

Misallocation or Mismeasurement?

Mark Bills

University of Rochester and NBER

Peter J. Klenow

Stanford University and NBER

Cian Ruane *

International Monetary Fund

January 13, 2020

Abstract

The ratio of revenue to inputs differs greatly across plants within countries such as the U.S. and India. Such gaps may reflect misallocation which hinders aggregate productivity. But differences in measured average products need not reflect differences in true marginal products. We propose a way to estimate the gaps in true marginal products in the presence of measurement error. Our method exploits how revenue growth is less sensitive to input growth when a plant's average products are overstated by measurement error. For Indian manufacturing from 1985–2013, our correction lowers potential gains from reallocation by 20%. For the U.S. the effect is even more dramatic, reducing potential gains by 60% and eliminating 2/3 of a severe downward trend in allocative efficiency over 1978–2013.

*We are grateful to numerous seminar participants and, especially, Joel David for comments. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. This research was performed at a Federal Statistical Research Data Center under FSRDC Project 1440. All results have been reviewed to ensure that no confidential information is disclosed. The views expressed in this paper are those of the author and should not be attributed to the International Monetary Fund, its Executive Board, or its management.

1. Introduction

The ratio of revenue to inputs differs substantially across establishments within narrow industries in the U.S. and other countries. See the survey by Syverson (2011). One interpretation of such gaps is that they reflect differences in the value of marginal products for capital, labor, and intermediate inputs. Such differences may imply misallocation, with negative consequences for aggregate productivity. This point has been driven home by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). See Hopenhayn (2014) and Restuccia and Rogerson (2017) for surveys of the growing literature surrounding this topic.

Differences in measured average products need not imply differences in true marginal products. First, marginal products are proportional to average products only under Cobb-Douglas production. Second, and to our point of emphasis, measured differences in revenue per inputs could simply reflect poor measurement of revenue or costs. For example, the capital stock is typically a book value measure that need not closely reflect the market value of physical capital. Misstatement of inventories will contaminate and distort measures of gross output and intermediates, since these are inferred in part based on the change in finished, work in process, and materials inventories.¹

We propose and implement a method to quantify the extent to which measured average products reflect true marginal products in the presence of measurement error and overhead costs. Our method is able to detect measurement error in revenue and inputs which is additive but whose variance can scale up with the plant's true revenue and inputs. Our method cannot identify proportional measurement error, and therefore may yield a lower bound on the magnitude of measurement error.

¹See White, Reiter and Petrin (2018) for how the U.S. Census Bureau tries to correct for measurement errors in its survey data on manufacturing plants. Rotemberg and White (2019) argue that the use of imputation in the U.S. but (perhaps) not in India could account for why allocative efficiency seems higher in the U.S. than in India. Bartelsman, Haltiwanger and Scarpetta (2013) and Asker, Collard-Wexler and De Loecker (2014) discuss why revenue productivity need not reflect misallocation even aside from measurement error, due to overhead costs and adjustment costs, respectively.

The intuition for our method is as follows. Imagine a world with constant (proportional) differences in true marginal products. The only shocks are to idiosyncratic plant productivity. Productivity shocks will move true revenue and inputs around across plants in the same proportion.² Thus, in the absence of measurement error, revenue growth will be proportional to input growth across all plants. Now suppose, instead, that revenue is overstated for a given plant. If this measurement error is additive and fixed over time, then the plant's measured revenue will move by less in percentage terms in response to a change in its productivity. Similarly, if a plant has overstated inputs in an additive and fixed way, its measured inputs will move less than proportionately in response to productivity shocks. Thus, if a plant's revenue/inputs are overstated by measurement error, its measured revenue growth will be less responsive to its measured input growth. We can then gauge the importance of measurement error in the cross-section by the degree to which high average product plants exhibit a low elasticity of revenue with respect to inputs over time.

Our method applies to less stark environments with changing true marginal products and measurement error over time for plants. A key restriction we do require is that the measurement errors be orthogonal to the true marginal products. As we will show, our approach involves regressing revenue growth on input growth, on average products, and on their interaction. The coefficient on the interaction term speaks to how much measurement error is contributing to the dispersion in measured average products in the cross-section.

We apply our methodology to panel data on U.S. manufacturing plants from 1978–2013 and formal Indian manufacturing plants from 1985–2013. The U.S. data is from the Annual Survey of Manufacturers (ASM) plus ASM plants in Census years, both from the Longitudinal Research Database (LRD). The Indian data is from the Annual Survey of Industries (ASI). The LRD contains about 50,000 ASM plants per year, and the ASI about 43,000 plants per year.

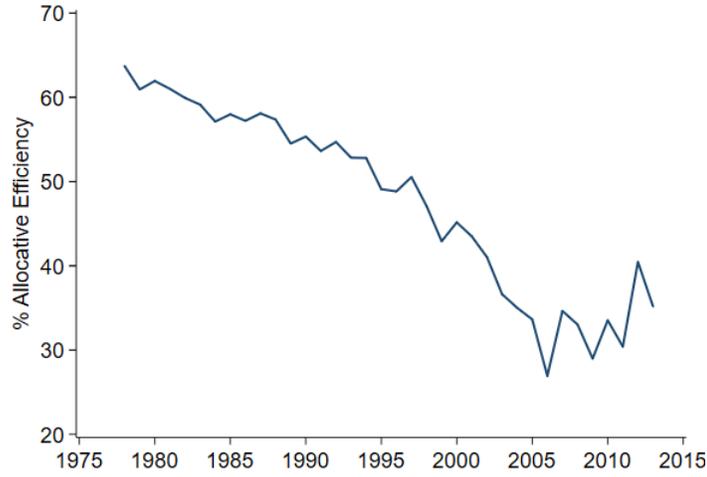
²Output increases more than inputs in response to a productivity shock, of course. But a plant's relative output price will decline with productivity so that its *revenue* will rise by the same proportion as its inputs. This is true if the plant's price-cost markup, or true ratio of revenue to inputs, does not change with its productivity.

We first report estimates of allocative efficiency without correcting for measurement error. The U.S. exhibits a severe decline, seemingly going from producing $3/5$ as much as it could by equalizing marginal products across plants to producing only $1/3$ as much as it could. Figure 1 displays this pattern. If true, this plunge reduced annual TFP growth rate by 2.5 percent per year over from 1978–2007. By comparison, we estimate that Indian manufacturing operated at about $1/2$ efficiency, with a fair bit of volatility from year to year but no clear trend despite major policy reforms. Thus by the end of the sample the U.S. appears to have *lower* allocative efficiency than India.

Once we correct for measurement error, U.S. allocative efficiency is much higher (above $2/3$) with a modest downward trend and much less volatility. Measurement error appears to be a growing problem in Census LRD plant data. In the Indian ASI, correcting for measurement error has a less dramatic effect. As a result, corrected allocative efficiency appears consistently higher in the U.S., raising manufacturing productivity by 10 to 50 percent relative to that in India (in all but one year).

The rest of the paper proceeds as follows. Section 2 presents a simple model wherein both measurement error and distortions are fixed over time. Section 3 presents the full model, which allows both measurement error and distortions to change over time. Section 4 describes the U.S. and Indian datasets, and raw allocative efficiency patterns in the absence of our correction for measurement error. Section 5 lays out our method for quantifying measurement error, and applies it to the panel data on manufacturing plants in the U.S. and India. As stated, these estimates impose the strong assumption that measurement error and true productivity are uncorrelated. We also rely on local approximations, so in Section 6 we examine how well our measure performs under alternative assumptions on the properties of shocks to productivity, distortions, and measurement error. Section 7 shows how correcting for measurement error affects the picture of allocative efficiency in the U.S. and India.

Figure 1: Allocative Efficiency in the U.S.



Source: U.S. LRD. The figure shows the % allocative efficiency in U.S. manufacturing between 1978 and 2013.

2. An Illustrative Model

In order to convey intuition for our methodology, we first present a simple model. We assume the economy has a fixed number of workers L and a single, competitive final goods sector producing aggregate output Y . Aggregate output is, in turn, produced by CES aggregation of the output Y_i of N intermediate goods producers with elasticity of substitution ϵ :

$$Y = \left(\sum_{i=1}^N Y_i^{1-\frac{1}{\epsilon}} \right)^{\frac{1}{1-\frac{1}{\epsilon}}}$$

The price index of the final good is given by $P = \left(\sum_i P_i^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}$ and is normalized to 1. Intermediate firms produce output using a linear production technology in labor under heterogeneous productivities: $Y_i = A_i L_i$. These firms are monopolistically competitive and face a downward sloping demand curve: $Y_i \propto P_i^{-\epsilon}$. They maximize profits taking as given Y , P , the wage w , and

an idiosyncratic revenue distortion τ_i :

$$\Pi_i = \frac{1}{\tau_i} P_i Y_i - w L_i$$

The researcher observes only *measured* revenue $\widehat{P}_i \widehat{Y}_i \equiv P_i Y_i + g_i$ and *measured* labor $\widehat{L}_i \equiv L_i + f_i$. Given the assumed CES demand structure, firms charge a common markup over their marginal cost (gross of the distortion):

$$P_i = \left(\frac{\epsilon}{\epsilon - 1} \right) \times \left(\tau_i \cdot \frac{w}{A_i} \right)$$

True revenue is therefore proportional to the product of true labor times and the idiosyncratic distortion:

$$P_i Y_i \propto \tau_i \cdot L_i \tag{1}$$

Thus variation across firms in true average revenue products $\left(\frac{P_i Y_i}{L_i} \right)$ is solely due to the distortion. Variation in *measured* average revenue products (TFPR), however, reflects both the distortion and measurement errors:

$$\text{TFPR}_i \equiv \frac{\widehat{P}_i \widehat{Y}_i}{\widehat{L}_i} \propto \left[\tau_i \times \frac{1 + g_i / (P_i Y_i)}{1 + f_i / L_i} \right].$$

While our methodology will allow both the true distortions and measurement errors to vary over time, to convey intuition we make the a number of simplifying assumptions in this section:

1. The true distortions τ_i are fixed over time
2. The additive measurement error terms g_i and f_i are fixed over time
3. The idiosyncratic productivities A_{it} are time-varying

Under these assumptions,

$$\Delta P_i Y_i = \Delta L_i = (\epsilon - 1) \Delta A_i.$$

Thus true revenue growth and true input growth are equal for a given plant. If we regress revenue growth on input growth, $\ln(\text{TFPR})$, and the interaction of input growth and $\ln(\text{TFPR})$ across plants in this economy, we should get a coefficient of 1 on input growth and 0 coefficients on both $\ln(\text{TFPR})$ and the interaction term if there is no measurement error whatsoever.

In the presence of measurement error, however, the relation between measured revenue and input growth in this economy is

$$\Delta \widehat{PY}_i = \Delta \widehat{L}_i \cdot \frac{\tau_i}{\text{TFPR}_i}.$$

The higher in TFPR relative to the true distortion τ , the lower is revenue growth relative to input growth. If we regress revenue growth on input growth, $\ln(\text{TFPR})$ and their interaction, we now expect a negative coefficient on the interaction term. The greater the variation in TFPR relative to τ across plants, the it should depress the responsiveness of revenue growth to input growth.

In Section 5 below we will generalize this logic to allow for shocks to both measurement error and distortions. The intuition from this simple example will remain: the extent to which high TFPR plants exhibit a low elasticity of revenue with respect to inputs will help us to estimate the role of measurement error in TFPR dispersion. We will use this, in turn, to estimate the variance of the true distortion and therefore the true level of allocative efficiency. In the next section we present the full model and a decomposition of aggregate and sectoral TFP into allocative efficiency vs. other terms.

3. Model

3.1. Economic Environment

We consider an economy with S sectors, L workers and an exogenous capital stock K . There are an exogenous number of firms N_s operating in each sector. A representative firm produces a single final good Q in a perfectly competitive final output market. This final good is produced using gross output Q_{st} from each sector s with a Cobb-Douglas production technology:

$$Q = \prod_{s=1}^S Q_s^{\theta_s} \quad \text{where} \quad \sum_{s=1}^S \theta_s = 1$$

We normalize P , the price of the final good, to 1. The final good can either be consumed or used as an intermediate input:

$$Q = C + X.$$

All firms use the same intermediate input, with the amounts denoted X_{si} so that $X \equiv \sum_{s=1}^S X_s = \sum_{s=1}^S \sum_{i=1}^{N_s} X_{si}$. Sectoral output Q_s is a CES aggregate of the output produced by the N_s firms in sector s :

$$Q_s = \left(\sum_{i=1}^{N_s} Q_{si}^{1-\frac{1}{\epsilon}} \right)^{\frac{1}{1-\frac{1}{\epsilon}}}.$$

We denote P_s the price index of output from sector s . Firms have idiosyncratic productivity draws A_{si} , and produce output Q_{si} using a Cobb-Douglas technology in capital, labor and intermediate inputs:

$$Q_{si} = A_{si} (K_{si}^{\alpha_s} L_{si}^{1-\alpha_s})^{\gamma_s} X_{si}^{1-\gamma_s} \quad \text{where} \quad 0 < \alpha_s, \gamma_s < 1.$$

The output elasticities α_s and γ_s are sector-specific, but time-invariant and common across firms within a sector. Firms are monopolistically competitive and

face a downward sloping demand curve given by $Q_{si} = Q_s \left(\frac{P_{si}}{P_s} \right)^{-\epsilon}$.³ Firms also face idiosyncratic labor distortions τ_{si}^L , capital distortions τ_{si}^K and intermediate input distortions τ_{si}^X . They maximize profits Π_{si} taking input prices as given.

$$\Pi_{si} = R_{si} - (1 + \tau_{si}^L)wL_{si} - (1 + \tau_{si}^K)rK_{si} - (1 + \tau_{si}^X)PX_{si}$$

where $R_{si} \equiv P_{si}Q_{si}$ is firm revenue.

3.2. Aggregate TFP

We define aggregate TFP as aggregate real consumption (or equivalently value-added) divided by an appropriately weighted Cobb-Douglas bundle of aggregate capital and labor:

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

We show in the [Model Appendix](#) that

$$TFP = \bar{T} \times \prod_{s=1}^S TFP_s^{\frac{\theta_s}{\sum_{s=1}^S \gamma_s \theta_s}}.$$

\bar{T} captures the effect of the *sectoral* distortions τ_s^L , τ_s^K and τ_s^X , which are the revenue-weighted harmonic means of the idiosyncratic firm-level distortions.⁴

Sectoral TFP is then:

$$TFP_s \equiv \frac{Q_s}{(K_s^{\alpha_s} L_s^{1-\alpha_s}) \gamma_s X_s^{1-\gamma_s}}$$

Within-sector misallocation lowers TFP_s . The sectoral distortions will induce a *cross-sector* misallocation of resources, which will show up in \bar{T} . While

³We assume that the firms treat P_s and Q_s as exogenous.

⁴ $(1 + \tau_s^L) \equiv \left[\sum_{i=1}^{N_s} \frac{R_{si}}{R_s} \frac{1}{1 + \tau_{si}^L} \right]^{-1}$ and similarly for $(1 + \tau_s^K)$ and $(1 + \tau_s^X)$. The [Model Appendix](#) expresses sectoral distortions as a function only of firm distortions and productivities.

cross-sector misallocation is of interest, it is not the focus of this paper. We therefore leave it to future research to determine how important this could be in determining cross-country aggregate TFP gaps.

3.3. Sectoral TFP Decomposition

Sector-level TFP is a function of firm-level productivities and distortions:

$$TFP_s = \left[\sum_{i=1}^{N_s} A_{si}^{\epsilon-1} \left(\frac{\tau_{si}}{\tau_s} \right)^{1-\epsilon} \right]^{\frac{1}{\epsilon-1}}$$

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

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

We can go one step further, and decompose sectoral TFP into the product of four terms: allocative efficiency (AE_s), a productivity dispersion term (PD_s), average productivity (\bar{A}_s) and a variety term ($N_s^{\frac{1}{\epsilon-1}}$).

$$TFP_s = \underbrace{\left[\frac{1}{N_s} \sum_i \left(\frac{A_{si}}{\tilde{A}_s} \right)^{\epsilon-1} \left(\frac{\tau_{si}}{\tau_s} \right)^{1-\epsilon} \right]^{\frac{1}{\epsilon-1}}}_{AE_s = \text{Allocative Efficiency}} \times \underbrace{\left[\frac{1}{N_s} \sum_i \left(\frac{A_{si}}{\bar{A}_s} \right)^{\epsilon-1} \right]^{\frac{1}{\epsilon-1}}}_{PD_s = \text{Productivity Dispersion}} \times \underbrace{N_s^{\frac{1}{\epsilon-1}}}_{\text{Variety}} \times \underbrace{\bar{A}_s}_{\text{Average Productivity}}$$

\tilde{A} is the power mean of idiosyncratic productivities, $\left[\frac{1}{N_s} \sum_{i=1}^{N_s} (A_{si})^{\epsilon-1} \right]^{\frac{1}{\epsilon-1}}$, and \bar{A} is the geometric mean of idiosyncratic productivities $\prod_{i=1}^{N_s} A_{si}^{\frac{1}{N_s}}$. AE_s is maximized and equal to 1 when there is no variation in the distortions across firms ($\tau_{si} = \tau_s \forall i$). The productivity dispersion term (PD_s) is the ratio of the power mean to the geometric mean. Because $\epsilon > 1$, greater dispersion in firm-level productivities induces a reallocation of labor towards the most productive firms, thereby increasing sectoral TFP. $N_s^{\frac{1}{\epsilon-1}}$ captures the productivity gains from expanding the set of varieties available to sectoral goods producers. Finally, it is clear why

increases in average productivity (\bar{A}_s) should increase sectoral TFP.

The goal of this paper is to present a methodology for inferring allocative efficiency (AE_s) from plant-level data while allowing for measurement error. In the next section we briefly describe the U.S. and Indian datasets we use, present the model-based approach to inferring allocative efficiency in the *absence* of measurement error, and show raw allocative efficiency patterns in the data.

4. Inferring Allocative Efficiency

4.1. The Datasets

We use two datasets of manufacturing plants in this paper: the Indian Annual Survey of Industries (ASI) from 1985 to 2013 and the U.S. Longitudinal Research Database (LRD) from 1978 to 2013.

The ASI is a nationally representative survey of formal manufacturing plants in India. The coverage includes plants with at least 10 workers using power, and plants with at least 20 workers not using power. Plants fall into two categories: *Census* and *Sample*. *Census* plants are surveyed every year, and consist of plants with at least 100 workers (the threshold increases to 200 workers in some survey years) as well as all plants in 12 of the “industrially backwards” states. *Sample* plants are randomly sampled each year within state \times industry cells. Official panel identifiers are available from 1998 on, and we use panel identifiers from an old version of the publicly available ASI prior to 1998. We construct an industry classification consisting of 50 manufacturing industries which are consistently defined throughout our time period.⁵

The LRD is a database of U.S. manufacturing plants put together by the U.S. Census Bureau. The coverage is all manufacturing plants with at least one employee. The database includes information from the Annual Survey of Manufactures (ASM) and the Census of Manufactures (CMF), augmented with

⁵The official sectoral classification (NIC) changed in 1987, 1998, 2004 and 2008. We use official NIC concordances to construct our harmonized classification.

establishment identifiers from the Longitudinal Business Database (LBD). The CMF is a census which is conducted in years ending in 2 or 7. The ASM is a survey which is conducted in all other years. The ASM covers large plants with certainty (typically plants with at least 100 workers, though the threshold varies by survey year) and randomly samples smaller plants. The ASM sample is redrawn in years ending in 4 and 9. In order to avoid any large changes in sample size over time, we use only the ‘ASM’ sample plants in CMF years. From here on, we refer to our U.S. dataset as the LRD. We use the harmonized sectoral classification from Fort and Klimek (2016) at the NAICS 3-digit level (86 sectors).

The main variables we use are gross output, labor, labor costs, capital, inventories, and intermediate inputs. We construct gross output as the sum of shipments, changes in finished and semi-finished good inventories, and other revenues. We construct intermediate inputs as the sum of materials, fuels and other expenditures. We include unpaid family workers in our measure of labor in India. We construct labor costs as the sum of wages, salaries, bonuses and supplemental labor costs. We set the capital stock as the sum of fixed assets and the stock of inventories. Because of data availability, we use the nominal book value of fixed assets in India, and the real market value of fixed assets in the U.S.⁶ We do not deflate any nominal variables. Industry-level deflators would get differenced out because all of our analyses focus on within-industry differences across plants. Official sampling weights are used in all of our calculations. We discuss more details about variable construction in our [Data Appendix](#).

We clean the ASI and LRD using the same approach. We drop plants with missing or negative values for any of the variables described above. We then trim the 1% tails of TFPR and TFPQ deviations from the industry average in each year (TFPR and TFPQ are defined in the next section). We describe these steps in more detail in our [Data Appendix](#). Our final sample sizes are 1,806,000 plant-years for the U.S. and 943,186 plant-years for India.⁷

⁶Book value capital stocks are not reported every year in the U.S., unlike investment in fixed assets. Our real capital stock measure is constructed using the perpetual inventory method.

⁷We round U.S. observation counts in accordance with Census data disclosure rules.

4.2. Inferring Allocative Efficiency

Continuing to use $\hat{\cdot}$'s to denote *measured* values, TFPR and TFPQ are:

$$\text{TFPR}_{si} \equiv \frac{\hat{R}_{si}}{(\hat{K}_{sit}^{\alpha_s} \hat{L}_{si}^{1-\alpha_s})^{\gamma_s} \hat{X}_{si}^{1-\gamma_s}} \quad (2)$$

$$\text{TFPQ}_{si} \equiv \frac{(\hat{R}_{sit})^{\frac{\epsilon}{\epsilon-1}}}{(\hat{K}_{sit}^{\alpha_s} \hat{L}_{si}^{1-\alpha_s})^{\gamma_s} \hat{X}_{si}^{1-\gamma_s}}$$

In the absence of measurement error, TFPR would be proportional to the distortion and TFPQ would be proportional to productivity:

$$\frac{R_{si}}{(\hat{K}_{sit}^{\alpha_s} \hat{L}_{si}^{1-\alpha_s})^{\gamma_s} \hat{X}_{si}^{1-\gamma_s}} \propto \tau_{si} \quad \text{and} \quad \frac{(R_{si})^{\frac{\epsilon}{\epsilon-1}}}{(\hat{K}_{sit}^{\alpha_s} \hat{L}_{si}^{1-\alpha_s})^{\gamma_s} \hat{X}_{si}^{1-\gamma_s}} \propto A_{si}$$

We infer sectoral allocative efficiency using the following expression:

$$\widehat{AE}_s = \left[\sum_{i=1}^{N_s} \left(\frac{\text{TFPQ}_{si}}{\text{TFPQ}_s} \right)^{\epsilon-1} \left(\frac{\text{TFPR}_{si}}{\text{TFPR}_s} \right)^{1-\epsilon} \right]^{\frac{1}{\epsilon-1}}$$

$$\text{where } \text{TFPQ}_s = \left[\sum_{i=1}^{N_s} \text{TFPQ}_{si}^{\epsilon-1} \right]^{\frac{1}{\epsilon-1}}$$

$$\text{and } \text{TFPR}_s = \left(\frac{\epsilon}{\epsilon-1} \right) \left[\frac{\text{MRPL}_s}{(1-\alpha_s)\gamma_s} \right]^{(1-\alpha_s)\gamma_s} \left[\frac{\text{MRPK}_s}{\alpha_s\gamma_s} \right]^{\alpha_s\gamma_s} \left[\frac{\text{MRPX}_s}{1-\gamma_s} \right]^{1-\gamma_s}$$

MRPL_s , MRPK_s and MRPX_s are the revenue-weighted harmonic mean values of the marginal products of labor, capital and intermediates, respectively. E.g.,

$$\text{MRPK}_s = \left[\sum_i \frac{\hat{R}_{si}}{\hat{R}_s} \frac{1}{\text{MRPK}_{si}} \right]^{-1}$$

$$\text{MRPK}_{si} = \left(\frac{\epsilon-1}{\epsilon} \right) \alpha_s \gamma_s \frac{\hat{R}_{st}}{\hat{K}_{si}}$$

Aggregating across sectors we obtain inferred aggregate allocative efficiency,

which is equal to true allocative efficiency when there is no measurement error:

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

In order to obtain estimates of allocative efficiency over time for the U.S. and India we need to pin down a number of parameters in the model. Based on evidence from Redding and Weinstein (2019), we pick a value of $\epsilon = 4$ for the elasticity of substitution across plants. Allocative *inefficiencies* are amplified under higher values of this elasticity. We infer α_s and γ_s based on country-specific average sectoral cost shares.⁸ We allow the aggregate output shares θ_{st} to vary across years, and base them on country-specific sectoral shares of manufacturing output.⁹ We use labor costs as our measure of labor input because it captures variation in human capital and hours worked across plants.

4.3. Time-Series Results

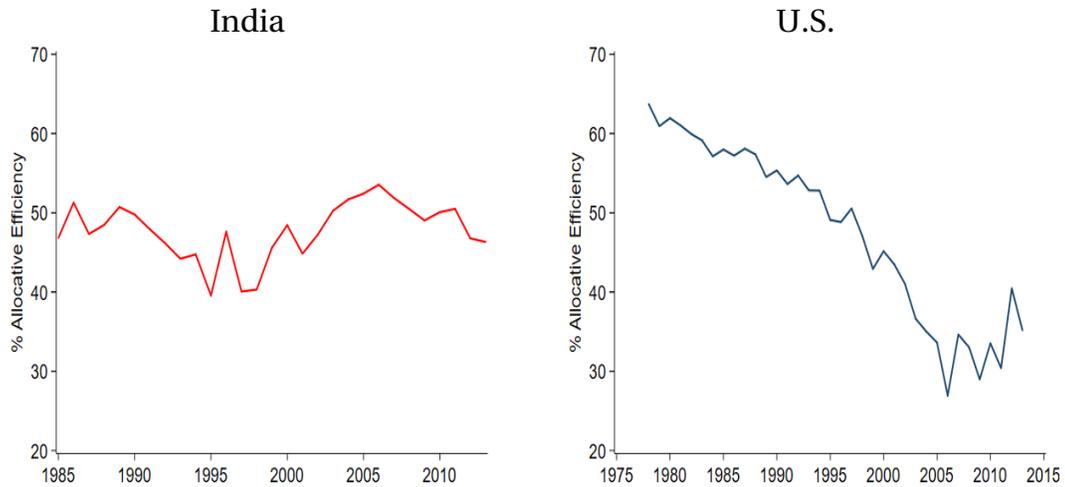
Figure 2 plots inferred allocative efficiency for India and the U.S. over their respective samples. While allocative efficiency exhibits no clear trend in India, there is a remarkable decrease in allocative efficiency in the U.S. from 1978 to 2006. As a result, over the entire sample allocative efficiency surprisingly averages 48% for both India and the United States.¹⁰ Figure 3 plots the ratio of U.S. to Indian allocative efficiency for the overlapping years of the samples. Allocative efficiency is *lower* in the U.S. than in India by around the year 2000. U.S. allocative efficiency averages only two-thirds of Indian allocative efficiency from 2003 to 2013.

⁸We assume a rental rate for fixed assets of 20% and a rental rate of 10% for inventories.

⁹Our results are not sensitive to the choice of constant or time-varying sectoral shares.

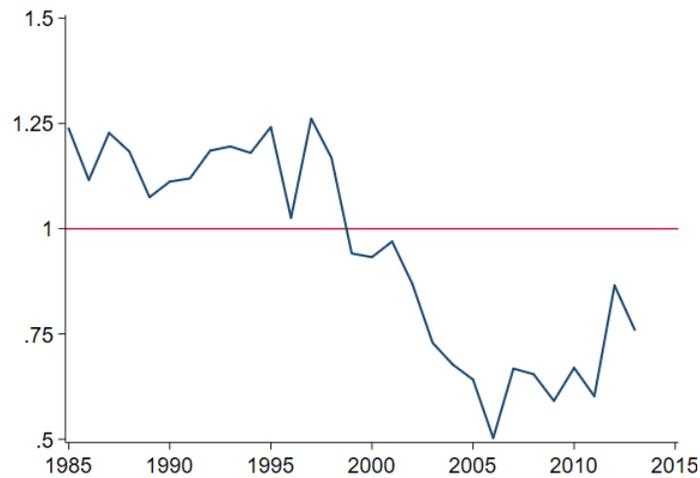
¹⁰Average gains from full reallocation are 123% for the U.S. versus 111% for India. In contrast, Hsieh and Klenow (2009) found 40-60% higher potential gains from reallocation in India than in the United States. Our estimates diverge from Hsieh and Klenow's for a number of reasons: gross output (us) vs. value added (them); 1978–2013 LRD plants vs. 1987, 1992 and 1997 Census plants; and trimming of 1% tails in the U.S. vs. 2% (they inconsistently trimmed 2% for the U.S. and only 1% for India).

Figure 2: Allocative Efficiency in India and the U.S.



Source: Indian ASI and U.S. LRD. The figure shows the % allocative efficiency for both countries. Average allocative efficiency is 49% in India and 54% in the U.S. over the respective sample periods.

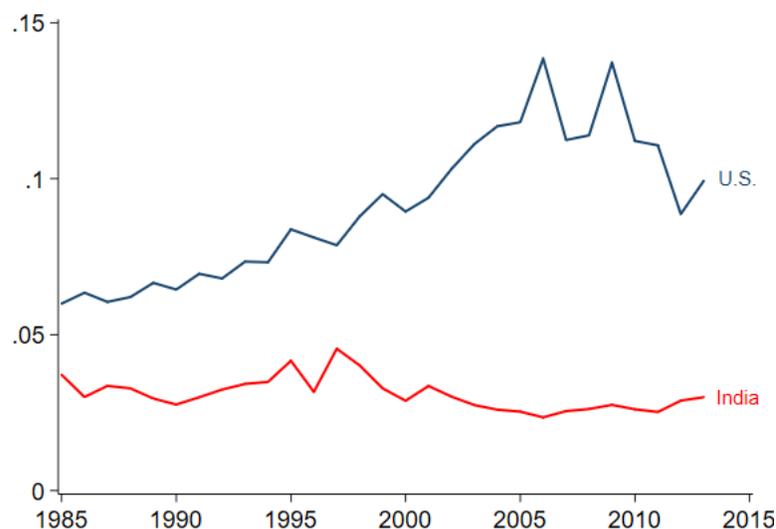
Figure 3: Allocative Efficiency, U.S. Relative to India



Source: Indian ASI and U.S. LRD. The figure plots the ratio of U.S. allocative efficiency to Indian allocative efficiency for the years 1985 to 2007 (years in which the datasets overlap).

The dramatic decline in allocative efficiency for the U.S. mirrors a sharp rise in the dispersion of its TFPR. Figure 4 displays the variance of $\ln(\text{TFPR})$ for both India and the U.S. for the years 1985 to 2013.¹¹ While the Indian data display little or no trend, the variance of $\ln(\text{TFPR})$ essentially doubles for the U.S. from 1985 to 2006–2009, before retreating part way by the end of the sample.¹² Figure 14 in the Appendix plots the 90:50 and 50:10 percentile ratios of $\ln(\text{TFPR})$ for the U.S., and shows that the increasing variance is mostly coming from the right tail.

Figure 4: Variance of $\ln(\text{TFPR})$



Source: Indian ASI and U.S. LRD. The figure plots the gross output share weighted variance of $\ln(\text{TFPR})$.

¹¹Figure 4 plots the output share weighted variance of $\ln(\text{TFPR})$, which maps more directly into allocative efficiency than the unweighted variance. The unweighted variance of $\ln(\text{TFPR})$ exhibits the same trend in the U.S., with the variance increasing from 0.06 in 1978 to 0.16 in 2009, before falling slightly to 0.13 in 2013.

¹²As noted above, allocative efficiency averaged 48% in both the Indian and U.S. samples. This may seem inconsistent with the markedly higher TFPR dispersion for the U.S. displayed in Figure 4. But India has a much higher intermediate input share than the U.S., which magnifies the impact of distortions on allocative efficiency.

Figure 15 in the Appendix shows that TFPQ dispersion rose in both the U.S. and India from 1985–2013. The variance in logs rose from about 0.35 to 0.45 or more in both countries. At the same time, the elasticity of TFPR with respect to TFPQ rose in the U.S. while falling in India — see Figure 16 in the Appendix. The elasticity rose from around 0.27 to 0.37 in the U.S. from 1985 to 2013. Gouin-Bonenfant (2019) formulates a theory in which rising TFPQ dispersion is responsible for both rising TFPR dispersion (and falling labor share of income) in the U.S. in recent decades. In addition to looking at trends in allocative efficiency, we would like to know the extent to which these patterns persist after adjusting for measurement error.

In the next section we present our methodology to correct TFPR for measurement errors with the goal of obtaining measures of allocative efficiency that are more robust to such errors.

5. Measurement Error

Calculations of misallocation, including those just presented, interpret plant differences in measured average revenue products (TFPR) as differences in true marginal products. In many of these studies the underlying plant data are longitudinal. We will show that, using such data, one can project the elasticity of revenue with respect to inputs on TFPR to answer the question: to what extent do plants with higher measured average products have higher true marginal products? The logic is similar to using the covariance of two noisy measures of a variable, here noisy measures of a plant's marginal revenue product, to recover a truer measure of the variable.

5.1. Measurement Error and TFPR

Consider the following description of measured inputs \widehat{I} and measured revenue \widehat{R} for plant i (year subscripts implicit):

$$\begin{aligned}\widehat{I}_i &\equiv \phi_i \cdot I_i + f_i \\ \widehat{R}_i &\equiv \chi_i \cdot R_i + g_i\end{aligned}$$

where I_i and R_i denote true inputs and revenues, f_i and g_i reflect additive measurement errors, and ϕ_i and χ_i are multiplicative errors.¹³ For simplicity, we treat the impact of measurement error in inputs as common across different inputs (capital, labor, intermediates).

In the setting of Section 2, profit maximization by each plant implies

$$\text{TFPR}_i \equiv \frac{\widehat{R}_i}{\widehat{I}_i} \propto \tau_i \left(\frac{\widehat{R}_i}{R_i} \frac{I_i}{\widehat{I}_i} \right).$$

Absent measurement error, TFPR_i provides a measure of plant i 's distortion τ_i . But, to the extent revenue is overstated or inputs are understated, TFPR_i will overstate τ_i . In that circumstance, the plant's marginal revenue product is *less* than implied by its TFPR_i .

The growth rate of a plant's TFPR will reflect the growth rate of its measurement error as well as the growth rate of its τ_i :

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

Δ denotes the growth rate of a plant variable relative to the mean in its sector.

If there are only *additive* measurement errors, then TFPR growth is

$$\Delta \text{TFPR}_i = \frac{\Delta \tau_i}{\widehat{R}_i/R_i} - \left(\frac{\widehat{R}_i - R_i}{\widehat{R}_i} - \frac{\widehat{I}_i - I_i}{\widehat{I}_i} \right) \Delta I_i + \frac{dg_i}{\widehat{R}_i} - \frac{df_i}{\widehat{I}_i}.$$

¹³Note that the additive terms f_i and g_i could alternatively reflect deviations from Cobb-Douglas production. For instance, positive values for f_i (such as overhead inputs), or negative for g_i , would imply marginal revenue exceeds average revenue per input.

As above, dx denotes the level change in x , as opposed to Δx , which denotes its percentage change. The response of TFPR_i to inputs speaks to the size of additive measurement error in revenue versus that in inputs. TFPR_i decreases when inputs rise if revenue is overstated relative to inputs ($\frac{\widehat{R}_i - R_i}{\widehat{R}_i} > \frac{\widehat{I}_i - I_i}{\widehat{I}_i}$), and TFPR_i increases when inputs rise when the reverse is true. Because relative measurement error, $\frac{\widehat{R}_i - R_i}{\widehat{R}_i}$ versus $\frac{\widehat{I}_i - I_i}{\widehat{I}_i}$, causes TFPR_i to mismeasure τ_i , the response of TFPR_i to inputs can identify the role of such errors in observed TFPR_i .

By contrast, if there are only *multiplicative* measurement errors, then the percentage change in TFPR equals:

$$\Delta \text{TFPR}_i = \Delta \tau_i + \Delta \chi_i - \Delta \phi_i.$$

Here TFPR_i growth provides no information on measurement error in the level of TFPR_i , except to the extent $\Delta \tau_i$, $\Delta \chi_i$, and $\Delta \phi_i$ project onto those errors. With proportional measurement errors, any increase in true inputs or revenue at plant i will scale up its measurement errors. Here errors that plague TFPR_i also contaminate the change in revenue relative to the change in inputs.

Going forward, we focus on purely additive measurement error. For this reason, our estimates should be viewed as a conservative assessment of the role of measurement error in generating differences in TFPR_i . We find that even this conservative assessment dramatically reduces the size and volatility of inferred misallocation. We further assume that measurement errors are mean zero.¹⁴ Finally, we assume that measurement errors are uncorrelated with the distortion τ_i across plants.

We next show how the relation between the responsiveness of a plant's revenue growth to its input growth and the level of its TFPR can address the role of measurement error in TFPR . We then present results for both U.S. and Indian manufacturing.

¹⁴We allow the variance of innovations to measurement error to scale with a plant's productivity A_i and distortion τ_i . For this reason, we do not predict that measurement errors become less important with trend growth or systematically differ large and small plants.

5.2. Identifying Measurement Error

Our focal point is the elasticity of measured revenue with respect to measured inputs conditional on the value of TFPR_k:

$$\beta_k \equiv \frac{Cov_k(\Delta \widehat{R}_i, \Delta \widehat{I}_i)}{Var_k(\Delta \widehat{I}_i)}$$

For exposition we first assume no measurement error, then allow for errors in both revenue and inputs. Absent measurement error, changes in revenue and inputs simply reflect changes in the plant's productivity and distortion:

$$\Delta \widehat{I}_i = \Delta I_i = (\epsilon - 1) \Delta A_i - \epsilon \Delta \tau_i, \tag{3}$$

$$\Delta \widehat{R}_i = \Delta R_i = (\epsilon - 1)(\Delta A_i - \Delta \tau_i).$$

So β_k is given by:

$$\beta_k = 1 + \phi_k, \tag{4}$$

$$\text{where } \phi_k \equiv \frac{Cov_k(\Delta \tau_i, \Delta I_i)}{Var_k(\Delta I_i)} = \frac{-\epsilon \cdot Var_k(\Delta \tau_i) + (\epsilon - 1) \cdot Cov_k(\Delta \tau_i, \Delta A_i)}{Var_k(\Delta I_i)}.$$

ϕ_k is the elasticity of $\Delta \tau$ with respect to ΔI .

β_k reflects a standard inference problem: a given increase in inputs creates a larger response in revenue if it is driven by A than if it is driven by a decline in τ . If $Var_k(\Delta \tau) = 0$ then $\beta_k = 1$, whereas if $Var_k(\Delta A) = 0$ then $\beta_k = \frac{\epsilon-1}{\epsilon} < 1$. If τ follows a random walk so that $\Delta \tau$ is i.i.d., then β_k reduces to $1 + \phi$ regardless of TFPR. For stationary τ , its conditional volatility is greater at extreme τ 's, reflecting τ 's regression back to its mean. Thus the $Var_k(\Delta \tau)$ is greater at extremes for TFPR, implying smaller values for β_k .

With measurement error in plant revenues and inputs, (4) becomes:

$$\beta_k \approx E_k \left(\frac{R_i \widehat{I}_i}{\widehat{R}_i I_i} \right) (1 + \phi_k) + \psi_k, \quad (5)$$

$$\text{where } \phi_k \equiv \frac{\text{Cov}_k \left(\frac{I_i}{I_i} \Delta \tau_i - \frac{df_i}{I_i}, \Delta \widehat{I}_i \right)}{\text{Var}_k \left(\Delta \widehat{I}_i \right)} \quad \text{and}$$

$$\begin{aligned} \psi_k \equiv \frac{1}{\text{Var}_k \left(\Delta \widehat{I}_i \right)} & \left(\text{Cov}_k \left(\frac{dg_i}{\widehat{R}_i}, \Delta \widehat{I}_i \right) + \text{Cov}_k \left(\frac{R \widehat{I}_i}{\widehat{R}_i I_i}, \Delta \widehat{I}_i \left(\Delta \widehat{I}_i + \frac{I_i}{\widehat{I}_i} \Delta \tau - \frac{df_i}{\widehat{I}_i} \right) \right) \right. \\ & \left. - E_k \left(\Delta \widehat{I}_i \right) \text{Cov}_k \left(\frac{R_i \widehat{I}_i}{\widehat{R}_i I_i}, \left(\Delta \widehat{I}_i + \frac{I_i}{\widehat{I}_i} \Delta \tau - \frac{dg_i}{\widehat{I}_i} \right) \right) \right). \end{aligned}$$

The approximate equality in (5) means it is a good approximation for relatively small changes in A , τ , f , and g .

Comparing equations (4) and (5), with measurement error the factor $1 + \phi_k$ in β_k scales by $E_k \left(\frac{R \widehat{I}}{R I} \right)$. For instance, if revenue is overstated, then any implied increase in revenue is only manifested to the proportion R/\widehat{R} . If inputs are over measured, then any true increase in inputs is scaled down by \widehat{I}/I , so the response in revenues is scaled *up* by \widehat{I}/I . The same factor $\frac{R \widehat{I}}{R I}$ confounds TFPR as a measure of the true distortion τ , with the expectation of τ at a particular TFPR level k given by:

$$E_k (\tau_i) = E_k \left(\frac{R_i \widehat{I}_i}{\widehat{R}_i I_i} \right) \cdot \text{TFPR}_k.$$

Using the definition of β_k from (5), we have:

$$E_k (\tau_i) = E_k \left(\frac{R_i \widehat{I}_i}{\widehat{R}_i I_i} \right) \cdot \text{TFPR}_k = \left(\frac{\beta_k - \psi_k}{1 + \phi_k} \right) \cdot \text{TFPR}_k. \quad (6)$$

Measurement error affects the interpretation of ϕ_k and potentially introduces the factor ψ_k . ϕ_k still reflects the inference problem that the elasticity of measured revenue with respect to measured inputs depends on the source of the change in inputs. There are now three sources of change in measured inputs: ΔA , $\Delta\tau$ and df . Increases in measured inputs driven either by a decrease in τ or an increase in f result in smaller responses in revenue. Turning to ψ_k , its first element allows for the possibility that changes in measurement errors in revenue are potentially correlated with measured changes in inputs. The latter terms are more subtle, reflecting any correlations between the measurement error component of TFPR and changes in inputs or measurement errors.

If $\Delta\tau$, ΔA , $\frac{df}{\widehat{I}}$ and $\frac{dg}{\widehat{R}}$ are each i.i.d., then $\psi_k = 0$ and ϕ_k reduces to ϕ , independent of the level of TFPR. In this case $\beta_k \approx E_k(\frac{\widehat{RI}}{RI})(1 + \phi)$ and (6) yields:

$$E_k(\tau_i) \propto \beta_k \cdot \text{TFPR}_k. \quad (7)$$

Thus, given estimates of the β_k 's, we can answer the question: If two plants differ in TFPR, what is the expected difference in their actual marginal revenue products due to differences in their τ 's. If differences in TFPR partially reflect errors f or g , then $\widehat{\beta}_k$'s will be systematically lower at higher TFPR's.

We will use equation (7) as a benchmark to correct TFPR for measurement error. In general, $\psi_k \neq 0$ and ϕ_k will depend on the level of lagged TFPR. To address these possibilities, we will simulate a model economy under plausible scenarios to see if our simple correction using (7) overstates or understates the role of measurement error in TFPR dispersion. In particular, we will simulate models where $\Delta\tau$, ΔA , $\frac{df}{\widehat{I}}$ and $\frac{dg}{\widehat{R}}$ are not i.i.d., for instance due to τ , A , or measurement errors being stationary. To anticipate, we will find that our correction based on (7) is quite accurate for a wide range of parameter values, especially if measurement errors are only moderately large. For very large measurement errors, we find (7) tends to *under* estimate the role of measurement, rendering our corrections somewhat conservative.

From (7), we can capture the dispersion in τ 's that is predicted by TFPR's. But in the presence of measurement error there will also be differences in τ 's that are orthogonal to TFPR. For instance, a plant with a high value for τ but understated revenue might display purely average TFPR. To capture this component, we "add back" variation in τ 's that is orthogonal to TFPR, under an assumption that true τ is orthogonal to the measurement error component of TFPR.

From the relation between TFPR, τ , and measurement error, we have:

$$Var\left(\ln \tau_i\right) = Var\left(\ln TFPR_i\right) - Var\left(\ln\left(\frac{R_i \hat{I}_i}{\widehat{R}_i I_i}\right)\right) + 2Cov\left(\ln \tau_i, \ln\left(\frac{R_i \hat{I}_i}{\widehat{R}_i I_i}\right)\right) \quad (8)$$

We assume that the two components of TFPR, namely the true distortion τ and the measurement error component $\frac{\hat{R}I}{\widehat{R}I}$, are orthogonal to each other in their natural logs. Then the last term in equation (8) is eliminated. Furthermore, using that $\ln\left(\frac{R\hat{I}}{\widehat{R}I}\right)$ equals $(-\ln TFPR + \ln \tau)$ and can also be broken into $E_k\left(\ln\left(\frac{R\hat{I}}{\widehat{R}I}\right)\right)$ and $\left(\ln\left(\frac{R\hat{I}}{\widehat{R}I}\right) - E_k\left(\ln\left(\frac{R\hat{I}}{\widehat{R}I}\right)\right)\right)$, the middle term can be written as the two terms:

$$-Var\left(\ln\left(\frac{R_i \hat{I}_i}{\widehat{R}_i I_i}\right)\right) = Cov\left(\ln TFPR_k, E_k\left(\ln\left(\frac{R_i \hat{I}_i}{\widehat{R}_i I_i}\right)\right)\right) - Cov\left(\ln \tau_i, \ln\left(\frac{R_i \hat{I}_i}{\widehat{R}_i I_i}\right)\right),$$

which reduces to its first term, given the assumption τ and $\frac{\hat{R}I}{\widehat{R}I}$ are orthogonal.

Thus equation (8) can be reduced to:

$$Var\left(\ln \tau_i\right) = Var\left(\ln TFPR_i\right) + Cov\left(\ln TFPR_k, E_k\left(\ln\left(\frac{R_i \hat{I}_i}{\widehat{R}_i I_i}\right)\right)\right). \quad (9)$$

The first term here is data. The second is provided by how the estimates of β_k , discussed earlier in this subsection, covary with TFPR. Based on (9), we add back dispersion in τ that is orthogonal to a plant's TFPR. We assume this dispersion is drawn from a lognormal distribution, with a variance as dictated by (9).

5.3. Estimates for India and the U.S.

For both India and the U.S. we start with plants that are observed in consecutive years. We then divide each country's data into separate time windows. For India we have 25 years in growth rates (1985–1986 to 2013–2014) that we split into five windows of approximately six years each.¹⁵ For the U.S. we divide 36 years in growth rates (1978–1979 to to 2013–2014) into five windows of approximately seven years each. We then regress plant revenue growth on plant input growth by decile of TFPR in each window:

$$\Delta \widehat{R}_i = \widehat{\lambda}_k + \widehat{\beta}_k \Delta \widehat{I}_i + e_i, \quad (10)$$

Here i denotes the plant and k the decile of TFPR. $\Delta \widehat{R}_i$, $\Delta \widehat{I}_i$ and TFPR are each deviations from the sector-year average for that plant. The individual decile-windows contain on average 13,000 plants per year for India and 41,000 for the U.S. Within each decile, plants are weighted by their share of total input costs.¹⁶

Given the estimates $\widehat{\beta}_k$, we can then construct a corrected cross-sectional distribution of distortions τ 's for each time frame in each country according to:

$$\ln(\widehat{\tau}) = \ln(\text{TFPR}) + \ln(\widehat{\beta}_k) + \varepsilon. \quad (11)$$

That is, each plant is assigned the β_k estimated for its TFPR decile.¹⁷ ε is drawn from a log normal distribution with variance dictated by equation (9), together with the estimates from equation (10) for the plant's time window.

¹⁵There are breaks in the panel identifiers, which means that we don't have growth rates for every set of consecutive years.

¹⁶More precisely, a plant's share of input costs reflects the (Tornqvist) average of its share across the two annual observations being differenced. Similarly, a plant's relative TFPR reflects the average of its relative TFPR's across the two years. In constructing a plant's input growth rate, growth in the plant's individual inputs (intermediates, labor, and capital) are weighted by its sector's average input shares. Observations where TFPR increases or decreases by a factor greater than five are excluded.

¹⁷Some years in the cross-sectional sample are reflected in two separate panel samples on growth rates. For instance the year 2002 for India is spanned by the panel window covering growth rates from 1997-98 to 2001-02 as well as the window for growth rates from 2002-03 to 2007-08. For such years we average the $\widehat{\beta}$'s for its backward and forward-looking windows.

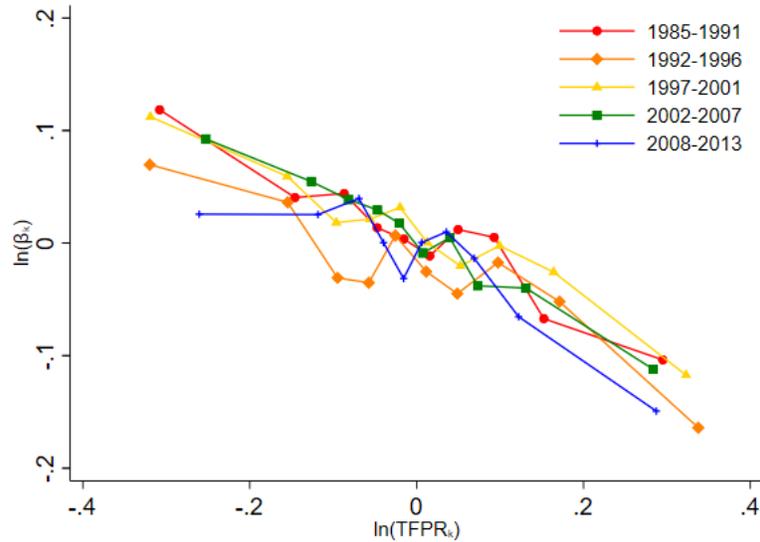
Measurement error is manifested in a lower β at higher levels of TFPR. Before looking across time windows, we first report $\hat{\beta}$ by decile of TFPR, pooling all years for each country. The estimates are in Table 1. Looking first at India, we see a clear tendency for $\hat{\beta}$ to decline with the level of TFPR. This decline is most pronounced in the top two deciles of TFPR. The difference in $\hat{\beta}$ going from the bottom to the top decile is -0.205 with a standard error of 0.020 . For the U.S. the negative relationship between TFPR and $\hat{\beta}$ is even more striking. $\hat{\beta}$ declines monotonically with TFPR and, as in India, most sharply in the top two deciles of TFPR. The difference in $\hat{\beta}$ between the top and bottom deciles is -0.505 with a standard error 0.025 .

Table 1: Coefficients $\hat{\beta}_k$ for revenue growth on input growth by TFPR decile

Decile $k \rightarrow$	1	2	3	4	5	6	7	8	9	10
India	1.087	1.040	1.027	0.996	1.007	0.991	0.996	0.989	0.951	0.882
1985–2013	(.017)	(.012)	(.010)	(.014)	(.011)	(.011)	(.009)	(.014)	(.010)	(.011)
U.S.	1.046	0.995	0.969	0.961	0.931	0.896	0.863	0.837	0.704	0.541
1978–2013	(.013)	(.019)	(.015)	(.019)	(.020)	(.016)	(.014)	(.019)	(.021)	(.021)

Source: Indian ASI and U.S. LRD. Entries are coefficients from regressing revenue growth on input growth by decile of $\ln(\text{TFPR})$, as shown in equation (10). TFPR deciles are constructed as Tornqvist deviations from the (cost-weighted) sector-year average. Decile 1 corresponds to the lowest decile of TFPR, and Decile 10 the highest decile of TFPR. Regressions are weighted by each plant's (Tornqvist) share of all input costs. Standard errors are clustered at the industry level.

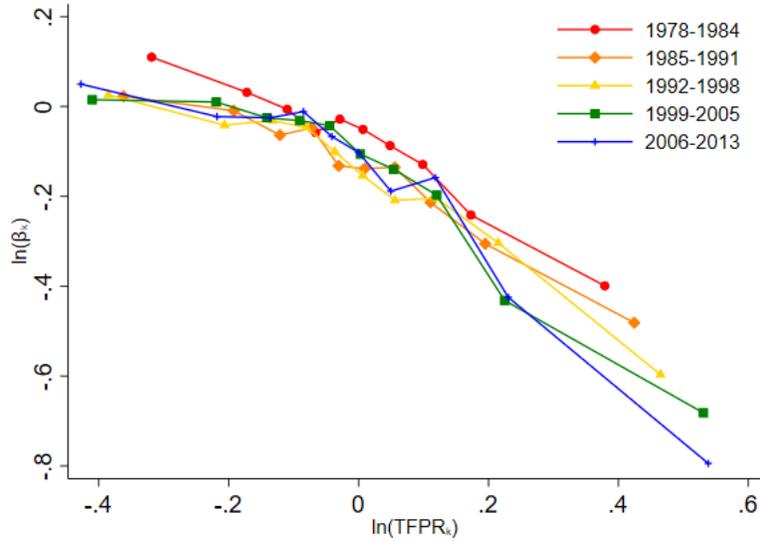
How does the relation between TFPR and $\hat{\beta}$ change through time? Figures 5 and 6 display these estimates for India and the U.S., with $\ln(\hat{\beta})$ plotted against $\ln(\text{TFPR})$ by sample period. For India in Figure 5, $\hat{\beta}$ decreases with TFPR to a similar degree in all periods. Figure 6 reveals the increase in U.S. TFPR dispersion over time, manifested by the increasing spread with respect to TFPR in later periods. Over time $\hat{\beta}$ becomes much more negatively related to TFPR in the United States. Thus the increased dispersion in TFPR in the U.S. is associated with an increased role for measurement error in that dispersion.

Figure 5: India β Slopes

Source: Indian ASI. The figure plots the $\hat{\beta}_k$ coefficients (recovered from running the regressions in equation 10) against deciles of TFPR. In each time window and for each TFPR decile k , revenue growth is regressed on input growth to obtain $\hat{\beta}_k$. The relationship is plotted separately for five different time windows.

From these $\hat{\beta}_k$ estimates, together with (11), we can estimate dispersion in $\ln(\tau)$'s over time. We report these in Table 2, expressed in terms of the variance of $\ln(\hat{\tau})$ relative to the variance of $\ln(\text{TFPR})$. For India, the variance of $\hat{\tau}$ is consistently about 30% smaller than the variance of TFPR. For the U.S. we estimate a larger and growing role for measurement error. In the U.S. the variance of $\hat{\tau}$ is about 70% lower than that for TFPR from 1999 onward. The implication, we shall see, is that measurement error exaggerates misallocation in both countries, but especially in the U.S. in recent years. But first we use simulations of a wider range of models to examine the robustness of our approach to identifying measurement error in the next section.

Figure 6: U.S. β Slopes



Source: U.S. LRD. The figure plots the $\hat{\beta}_k$ coefficients (recovered from running the regressions in equation 10) against deciles of TFPR. In each time window and for each TFPR decile k , revenue growth is regressed on input growth to obtain $\hat{\beta}_k$. The relationship is plotted separately for five different time windows.

Table 2: Dispersion in Marginal Products vs. TFPR Dispersion

Panel A: India					
	1985–1991	1992–1996	1997–2001	2002–2007	2008–2013
$\frac{\sigma_{\hat{\tau}}^2}{\sigma_{TFPR}^2}$	0.74	0.76	0.75	0.70	0.73
Panel B: U.S.					
	1978–1984	1985–1991	1992–1998	1999–2005	2006–2013
$\frac{\sigma_{\hat{\tau}}^2}{\sigma_{TFPR}^2}$	0.40	0.43	0.38	0.32	0.27

Source: Indian ASI and U.S. LRD. The table shows the ratio of the variance of $\ln(\hat{\tau})$ to the variance of $\ln(\text{TFPR})$, for both India and the U.S. for five different time periods. Variances are output share weighted. $\hat{\tau}$ is constructed as in equation (11).

6. Robustness in Simulations

Our approach to estimating the share of true τ dispersion in TFPR dispersion rests on a set of assumptions, most notably that measurement errors are additive and that their role in a plant's TFPR is orthogonal to its true τ distortion.¹⁸ Furthermore, if innovations to A , τ , or measurement errors are not i.i.d., then our corrections based on equation (6) are potentially clouded as terms ϕ_k and ψ_k may influence the projection of β_k on TFPR. Finally, our derivations rely on first-order approximations which may not perform well for large shocks to productivity, distortions or measurement errors.

For these reasons, we explore the performance of our estimator in simulations. We match simulated moments to data moments for India and the United State. We find that our estimator performs well even if measurement error is sizable, such as we estimate for India. If measurement error is enormously important, as we find for the U.S., then our approach is conservative, as it tends to understate the role of measurement error in TFPR dispersion.

We assume that plant i 's idiosyncratic productivity in period t is given by:

$$A_{it} = A_i \cdot \tilde{A}_{it}.$$

A_i is the permanent component of a plant's productivity, which we assume is lognormally distributed: $\ln(A_i) \sim N(0, \sigma_a^2)$. \tilde{A}_{it} is the transitory component of plant productivity. Plants also face an idiosyncratic, time-varying distortion τ_{it} . \tilde{A}_{it} and τ_{it} follow:

$$\ln(\tilde{A}_{it}) = \rho_a \cdot \ln(\tilde{A}_{it-1}) + \eta_{it}^a \text{ where } \eta_{it}^a \sim N(0, \sigma_a^2) \tag{12}$$

$$\ln(\tau_{it}) = \rho_\tau \cdot \ln(\tau_{it-1}) + \eta_{it}^\tau \text{ where } \eta_{it}^\tau \sim N(0, \sigma_\tau^2).$$

¹⁸Assuming that a firm's τ is orthogonal to the size of its measurement errors f and g does not translate directly to orthogonality of τ and the measurement error component of TFPR, because a plant's τ affects its scale and thereby the relative importance of its measurement errors.

Measurement errors in inputs and revenues follow AR(1) processes, with the variance of the shocks scaling with the size of the plant:

$$f_{it} = \rho_f \cdot f_{it-1} + \eta_{it}^f \cdot I_{it} \text{ where } \eta_{it}^f \sim N(0, \sigma_f^2) \quad (13)$$

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

As a baseline, we consider the case with measurement error in inputs only. We set $\epsilon = 4$ and $\rho_a = \rho_r = \rho_f = 0.9$ and use the simulated method of moments (simulating 30,000 plants over 50 years) to calibrate σ_τ , σ_a , $\sigma_{\tilde{a}}$ and σ_f .¹⁹ We target four moments from the data: the (output share weighted) variance of $\ln(\text{TFPR})$, the unweighted variance of $\ln(\text{TFPQ})$, the slope of $\ln(\hat{\beta})$ vs. Tornqvist $\ln(\text{TFPR})$ across deciles, and the variance of input growth.²⁰

The parameters are jointly calibrated, but they are differentially important for certain moments. The variance of $\ln(\text{TFPR})$ and the $\ln(\hat{\beta})$ slope are particularly important in disciplining the values of σ_τ and σ_f . The variance of $\ln(\text{TFPQ})$ is sensitive to the variance of the permanent component of firm productivity σ_a , while the variance of input growth relates strongly to the variance of productivity shocks $\sigma_{\tilde{a}}$. Table 3 shows our estimated parameter values. The targeted data moments and simulated moments for each time period in India and the U.S. are in Table 4. The simulated moments are always close to the targeted moments — though not exactly the same because of non-linearities.

The main outcome of interest is how accurately our estimator captures the variance of distortions relative to TFPR dispersion, $\sigma_\tau^2 / \sigma_{\text{TFPR}}^2$. We compare our estimated share $\sigma_{\tilde{\tau}}^2 / \sigma_{\text{TFPR}}^2$ to the true share for India and the U.S. in Table 5. Our estimator performs remarkably well for India in all time periods. While

¹⁹We clean the simulated data in the same way we do the actual micro data: dropping observations with negative revenues or inputs and those where TFPR changes by a factor of five or more. We simulate all plants for 50 years, and then construct the simulated moments as the average value over the last 30 years.

²⁰We minimize the sum of the absolute \ln differences between the targeted data moments and the simulated moments.

Table 3: Calibrated Parameter Values

	σ_τ	σ_a	$\sigma_{\tilde{a}}$	σ_f
Panel A: India				
1985–1991	0.060	0.637	0.181	0.030
1992–1996	0.074	0.711	0.102	0.040
1997–2001	0.073	0.712	0.088	0.042
2002–2007	0.056	0.733	0.110	0.039
2008–2013	0.060	0.744	0.113	0.035
Panel B: U.S.				
1978–1984	0.036	0.520	0.103	0.079
1985–1991	0.053	0.517	0.075	0.088
1992–1998	0.050	0.542	0.071	0.103
1999–2005	0.044	0.500	0.070	0.123
2006–2013	0.017	0.494	0.103	0.130

Source: This table shows the parameter values recovered from the model calibration. The model is calibrated for both India and the U.S. separately for five different time periods. 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, $\sigma_{\tilde{a}}$ is the standard deviation of the time-varying component of plant productivity, and σ_f is the standard deviation of shocks to measurement error in inputs.

conservative, in that it understates the role of measurement error, it deviates from the true share of distortions in TFPR dispersion by less than 1.3 percentage points on average. For the U.S. the discrepancy is larger. We overestimate the share of true distortions in TFPR dispersion for all periods, particularly when there is a lot of measurement error. For the 2006–2013 window, our estimator says the variance of distortions is one fourth the variance of TFPR, when in fact it is only one fiftieth. Despite this, our estimator does capture reasonably well the movements over time in the share of distortions in TFPR dispersion.

We obtain similar results when we add the autocorrelation of TFPR as a targeted moment, with $\rho = \rho_\tau = \rho_f$ as an additional parameter to calibrate. Likewise when we consider measurement error in revenues rather than inputs. Although we have more difficulty simultaneously hitting all targeted data moments, our estimator continues to be a conservative gauge of the importance of measurement error. We provide these results in our [Simulation Appendix](#).

Table 4: Data Moments versus Simulated Moments

	σ_{TFPR}^2		σ_{TFPQ}^2		$\ln(\hat{\beta})$ slope		$\sigma_{\Delta I}^2$	
	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_{TFPR}^2 is the output share weighted variance of $\ln(TFPR)$. σ_{TFPQ}^2 is the variance of $\ln(TFPQ)$. $\ln(\hat{\beta})$ slope is the slope of $\hat{\beta}$ against mean Tornqvist $\ln(TFPR)$ across deciles. $\sigma_{\Delta I}^2$ is the variance of input growth.

In our [Simulation Appendix](#) we also check how adjustment costs and multiplicative measurement error affect our estimator. As expected, if all measurement error is multiplicative we find that our estimator predicts that TFPR dispersion stems entirely from distortions. So, in the presence of multiplicative errors, our estimated share of measurement error in TFPR dispersion is conservative. We also consider a model with adjustment costs whereby a plant chooses its inputs one period ahead, before its productivity shock is observed. Our estimator interprets TFPR dispersion due to adjustment costs as if this dispersion was due to true distortions. This would suggest that, if adjustment costs are important in the U.S. and India, the share of TFPR dispersion due to τ dispersion may be even lower than what our estimator finds.

Table 5: Simulations: Our Estimator vs. Truth

Panel A: India					
	1985–1991	1992–1996	1997–2001	2002–2007	2008–2013
$\frac{\sigma_{\hat{\tau}}^2}{\sigma_{\text{TFPR}}^2}$ (our estimator)	0.678	0.712	0.699	0.640	0.692
$\frac{\sigma_{\tau}^2}{\sigma_{\text{TFPR}}^2}$ (truth)	0.604	0.693	0.682	0.581	0.656
Panel B: U.S.					
	1978–1984	1985–1991	1992–1998	1999–2005	2006–2013
$\frac{\sigma_{\hat{\tau}}^2}{\sigma_{\text{TFPR}}^2}$ (our estimator)	0.334	0.407	0.366	0.304	0.236
$\frac{\sigma_{\tau}^2}{\sigma_{\text{TFPR}}^2}$ (truth)	0.143	0.240	0.181	0.107	0.015

Source: This table shows, for simulated data, the ratio of the variance $\ln(\hat{\tau})$ to the variance of $\ln(\text{TFPR})$, and the ratio of the true variance of $\ln(\tau)$ to the variance of $\ln(\text{TFPR})$. Results are for both India and the U.S., and for five different time-periods each. The parameters used to generate each of these results are shown in Table 3.

7. Revisiting Misallocation

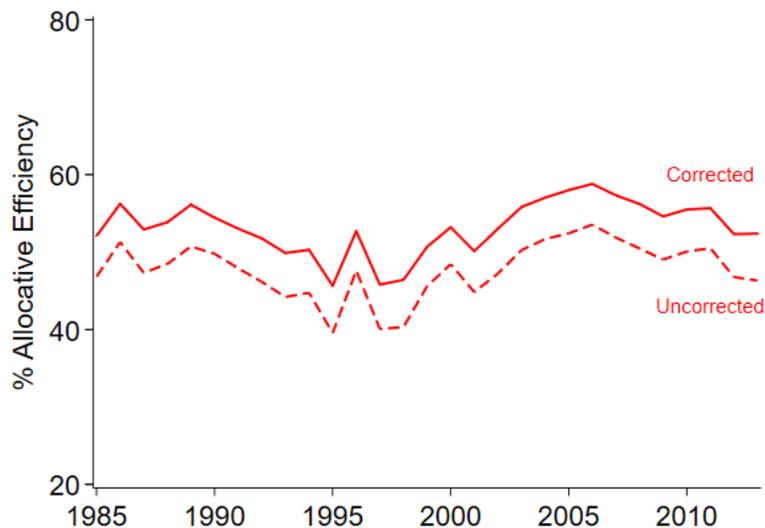
We now compare the “raw” measures of allocative efficiency for Indian and U.S. manufacturing to our estimates purging the impact of measurement error. This is achieved by replacing TFPR as an estimate of τ with the estimated dispersion of τ 's implied by equation (11), which we repeat here for convenience:

$$\ln(\hat{\tau}) = \ln(\text{TFPR}) + \ln(\hat{\beta}_k) + \varepsilon.$$

To construct allocative efficiency we also need measures of plant TFPQ and sectoral TFPR. Measurement error in inputs affects TFPQ and TFPR in the same way, so we get that $\ln(\hat{A}) = \ln(\text{TFPQ}) + \ln(\hat{\beta}_k) + \varepsilon$.²¹ To construct sectoral TFPR we need corrected measures of sectoral MRPL, MRPX and MRPK. We assume that measurement error is common to all inputs, and therefore affects each of

²¹Measurement error in revenues is amplified by $\frac{\sigma}{\sigma-1}$, but this is not quantitatively important.

Figure 7: Allocative Efficiency in India



Source: Indian ASI. The figures show uncorrected and corrected allocative efficiency (AE) for years 1985 to 2013. Average uncorrected AE is 47.7% while average corrected AE is 53.4%.

these in the same way.²² Our estimates for $\ln(\hat{\beta}_k)$ are in Figures 5 and 6.

We display corrected vs. uncorrected allocative efficiency for India in Figure 7. Averaging across years, the correction increases allocation efficiency modestly from 48% to 53%. The impact of the correction is fairly stable for India across the 29 years. Another way to express misallocation is in terms of the increase in productivity that can be reaped by attaining perfect (100%) allocative efficiency. This is reported for India in Table 6. With distortions measured by TFPR dispersion, the potential increase in productivity is 111%. Based on the corrected numbers it is 21 percent lower, at 89% — still quite massive.

In Figure 8 we see a far greater impact in the United States. Our correction eliminates more than half of potential gains from reallocation and the lion's share of the conspicuous downward trend in allocative efficiency. Allocative efficiency declines by 15% over the 35 years according to our corrected series. While significant, this is only a third of the 45% decline with no corrections.

²²We have that $\ln(\widehat{\text{MRPL}}) = \ln(\text{MRPL}) + \ln(\hat{\beta}_k) + \varepsilon$, and similarly for MRPK and MRPX.

Table 6: Uncorrected and Corrected Gains from Reallocation

	India		U.S.	
	Mean	S.D.	Mean	S.D.
Uncorrected gains	110.8%	17.1%	123.1%	59.7%
Corrected gains	89.0%	12.9%	49.2%	12.1%
Shrinkage	21%	20%	60%	80%

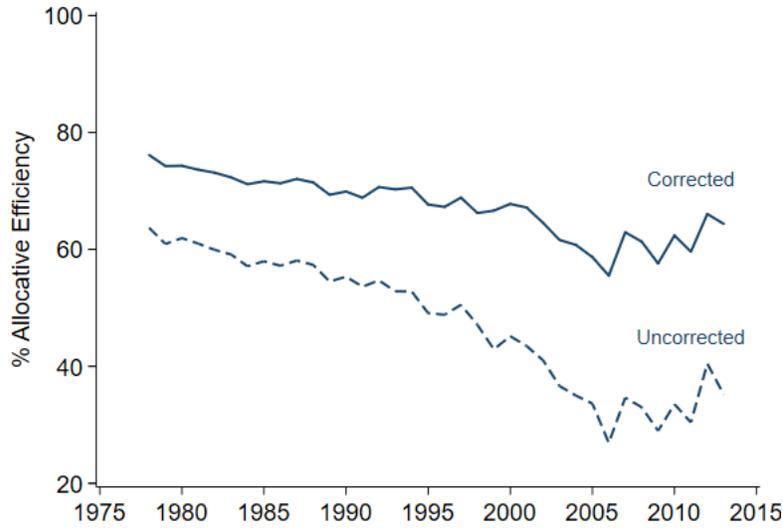
Source: Indian ASI and U.S. LRD. This table reports the average and standard deviation of uncorrected and corrected gains from improving allocative efficiency to 100% in India (1985–2013) and the U.S. (1978–2013). The shrinkage is the percent reduction in the average or standard deviation of gains after our corrections.

Table 6 reports the potential percent gains from going to 100% allocative efficiency in the United States. The correction reduces potential gains, averaged across years, from 123 percent down to 49 percent. In addition to dampening the trend in misallocation, the correction moderates its higher frequency vagaries. As a result, volatility for the time series for gains, as measured by its standard deviation, plummets from 60 percent down to 12 percent.

Our corrections dramatically affect comparisons of allocative efficiency in the U.S. versus India. Figure 9 displays the allocative efficiency for the U.S. relative to that in India. Without correcting, the U.S. averages a 16 percent advantage in allocative efficiency for the first ten years (1985 to 1994). But in the last ten years, U.S. efficiency collapses relative to India. Over those years U.S. efficiency averages only two-thirds that for India.

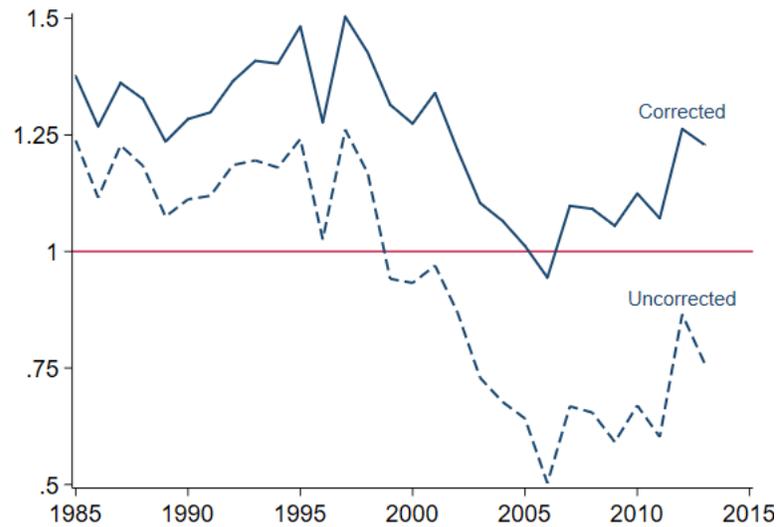
Our corrected series, however, looks entirely different. The U.S. advantage relative to India is much higher, averaging 25 percent, compared to –6 percent uncorrected. The U.S. advantage remains positive throughout, with a modest exception in 2006.

Figure 8: Allocative Efficiency in the United States

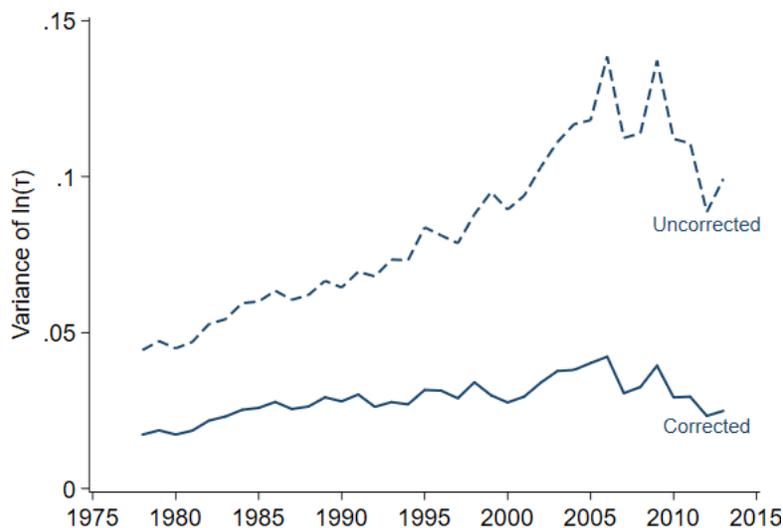


Source: U.S. LRD. The figures show uncorrected and corrected allocative efficiency (AE) for years 1978 to 2013. Average uncorrected AE is 47.6% while average corrected AE is 67.4%.

Figure 9: Allocative Efficiency: U.S. relative to India



Sources: Indian ASI and U.S. Census LRD. The figures show uncorrected and corrected allocative efficiency for the U.S. relative to India for years 1985 to 2013.

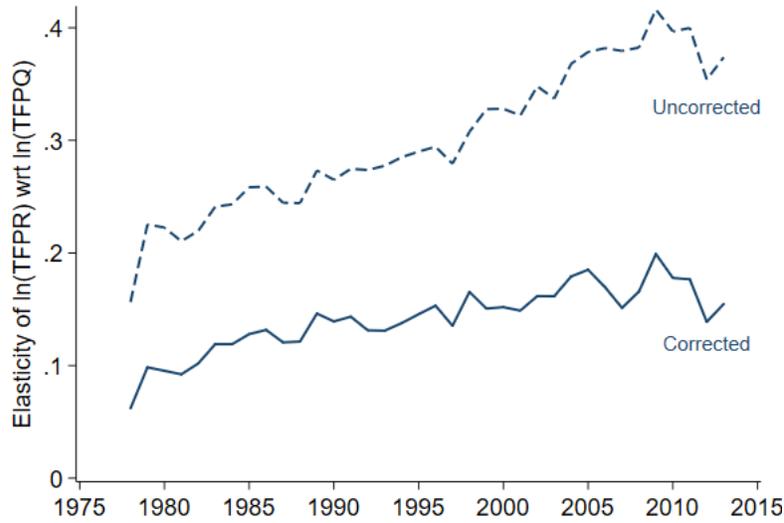
Figure 10: Variance of $\ln(\tau)$ for U.S.

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

We demonstrated in Section 4.3. above that the perceived decline in allocative efficiency for the U.S. mapped directly to a sharp rise in the variance of $\ln(\text{TFPR})$ in the United States. It should come as no surprise, then, that our correction to allocative efficiency reflects much less estimated dispersion in τ 's. Figure 10 shows this to be the case. Dispersion in our corrected TFPR series is much lower, trends less, is generally less volatile.

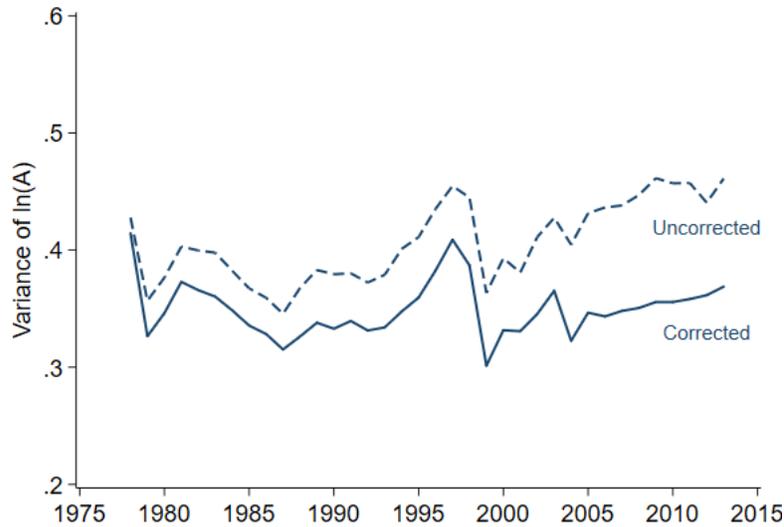
Our adjustments for measurement error also alter the implied elasticity of TFPR with respect to TFPQ in the United States. As displayed in Figure 11, our corrections lower the elasticity and its upward drift. Bento and Restuccia (2017) and Decker, Haltiwanger, Jarmin and Miranda (January 2018) highlight this elasticity as reflecting potential barriers to investing in or benefiting from higher TFPQ at the plant level. Related, our corrections undermine the case that TFPQ dispersion has risen across plants in the U.S. (see Figure 12), as emphasized for example by Gouin-Bonenfant (2019).

Figure 11: 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 12: 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.

8. Conclusion

We proposed a way to estimate the true dispersion of marginal products across plants in the presence of additive measurement errors in revenue and inputs. We showed that the response of revenue *growth* to input *growth* should be lower for high-TFPR plants in the presence of measurement error. And then used the projection of that response on TFPR to correct for measurement error. While our method employs several assumptions, we used simulations to demonstrate that our approach is robust or at least conservative.

We implemented our method on data from the Indian *Annual Survey of Industries* from 1985–2013 and the US. *Annual Survey of Manufacturing* from 1978–2013. In India, we estimated that true marginal products were significantly less dispersed than average products. As a result, potential gains from reallocation fell 21% and the volatility of those gains across years fell by 20%. In the U.S. our correction had even more bite. Average potential gains from reallocation fell by 60%, while time-series volatility fell by 80%. Our correction eliminated 2/3 of a severe downward trend in allocative efficiency for the U.S. Even corrected, allocative efficiency declined by 15% for U.S. manufacturing over the 35 years. Based on uncorrected data, allocative efficiency was 6% *lower* in the U.S. than in India for 1985 to 2013. In contrast, our corrected series implies consistently higher allocative efficiency in the U.S. than in India.

We hope our method provides a useful diagnostic for measurement errors that can be applied when researchers have access to panel data on plants and firms. For example, David and Venkateswaran (2019) and Bai, Jin and Lu (August 2019) apply our correction to firm-level data for China.

Our findings leave many open questions for future research. Why did measurement error worsen considerably over time in the U.S.? Why, even after our corrections, does ample misallocation remain in the U.S. and India? Is this real or due to some combination of model misspecification and proportional measurement error? If it is real, can it be traced to specific government policies or market failures (e.g. markup dispersion or capital/labor market frictions)?

9. Model Appendix

9.1. 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:

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

9.2. Aggregating to 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 distortions* 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 .

9.3. Aggregate Consumption

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

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

where

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

$$\tau^L \equiv \frac{1}{\sum_i^S \frac{R_{si}}{Q} \frac{1}{1 + \tau_{si}^L}}$$

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.

10. Data Appendix

10.1. Indian Annual Survey of Industries

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 7.

10.1.1. 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 7: 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 ≥ 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.

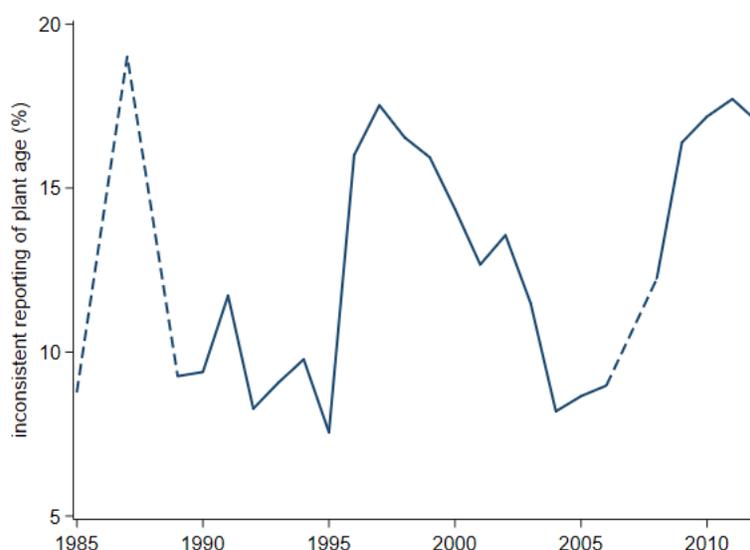
Intermediates: We construct intermediates as the sum of the value of materials consumed, fuels consumed and other intermediate expenses. Other intermediate expenses include repair and maintenance costs (plant/machinery, building, etc...), costs of contract and commission work, operating expenses (freight 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, Collard-Wexler and O'Connell (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 13 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 13: 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 $t+1$.

10.2. U.S. Annual Survey of Manufactures

The ASM and CMF are restricted access database of U.S. manufacturing plants put together by the U.S. Census Bureau. Both the ASM and CMF mail-back surveys. We only use data for plants in the ASM sample, which 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.

10.2.1. 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 payments for contract work performs, installation, repair work, etc...

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, etc...

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 have a similar number of sectors in the U.S. and 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.

10.3. 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.

11. Simulations Appendix

In this appendix we explore how our methodology behaves in the presence of measurement error in revenues, multiplicative measurement error and adjustment costs. In Table 9 we show the model moments, inferred share of distortions in TFPR dispersion and true share of TFPR dispersion for each simulation. In Table 8 we show the parameter values underlying these simulations. We treat the calibration to the U.S. economy in 2006–2013 as a baseline.

The second row of Table 9 shows 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 we calibrate σ_τ , σ_a , $\sigma_{\bar{a}}$ and σ_g . The calibrated parameter values are shown in Table 8. With measurement error entirely in revenues we are not able to exactly match all four moments, and

Table 8: Simulation Parameters: Multiplicative Meas. Err. and Adjustment Costs

	σ_τ	σ_a	$\sigma_{\tilde{a}}$	σ_f	σ_g	σ_{fm}
1. Baseline (U.S. 2006–2013)	0.017	0.494	0.103	0.130	0	0
2. Measurement Error in R	0.025	0.228	0.098	0	0.143	0
3. Multiplicative Meas. Err.	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 11. 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. $\sigma_{\tilde{a}}$ 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_{fm} is the standard deviation of shocks to multiplicative measurement error in inputs.

we miss slightly on the $\ln(\hat{\beta})$ slope. However, similarly to the results found in Section 6, we find that our correction underestimates the share of measurement error in TFPR dispersion.

The third row of Table 9 shows our simulations with multiplicative measurement error in inputs instead of additive measurement error in inputs. We keep σ_τ , σ_a and $\sigma_{\tilde{a}}$ at the same values as in the baseline calibration. We then calibrate σ_{fm} to match the variance of $\ln(\text{TFPR})$ in the U.S. in 2006–2013. In this case, the variance of $\ln(\tau)$ is only 1.5% of the variance of $\ln(\text{TFPR})$, 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 row of Table 9 shows 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.²³ We keep σ_τ , σ_a and $\sigma_{\tilde{a}}$ 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 $\ln(\text{TFPR})$ 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 $\ln(\hat{\tau})$ accounts for all of $\ln(\text{TFPR})$ dispersion. This confirms that adjustment costs (of this type), similarly to multiplicative measurement error, leads us to overestimate the dispersion of true marginal products.

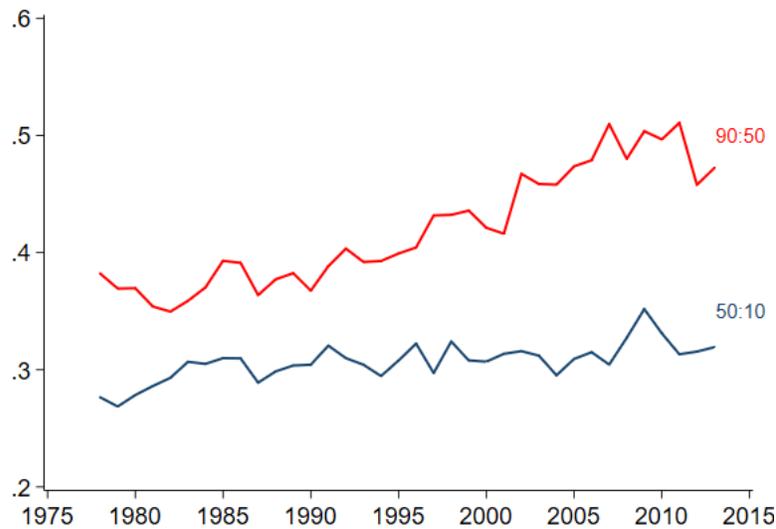
²³Plants know the true distribution of productivity shocks and form their expectations accordingly.

Table 9: Simulation Moments: Multiplicative Meas. Err. and Adjustment Costs

	σ_{TFPR}^2	σ_{TFPQ}^2	$\ln(\hat{\beta})$ slope	$\sigma_{\Delta I}^2$	$\sigma_{\hat{\tau}}^2/\sigma_{TFPR}^2$	$\sigma_{\tau}^2/\sigma_{TFPR}^2$
1. Baseline (U.S. 2006–2013)	0.102	0.41	-0.90	0.16	0.236	0.015
2. Measurement Error in R	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

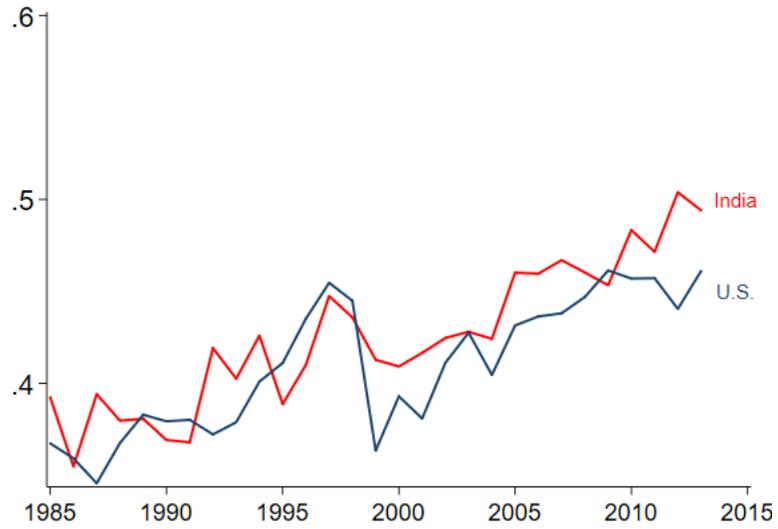
Source: This table shows the data moments generated by the model for each simulation. Simulation 1. is the same as that shown in Table 5. Simulation 2. has additive measurement error only in revenues. Simulation 3. has no additive measurement error, but has multiplicative measurement error in inputs. Simulation 4. has no measurement error, but has adjustment costs. σ_{TFPR}^2 is the output share weighted variance of $\ln(TFPR)$. σ_{TFPQ}^2 is the variance of $\ln(TFPQ)$. $\ln(\hat{\beta})$ slope is the slope of $\hat{\beta}$ against mean Tornqvist $\ln(TFPR)$ across deciles. $\sigma_{\Delta I}^2$ is the variance of input growth. $\sigma_{\hat{\tau}}^2/\sigma_{TFPR}^2$ is the ratio of the variance of $\ln(\hat{\tau})$ to the variance of $\ln(TFPR)$. $\sigma_{\tau}^2/\sigma_{TFPR}^2$ is the ratio of the variance of $\ln(\tau)$ to the variance of $\ln(TFPR)$.

12. Additional Figures Appendix

Figure 14: 90:50 and 50:10 Percentile Ratios of $\ln(TFPR)$ 

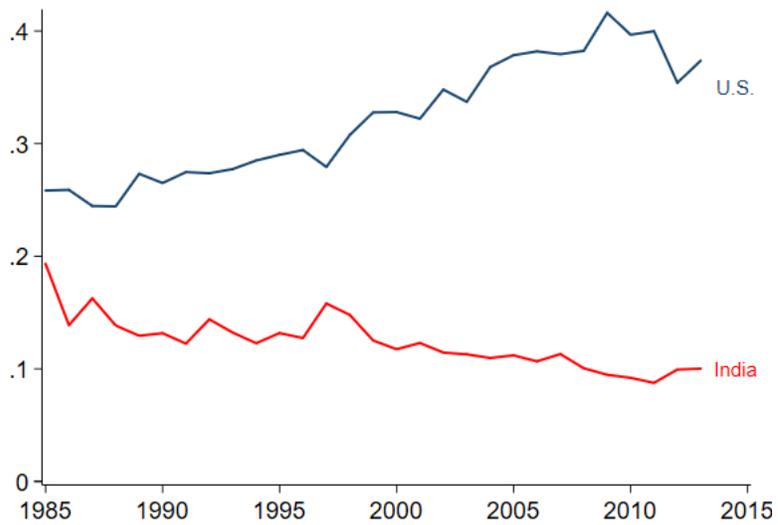
Source: U.S. LRD. The figure shows the 90:50 percentile ratio and the 50:10 percentile ratio of $\ln(TFPR)$ for the U.S. between 1978 and 2013. The percentiles are of the epanechnikov kernel density estimates of the distribution of $\ln(TFPR)$ (with a bandwidth of 0.05) so as to avoid disclosure of confidential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

Figure 15: Variance of $\ln(\text{TFPQ})$



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

Figure 16: Elasticity of TFPR with respect to TFPQ



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

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.
- Asker, John, Allan Collard-Wexler, and Jan De Loecker, "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 2014, 122 (5), 1013–1063.
- Bai, Yan, Keyu Jin, and Dan Lu, "Misallocation under Trade liberalization," *NBER working paper 26188*, August 2019.
- Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, "Cross-country Differences in Productivity: The Role of Allocation and Selection," *American Economic Review*, 2013, 103 (1), 305–334.
- Bento, Pedro and Diego Restuccia, "Misallocation, Establishment Size, and Productivity," *American Economic Journal: Macroeconomics*, 2017, 9 (3), 267–303.
- David, Joel M and Venky Venkateswaran, "The Sources of Capital Misallocation," *American Economic Review*, 2019, 109 (7), 2531–2567.
- Decker, Ryan A., John C. Haltiwanger, Ron S. Jarmin, and Javier Miranda, "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness," *NBER working paper 24236*, January 2018.
- 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.
- Gouin-Bonenfant, Emilien, "Productivity Dispersion, Between-Firm Competition and the Labor Share," 2019.
- Hopenhayn, Hugo A., "Firms, Misallocation, and Aggregate Productivity: A Review," *Annual Review of Economics*, 2014, 6 (1), 735–770.

- Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, 124 (4), 1403–1448.
- Redding, Stephen J. and David E. Weinstein, "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences," *Quarterly Journal of Economics*, 2019, 135 (1), 503–560.
- Restuccia, Diego and Richard Rogerson, "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Review of Economic Dynamics*, October 2008, 11, 707–720.
- and —, "The Causes and Costs of Misallocation," *Journal of Economic Perspectives*, 2017, 31 (3), 151–174.
- Rotemberg, Martin and T. Kirk White, "Measuring Cross-Country Differences in Misallocation.," 2019.
- Syverson, Chad, "What Determines Productivity?," *Journal of Economic literature*, 2011, 49 (2), 326–365.
- White, T. Kirk, Jerome P Reiter, and Amil Petrin, "Imputation in U.S. Manufacturing Data and Its Implications for Productivity Dispersion," *Review of Economics and Statistics*, 2018, 100 (3), 502–509.