR}		σ^2_{TFPQ}		$\ln(\widehat{\beta})$ slope		$\sigma^2_{\Delta I}$	
	Data	Model	Data	Model	Data	Model	Data	Model
Panel A: India								
1985–1991	0.032	0.032	0.47	0.54	-0.35	-0.35	0.38	0.38
1992–1996	0.038	0.038	0.52	0.52	-0.31	-0.30	0.20	0.20
1997–2001	0.038	0.038	0.50	0.50	-0.33	-0.32	0.17	0.17
2002–2007	0.027	0.027	0.54	0.54	-0.39	-0.38	0.17	0.17
2008–2013	0.027	0.027	0.56	0.56	-0.33	-0.32	0.18	0.18
Panel B: U.S.								
1978–1984	0.047	0.047	0.35	0.35	-0.73	-0.75	0.14	0.14
1985–1991	0.060	0.060	0.33	0.33	-0.67	-0.65	0.12	0.12
1992–1998	0.072	0.073	0.37	0.37	-0.74	-0.72	0.12	0.12
1999–2005	0.096	0.096	0.36	0.36	-0.82	-0.80	0.12	0.12
2006–2013	0.102	0.102	0.41	0.41	-0.90	-0.90	0.16	0.16

Source: This table shows the data moments to which the model described in Section 6 is calibrated, and the model-generated moments from simulations with the calibrated parameter values reported in Table 3. The model is calibrated for both India and the U.S. separately for five different time periods in each. σ^2_{TFPR} is the output share weighted variance of ln(TFPR). σ^2_{TFPQ} is the variance of ln(TFPQ). $\ln(\widehat{\beta})$ slope is the slope of $\widehat{\beta}$ against mean Tornqvist ln(TFPR) across deciles. $\sigma^2_{\Delta I}$ is the variance of input growth.

measurement error just in revenues, we are not able to exactly match all four moments, missing slightly on the $\ln(\widehat{\beta})$ slope. However, similar to the results found in Section 6 of the main text, we find that our correction underestimates the share of measurement error in TFPR dispersion.

The third row of Table A4 shows our simulations with multiplicative measurement error in inputs instead of additive measurement error in inputs. We keep σ_τ, σ_A, and σ_a at the same values as in the baseline calibration. We then calibrate σ_{f^m} to match the variance of ln(TFPR) in the U.S. in 2006–2013. In this case, the variance of $\ln(\tau)$ is only 1.5% of the variance of ln(TFPR), but our correction infers that it is 95.8%. This confirms that multiplicative measurement error leads us to overestimate the dispersion of true marginal products.

The fourth rows of the tables show our simulations with adjustment costs. In this setup, plants have to choose their inputs one period in advance, before they observe the realization of their productivity shock. (Plants form expectations

Table A3: Simulation Parameters in Extra Cases

	σ_τ	σ_A	σ_a	σ_f	σ_g	σ_{f^m}
1. Baseline (U.S. 2006–2013)	0.017	0.494	0.103	0.130	0	0
2. Measurement Error in Revenue	0.025	0.228	0.098	0	0.143	0
3. Multiplicative Measurement Error	0.017	0.494	0.103	0	0	0.152
4. Adjustment Costs	0.017	0.494	0.103	0	0	0

Source: This table shows the parameter values used for the simulations in Appendix 3 σ_τ is the standard deviation of the shocks to the distortions. σ_A is the standard deviation of the permanent component of plant productivity. σ_a is the standard deviation of the time-varying component of plant productivity. σ_f is the standard deviation of shocks to additive measurement error in inputs. σ_g is the standard deviation of shocks to additive measurement error in revenues. σ_{f^m} is the standard deviation of shocks to multiplicative measurement error in inputs.

Table A4: Simulation Moments in Extra Cases

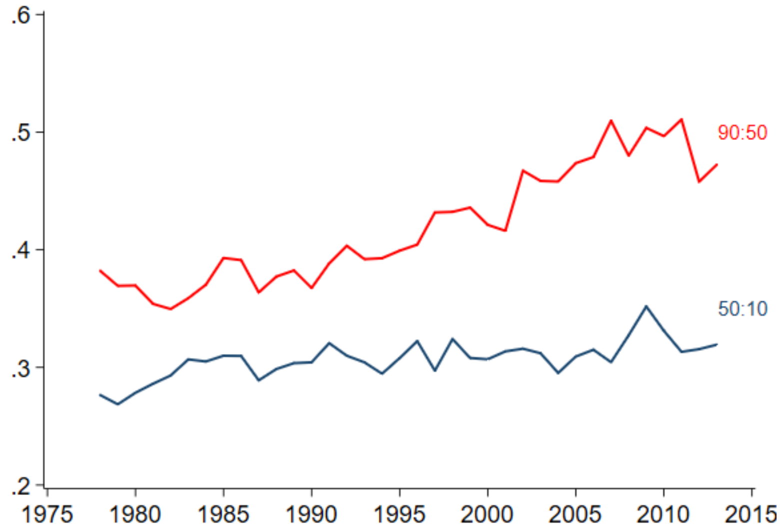
	σ^2_{TFPR}	σ^2_{TFPQ}	$\ln(\widehat{\beta})$ slope	$\sigma^2_{\Delta I}$	$\sigma^2_{\widehat{\tau}}/\sigma^2_{\text{TFPR}}$	$\sigma^2_{\tau}/\sigma^2_{\text{TFPR}}$
1. Baseline (U.S. 2006–2013)	0.102	0.41	-0.90	0.16	0.236	0.015
2. Meas. Err. in Revenue	0.102	0.41	-0.79	0.16	0.238	0.033
3. Multiplicative Meas. Err.	0.102	0.37	-0.028	0.13	0.958	0.015
4. Adjustment Costs	0.006	0.26	-0.001	0.078	1.042	0.240

Source: This table shows the data moments generated by the model for each simulation. Simulation 1. is the same as that shown in Table 4. Simulation 2. has additive measurement error only in revenues. Simulation 3. has no additive measurement error, but has multiplicative measurement error in inputs. Simulation 4. has no measurement error, but has adjustment costs. σ^2_{TFPR} is the output share weighted variance of ln(TFPR). σ^2_{TFPQ} is the variance of ln(TFPQ). $\ln(\widehat{\beta})$ slope is the slope of $\widehat{\beta}$ against mean Tornqvist ln(TFPR) across deciles. $\sigma^2_{\Delta I}$ is the variance of input growth. $\sigma^2_{\widehat{\tau}}/\sigma^2_{\text{TFPR}}$ is the ratio of the variance of $\ln(\widehat{\tau})$ to the variance of ln(TFPR). $\sigma^2_{\tau}/\sigma^2_{\text{TFPR}}$ is the the ratio of the variance of $\ln(\tau)$ to the variance of ln(TFPR)

based on knowledge of the true distribution of productivity shocks.) We keep σ_τ, σ_A and σ_a at the same values as in the baseline calibration, and set $\sigma_f = 0$. Because there is no measurement error, all of the variance of ln(TFPR) reflects true dispersion of marginal products. The variance of $\ln(\tau)$ accounts for 24.0% of this dispersion. Our correction infers that the variance of $\ln(\widehat{\tau})$ accounts for all of ln(TFPR) dispersion. This confirms that adjustment costs (of this type), similarly to multiplicative measurement errors, lead us to overestimate the dispersion of true marginal products.

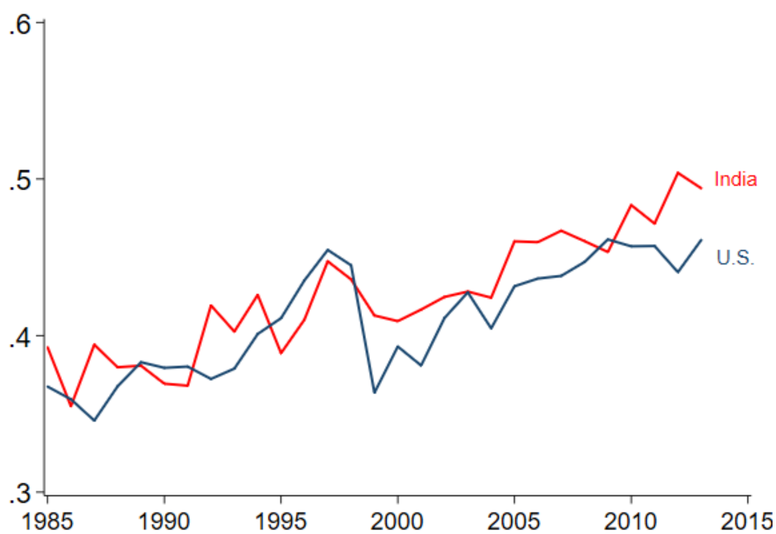
4 Additional Figures

Figure A2: 90:50 and 50:10 Percentile Ratios of ln(TFPR)



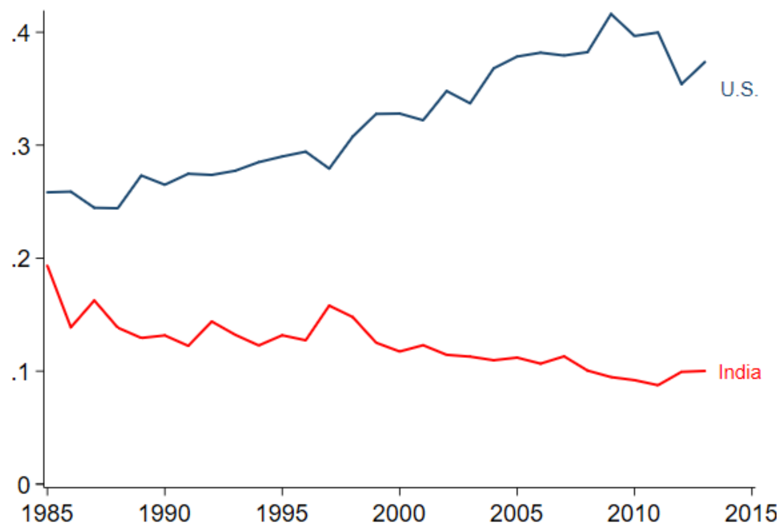
Source: U.S. LRD. The figure shows the 90:50 percentile ratio and the 50:10 percentile ratio of ln(TFPR) for the U.S. between 1978 and 2013. The percentiles are of the epanechnikov kernel density estimates of the distribution of ln(TFPR) (with a bandwidth of 0.05) so as to avoid disclosure of confidential information.

Figure A3: Variance of ln(TFPQ)



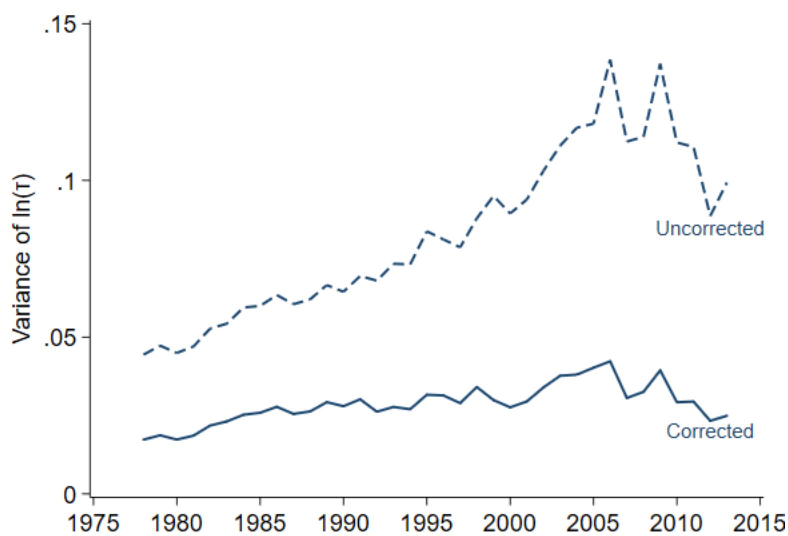
Source: Indian ASI and U.S. LRD. The figure shows the uncorrected variances of ln(TFPQ) for India and the U.S. between 1985 and 2013.

Figure A4: Elasticity of TFPR with respect to TFPQ



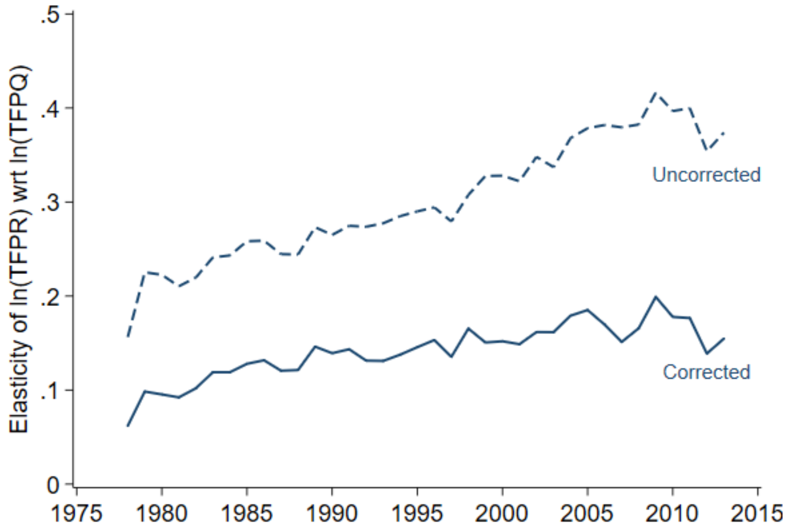
Source: Indian ASI and U.S. LRD. The figure shows the uncorrected elasticity of ln(TFPR) with respect to ln(TFPQ) for India and the U.S. between 1985 and 2013.

Figure A5: Variance of $\ln(\tau)$ for U.S.



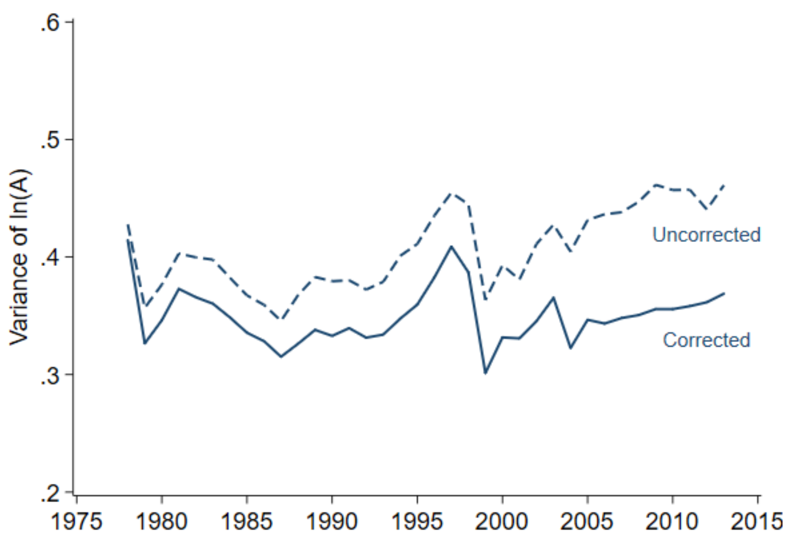
Source: U.S. LRD. The figure shows uncorrected and corrected variance of $\ln(\tau)$ for the U.S., 1978 to 2013.

Figure A6: Elasticity of TFPR with respect to TFPQ in the U.S.



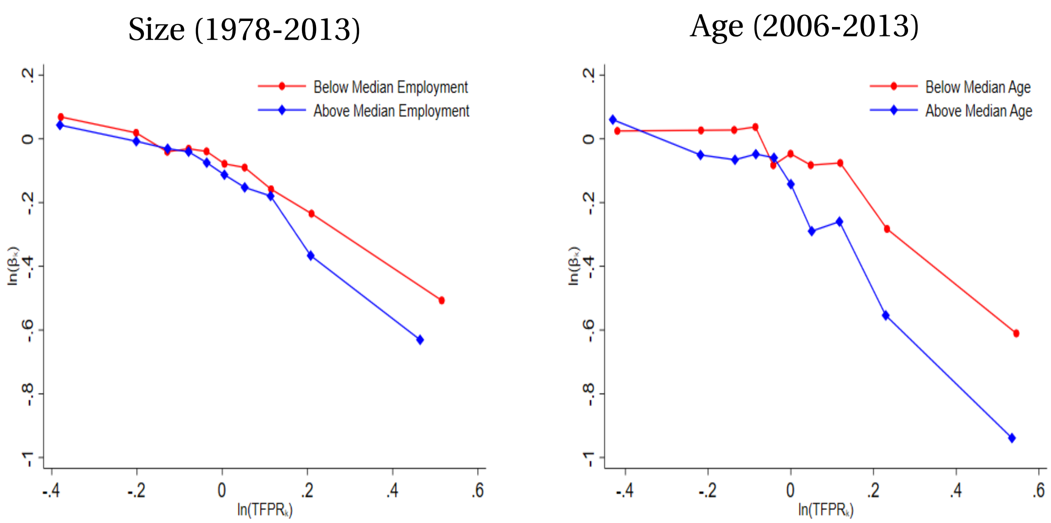
Source: U.S. LRD. The figure shows uncorrected and corrected elasticity of TFPR with respect to TFPQ for the U.S., 1978 to 2013.

Figure A7: Variance of $\ln(A)$ in the U.S.



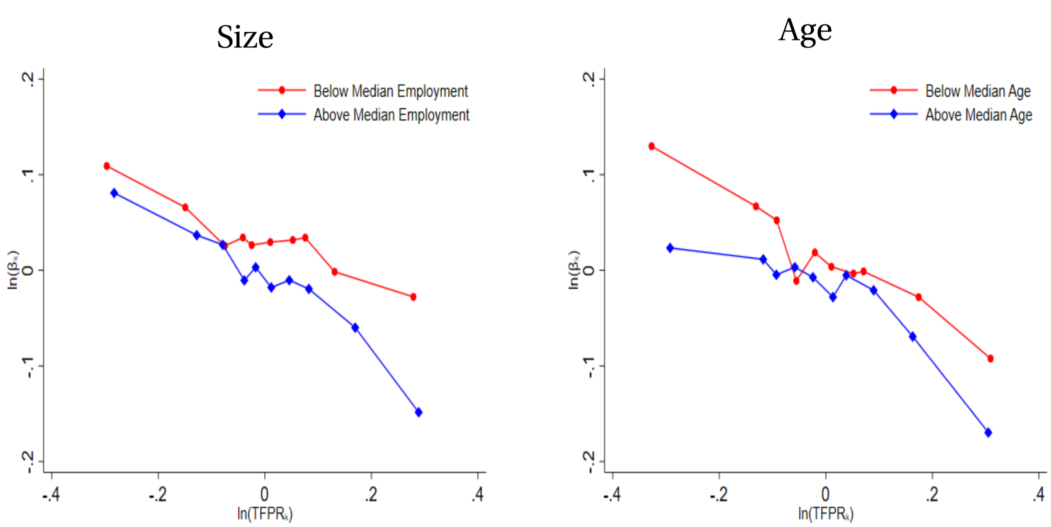
Source: U.S. LRD. The figure shows uncorrected and corrected variance of $\ln(A)$ for the U.S., 1978 to 2013.

Figure A8: U.S. β Slopes by Size and Age



Source: U.S. LRD. The figure plots the $\widehat{\beta}_k$ coefficients recovered from running the regressions in (12) against deciles of TFPR, separately for size and age groups. Median employment is calculated as the median (Tornqvist) employment of panel plants in our full sample (1978-2013). We calculate age based on the first year in which a plant appears in the LBD. Given this censors age in the early part of our sample, we only estimate $\widehat{\beta}_k$ by age group for the last window: 2006-2013.

Figure A9: India β Slopes by Size and Age (1985-2013)



Source: Indian ASI. The figure plots the $\widehat{\beta}_k$ coefficients recovered from running the regressions in (12) against deciles of TFPR, separately for size and age groups. Age is calculated based on reported year of incorporation of the plant. Median (Tornqvist) employment and age are calculated across our full sample of panel plants, from 1985 to 2013.

5 Steps to correct for measurement error

Here we provide more detail regarding the steps to follow in order to implement our procedure. While proxies for allocative efficiency can be constructed using data on revenues and inputs from a single cross-section of plants (or firms), our measurement error correction requires panel data for at least a subset of plants. We implement the following steps (after basic cleaning of the data, including trimming the 1% tails of TFPR and TFPQ in each year):

1. For a sample of panel plants, construct output growth, composite input growth, and the deciles of TFPR. We construct composite input growth as the weighted average of labor, capital, and intermediate input growth, with the weights given by sectoral factor shares (α_s and γ_s). To arrive at deciles of TFPR, for each plant we first calculate the deviation of ln(TFPR) from the weighted average ln(TFPR) at the sector-year level. We construct these ln(TFPR) deviations both in period t and $t+1$. We average these to obtain Tornqvist ln(TFPR) deviations from the sector-year average at the plant-level. We then place these Tornqvist ln(TFPR) deviations into deciles, with an equal share of total costs in each decile (rather than an equal number of plants). We trim extreme values of TFPR growth, namely when TFPR (relative to the sector-year average) increases or decreases by a factor of 5 or more from one year to the next.

2. We subtract sector-year fixed effects for output growth and input growth, respectively. We regress (residualized) output growth on (residualized) input growth separately for each decile k of TFPR, with each observation weighted by the plant's share of aggregate costs (averaged across period t and $t+1$). The coefficients on input growth are the $\widehat{\beta}_k$ estimates. For our $\ln(\widehat{\beta}_k)$ vs. ln(TFPR) plots, we construct the (weighted) average of Tornqvist ln(TFPR) within each decile.

3. We merge the $\widehat{\beta}_k$ estimates into the full unbalanced panel of plants using the cutoffs between each Tornqvist ln(TFPR) decile from the panel.

The distribution of Tornqvist ln(TFPR) is more compressed than that for ln(TFPR) because of mean-reversion, so we adjust for this when merging in our $\widehat{\beta}_k$ estimates. To do this adjustment, we estimate the elasticity of Tornqvist ln(TFPR) with respect to ln(TFPR) using our panel data. We use this elasticity to construct expected Tornqvist ln(TFPR) conditional on observed ln(TFPR) in our unbalanced panel of plants. We merge our $\widehat{\beta}_k$ estimates into the unbalanced panel of of plants based on the value of each plant's expected Tornqvist ln(TFPR) and the cutoffs between the Tornqvist ln(TFPR) deciles from the panel estimation.

4. We estimate the variance of $\ln(\tau)$ to be the variance of ln(TFPR) plus the covariance of ln(TFPR) and ln($\widehat{\beta}_k$), in the unbalanced panel of plants. Each plant within the same decile has the same value of ln($\widehat{\beta}_k$). We focus on the output-share weighted variance of ln(TFPR) and its corrected version, though our correction works with unweighted variances as well.

5. To calculate a corrected measure of misallocation, we first construct a plant-level estimate of $\widehat{\tau}$ as ln $(\widehat{\tau}) = \ln(\text{TFPR}) +$ ln($\widehat{\beta}_k$) $+ \varepsilon$. We assume ε is lognormally distributed with a mean of zero and a variance given by – Cov[ln(TFPR), ln($\widehat{\beta}_k$)] – Var(ln($\widehat{\beta}_k$). Adding back the term ε produces a variance for ln $(\widehat{\tau})$ equal to the variance of ln(TFPR) plus its covariance with ln($\widehat{\beta}_k$), as described in the previous bullet. In this way we obtain a corrected estimate of τ for each plant in each sector — call this τ_{si}. To arrive at allocative efficiency requires a corrected estimate of τ for each sector — call this τ_s. τ_s is a geometric weighted average of the harmonic means of τ_{si}^K, τ_{si}^L and τ_{si}^X. We assume that the measurement error in TFPR affects MRPL, MRPK and MRPX in the same proportion. So ln $(\widehat{\tau}^K) = \ln(\text{MRPK}) +$ ln($\widehat{\beta}_k$) $+ \varepsilon$, and similarly for $\widehat{\tau}^L$ and $\widehat{\tau}^X$. Finally, correcting allocative efficiency also requires a corrected estimate of A. If measurement error is in inputs, then the error affects A in the same proportion it affects τ. If measurement error is in revenues, then we have that ln($\widehat{A}$) =

$\ln(A) + \frac{\sigma}{\sigma - 1} (\ln(\widehat{\beta}_k) + \varepsilon)$. We find that our corrected allocative efficiency estimates are not sensitive to assuming measurement error is in inputs or in revenues. Our baseline in the text, Section 7, assumes that the measurement error is in inputs.

References

Allcott, Hunt, Allan Collard-Wexler, and Stephen D. O'Connell, "How Do Electricity Shortages Affect Industry? Evidence from India," *American Economic Review*, 2016, *106* (3), 587–624.

Fort, Teresa C. and Shawn D. Klimek, "The Effect of Industry Classification Changes on U.S. Employment Composition," 2016.

Foster, Lucia, Cheryl Grim, and John Haltiwanger, "Reallocation in the Great Recession: Cleansing or Not?," *Journal of Labor Economics*, 2016, *24* (1), 293–331.