Macro and Micro Dynamics of Productivity: From Devilish Details to Insights

G. Jacob Blackwood, Lucia S. Foster, Cheryl A. Grim, John Haltiwanger, Zoltan Wolf*

Abstract

Firm-level revenue-based productivity measures have become ubiquitous in studies of firm dynamics and aggregate outcomes. One commonly used measure has increasingly been interpreted as reflecting "distortions" since in their absence equalization of marginal revenue products should yield no dispersion in this measure. A commonly used, but distinct, measure is the residual of the firm-level revenue function which reflects "fundamentals". Using plant-level U.S. manufacturing data, we find these alternative measures are highly correlated, exhibit similar dispersion, and have similar relationships with growth and survival. However, the distinction between these alternative measures is important for quantitative assessment of aggregate allocative efficiency.

*Blackwood: Amherst College, jblackwood@amherst.edu, Foster: U.S. Census Bureau, lucia.s.foster@census.gov, Grim: U.S. Census Bureau, cheryl.ann.grim@census.gov, Haltiwanger: University of Maryland, haltiwan@econ.umd.edu, Wolf: New Light Technologies, zoltan.wolf@census.gov. We thank Jan DeLoecker, Ron Jarmin, Kirk White and conference participants at the 2013 Comparative Analysis of Enterprise Data in Atlanta, 2014 Research Data Center Annual Conference, the 2015 NBER CRIW workshop and two anonymous referees for valuable comments. Any remaining errors are our own. Any conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The statistics reported in this paper have been reviewed and approved by the Census Bureau's Disclosure Review Board (#DRB-B0026-20190205).

1 Introduction

The ubiquitous finding in the empirical literature on firm dynamics that there are large differences in measured productivity across establishments within narrowly defined industries has generated much analysis of the causes and consequences of this dispersion.[1] Potential causes put forth include curvature in the profit function that prevents the most productive firm from taking over an industry, adjustment frictions, and distortions that drive wedges in the forces that equalize marginal products across establishments. In terms of consequences, many papers have found that more productive businesses are more likely to grow and survive, implying that the reallocation of resources increases economic performance by enhancing aggregate productivity. Related to this finding, there is increased attention to reasons why reallocation dynamics may vary over the business cycle or across countries. In this context, a recent theoretical and empirical literature relies extensively on productivity dispersion measures as indicators of misallocation in the evaluation of economic performance.[2]

While there is consensus that accounting for dispersion and its connection to the allocation of activity are important for understanding economic performance, there is no consensus about the basics of estimating firm-level productivity. Since most micro datasets do not contain information on firm-level output prices, the majority of results are based on revenue productivity measures. How these are related to the standard concept of technical efficiency in theoretical models (which following the recent literature we denote as TFPQ) depends on the assumptions about the economic environment. For example, it has become increasingly recognized that there is price heterogeneity within narrow sectors reflecting, at least in part, product differentiation and thus likely some degree of market power. In the absence of price data, the way this endogeneity is modeled has critical implications for the relationship between revenue productivity measures and underlying fundamentals.[3]

A variety of methods are available to researchers to estimate firm-level revenue productivity in the absence of direct measures of prices and quantities. All of them start with a measure of revenue per composite input, often denoted as TFPR. The interpretation of such revenue productivity measures depends critically on the weights used to create the composite input. One common approach uses the shares of input expenditures of total costs as weights. We call the productivity measure implied by this approach TFPR^{cs}, where "cs" denotes "cost-shares." If plants are cost-minimizing, the assumption of constant returns to scale holds (CRS), and first order conditions for inputs hold at least in the long run (or averaged across firms), then cost-shares identify output elasticities of the production function. An alternative commonly used approach is to estimate the relationship between revenue and inputs using regression techniques. To overcome endogeneity issues, proxy estimation methods have been

[1] See the survey in Syverson (2011) for the relevant theoretical and empirical literature on productivity dispersion and its relationship to reallocation, growth, and aggregate productivity.

[2] See Restuccia and Rogerson (2008), Hsieh and Klenow (2009), and Bils et al. (2017).

[3] As will become apparent below and has become increasingly recognized in the literature, the relevant fundamentals for accounting for plant-level heterogeneity include technical efficiency and product quality/appeal components. This recognition is especially important for the empirical measures of fundamentals that emerge when using revenue and input data which is most of the literature (that is in the absence of price and quantity data). Such measures inherently reflect both technical efficiency and demand components. As such, we use the term TFPQ loosely in this context.

developed and become widely used. We call the implied revenue productivity measure TFPR^{rr} that emerges from estimation of the revenue function, where "rr" is short for "revenue function residual". Econometric issues aside, the most important distinction between cost-share-based methods and projection-based techniques is that the latter yields revenue elasticities, while the former yields output elasticities (up to the returns to scale). It is important to emphasize that estimation of the revenue elasticities and the revenue function residual need not rely on the projection-based proxy methods. If, for example, CRS holds so that the output elasticities can be estimated from cost shares, CES demand is assumed, and there is an independent estimate of the demand elasticity, then the revenue elasticity can be computed along with the implied revenue function residual. However, as we discuss below, estimation of demand elasticities in the absence of price and quantity data is a challenge.

Although the distinction between revenue and output elasticities is known, the implications for the interpretation of implied revenue productivity measures have not been widely recognized.[4] In an influential paper that has led to a burgeoning literature on misallocation, Hsieh and Klenow (2009) show, under specific assumptions about demand and production technology, that TFPR^{cs} exhibits no dispersion if marginal revenue products are equalized.[5] The insight of their model is that, in the absence of other frictions, TFPR^{cs} is proportional to idiosyncratic firm-level distortions. Thus, dispersion in such distortions can be used to make inference about the extent of misallocation – more dispersion in TFPR^{cs} translates into lower allocative efficiency (AE) and aggregate productivity. Under the same assumptions that yield this interpretation of TFPR^{cs}, TFPR^{rr} is conceptually different. Specifically, TFPR^{rr} is a measure reflecting fundamentals – i.e., technical efficiency and (if present) demand/product appeal shocks. As we show below, under the assumptions made by Hsieh and Klenow (2009) the log of TFPR^{rr} is proportional to the log of the indirect measure of TFPQ that they compute in their empirical analysis. Crucially, the joint distribution of measured TFPQ and distortions is important for measures of misallocation. Thus, an interesting implication of our analysis is that these distinct measures are critical to the construction of the AE measure that has become the benchmark in the literature.

This discussion highlights the importance of the conceptual, measurement, and estimation issues associated with distinguishing between the elasticities of production and revenue, and in turn the relationship between them. Under common assumptions in the literature (CES demand and Cobb-Douglas technology) there is a simple relationship between output and revenue elasticities. Specifically, for the composite input, the elasticity of the revenue function is equal to the returns to scale of the production technology divided by the (constant CES) markup. For a specific input, the revenue elasticity is equal to the output elasticity divided by the markup. In practice then, it

[4] A few recent papers provide some discussion and evidence on these issues. De Loecker (2011) notes the output-revenue elasticity distinction. Foster et al. (2016b) focus on the relationship between TFPR^{cs} and TFPR^{rr}. Haltiwanger (2016) discusses the theoretical relationship between TFPR^{cs} and TFPR^{rr}. Decker et al. (2017) analyse possible causes of declining business dynamism and find their results are robust to using either TFPR^{cs} or TFPR^{rr}.

[5] The connection between idiosyncratic distortions and misallocation has been pursued by many researchers. The framework itself is developed by Restuccia and Rogerson (2008). Hsieh and Klenow (2009) propose the result that TFPR^{cs} reflects distortions. More recent examples include Bartelsman et al. (2013), Gopinath et al. (2015) and Bils et al. (2017).

is crucial whether the revenue productivity measure uses output elasticities or revenue elasticities to compute the composite input, and in turn, how those elasticities and markups are estimated.

In order to gain insight about the empirical relevance of such conceptual differences, we implement both the cost share and revenue function estimation approaches. The findings indicate that TFPR^{cs} and TFPR^{rr} are highly correlated and exhibit similar dispersion. In addition, the relationship between productivity, growth, and survival holds equally well for both measures.[6] This is important because theory implies that, all else equal, firms with higher fundamentals should be more likely to grow and survive and therefore this prediction should hold for TFPR^{rr}. However, under the distortion interpretation, there is no reason why TFPR^{cs} should be positively correlated with growth and survival.[7] The positive empirical relationship between TFPR^{cs} and TFPR^{rr} helps explain why much of the empirical literature has used one or the other without drawing out their conceptual differences.

Focusing on the conceptual distinction and building on the existing literature, we show that aggregate productivity depends critically on the relationship between these two measures. We derive a generalization of the AE framework that simultaneously permits non-CRS, downward sloping demand, technical efficiency shocks, and demand/product appeal shocks. This generalization is useful because it highlights the importance of decomposing revenue elasticities into their demand elasticity and output elasticity components. Ideally what is required are internally consistent methods to jointly estimate the demand and output elasticities. With price, quantity, and input data this is more readily feasible.[8] In the absence of price and quantity data at the micro level (which characterizes the vast majority of available datasets including the one we use for this paper), we explore and evaluate two alternatives. First, we implement the approach of De Loecker and Warzynski (2012) to estimate markups at the industry-level under the assumption of CRS so that the output elasticities are readily estimated using cost shares. In this approach, demand elasticities/markups are computed using cost shares of revenue along with the output elasticities.[9] This method yields average markups of about 25 percent, consistent with much of the calibrated values in the recent literature but considerable cross-industry dispersion. We then explore a second approach which exploits the relationship between plant-level and industry-level variation as in Klette and Griliches (1996) to decompose the revenue elasticity into its output and demand elasticity components. We implement this approach in the context of the projection-based proxy methods as we discuss below. An advantage of this approach is that it does not require imposing CRS. We find lower markups and mildly increasing returns to scale on average but also considerable cross-industry dispersion in both.

The question is: what are the implications of these elasticity differences? Perhaps surprisingly, the conceptual differences in elasticities are not critical for inferences about basic variance and covariance properties of the revenue productivity measures (e.g., revenue productivity dispersion, relationship

[6] A list of antecedent papers for our findings is provided in the working paper version.

[7] Haltiwanger et al. (2018) take a direct approach to measure TFPQ and demand shocks using price and quantity data. Direct measurement has many advantages but can only be explored for a limited set of products in the U.S. Their findings, like ours, raise questions about interpreting TFPR^{cs} as a measure of distortions.

[8] See, e.g., Foster et al. (2008), Haltiwanger et al. (2018), and Eslava and Haltiwanger (2018).

[9] Unlike De Loecker and Warzynski (2012) we only estimate markups at the industry level rather than the plant-level.

with survival and growth). It might thus be tempting to conclude that these details are not so devilish after all. However, we find that these details matter critically for quantifying measures of AE. For example, we find that the same underlying data imply very different average sectoral AE (by more than a factor of two) due to differences in elasticity estimates. Moreover, the variation in elasticity estimates yields declines in AE for the average industry from the 1970s to the post-2000 period that vary between less than 20 and more than 40 percent.

Our generalized AE framework highlights that the demand elasticity and returns to scale do not enter AE symmetrically. That is, we show that there is not a simple relationship between the overall curvature of the revenue function and AE. The simple intuition relating curvature and AE is that as the revenue function approaches linearity (e.g., converging on perfect competition and CRS) then AE will decline for a given distribution of fundamentals and distortions since it becomes increasingly costly to not allocate resources to the firm with the highest composite fundamentals. The limitation of this intuition in practice is that the measured distribution of fundamentals and distortions (both in terms of variances and covariances) depends on the estimated curvature parameters implying a complex relationship between those parameters and AE. The generalized AE concept also highlights the importance of distinguishing between TFPR^{cs} and TFPR^{rr}. Although these measures exhibit similar and rising dispersion and correlation, the differences are highly informative for the level and trends in AE.

Our findings suggest that both the interpretation of revenue productivity measures and their implications for structural analyses, such as AE, depend non-trivially on the assumptions about the economic environment in which establishments operate. Our results indicate that changing the properties of the parameter distributions by changing the estimation method accounts for these differences. We explore sensitivity within the standard framework of Cobb-Douglas production, competitive input markets, and isoelastic product demand. More research is needed in order to assess the importance of more general functional forms, as well as the heterogeneity in key variables such as input/output prices, and markups or elasticities. Such analyses will likely have to rely on comparing and contrasting methods based on direct measures of prices and quantities with approaches that impose structural restrictions. This paper is not intended to be prescriptive, it is intended to provide a framework for understanding these issues, highlight implications of commonly used assumptions, and set the stage for future work.

The paper is organized as follows. We discuss our methodology and data in Sections 2 and 3. Section 4 describes the effect of estimation methods on the distribution of elasticity estimates. Section 5 describes the implications of the differences in elasticity distributions on productivity dispersion, plant growth and survival, and allocative efficiency. Section 6 concludes.

2 Methodology

2.1 Revenue productivity measures

We specify a Cobb-Douglas production function and a CES demand structure which are common in the literature.[10] The inverse demand function is given by $P_{is} = P_s Q_s^{1-\rho_s} Q_{is}^{\rho_s - 1} \xi_{is}$ for plant i in industry s where $\rho_s - 1$ is the inverse of the price elasticity of demand with $\rho_s < 1$, ξ_{is} denotes an idiosyncratic demand shifter, P_{is} and Q_{is} denote plant-level prices and quantities and P_s and Q_s denote industry level prices and quantities. The plant-level production function is given by $Q_{is} = A_{is} \Pi_{js} X_{ijs}^{\alpha_{js}}$, where A_{is}=TFPQ$_{is}$, X_{ijs} are the plant-level factor inputs (e.g., capital, labor, materials, and energy) and α_{js} is the elasticity of Q_{is} with respect to X_{ijs}.[11] The log of the revenue function is given by:

$$\mathrm{p}_{is} + q_{is} = \sum_j \beta_{js} x_{ijs} + \rho_s a_{is} + \ln \xi_{is} + (1-\rho) q_s + p_s, \tag{1}$$

where the revenue elasticities satisfy β_{js}=$\rho_s \alpha_{js}$ and lower case indicates logs. Various revenue productivity measures have been used in the theoretical and empirical literature. One typical measure is tfpr, given by (see Foster et al. (2008)):

$$\mathrm{tfpr}_{is} = p_{is} + q_{is} - \sum_j \alpha_{js} x_{ijs} = p_{is} + \mathrm{tfpq}_{is} \tag{2}$$

Equation (2) makes explicit that tfpr$_{is}$ confounds the effect of output prices and technical efficiency or tfpq$_{is}$. tfpq$_{is}$ is unobserved because most micro datasets only contain information about costs and revenues but not plant-level prices implying that the majority of results in the empirical productivity literature are based on revenue productivity measures. An important special case emerges under the assumption that plants minimize total costs and have a CRS technology: the share of the jth input expenditure in total costs equals α_{js}. Formally:

$$\mathrm{tfpr}_{is}^{cs} = p_{is} + q_{is} - \sum_j cs_{js} x_{ijs} = \mathrm{tfpr}_{is} + \sum_j \left(\alpha_{js} - cs_{js}\right) x_{ijs}, \tag{3}$$

where cs_{js} denotes the cost share of the jth input. Note the equivalence between tfpr$_{is}$ from equation (2) and tfpr$_{is}^{cs}$ does not hold without CRS. Still, tfpr$_{is}^{cs}$ is of interest in and of itself, even without CRS, since it is indicative of distortions under certain assumptions, as we demonstrate below.

The revenue productivity measures above are distinct from the revenue function residual which

[10] CES demand is a standard assumption in the productivity literature, see for example, Hsieh and Klenow (2009), Bartelsman et al. (2013), Foster et al. (2016b), and Bils et al. (2017). Our formulation is consistent with the final good being a CES aggregate of perfectly competitive intermediate goods producers. See appendix A.1 for more details. Time subscripts are omitted in this section in the notation and equations for expositional convenience.

[11] There is some abuse of notation in our conceptual framework. We use A_{is}=TFPQ$_{is}$ to capture technical efficiency of the production technology. We then show that tfpr$_{is}^{rr}$ is a function of fundamentals including both technical efficiency and demand. Then, consistent with the literature, we observe that tfpr$_{is}^{rr}$ is proportional to the indirect empirical measure of lnTFPQ$_{is}$. The latter is a composite shock reflecting technical efficiency and demand effects.

is given by:

$$\text{tfpr}_{is}^{rr} = p_{is} + q_{is} - \sum_j \beta_{js} x_{ijs} = \rho_s a_{is} + \ln \xi_{is} + (1-\rho_s) q_s + p_s, \tag{4}$$

which says that the revenue function residual depends on technical efficiency, idiosyncratic demand shocks and aggregate prices and quantities. In addition, these assumptions imply $\gamma_s = \sum_j \alpha_{js} = \rho_s^{-1} \sum_j \beta_{js}$, where γ_s denotes returns to scale. The implication is that tfpr_{is}^{rr} is different from both tfpr_{is} and its estimate tfpr_{is}^{cs}:

$$\text{tfpr}_{is}^{rr} = p_{is} + q_{is} - \sum_j \beta_{js} x_{ijs} \neq \begin{cases} \text{tfpr}_{is} \\ \text{tfpr}_{is}^{cs}. \end{cases} \tag{5}$$

CES demand is critical for the interpretation of $\text{tfpr}_{is}^{cs}$and tfpr_{is}^{rr}. Without idiosyncratic frictions or distortions, marginal revenue products are equalized across production units and there is no within-industry dispersion in tfpr_{is}^{cs}. Since this outcome is counterfactual, Hsieh and Klenow (2009) posit the presence of distortions that account for such dispersion. To illustrate this point, we consider the decision problem of firms who maximize static profits with input distortions,[12] which imply:

$$\text{TFPR}_{is}^{cs} \propto \tau_{is}, \tag{6}$$

where $\tau_{is} = \prod_j (1+\tau_{ijs})^{\alpha_{js}/\gamma_s}$ denotes a plant-specific weighted geometric average of input distortions and the weights are given by cost shares. Note the proportionality result (6) is obtained equivalently when there are only scale distortions $\tau_{Q_{is}}$, by using the substitution $\tau_{is} = (1-\tau_{Q_{is}})^{-1}$. In contrast, tfpr_{is}^{rr} is proportional to tfpq_{is} and demand shocks under the same assumptions:[13]

$$\text{TFPR}_{is}^{rr} \propto A_{is}^{\rho_s} \xi_{is}. \tag{7}$$

The key implication for the objective of this paper is that TFPR_{is}^{cs} is proportional to idiosyncratic distortions while TFPR_{is}^{rr} is proportional to fundamentals. This conceptual difference, due to using output vs. revenue elasticities, is what motivates the empirical analysis below.

Estimating tfpr_{is}^{rr} to measure fundamentals is not novel to this paper. Cooper and Haltiwanger (2006) used revenue function residuals as measures of plant-level fundamentals. The empirical measure of fundamentals used by Hsieh and Klenow (2009) is also tightly linked to this revenue residual approach. To see this, note that their empirical measure of log-TFPQ is equivalent to a composite shock given by $a_{is} + \frac{1}{\rho_s} \ln \xi_{is}$. That is, their empirical measure of tfpq_{is} is given by: $(p_{is}+q_{is})/\rho_s - \sum_j \alpha_{js} x_{ijs}$ which given the above implies that the indirect measure of $\text{tfpq}_{is} = \frac{1}{\rho_s} \text{tfpr}_{is}^{rr}$. Measurement of TFPQ in this indirect fashion is more challenging than measuring tfpr_{is}^{rr} since doing so requires decomposing revenue elasticities into the output elasticity and demand elasticity components. As will become clear, such a decomposition is important conceptually and empirically.

Under the assumptions made in this section, there need be no systematic relationship between

[12] The profit function in this case is given by $P_{is} Q_{is} - \sum_j w_{js}(1+\tau_{ijs}^*) x_{ijs}$, where w_{js} denotes the jth input price.

[13] Here we abstract away from industry-level shifters that can be captured by industry-year effects.

tfpr_{is}^{cs} and tfpr_{is}^{rr}. However, it is important to emphasize that the assumptions under which tfpr_{is}^{cs} reflects only distortions are restrictive. For example, if demand is not CES, there are overhead factors of production, or there are adjustment costs, tfpr_{is}^{cs} is determined by the cumulative effect of all these factors and therefore the proportionality result relating tfpr_{is}^{cs} *only* to distortions does not hold. In other words, distortions are not identified by tfpr_{is}^{cs} without additional information, and as a consequence, a systematic relationship between tfpr_{is}^{cs} and tfpr_{is}^{rr} may emerge.[14] A broader view is to treat the distortions identifed in this manner as a reduced form capturing all of these alternative factors. Under this view, there are many reasons for a correlation between fundamentals and (reduced-form) distortions. This encompasses distortions from the business climate (e.g., size-related distortions) as well as adjustment costs that yield such a correlation.

Our contribution is to explore the systematic relationship between these conceptually distinct measures. We do not investigate formally non-CES demand structures, overhead labor, or adjustment costs but discuss our findings below in light of the studies that consider these possibilities. A subtle point in our analysis is that when production exhibits non constant returns to scale (NCRS), tfpr_{is}^{cs} is not equal to tfpr_{is}. In this case, tfpr_{is}^{cs} will still only reflect distortions while tfpr_{is} will exhibit dispersion in the absence of distortions. In contrast, the finding that tfpr_{is}^{rr} is only a function of fundamentals is robust to deviations from CRS.

2.2 Allocative Efficiency

Recent literature, beginning with Hsieh and Klenow (2009) and more recently Bils, Klenow, and Ruane (2017) (hereafter HK and BKR, respectively), build on the distinction between TFPQ and TFPR using the assumptions made in the prior section to construct a measure of misallocation which they term allocative efficiency (AE). We revisit these issues since they help highlight the importance of distinguishing between tfpr_{is}^{cs} and tfpr_{is}^{rr}.

Following this literature, we initially assume in this subsection that the firm-level production technology exhibits CRS to illustrate the properties of AE. At the sectoral level, AE is a ratio of sectoral productivity to undistorted sectoral productivity. Sectoral productivity is defined as sectoral output per composite input: $\text{TFPQ}_s = Q_s / \prod_j X_{js}^{\alpha_{js}}$.[15] Using CES demand and Cobb-Douglas production with CRS, BKR show that TFPQ_s can be expressed as a power sum of A_{is} weighted by relative distortions:

$$\text{TFPQ}_s = \left(\sum_i A_{is}^{\frac{\rho_s}{1-\rho_s}} \left(\frac{\tau_{is}}{\widetilde{\tau}_s} \right)^{\frac{-\rho_s}{1-\rho_s}} \right)^{\frac{1-\rho_s}{\rho_s}}, \tag{8}$$

where $\widetilde{\tau}_s$ is a function of idiosyncratic physical productivities and distortions, and can be thought of

[14] See, e.g., Bartelsman et al. (2013), Asker et al. (2014), Haltiwanger et al. (2018), Haltiwanger (2016), and Foster et al. (2016b).

[15] We also have the same abuse of notation as above with $\text{TFPQ}_{is}=\text{A}_{is}$ for notational convenience, even though the empirical measure of TFPQ is an indirect composite that is proportional to tfpr_{is}^{rr}. That is, as in HK, we think of the empirical measure of fundamentals as a composite of $\ln\text{A}_{is}$ and $\ln \xi_{is}$, as discussed above.

as the average distortion in the sector (see Appendix A.2.2). TFPQ$_s$ is maximized when $\tau_{is}=\widetilde{\tau}_s$,[16] in which case, following from equation (8), TFPQ$_s$ is given by $A_s^* = \left(\sum_i A_{is}^{\rho_s/(1-\rho_s)}\right)^{(1-\rho_s)/\rho_s}$. AE is defined as the ratio of TFPQ$_s$ to the maximized, counterfactual TFPQ$_s$. Multiplying and dividing by $N_s^{(1-\rho_s)/\rho_s}$, where N_s is the number of plants in the sector, sectoral AE can be expressed as

$$AE_s = \left(\frac{1}{N_s}\sum_i \left(\frac{A_{is}}{\widetilde{A}_s}\right)^{\frac{\rho_s}{1-\rho_s}} \left(\frac{\tau_{is}}{\widetilde{\tau}_s}\right)^{\frac{-\rho_s}{1-\rho_s}}\right)^{\frac{1-\rho_s}{\rho_s}}, \tag{9}$$

where $\widetilde{A}_s = \left(N_s^{-1}\sum_i A_{is}^{\rho_s/(1-\rho_s)}\right)^{(1-\rho_s)/\rho_s}$ is the power mean analogue to A_s^*.

Given our interest in alternative estimation methods that do not impose CRS on plant-level technology, we generalize (9) to be robust to deviations from CRS but otherwise maintain the CES demand and Cobb-Douglas production. Our approach, described in more detailed in Appendix A.2, builds on the appendix of HK–who derive AE$_s$ using a single-input production technology that exhibits decreasing returns to scale–but allows for multiple inputs and NCRS. This generalization is useful as it helps us draw out the implications of the alternative estimation approaches for AE.[17] Under these assumptions TFPQ$_s$ is given by:[18]

$$\mathrm{TFPQ}_s = \frac{\left(\sum_i A_{is}^{\frac{\rho}{1-\rho_s\gamma_s}} \left(\frac{\tau_{is}}{\widetilde{\tau}_s}\right)^{\frac{-\rho_s\gamma_s}{1-\rho_s\gamma_s}}\right)^{\frac{1-\rho_s\gamma_s}{\rho_s}}}{\left(\prod_j X_{js}^{\alpha_j/\gamma_s}\right)^{1-\gamma_s}}. \tag{10}$$

Denote aggregate input j corresponding to max{TFPQ$_s$} as X_{js}^*. For ease of exposition, we refer to this as the "distortionless" case, where the label "distortionless" means distortions are equalized across establishments. Dividing and multiplying by N_s appropriately, AE$_s$ can then be obtained as:

$$AE_s = \underbrace{\left(\frac{1}{N_s}\sum_i \left(\frac{A_{is}}{\widetilde{A}_s}\right)^{\frac{\rho_s}{1-\rho_s\gamma_s}} \left(\frac{\tau_{is}}{\widetilde{\tau}_s}\right)^{\frac{-\rho_s\gamma_s}{1-\rho_s\gamma_s}}\right)^{\frac{1-\rho_s\gamma_s}{\rho_s}}}_{AE_s^{COV}} \underbrace{\left(\frac{\prod_j X_{js}^{*\alpha_{js}}}{\prod_j X_{js}^{\alpha_{js}}}\right)^{\frac{1-\gamma_s}{\gamma_s}}}_{\text{Sectoral term}}. \tag{11}$$

Equation (11) is a generalization of (9) that fully accounts for the effect of NCRS production technology.[19] The first term – labeled as AE$_s^{COV}$ in order to emphasize that it resembles a covariance term– shows the effect of NCRS on the within-industry component of AE. The second term in (11) captures the effect of NCRS via sectoral inputs. Importantly, this term equals 1 when all production factor supplies are exogenous, implying that only AE$_s^{COV}$ is relevant in this case.

[16] More details are available in Appendix A.2.4, which also shows that the sufficient condition is satisfied only if $\rho_s\gamma_s<1$. This is an intuitive restriction, as a market equilibrium with increasing returns in the revenue function would imply one firm taking over the market. However, not all estimation methods restrict the parameter space to ensure this is the case.

[17] Even though we permit NCRS at the firm-level, we follow HK in using cost shares to weight sectoral inputs when calculating the composite input in the definition of sectoral productivity. That is, the aggregate technology exhibits CRS while the plant-level technology exhibits NCRS.

[18] This equation is a generalization of the equation on page 1445 of the appendix in HK.

[19] This can be seen by noting that (11) simplifies to (9) under CRS ($\gamma_s = 1$).

For the most part, we focus on AE at the sectoral level. However, it is helpful to explore various methods aggregating AE across sectors. Following the literature (e.g, HK and BKR) we treat supply of aggregate primary factors as fixed and we assume a Cobb-Douglas CRS aggregator for output across sectors into a final good. In addition, we assume a representative perfectly competitive firm that produces this final output. It can be shown that these two assumptions imply that primary sectoral inputs are constant so long as average industry distortions are unchanged. Thus their contribution to AE drops out of (11).

Two alternative approaches have been used with respect to treatment of intermediate inputs. One approach (e.g., HK) is to assume there are only primary inputs. This is consistent with treating intermediates as raw materials that are in fixed supply. Sectoral AE simplifies to AE_s^{COV} in this case and overall AE is given by:[20]

$$AE = \prod_{s}^{S} AE_s^{\theta_s} = \prod_{s}^{S} \left(AE_s^{COV}\right)^{\theta_s}, \tag{12}$$

where θ_s denotes the revenue share of industry s. A second approach (e.g., BKR), labeled roundabout production, endogenizes intermediate input production, which amounts to saying intermediate inputs are all produced within the sectors under consideration. In this case, while inputs do not completely cancel, we can still express AE in industry s as a function of AE_S^{COV}:[21]

$$AE_s = \left(AE_s^{COV}\right)^{\frac{\sum_{k=1}^{S}\theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)}{\sum_{k=1}^{S}\theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)+\theta_s\frac{\alpha_{Ms}}{\gamma_s}(1-\gamma_s)}}, \tag{13}$$

where α_{Ms}/γ_s denotes the cost share of intermediate inputs in industry s and AE_s^{COV} is defined in equation (11).[22] Aggregation across sectors implies the following expression:

$$AE = \prod_{s=1}^{S} AE_s^{\frac{\theta_s}{\sum_s \theta_s\left(1-\frac{\alpha_{Ms}}{\gamma_s}\right)}} = \prod_{s=1}^{S}\left(AE_s^{COV}\right)^{\frac{\theta_s}{\sum_{k=1}^{S}\theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)+\theta_s\frac{\alpha_{Ms}}{\gamma_s}(1-\gamma_s)}}. \tag{14}$$

These two aggregate concepts are relevant for the empirical analysis because they act as upper and lower bounds on true AE. Equation (12) overstates true AE if some industries are characterized by roundabout production. On the other hand, equation (14) understates true AE if there are industries where inputs are in fixed supply. In order to remain conservative and given the properties of the sample,[23] we follow on the first approach in the detailed empirical analysis and briefly touch on the role of roundabout production.

Returning to our primary focus on sectoral AE, it is useful to highlight the role of ρ_s, γ_s, and α_{js}

[20] See appendix A.2.6 for details.

[21] See appendix A.2.6.

[22] Under CRS, the outer exponent simplifies to 1, and therefore (13) collapses to (9). When returns to scale are increasing, the exponent is greater than one and AE_s^{COV} is smaller relative to when inputs are exogenous, because the influence of intermediates serves to amplify the effect of the inner term. On the other hand, decreasing returns increases measured AE_s^{COV}.

[23] Our sample contains the 50 most populous 4-digit SIC industries that were mapped one-to-one between different industry classification systems, see section 3. This property implies that our AE results are estimates of true AE.

for sectoral AE. Equations (11) and (13) show that these parameters affect AE_s^{COV} via the exponents. In addition, ρ_s and γ_s affect relative technical efficiencies and distortions since both A_{is} and tfpr_{is}^{cs} are constructed using an input index as the denominator where the input index depends directly on α_{js} estimates. The implication is that the joint distribution of these variables is a key determinant of AE. In order to formalize this result, we express AE_s^{COV} as a function of the covariance between transformations of τ_{is} and A_{is}:[24]

$$\ln\left(AE_s^{COV}\right) = \ln\left(\frac{\widetilde{\tau}_s}{\overline{\tau}_s}\right) + \frac{1-\rho_s\gamma_s}{\rho_s}\ln\left[\text{cov}\left(\left(\frac{A_{is}}{\widetilde{A}_s}\right)^{\frac{\rho_s}{1-\rho_s\gamma_s}}, \left(\frac{\tau_{is}}{\overline{\tau}_s}\right)^{\frac{-\rho_s\gamma_s}{1-\rho_s\gamma_s}}\right)+1\right]. \tag{15}$$

Equation (15) says that AE_s^{COV} depends on sectoral distortions (term 1), and a function of the covariance between exponentiated relative technical efficiencies and distortions (term 2).[25] By definition, the covariance depends on the dispersion of these two variables and the correlation between them. This relationship creates a useful link between AE_s and the properties of the within-industry productivity distribution. For example, if the correlation between fundamentals and distortions is negative, then increases in dispersion of either will increase allocative efficiency (note the negative exponent on distortions in (15)). However, if distortions are positively correlated with fundamentals, as is increasingly assumed in the literature, then allocative efficiency is decreasing in dispersion of either A_{is} or τ_{is}. Furthermore, changes in the correlation could also account for observed patterns.

We use equation (15) in our empirical analysis below to provide guidance about the sensitivity of the measures of AE to the estimates of ρ_s, γ_s, and α_{js} . The second term of equation (15) highlights the complex relationship between curvature parameters and AE. Analytic derivatives of this second term (available upon request) with respect to ρ_s and γ_s deliver an ambiguous sign that depends on the sign of the covariance between fundamentals and distortions. In the empirical analysis, this is further complicated by the fact that the measured distributions of fundamentals and distortions depend on these estimated parameters.[26]

Before proceeding it is useful to highlight the connection between $\ln\left(\text{AE}_s^{COV}\right)$ and tfpr_{is}^{cs} and tfpr_{is}^{rr}. Doing so makes it more transparent that TFPQ_{is} is a composite shock. That is, we formally recognize that A_{is} in equation (15) and earlier expressions is measured as $\ln A_{is} = \frac{1}{\rho_s}\text{tfpr}_{is}^{rr}$. Since $\tau_{is} \propto \text{TFPR}_{is}^{cs}$ in practice, the empirical equivalent of equation (15) can be re-written as:[27]

$$\ln\left(AE_s^{COV}\right) = \ln\left(\frac{\widetilde{\text{TFPR}}_s^{cs}}{\overline{\text{TFPR}}_s^{cs}}\right) + \frac{1-\rho_s\gamma_s}{\rho_s}\ln\left[\text{cov}\left(\left(\frac{\text{TFPR}_{is}^{rr}}{\widetilde{\text{TFPR}}_s^{rr}}\right)^{\frac{1}{1-\rho_s\gamma_s}}, \left(\frac{\text{TFPR}_{is}^{cs}}{\overline{\text{TFPR}}_s^{cs}}\right)^{\frac{-\rho_s\gamma_s}{1-\rho_s\gamma_s}}\right)+1\right]. \tag{16}$$

[24]Equation (15) can be obtained by multiplying and dividing AE_s^{COV} by $\overline{\tau}_s = \left(\frac{1}{N_s}\sum_i \tau_{is}^{(\rho_s\gamma_s)/(\rho_s\gamma_s-1)}\right)^{(\rho_s\gamma_s-1)/(\rho_s\gamma_s)}$ and rearranging the resulting expression.

[25]This formulation also highlights the role of ρ_s and γ_s via the exponents.

[26]We show in appendix (A.2.7) that while both ρ_s and γ_s directly determine plant-level responses to productivity shocks, ρ_s also impacts sectoral responses through the CES aggregator, generating an asymmetry in the influence of the two parameters.

[27]In this expression we use the same type of aggregation as in equation (15).

In other words, measured AE_s^{COV} is a function of the two distinct revenue productivity measures derived above. We return to this relationship in our empirical analysis below.

2.3 Estimation methods

We begin by reviewing the cost-share-based method for estimating output elasticities that we use in our estimates of tfpr_{is}^{cs}. Cost-share-based methods (CS) exploit first order conditions from the firm's cost-minimization problem. This framework implies that under CRS, the share of input expenditures in total costs identify output elasticities even without data on prices and quantities. This is a useful property because it makes CS robust to alternative demand structures and also imperfect competition in output markets. A potential caveat is that an estimate of, or an assumption about, returns to scale is necessary. Since the requirement that the first order conditions of cost minimization hold for all businesses in every time period may be considered too restrictive, most studies (see Syverson (2011)) average them across plants in an industry and/or over time. This approach is equivalent to assuming that elasticities are homogenous within an industry or over time, a restriction that alternative estimation methods also typically require.

In order to estimate the revenue elasticities and tfpr_{is}^{rr}, we use a projection-based proxy method.[28] We include OLS in our analysis as a point of reference despite the fact that OLS estimates are inconsistent because inputs are correlated with unobserved productivity, cost and demand shocks.[29] Proxy methods were developed to address this issue and are commonly used in the applied literature. Since prices are typically unobserved, these methods (without further structure) estimate revenue elasticities and not output elasticities.[30] This point – often neglected in the applied literature[31] – is critical for our purposes since it implies that the estimated coefficients are revenue elasticities and the residual is a measure of fundamentals: tfpr_{is}^{rr}.

The original two-step proxy method procedure, developed in Olley and Pakes (1996) (OP), is based on the assumption that investment is monotonically increasing in the composite revenue shock (in our setting from TFPQ and demand shocks) that is the only unobserved state variable.[32] Under this assumption, the control function is invertible and plant-level composite revenue shock can be recovered. As a consequence, variable input elasticities can be identified in an OLS step because the influence of the composite revenue shock on variable inputs is accounted for by the investment proxy. In contrast, quasi-fixed input decisions have dynamic consequences, and therefore their elasticities are determined in an additional step using the residual revenue variation that is left after removing the contribution of variable inputs.[33]

[28]The NBER working paper version of this paper considers a wider range of proxy methods.

[29]See Marschak and Andrews (1944).

[30]An exception is Eslava and Haltiwanger (2018) which uses plant-level price and quantity data to jointly estimate the production and demand functions using GMM procedures motivated by the proxy method approach.

[31]It is common in the applied literature to state that there may be biases in the estimates of output elasticities since plant-level prices are not known.

[32]For estimation purposes, we are treating the composite shock as a single unobserved state variable.

[33]More details on OP and its mutations can be found in the earlier NBER working paper version. We note that we found sensitivity to the alterative proxy methods but without the hybrid approach we implement here where only the

The identifying assumptions that underlie proxy methods have been criticized. For example, Ackerberg, Caves, and Frazer (2015) (ACF) argue that variable inputs are deterministic functions of revenue shocks. This means that the control function is not invertible and therefore variable input elasticities are not identified in the first step. They recommend not identifying any of the elasticities in the first stage. In this sense the approach is similar to Wooldridge (2009) who proposed to resolve the issue by estimating all coefficients in a single GMM step and using earlier outcomes of both capital and variable inputs as instrumental variables. His approach is useful because it is robust to the ACF critique and because the efficiency loss that arises from two-step estimation is eliminated. A related but separate point is raised by Gandhi, Navarro, and Rivers (2012), who argue that once intermediate input expenditures are used as a proxy, as proposed by Levinsohn and Petrin (2003), no independent variation remains that could be used to identify the elasticity of intermediate inputs, whether or not elasticities are estimated structurally or using feasible GMM.

In order to circumvent these issues, we use a hybrid approach.[34] The basic idea is to combine constrained optimization and OP, which is why this approach is labeled as OPH where "H" denotes hybrid. The main difference relative to OP is that first order conditions of profit maximization are used to identify the revenue elasticities of variable inputs (that is, variable input revenue elasticities are determined from average revenue cost shares across plants).[35] Specifically, we estimate industry-specific revenue elasticities for variable factors as the mean of the plant-level cost share of total revenue.[36] This approach can be interpreted as a non-parametric alternative to Gandhi, Navarro, and Rivers (2012), which implies that the issues raised in the recent literature related to proxy methods are overcome.

Part of our focus is on the empirical relationship between tfpr_{is}^{cs} and tfpr_{is}^{rr} and the above described approaches are sufficient to construct these measures. However, further insights can be gained by exploring the relationship between the revenue and output elasticities, which requires estimates of demand elasticities. We use two approaches. First, we estimate the demand elasticity ρ_s using the De Loecker and Warzynski (2012) approach at the industry level. Specifically, the first order condition for a variable factor like materials yields that the markup $(1/\rho_s)$ is equal to the ratio of the output elasticity to the cost share of revenue of the variable factor. The challenge then is how to estimate the output elasticity. We use the cost share of total costs for materials at the industry level to obtain the output elasticity associated with our CS methodology. This approach, which we denote by DW, yields estimates of markups (and thus ρ_s) that vary across detailed industries.[37]

quasi-fixed factor elasticiites are estimated by proxy method. In unreported results, we have found that in applying alternative proxy methods where the variable factor revenue elasticities are estimated non-parameterically as we do here, the alternative proxy methods yield more similar revenue elasticilty estimates for the quasi-fixed factors.

[34] See Haltiwanger and Wolf (2018) for more discussion.

[35] The FOCs also act as regularization constraints during estimation of the quasi-fixed revenue elasticities.

[36] Critically this differs from the CS approach above which uses cost shares of total costs. For variable inputs, we use the cost share of total revenue.

[37] As we note above, estimating output elasticities in this manner requires strong assumptions including CRS. De Loecker and Warzynski (2012) use a different approach estimating the output elasticities using a proxy method that does not impose CRS. However, since they do not have plant-level prices their estimates of output elasticities are

For our second approach, we consider a variant of OPH in which revenue elasticities and the demand elasticity are jointly estimated using the method described in Klette and Griliches (1996). This latter specification is labeled OPHD. This approach has the advantage of estimating the revenue elasticity and demand elasticity in an internally consistent manner without imposing restrictions such as CRS for the output elasticities. This approach is based on the idea that CES demand implies a specification in which plant-level revenues are regressed on inputs and an indicator of industry-level revenues. The coefficient of the latter identifies ρ_s.

3 Data

3.1 Source data

Our industry-level data, including deflators, capital rental prices, and depreciation rates, are taken from the NBER-CES Manufacturing database[38], the Bureau of Labor Statistics and the Bureau of Economic Analysis. We use establishment-level information from the Annual Survey of Manufactures (ASM), Census of Manufactures (CM), and the Longitudinal Business Database (LBD).

We use the ASM and CM to construct plant-level measures of inputs and output. Output is measured as the deflated total value of shipments, corrected for the change in finished goods and work-in-process inventories. Total hours worked is constructed as the product of production worker hours and the ratio of the total wage bill to production worker wages. Our intermediate input variable is given by the the sum of three items: cost of parts, contracted work, and goods resold. The energy input consists of deflated electricity and fuel costs. We create establishment-level capital stock measures using a version of the Perpetual Inventory Method, which calculates current capital as a sum of the depreciated stock and current investment.

The LBD serves two purposes in our analysis. First, high-quality longitudinal identifiers help us determine the accurate time of establishments' exit which is needed to estimate the relationship between productivity, growth, and exit. Second, the LBD acts as a universe file; we use employment and establishment age data from the LBD to construct inverse propensity score weights that control for selection to the ASM. More details about the data can be found in the working paper version of this study and Appendix A in Foster, Grim, and Haltiwanger (2016a) (FGH, hereafter). These descriptions include how cost shares of inputs are measured.

3.2 Analysis samples

Our initial sample includes approximately 3.5 million plant-year observations between 1972 and 2010. Although this is a large dataset, we restrict the sample in the empirical analysis because it needs to fulfill two potentially contradicting requirements. First, industries should be defined narrowly enough that we can plausibly assume elasticities are constant across establishments within an industry. To

more appropriately interpreted as revenue elasticities. De Loecker and Warzynski (2012) also estimate markups at the plant-level using the first-order condition for variable factors at the plant-level.

[38]The NBER-CES Manufacturing Industry database is available at http://www.nber.org/nberces. An earlier version is documented in Bartelsman and Gray (1996).

fulfill this requirement, we choose a 4-digit SIC grid, which is considered narrowly defined, and roughly corresponds to a 6-digit NAICS grid. Second, the number of plant-year observations within each industry should be large enough that elasticities can be estimated by all reviewed methods. Changes in industry classification systems over time create complications because these changes entail spurious breaks in plant-level time series and a drop in sample size. We address this issue by selecting 4-digit SIC industries which were either not affected by classification changes or mapped one-to-one into another SIC category (in 1987) or NAICS category (in 1997). There are 292 such industries of which we selected the first 50 based on the number of plant-year observations.[39] We focus on the industries with a large number of plant-year observations because the proxy method estimation methods use high-order polynomials making estimates sensitive to small samples. An implication (and thus a limitation in practice) of the proxy method approach is that often empirical studies using this method classify industries at the 2 or 3-digit level in order to be able to generate sensible elasticities for all industries. We wanted to avoid this so that we could compare alternative methods of estimating elasticities using detailed industries.

4 Elasticity distributions

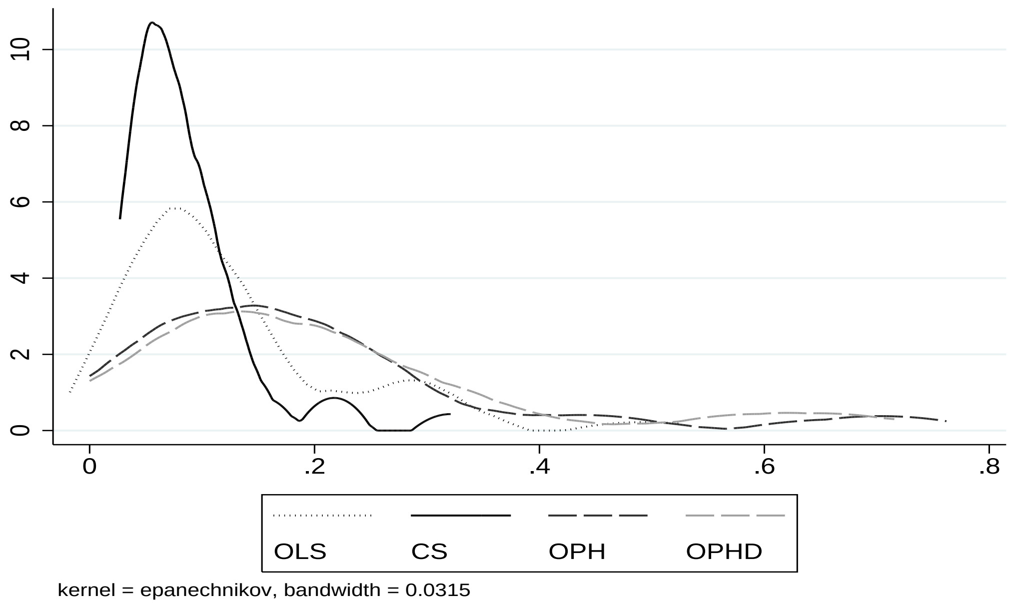
We start by estimating output and revenue elasticities for each of our four inputs: capital, labor, energy, and materials. Figure 1 plots the estimated elasticity of output (CS) and revenue (OLS, OPH, OPHD) with respect to capital and labor, distributed across industries (discussion of variable factor elasticities is below). There are non-trivial differences in both the location and the shape of the distributions. Most notably, the CS-based capital and labor elasticities tend to be smaller than regression-based estimates. At first glance, this is contrary to expectations since under CRS technology and CES demand the output elasticities should exceed revenue elasticities. We explore this issue further below.

We now turn our attention to revenue elasticity distributions. The direction of the bias in OLS-estimates is determined by several factors. First, since input demand functions are increasing in productivity, OLS estimates are biased upward. If this is important in our data and proxy methods correct for it, then we should see proxy-based distributions to the left of OLS. However, the direction of the bias depends on additional factors, for example selection. OP argue that since plants' profit and value functions are increasing in capital, larger establishments anticipate larger future returns and therefore can operate at lower current productivity levels, which entails a negative bias in OLS. If proxy methods correct for such selection-induced bias, and this effect is important in our data, then the distributions of quasi-fixed input elasticities (β_{ks} and β_{ls}) should be located to the right of OLS. Figures 1(a)-1(b) show that hybrid estimates of β_{ks} and β_{ls} are to the right of OLS, which is consistent with the predictions above. In addition, the estimates under OPH and OPHD are very close.[40] In contrast, variable input elasticities β_{es} and β_{ms} tend to be lower than cost shares, see

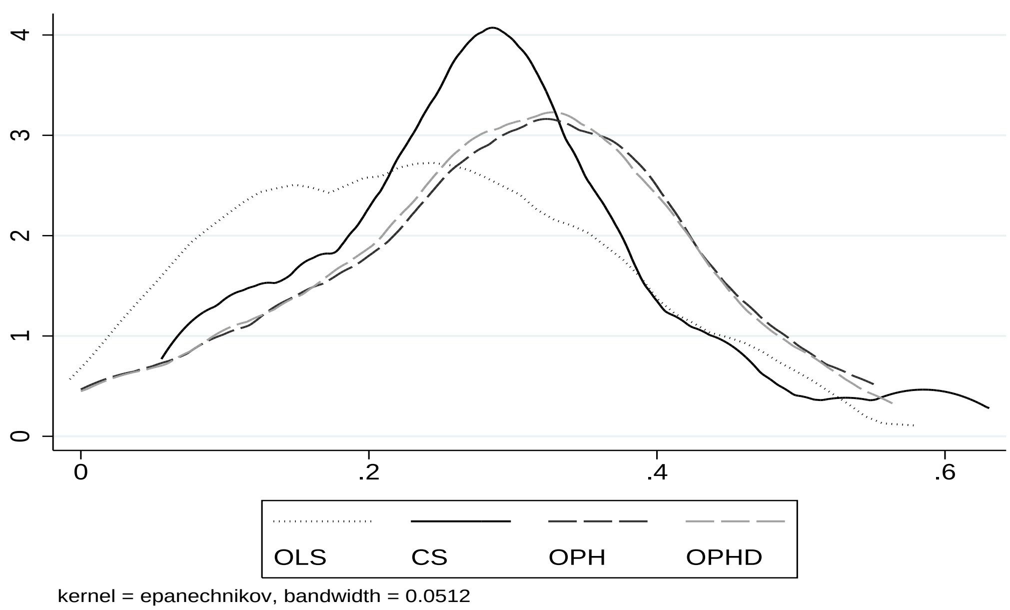
[39] By the NBER-CES database, this industry set accounts for about 36% percent of total Manufacturing value added between 1972 and 2010 with very little annual volatility.

[40] We tested the observed differences between OPH and OPHD for all elasticity distributions using the Kolmogorov-

figure A1. This is to be expected because β_{es} and β_{ms} are obtained as the revenue share of costs under OPH and OPHD, while they are calculated as the share of input expenditures in total costs under CS.



(a) Capital: $\hat{\beta}_{ks}$ and $\hat{\alpha}_{ks}$.



(b) Hours: $\hat{\beta}_{ls}$ and $\hat{\alpha}_{ls}$.

Figure 1: Cross-industry distributions of output (α) and revenue elasticities (β).

Next we pull together the estimates for the four inputs for our two preferred estimation methods, OPH and OPHD. Figures 2-3 summarize the overall implications of the differences between these

Smirnov test. Based on these tests, not shown here, we cannot reject the null hypothesis that OPH and OPHD yield the same elasticity distributions.

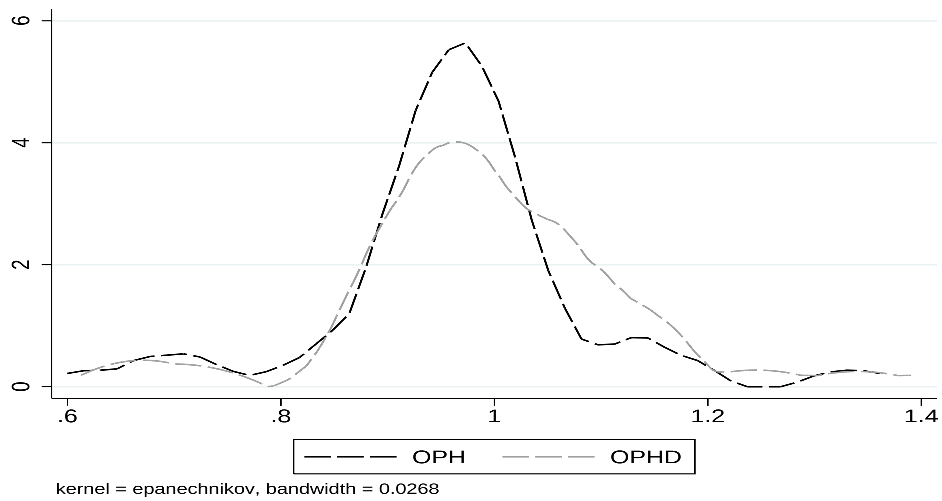


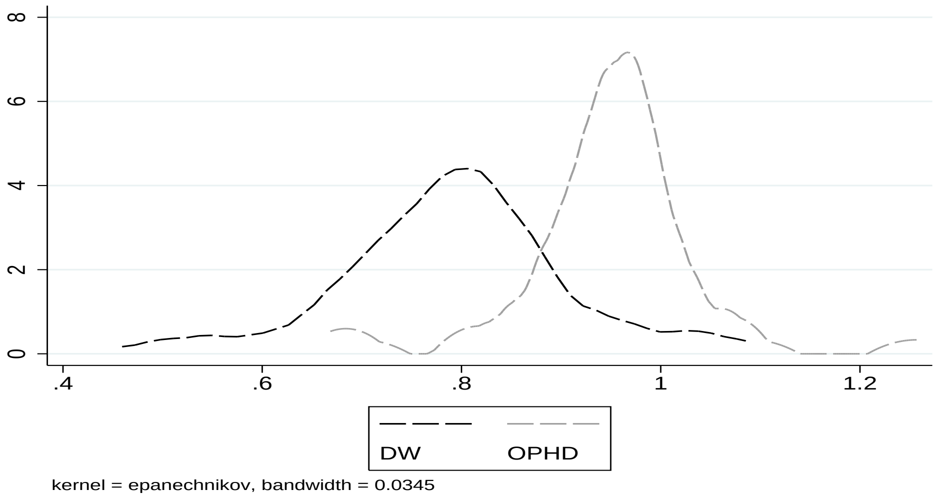
Figure 2: Revenue function curvature: $\widehat{\rho_s \gamma_s} = \sum_j \beta_{js}$.

estimation strategies. Figure 2 indicates that the similar revenue elasticities under OPH and OPHD (figures 1(a)-1(b)) imply consequently similar revenue curvature distributions. The sample averages of $\sum_j \widehat{\beta}_{js}$ under OPH and OPHD, respectively, are 0.94 and 0.95. The standard deviation is 0.24 for both distributions. The Kolmogorov-Smirnov test indicates that the two distributions are not significantly different. These patterns are broadly consistent with those in the literature. Estimated revenue curvature in the average industry was found to be close to 1 in many other papers, as well.[41] This common finding is important for the properties of tfpr_{is}^{rr} as we shall see below.

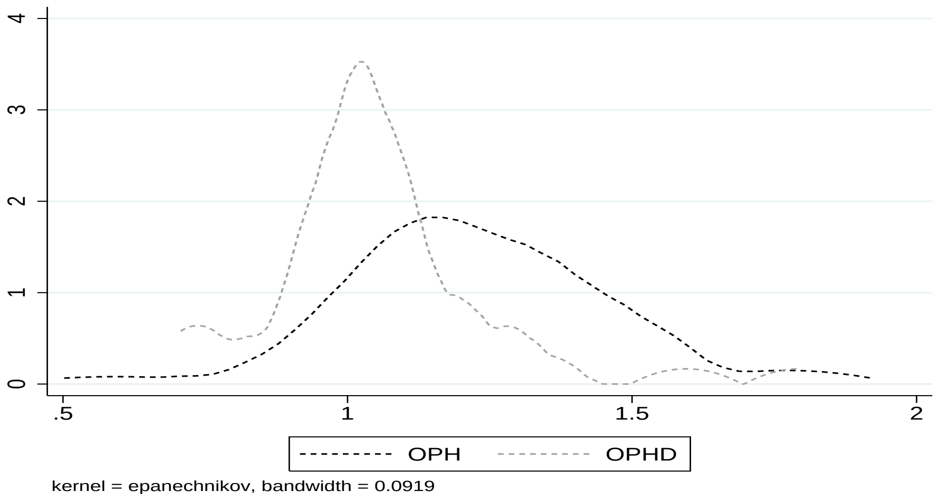
In our next two exercises, we decompose the curvature of the revenue function into its two component pieces: demand elasticities and returns to scale. We start with demand elasticities where, recall from our earlier discussion, we rely on two methods. The first is based on De Loecker and Warzynski and is denoted by DW. The second is based on Klette and Griliches and is denoted by OPHD. In order to compare the contributions of the demand parameter to revenue curvature, we plot the distributions of ρ_s estimates in Panel 3(a) using our two approaches. Although the heterogeneity is similar (standard deviations are 0.11 and 0.09), the ρ_s that are calculated using industry-level information are significantly lower than the OPHD estimates (means are 0.8 and 0.94). The Kolmogorov-Smirnov test confirms that the difference between the two distributions is statistically significant.

Finally, with these components, we can turn our attention to the implied returns to scale. We consider the implications of the disparate ρ_s estimates for the distribution of γ_s, shown in Figure 3(b). In the case of OPHD –where γ_s and ρ_s estimates are internally consistent–, mean returns to scale is close to one (mean γ_s is 1.01) although there is considerable variation across industries with some industries exhibiting decreasing while others showing increasing returns to scale. We also combine the β_{js} estimates from OPH with the ρ_s estimates from the DW procedure. This is mixing estimates from alternative procedures but instructive about their relationship. Taken at face value, the implied

[41] See, for example, Olley and Pakes (1996), Klette and Griliches (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2006), White et al. (2011), De Loecker (2011), Gandhi et al. (2012), and Gopinath et al. (2015).



(a) ρ_s estimates as in De Loecker and Warzynski (2012) and Klette and Griliches (1996).



(b) Implied returns to scale ($\widehat{\gamma}_s$).

Figure 3: Demand elasticity and returns to scale.

returns to scale by combining these estimates (labeled OPH in this figure) is significantly larger than 1 (mean 1.21). Such estimates are difficult to interpret given the mixing of the procedures but they do imply a tension between the estimates across the procedures (since the DW based estimates of markups rely on a CRS assumption).

Taking stock, the different estimation methods yield disparate elasticities and implied curvature of the revenue function. The proxy method yields substantial heterogeneity in revenue function curvature across industries with the average just below one. This heterogeneity may be further amplified into a wide range of returns to scale estimates, depending on the underlying demand elasticities. It is helpful to consider the implications of these findings for the firm dynamics literature that requires calibrating the curvature of the revenue function. To help preserve well-behaved optimization problems, it is typical to assume that $\rho_s \gamma_s \leq 1$. Our findings are not inconsistent with this assumption but indicate that estimated curvature and its sources may exhibit significant variation, depending on the estimation method. If the primary source of curvature is markups and they are substantial, i.e. in the 25 percent range (i.e., ρ_s=0.8), then the proxy methods imply that the returns to scale for production are well above one. This inference is not limited to our findings since, as noted above, it is a common finding in the empirical literature that the proxy estimates of the revenue function yield an estimate of the $\sum_j \widehat{\beta}_{js}$ close to or just below one. In short, the proxy method estimates imply that either markups are small or returns to scale of production are above one. Widespread recognition of this implication has been limited as often estimates of β_{js} have been interpreted as estimates of α_{js}.[42]

5 Implications of the differences in elasticity distributions

We now turn to exploring the effects of these differences in terms of basic properties of revenue productivity measures and allocative efficiency (AE). In particular, we explore the effect of these differences in terms of productivity dispersion, productivity correlations, and the relationship between productivity, growth and survival. We also investigate how these alternative productivity measures impact AE, and in turn, the sensitivity of AE to these different approaches to estimating factor and revenue elasticities.

5.1 Productivity dispersion and correlations

Does it matter whether one uses output or revenue elasticities to compute the composite input and in turn does it matter how one estimates the output and revenue elasticities? In spite of the large differences in elasticity estimates presented above, our results suggest that, at least on average, dispersion in revenue productivity is broadly similar across methods. The interquartile range, shown in the second column of table 1, indicates that the average productivity difference between establishments at the 75th and 25th percentiles in the average industry varies between 0.27 and 0.35 across the methods considered for the purposes of this paper. This narrow range amounts to a 31-42% productivity difference, indicating substantial within-industry dispersion in revenue productivity. When measuring dispersion using the standard deviation, the results are qualitatively the same (see the third column).[43]

We next investigate whether the choice of estimation method has consequences also for the productivity rank of establishments. The Pearson and Spearman correlations in Table 2 indicate that the association between proxy results (tfpr_{is}^{rr}) and CS (tfpr_{is}^{cs}) is generally weaker than between proxy methods themselves, but all correlations are higher than 0.7.

These findings suggest that tfpr_{is}^{cs} and tfpr_{is}^{rr} exhibit similar dispersion and are strongly correlated. The dispersion in tfpr_{is}^{rr} tends to be somewhat larger than tfpr_{is}^{cs}. Under the assumption of isoelastic (CES) demand, tfpr_{is}^{rr} is a measure of fundamentals. Consistent with the findings of the recent

[42] A notable exception is De Loecker (2011).

[43] Approximate 95% confidence intervals, constructed using bootstrapped standard errors of the interquartile range, not shown here, indicate that dispersion measures under proxy methods are higher than under CS (or OLS) but they are not significantly different from each other. More variants are explored in the working paper version, where we show that all of these methods imply similarly large productivity differences across establishments.

Table 1: Productivity dispersion implied by different methods.

	N (1000)	IQR	SDEV
OLS	589	0.27	0.31
CS	589	0.28	0.31
OPH	563	0.35	0.35
OPHD	563	0.35	0.36

All statistics are based on deviations of plant-level log-productivity from industry- and time-specific means and are calculated from a weighted distribution where the weights are based on the number of plant-year observations in an industry. The top and bottom 1% of the distributions are discarded in order to guard against the effect of outliers.

Table 2: Correlations among within-industry productivity distributions.

	OLS	CS	OPH	OPHD	OLS	CS	OPH	OPHD
	Pearson				Spearman			
OLS	1				1			
CS	0.86	1			0.88	1		
OPH	0.79	0.73	1		0.78	0.74	1	
OPHD	0.79	0.71	0.88	1	0.77	0.71	0.90	1

The top and bottom 1% of the distributions are discarded in order to guard against the effect of outliers.

literature, our results imply that whether or not tfpr_{is}^{cs} is an appropriate measure of distortions, it is positively correlated with, and similarly dispersed as, fundamentals.[44] An important part of what is driving this close correspondence between tfpr_{is}^{cs} and tfpr_{is}^{rr} is that they (at least on average) both generate the composite input using weights that sum close to one.

5.2 Growth and survival

In this sub-section, we explore whether one of the most important predictions from standard models of firm dynamics is robust to the aforementioned differences. For this purpose, we are interested in models in which firms face adjustment frictions on both the entry/exit and intensive margins.[45] In such a model where employment is the single production factor subject to adjustment frictions, incumbent firms have two key state variables each period: the prior period level of employment and the realization of productivity in the period. The standard prediction from this model is that, conditional on prior period level of employment, firms with sufficiently low productivity-draws exit and firms with higher realizations of productivity grow. Syverson (2011) highlights that the positive relationship between productivity, growth, and survival is a robust finding in the literature. If the productivity measure is tfpr_{is}^{rr}, this should not surprise us because it reflects fundamentals. However, there is no inherent reason the prediction should hold for tfpr_{is}^{cs}, given our motivating framework.

We consider this relationship for all establishments, exiters, and incumbents separately and esti-

[44] These findings are also consistent with studies that use price and quantity data to compute direct measures of tfpq_{is} and demand shocks. Foster et al. (2008), Foster et al. (2016c), Haltiwanger et al. (2018), and Eslava et al. (2013) provide evidence that tfpr_{is}^{cs} is highly correlated with direct measures of tfpq_{is} and positively correlated with demand shocks. Moreover, these studies find that tfpr_{is}^{cs} dispersion is slightly lower than tfpq_{is} dispersion, indicating that prices are inversely related to tfpq_{is} under downward sloping demand.

[45] For example, on the entry/exit margins see Hopenhayn (1992) and Hopenhayn and Rogerson (1993). For adjustment cost models at the firm-level on employment, see Cooper et al. (2007) and Elsby and Michaels (2013).

mate the following specification:

$$y_{it+1} = \beta_1 \omega_{it} + \beta_2 \theta_{size_{it}} + \mathbf{x}'_{it}\boldsymbol{\delta} + \epsilon_{it+1}, \tag{17}$$

where y_{it+1} denotes a plant-level outcome (either growth between periods t and $t+1$ or exit), ω is the idiosyncratic component of plant-level productivity (i.e., measures deviated from industry by year effects), $\theta_{size_{it}}$ is the control for initial size (employment) in the period, and $\mathbf{x}_{it}$ is a vector of additional controls including year effects, state effects, and the change in the unemployment rate at the state level that controls for cyclical effects.[46]

Table 3 shows $\widehat{\beta}_1$ from equation (17) for each outcome (column 1) and productivity estimator (columns 2-4). We focus on the two main alternatives: CS and OPH. All point estimates are statistically significant but these two are quite similar. A one standard deviation increase in tfpr^{cs}_{is} yields a roughly 4.3-log-point increase in growth and a 1.5-log-point decline in the probablity of exit. For tfpr^{rr}_{is}, the analogous estimates are 4.5 and 1.5 log points. The implication is that productivity and growth (exit) are positively (negatively) associated, irrespective of how productivity is measured. The similarity of conclusions is one of the reasons why Syverson (2011) states that the finding that high productivity plants are less likely to exit is one of the most ubiquitous findings in the literature.

The findings in this and the prior sub-section help explain why tfpr^{cs}_{is} remains a commonly used measure of firm performance in the empirical firm dynamics literature. Since Baily et al. (1992) and Foster et al. (2001), tfpr^{cs}_{is} has been commonly used for investigating a range of issues from the determinants of firm-level growth and survival, adjustment costs for capital and labor, and the relationship between firm performance, exporter status, and management practices (see, e.g., Syverson (2011) for a survey). Ten years after Hsieh and Klenow (2009) raised questions about the interpretation of this measure, tfpr^{cs}_{is} still remains commonly used in recent papers about firm dynamics.[47] This is relevant because one implication of our analysis is that, under the assumptions in this paper, this literature should use tfpr^{rr}_{is} instead of tfpr^{cs}_{is}. The continuing popularity of tfpr^{cs}_{is} may be partly due to the fact that, despite its theoretical appeal, tfpr^{rr}_{is} estimation using proxy methods has practical limitations (e.g., often requires defining industries at a 2 or 3-digit level to obtain sensible estimates for all industries) while calculating tfpr^{cs}_{is} at a detailed industry level is straightforward. In light of the results in this section, we believe the findings in the literature using tfpr^{cs}_{is} would likely be robust to the conceptually more appropriate index.[48] It remains an open question why these

[46] This is a simplified version of the specification considered by FGH. We follow them using integrated ASM-LBD data for this analysis. The ASM provides the distribution of plant-level productivity in any given year and the LBD provides the growth and survival outcomes for the full set of plants in the ASM in that year between t and $t+1$.

[47] For example, two recent prominent papers that use this as a measure of firm performance are Bloom et al. (2019) and Ilut et al. (2018). A main finding of the former paper is that plants with more structured management practices have higher tfpr^{cs}_{is}. The latter uses tfpr^{cs}_{is} to identify strikingly nonlinear responsiveness between hires and separations at the plant-level fundamentals.

[48] One reason that tfpr^{cs}_{is} might actually be preferred is unmeasured quality differences in materials inputs across plants. The lack of plant-level output prices extends to plant-level input prices. Inclusion of variation in output prices can help control for unmeasured input price variation since they tend to be positively correlated. See De Loecker et al. (2016).

distinct measures are so tightly linked empirically (although there are competing explanations such as adjustment costs and variable markups that increase with fundamentals). It might be tempting at this juncture to argue the details we have emphasized are not important. However, we now turn to analysis of allocative efficiency where these details are of critical importance.

Table 3: The productivity impact on outcomes for different estimation methods. Outcomes are employment growth among all establishments (row 1), exit (row 2), employment growth among continuers (row 3).

	OLS	CS	OPH	OPHD
overall growth	0.121***	0.140***	0.129***	0.079***
exit	-0.045***	-0.049***	-0.042***	-0.024***
conditional growth	0.033***	0.046***	0.049***	0.034***

The table shows estimates of $\widehat{\beta}_1$ in equation (17). *** denotes statistical significance at 1%. Standard errors are clustered at the state level. All regressions are based on trimmed productivity distributions (top and bottom 1% in each industry and year). Sample size information can be found in table A1.

5.3 Allocative efficiency

We now examine the importance of the distinction between tfpr_{is}^{cs} and tfpr_{is}^{rr} for allocative efficiency (AE). That is, we explore the sensitivity of AE to the differences in output and demand elasticities across estimation methods. AE is an ideal metric to assess these issues for multiple reasons. First, AE provides guidance on the aggregate consequences of plant-level heterogeneity in productivity. In addition, the exact shape of production technology and demand characteristics affect it directly, see equation (11). For example, A_{is} is measured using α_{js} as weights in the input index while τ_{is} is measured using α_{js}/γ_s as weights in the input index. Furthermore, AE depends on γ_s and ρ_s via the exponents in a non symmetric manner. Finally, (16) shows that AE can be expressed as a function of the two revenue productivity measures that are the focus of the above analysis.

We assess the effect of parameters on AE by comparing and contrasting the results implied by the aforementioned estimation methods. In CS, we measure output elasticities by cost shares of total costs. We consider a range of demand elasticities for this approach. First, we impose ρ_s=0.75 across industries, following the original methodology in BKR. Second, we impose ρ_s=0.8 for all industries, which is consistent with the mean of the ρ_s-distribution obtained using the DW method. In the third experiment, we use the non-degenerate ρ-distribution implied by DW. This exercise is interesting because it illustrates the effect of allowing for heterogeneity in ρ_s. In the fourth experiment, ρ_s=0.94 for all industries, which equals the mean of the ρ_s distribution obtained using OPHD. For OPH, we start with the revenue elasticities and then also consider multiple estimates of demand elasticities. Given the β_{js} and ρ_s, we can infer the output elasticity for factor j from the relationship $\alpha_{js}=\beta_{js}/\rho_s$. We consider ρ_s=0.75 and ρ_s=0.8 for all industries, and the two non-degenerate ρ_s-distributions implied by the DW and OPHD methods, respectively.[49]

[49]We also considered the CS case with the full distribution of demand elasticities from OPHD and the proxy method

In order to abstract from high-frequency variation in the empirical analysis, we calculate decade-specific timeseries averages of AE. Figure 4(a) shows the cross-industry average of AE_s^{COV} calculated under alternative parameter estimates discussed above. The different parameter values are indicated in the figure header. The first four sets of bars show average AE as implied by CS, the remaining four sets depict AE implied by OPH. Several observations emerge from Figure 4(a). First, AE is generally higher under CS in the average industry: for example, conditional on $\rho_s = 0.75$ (first set of bars), CS yields about twice the AE obtained under OPH (fifth set of bars). Second, AE's sensitivity to ρ_s depends on the estimation method. This can be seen by contrasting the change between the first and second bar sets against the change between the fifth and sixth sets, which reflects the effect of increasing ρ_s from 0.75 to 0.8. In particular, this increase yields lower AE under CS, which is consistent with the findings in BKR. However, the same increase results in higher AE under OPH. These conflicting findings indicate that the effect of the demand elasticity on AE depends on how factor elasticities are estimated. Allowing for heterogeneity around $E[\rho_s]$=0.8 yields similar results (second and third sets versus sixth and seventh sets). Third, increasing ρ_s significantly in CS results in lower AE (first and fourth), while increasing a heterogenous ρ_s yields higher AE under OPH (fifth and eighth). Fourth, the magnitude of the trend decline in AE depends critically on the elasticities. On average, AE under CS and ρ_s=0.75 declines by about 20 percent, while with the same ρ_s, average AE declines by about 40 percent under OPH.[50]

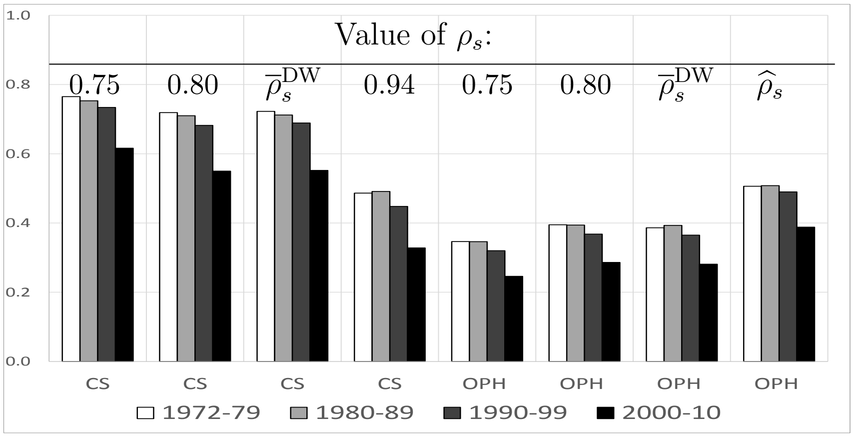
As discussed in section 2.2, the overall implications for AE depend on how sectoral AE measures are aggregated. When every production factor is deemed to be in fixed supply, weighting by revenue shares is appropriate, see equation (12). Figure 4(b) shows overall estimated AE in this case. The patterns for this case are broadly similar to those for the unweighted industry means in Figure 4(a). However, especially for OPH, the trend declines are larger when computing the revenue weighted geometric mean across industries. Figure 5 compares the AE concept of 4(b) with the one derived under the assumption of roundabout production technology in equation (14), i.e. assuming that intermediate inputs are endogenous.[51] The results indicate that this approach implies lower AE and a more pronounced trend decline. These findings highlight the fact that the way AE_s^{COV} is aggregated into an economy-wide estimate has an impact on both the level and dynamics of AE. The relationships here are complex since the elasticities enter the aggregate expressions for AE in a highly nonlinear manner. These aggregation issues are not our focus but the results show that the quantitative implications of the aggregation depend on the estimation methods for the elasticities.

To explore the determinants of sectoral AE further, we implement decomposition (15) empirically. The first step is to establish the relationship between AE_s^{COV} and the dispersion of tfpq_{is}, tfpr_{is}^{cs}, and

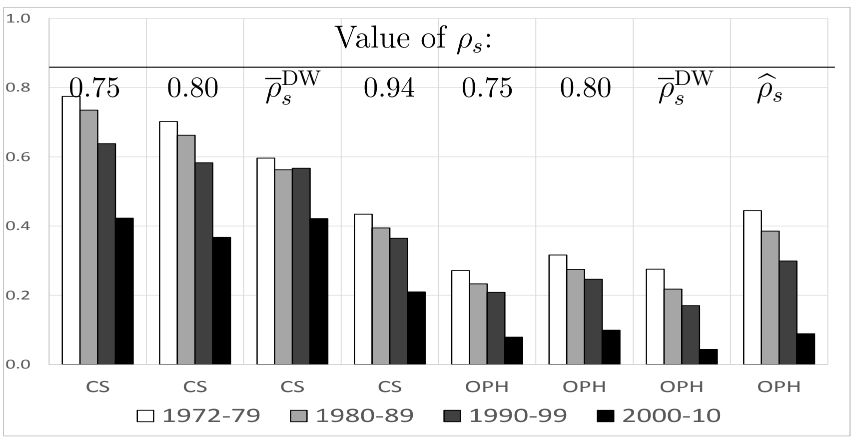
case using ρ_s=0.94 for all sectors. The results are consistent with those we report in this section and therefore we omit them for ease of presentation.

[50] These results are confirmed by the median AE_s^{COV}, not shown here.

[51] Fully adjusting AE_s^{COV} for NCRS as described in equation (14) has a negligible empirical effect relative to the case where cross-industry aggregation is based on weights defined as $\theta_s / \left(\sum_s \theta_s \left(1 - \alpha_{Ms}/\gamma_s\right)\right)$, as in BKR. Therefore, we focus on the latter case in the remainder of the analysis.



(a) Unweighted average, $N_S^{-1}\sum_s \widehat{\mathrm{AE}}_s^{COV}$.



(b) Weighted average: $\mathrm{AE}_s = \prod_s \left(\widehat{\mathrm{AE}}_s^{COV}\right)^{\theta_s}$ (fix K, L, M supply).

Figure 4: Descriptive statistics of AE_s^{COV} under different measures and demand elasticities.

Note: In industries where $1 < \rho_s \gamma_s$ at the 4-digit level, 2-digit estimates are used. $\overline{\rho}_s^{\mathrm{DW}}$ denotes industry-specific timeseries averages calculated as in De Loecker and Warzynski (2012).

the partial correlation between these two variables.[52] Interpreting the results of the decomposition, it is important to recall that the measure of tfpq_{is} is indirect and proportional to tfpr_{is}^{rr}. Figures 6(a)-6(c) show that while dispersion in tfpr_{is}^{cs} is very similar across estimation methods, dispersion in tfpq_{is} is lower and the partial correlation is higher under OPH. The latter finding provides guidance as to why OPH yields lower AE – the higher correlation between (measured) fundamentals (TFPQ) and distortions (TFPR^{cs}) under OPH is a drag on AE. Since it is the same underlying micro data, the difference in the correlation between fundamentals and distortions is driven entirely by differences

[52] tfpr_{is}^{cs} for OPH is computed in an internally consistent manner: the β_{js} and ρ_s are used to compute α_{js}, which then can be used to estimate cost shares as α_{js}/γ_s.

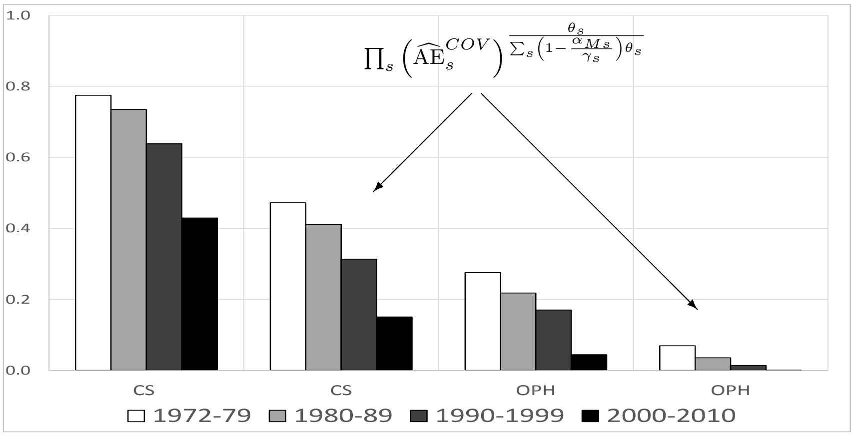


Figure 5: The effect of roundabout production on AE_s, see equation (14). Estimators: CS (ρ_s=0.75) and OPH($\overline{\rho}_s^{DW}$).

in the elasticity estimates. This finding highlights that the nature and empirical contribution of correlated distortions is being driven by estimated elasticities of the demand and production structure. In addition, inferences about the effect of curvature parameters depends on the distribution of the measured fundamentals and distortions.

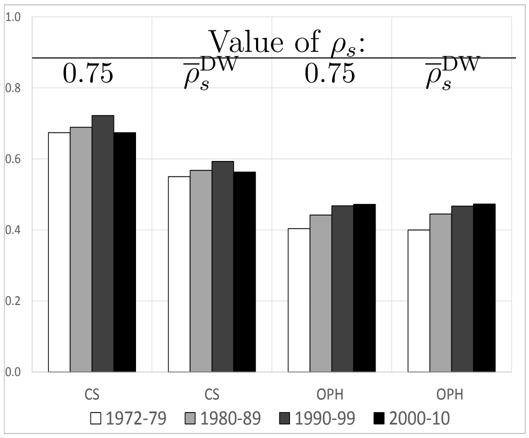
Over time, both productivity measures exhibit increasing dispersion and the correlation is also increasing. These findings are in line with the prediction that if distortions are positively correlated with productivity, then AE is decreasing in dispersion of either TFPQ_{is} or TFPR_{is}^{cs}, see section 2.2. However, these results also highlight that, in principle, falling AE may also be a consequence of rising dispersion in TFPQ while distortions themselves are not changing. These productivity indicators directly affect AE_s^{COV} via the covariance between relative productivities and distortions in equation (15). In order to assess their overall effect on AE_s^{COV}, Figure 6(d) shows the contributions of the two terms of the right hand side equation (15) evaluated under CS and OPH. The increasing dispersion in tfpq_{is} and tfpr_{is}^{cs} together with the increasing positive correlation yield a negative contribution for the second term in equation (15), which accounts for the majority of the decrease in measured AE.

One of the main implications of this analysis is that both factor and demand elasticities are crucial for inference about AE. Using the same underlying plant-level data, alternative estimation methods imply average level-differences that vary by more than a factor of two. In addition, the average industry trend decline from the 1970s to 2000s varies between about 20 percent and more than 40 percent. In interpreting this sensitivity, it is not simply that the different methods yield different average curvature of the revenue function. Instead, the different demand, output, and return to scale parameters interact with the underlying micro data in a highly nonlinear manner.

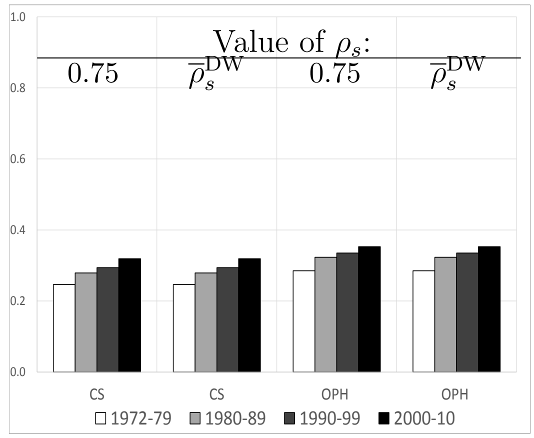
Another interesting insight from Figure 6 is that AE's sensitivity to parameters is reflected as the sensitivity of tfpq_{is} dispersion and the correlation between tfpq_{is} and tfpr_{is}^{cs}. In contrast, tfpr_{is}^{cs} dispersion does not seem to be as sensitive to these parameters. In other words, it is sensitivity of the dispersion of fundamentals and the correlation of fundamentals with distortions rather than sensitivity of the dispersion in distortions that matters. There is apparently more widespread agreement on distortions than on fundamentals. In spite of the sensitivity of the average and changing magnitude of AE to demand and output elasticities, there are some common messages from this analysis that help relate the findings in this section to the earlier basic facts. In all specifications, measured AE is declining over time, dispersion in tfpq_{is} and tfpr^{cs}_{is} are rising over time, and the correlation between these two alternative measures is rising over time.

We conclude this discussion by noting that since tfpq_{is} and tfpr^{rr}_{is} are proportional, many of the inferences above for tfpq_{is} also hold for tfpr^{rr}_{is}. However, this statement is holding the distribution of demand elasticities constant but our findings show that demand elasticitity estimates vary considerably across specifications. The magnitude and time series variation in AE depends critically on markups while properties of tfpr^{rr}_{is} (such as its dispersion) depends on revenue elasticities. This latter point reflects back on our earlier findings. There is more similarity in our findings –and more widespread agreement in the literature– about the sum of revenue elasticities than in its decomposition into markup and returns to scale (or demand elasticity and output elasticity) components. A message of our analysis is that for some purposes the decomposition of revenue elasticities into their components is not critical; but for structural analysis of AE, this decomposition is of critical importance.

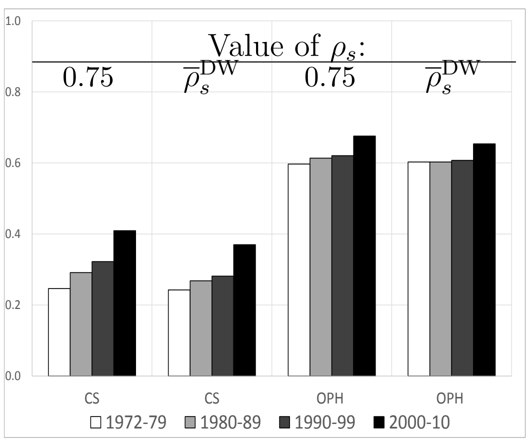
The sensitivity of the properties of tfpq_{is} to variation in ρ_s also helps reconcile some of our findings above. Table 2 shows that the dispersion of tfpr^{rr}_{is} is similar to tfpr^{cs}_{is} and they are highly correlated. In contrast, for specifications with a low ρ_s dispersion in tfpq_{is} is large and its correlation with tfpr^{cs}_{is} is low highlighting the sensitivity of (the indirect measure of) tfpq_{is} to ρ_s.



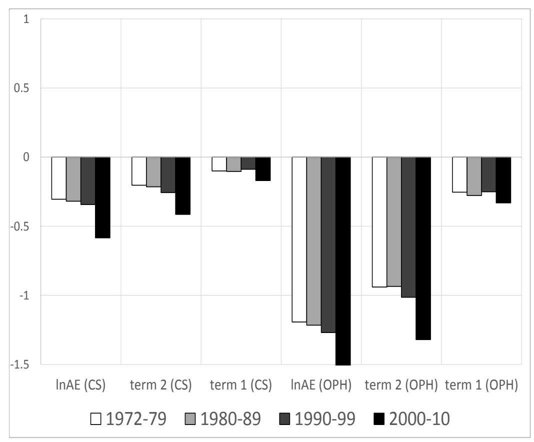
(a) Log-TFPQ dispersion



(b) Log-TFPR dispersion



(c) Partial correlation between log-TFPQ and log-TFPR



(d) Components of AE, CS(ρ_s=0.75) and OPH($\overline{\rho}_{DW}$)

Figure 6: AE and productivity moments.

Note: In industries where $1 < \rho_s \gamma_s$ at the 4-digit level, 2-digit estimates are used. $\overline{\rho}_s^{\mathrm{DW}}$ denotes industry-specific timeseries averages calculated as in De Loecker and Warzynski (2012).

6 Concluding remarks

Researchers have been using a variety of methods for estimating productivity at the firm-level. Absent data on prices and quantities, these methods yield what have become known as revenue productivity measures. It is perhaps less recognized that the differences across estimation methods have important consequences for interpretation since the alternative measures are different conceptually. The shares of input expenditures in total costs are equivalent to output elasticities under certain assumptions, while regression-based estimates are revenue elasticities absent data on prices and quantities. The revenue residuals implied by cost shares, or tfpr_{is}^{cs}, have increasingly become used as a measure of distortions. Under the same assumptions, the residual from revenue function estimation, or tfpr_{is}^{rr}, reflects fundamentals such as technical efficiency and demand shocks.

In spite of these conceptual differences, we find that these alternative measures are positively correlated, exhibit similar dispersion, and similar relationships with growth and survival suggesting that the effect of the differences in elasticity estimation are not significant. This helps explain why tfpr_{is}^{cs} remains a commonly used measure of firm performance in the empirical literature even though arguably tfpr_{is}^{rr} is the preferred measure. It remains an open question as to why these distinct measures are so tightly linked empirically although there are a number of competing explanations with empirical support (e.g., adjustment costs, variable markups that are increasing in fundamentals).

In contrast, the differences underlying these measures and estimation methods are critical for measuring allocative efficiency (AE) in a benchmark structural approach that has been developed in the recent literature. This benchmark AE is an ideal metric to assess the empirical importance of these issues because both production and demand parameters affect it directly. In addition, the joint distribution of fundamentals and distortions implied by these parameters is also critical. Our findings suggest that the properties of these parameter distributions are important for inferred AE, casting into doubt previous conclusions about relative levels of efficiency and productivity dynamics due to misallocation. An interesting implication of our analysis is that tfpr_{is}^{cs} and tfpr_{is}^{rr} are both critical inputs for benchmark AE. Put differently, since tfpr_{is}^{cs} reflects distortions while tfpr_{is}^{rr} reflects fundamentals, these revenue productivity measures contain important information for AE.

In an attempt to establish the relationship between AE and properties of the within-industry productivity distribution, we decompose industry-level AE into an aggregate term and a term that captures the covariance between measured relative productivities and distortions. The decomposition is interesting because the covariance is, by definition, a function of the dispersion in these two variables and the correlation between them. This relationship creates a useful link between industry-level AE and these indicators. The empirical implementation of the decomposition indicates that increasing dispersion in the alternative revenue productivity measures we consider and the increasing correlation between them account for the majority of the decrease in measured AE.

It remains to be seen whether the benchmark AE methodology that we build on in this paper is the most appropriate method for aggregating the properties of alternative measures of within-industry revenue productivity given the strong assumptions required to implement this methodology. Moreover, our findings on the sensitivity of the magnitude to the demand and output elasticities suggest caution in interpreting the quantitative variation in AE in alternative studies. But our finding that in all specifications, the alternative revenue productivity measures exhibit rising dispersion and increasing correlation suggests that distinguishing between these closely related but distinct revenue productivity measures is important. Understanding what is driving the cross sectional and time series relationship in these alternative measures should be a priority for future research.

Our analysis helps reconcile alternative views on the importance of distinguishing between these measures. On the one hand, their close correspondence in terms of dispersion, correlation, and relationship between firm outcomes like growth and survival (and other measures like management practices and exporters) stems from the finding that both measures yield similar composite inputs at the micro level. The reason behind this is that both are based on weights that sum either exactly to or close to one. In contrast, implementing AE empirically using only revenue and input data requires decomposing revenue elasticities into their demand elasticity and output elasticity components. The latter decomposition is quite sensitive to methods. Put differently, there is much less agreement about markups and returns to scale than there is about the overall curvature of the revenue function (although even here there is disagreement). We also show that this sensitivity of the decomposition of the revenue elasticity into its components translates into greater sensitivity of dispersion of the fundamentals and the correlation of the latter with distortions than in the dispersion of distortions themselves. It is interesting that there is more widespread agreement (across methods) on the dispersion of distortions/wedges than there is about fundamentals themselves.

In sum, it is important to understand when the devil is in the details. One remaining detail is the impact of heterogeneous and endogenous plant-level product and input prices. The results from our admittedly restrictive demand analysis can be used to make inferences about output elasticities and returns to scale; however, without plant-level data on prices and/or quantities, the effects of prices on other key stylized facts are difficult to quantify. We have commented on the likely impact of endogenous demand-side factors throughout but it would be of interest to consider this issue in more depth. We think that exploring the role of endogenous demand-side factors in the current context will require comparing approaches that include direct measures of prices and quantities (for the limited number of products with such information) with methods that impose strong functional form assumptions (e.g., isoelastic demand structures) to deal with these issues. We also neglect the impact of input price heterogeneity which may be another source of dispersion in micro-level productivity measures. A related issue is that we have relied on a framework where output and demand elasticities do not vary across plants. Variable markups across plants as well as differences in technologies across plants are likely important contributors to measured differences in productivity across plants. This paper is intended to provide some structure to thinking about these issues; it is our hope that it opens the door to future discussions.

References

Ackerberg, Daniel, Kevin Caves, and Garth Frazer. Structural identification of production functions. MPRA Paper 38349, University Library of Munich, Germany, December 2006.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. Identification Properties of Recent Production Function Estimators. *Econometrica*, 83:2411–2451, November 2015.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. Dynamic Inputs and Resource (Mis)Allocation. *Journal of Political Economy*, 122(5):1013 – 1063, 2014.

Baily, Martin N., Charles R. Hulten, and David Campbell. Productivity dynamics in manufacturing plants. In Baily, Martin N. and C. Winston, editors, *Brookings Papers on Economic Activity: Microeconomics*, volume 4. DC: Brookings Institute, 1992.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review*, 103(1):305–34, February 2013.

Bartelsman, Eric J. and Wayne Gray. The nber manufacturing productivity database. Working Paper 205, National Bureau of Economic Research, October 1996.

Bils, Mark, Peter J. Klenow, and Cian Ruane. Misallocation or Mismeasurement? 2017 Meeting Papers 715, Society for Economic Dynamics, 2017.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. What Drives Differences in Management Practices? *American Economic Review*, forthcoming 2019.

Cooper, Russell W. and John C. Haltiwanger. On the Nature of Capital Adjustment Costs. *Review of Economic Studies*, 73(3):611–633, 2006.

Cooper, Russell W., John C. Haltiwanger, and Jonathan Wiliis. Search frictions: Matching aggregate and establishment observations. *Journal of Monetary Economics*, 54:56–78, 2007.

De Loecker, Jan. Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, 79(5):1407–1451, 09 2011.

De Loecker, Jan and Frederic Warzynski. Markups and Firm-Level Export Status. *American Economic Review*, 102(6):2437–2471, October 2012.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. Prices, markups, and trade reform. *Econometrica*, 84(2):445–510, 03 2016.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. Changing business dynamics and productivity: Shocks vs. responsiveness. 2017.

Elsby, Michael and Ryan Michaels. Marginal Jobs, Heterogeneous Firms, and Unemployment Flows. *American Economic Journal: Macroeconomics*, 5(1):1–48, 2013.

Eslava, Marcela and John Haltiwanger. The Life-cycle Growth of Plants: The Role of Productivity, Demand and Wedges. Unpublished manuscript, 2018.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler. Trade and market selection: Evidence from manufacturing plants in colombia. *Review of Economic Dynamics*, 16(1):135–158, 2013.

Foster, Lucia, Cheryl Grim, and John Haltiwanger. Reallocation in the Great Recession: Cleansing or Not? *Journal of Labor Economics*, 34(S1):S293–S331, 2016a.

Foster, Lucia S., John C. Haltiwanger, and Cornell J. Krizan. Aggregate Productivity Growth. Lessons from Microeconomic Evidence. In *New Developments in Productivity Analysis*, NBER Chapters, pages 303–372. National Bureau of Economic Research, Inc, 2001.

Foster, Lucia S., John C. Haltiwanger, and Chad Syverson. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425, 2008.

Foster, Lucia S., Cheryl A. Grim, John C. Haltiwanger, and Zoltan Wolf. Firm-Level Dispersion in Productivity: Is the Devil in the Details? *American Economic Review*, 106(5):95–98, May 2016b.

Foster, Lucia S., John C. Haltiwanger, and Chad Syverson. The slow growth of new plants: Learning about demand? *Economica*, 83(1):91–129, 2016c.

Gandhi, Amit, Salvador Navarro, and David Rivers. On the Identification of Production Functions: How Heterogeneous is Productivity? 2012 Meeting Papers 105, Society for Economic Dynamics, 2012.

Gopinath, Gita, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. Capital Allocation and Productivity in South Europe. *NBER Working Paper Series*, (21453), 2015.

Haltiwanger, John. Firm Dynamics and Productivity: TFPQ, TFPR, and Demand Side Factors. *Economia*, 0:3–26, Fall 2016.

Haltiwanger, John and Zoltan Wolf. Reallocation and Productivity: Structural Estimation and Interpretation. Unpublished manuscript, 2018.

Haltiwanger, John, Robert Kulick, and Chad Syverson. Misallocation measures: The distortion that ate the residual. 2018.

Hopenhayn, Hugo A. Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica*, 60 (5):1127–50, September 1992.

Hopenhayn, Hugo A. and Richard Rogerson. Job turnover and policy evaluation: A general equilibrium analysis. *Journal of Political Economy*, 101(5):915–938, 1993.

Hsieh, Chang-Tai and Peter J. Klenow. Misallocation and manufacturing TFP in china and india. *Quarterly Journal of Economics*, 124(4):1403–1448, November 2009.

Ilut, Cosmin, Matthias Kehrig, and Martin Schneider. Slow to hire, quick to fire: Employment dynamics with asymmetric responses to news. *Journal of Political Economy*, 126(5):2011–2071, 2018.

Klette, Tor Jakob and Zvi Griliches. The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous. *Journal of Applied Econometrics*, 11(4):343–61, July-Aug. 1996.

Levinsohn, James A. and Amil Petrin. Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341, April 2003.

Marschak, Jacob and William H. Andrews. Random simultaneous equation and the theory of production. *Econometrica*, 12(3/4):143–205, 09 1944.

Olley, Steven G. and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, 1996.

Restuccia, Diego and Richard Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720, October 2008.

Syverson, Chad. What determines productivity. *Journal of Economic Literature*, 49(2):326–365, 2011.

White, T. Kirk, Jerome P. Reiter, and Amil Petrin. Plant-Level Productivity and Imputation of Missing Data in the Census of Manufactures. Working Papers 11-02, Center for Economic Studies, U.S. Census Bureau, January 2011.

Wooldridge, Jeffrey M. On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114, September 2009.

A Appendix

A.1 tfpr^{rr} and fundamentals

We derive the properties of tfpr$_{is}^{rr}$ under the more general specification of demand given by $P_{is}Q_{is} = P_s Q_s^{1-\rho_s} Q_{is}^{\rho_s - 1}\xi_{is}Q_{is} = P_s Q_s^{1-\rho_s} Q_{is}^{\rho_s}\xi_{is}$. In this case, log plant level revenues can be written as

$$p_{is} + q_{is} = \rho_s q_{is} + (1-\rho_s)q_s + p_s + \ln \xi_{is} = \rho_s \left(\sum_j \alpha_{js} x_{ijs} + a_{is} \right) + (1-\rho_s)q_s + p_s + \ln \xi_{is}.$$

This permits characterizing tfpr$_{is}^{rr}$ as:

$$p_{is} + q_{is} - \rho_s \sum_j \alpha_{js} x_{ijs} = p_{is} + q_{is} - \sum_j \beta_{js} x_{ijs} = \rho_s a_{is} + \ln \xi_{is} + (1-\rho_s)q_s + p_s,$$

which says that tfpr$_{is}^{rr}$ is a function of tfpq$_{is}$ (a_{is}), demand shocks (ξ_{is}), and sectoral level factors (q_s, p_s and ρ_s). In the main text, we abstract from idiosyncratic demand shocks and sectoral level factors for transparency. We estimate the β_{js} using regression (proxy) methods. In some robustness analysis, we also use the Klette and Griliches (1996) approach to jointly estimate β_{js} and ρ_s by including a measure of industry-level output as a regressor. This permits us to back out the α_{js} from the combined estimates and provides an alternative method to cost shares for estimating α_{js}. The advantage of this approach is that is does not impose CRS. The disadvantage of this approach is that in the absence of data on plant level prices and quantities, this is pushing the data quite hard. Foster, Grim, Haltiwanger, and Wolf (2016b) discuss the latter limitations in more depth.

A.2 AE under NCRS

A.2.1 Industry-level prices

Defining $\tau_{is} = \prod_j (1+\tau_{is}^j)^{\frac{\alpha_{js}}{\gamma_s}}$, we have:

$$P_s = Q_s^{\frac{1-\gamma_s}{\gamma_s}} \frac{1}{\kappa_s} \left(\sum_i \left(\frac{A_{is}}{\tau_{is}^{\gamma_s}} \right)^{\frac{\rho_s}{1-\rho_s\gamma_s}} \right)^{\frac{\rho_s\gamma_s - 1}{\rho_s\gamma_s}} \tag{18}$$

where $\kappa_s = \left(\rho_s \prod_j \left(\frac{\alpha_{js}}{w_{js}} \right)^{\alpha_{js}/\gamma_s} \right)^{-1}$ is a function of input prices and parameters.

A.2.2 Industry-level Distortions

It can be shown that industry-level distortions can be written as a function of TFPR$_s$ and a constant:[53]

$$\widetilde{\tau}_s = \kappa_s TFPR_s = \kappa_s \frac{P_s Q_s}{\prod_j X_{js}^{\alpha_{js}/\gamma_s}}. \tag{19}$$

As noted in Bils, Klenow, and Ruane (2017) (hereafter BKR), this can be expressed as a product of sectoral input distortions, which are in turn revenue-weighted harmonic means of plant-level input

[53] The formula can be obtained by writing TFPR$_s$ as a geometric average of sectoral marginal revenue product where the weights are based on the cost shares of respective inputs.

distortions. Note that $\widetilde{\tau}_s$ can be written as a function of idiosyncratic physical productivities and distortions. To do so, use that $\text{TFPR}_s^{cs} = \sum_i \frac{I_{is}}{\sum_i I_{is}} TFPR_{is}^{cs}$, where I_{is} denotes the establishment's cost-share based input index.[54] Expressing I_{is} as a function of A_{is}, τ_{is} and parameters implies

$$\widetilde{\tau}_s = \frac{\sum_i A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \left(\tau_{is}\right)^{\frac{-\rho_s\gamma_s}{1-\rho_s\gamma_s}}}{\sum_i A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \left(\tau_{is}\right)^{\frac{-1}{1-\rho_s\gamma_s}}} = \frac{S_1}{S_2}. \tag{20}$$

A.2.3 Industry-level TFP

We define the industry-productivity measure consistent with Hsieh and Klenow (2009) and BKR, where the denominator is the input index weighted by cost shares $\text{TFP}_s = \left(\prod_j X_{js}^{\alpha_{js}/\gamma_s}\right)^{-1} Q_s$. Multiplying and dividing by $P_s^{\gamma_s}$ yields

$$TFP_s = \frac{P_s^{\gamma_s} Q_s}{P_s^{\gamma_s} \prod_j X_{js}^{\frac{\alpha_{js}}{\gamma_s}}}. \tag{21}$$

Combining this expression with equation (18) yields $TFP_s = \left(\prod_j X_{js}^{\frac{\alpha_{js}}{\gamma_s}}\right)^{-1} \left(\sum_i \left(\frac{A_{is}}{\tau_{is}^{\gamma_s}}\right)^{\frac{\rho_s}{1-\rho_s\gamma_s}}\right)^{\frac{1-\rho_s\gamma_s}{\rho_s}} \times$ $\kappa_s^{\gamma_s} P_s^{\gamma_s} Q_s^{\gamma_s - 1} Q_s$, which can be rearranged as

$$TFP_s = \left(\sum_i \left(\frac{A_{is}}{\tau_{is}^{\gamma_s}}\right)^{\frac{\rho_s}{1-\rho_s\gamma_s}}\right)^{\frac{1-\rho_s\gamma_s}{\rho_s}} \frac{\left(\frac{\kappa_s P_s Q_s}{\prod_j X_{js}^{\frac{\alpha_{js}}{\gamma_s}}}\right)^{\gamma_s}}{\left(\prod_j X_{js}^{\frac{\alpha_{js}}{\gamma_s}}\right)^{1-\gamma_s}} = \left(\prod_j X_{js}^{\frac{\alpha_{js}}{\gamma_s}}\right)^{\gamma_s - 1} \left(\sum_i \left(\frac{A_{is}}{\tau_{is}^{\gamma_s}}\right)^{\frac{\rho_s}{1-\rho_s\gamma_s}}\right)^{\frac{1-\rho\gamma_s}{\rho}} \widetilde{\tau}_s^{\gamma_s}$$

$$= \left(\prod_j X_{js}^{\frac{\alpha_{js}}{\gamma_s}}\right)^{\gamma_s - 1} \left(\sum_i A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \left(\frac{\tau_{is}}{\widetilde{\tau}_s}\right)^{\frac{\rho_s\gamma_s}{1-\rho_s\gamma_s}}\right)^{\frac{1-\rho_s\gamma_s}{\rho_s}} \tag{22}$$

This is analogous to the expression obtained in Appendix 1 of HK.

A.2.4 Maximum of TFP$_s$

This section outlines the solution to the constrained optimization that yields AE_s under NCRS. Using the notation in (20), the Lagrangean of the problem is given by

$$\mathcal{L}(\tau_{is}) = \left(\sum_i A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \left(\frac{\tau_{is}}{\widetilde{\tau}_s}\right)^{-\frac{\rho_s\gamma_s}{1-\rho_s\gamma_s}}\right)^{\frac{1-\rho_s\gamma_s}{\rho_s}} + \lambda\left(\widetilde{\tau}_s - \sum_i \theta_i^I \tau_{is}\right) = \widetilde{\tau}_s^{\gamma} S_1^{\frac{1-\rho_s\gamma_s}{\rho_s}} + \lambda\left(\widetilde{\tau}_s - \frac{S_1}{S_2}\right), \tag{23}$$

where $S_1 = \sum_i A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \tau_{is}^{\frac{-\rho_s\gamma_s}{1-\rho_s\gamma_s}}$ and $S_2 = \sum_i A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \tau_{is}^{\frac{-1}{1-\rho_s\gamma_s}}$. The first derivative of $\mathcal{L}(\tau_{is})$ is shown below:

$$\frac{\partial \mathcal{L}}{\partial \tau_{is}} = -\gamma_s \widetilde{\tau}_s^{\gamma_s} S_1^{\frac{1-\rho_s\gamma_s}{\rho_s} - 1} A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \tau_{is}^{-\frac{1}{1-\rho_s\gamma_s}} + \lambda \frac{\rho_s\gamma_s A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \tau_{is}^{\frac{-1}{1-\rho_s\gamma_s}} S_2 - S_1 A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \tau_{is}^{\frac{-1}{1-\rho_s\gamma_s} - 1}}{(1-\rho_s\gamma_s) S_2^2}. \tag{24}$$

[54] This is because $\sum_i \frac{I_{is}}{\sum_i I_{is}} TFPR_{is}^{cs} = \sum_i \frac{I_{is}}{\sum_i I_{is}} \frac{R_{is}}{I_{is}} = \frac{\sum_i R_{is}}{\sum_i I_{is}} = TFPR_s^{cs}$.

Summing (24) over i yields a condition that can be solved for λ. Plugging the resulting expression back to (24) and rearranging implies $\tau_{is} = \widetilde{\tau}_s$. Differentiating (24) with respect to τ_{is} yields $-2\gamma_s(1-\rho_s\gamma_s)^{-1} S_2^{-2} S_1^{\frac{1-\rho_s\gamma_s}{\rho_s}} A_{is}^{\frac{\rho_s}{1-\rho_s\gamma_s}} \widetilde{\tau}_s^{\frac{-2}{1-\rho_s\gamma_s}-2+\gamma_s} \sum_{j\neq i} A_{js}^{\frac{\rho_s}{1-\rho_s\gamma_s}}$. Since $0 < A_{is}, \tau_{is}, \rho_s, \gamma_s$, the sign of the second derivative depends on the sign of $(1-\rho_s\gamma_s)$. If $\rho_s\gamma_s < 1$ then $\frac{\partial^2 TFP_s}{\partial \tau_i^2}\Big|_{\widetilde{\tau}_s} < 0$ and therefore $\tau_{is} = \widetilde{\tau}_s$ is a maximum point. However, if $1 < \rho_s\gamma_s$ then $0 < \frac{\partial^2 TFP_s}{\partial \tau_{is}^2}\Big|_{\widetilde{\tau}_s}$ and $\tau_{is} = \widetilde{\tau}_s$ cannot be a maximum point.

A.2.5 Aggregate and sectoral production

Aggregate output is assumed to be a Cobb Douglas CRS aggregate of sectoral output. This implies:

$$Q = \prod_s^S Q_s^{\theta_s} = \prod_s^S \left(A_s \prod_j X_{js}^{\frac{\alpha_{js}}{\gamma_s}} \right)^{\theta_s} \tag{25}$$

It can then be shown that aggregate output Q can be written as a product of the geometric averages of industry-level technical efficiencies, revenue shares, cost shares, distortions and inputs:

$$\begin{aligned} Q = &\prod_s^S A_s^{\delta_{s1}} \times \prod_S^S \theta_s^{\delta_{s1}} \times \prod_s^S \left[\prod_j \left(\frac{\alpha_{js}}{\gamma_s} \right)^{\frac{\alpha_{js}}{\gamma_s}} \right]^{\delta_{s1}} \times \prod_s^S \tau\tau_s^{-\delta_{s1}} \times (1+\tau^X)^{-\delta_{s2}} \\ &\times \prod_{j\neq M} \left[X_j \left(\sum_s \theta_s \rho_s \alpha_{js} \left(1 - \frac{\alpha_{Ms}}{\gamma_s} \right) \frac{1+\tau^j}{1+\tau_s^j} \right)^{-1} \right]^{\delta_{s3}}, \end{aligned} \tag{26}$$

where $\delta_{s1} = \frac{\theta_s}{\sum_s^S \theta_s \left(1-\frac{\alpha_{Ms}}{\gamma_s}\right)}$, $\delta_{s2} = \frac{\sum_s^S \theta_s \frac{\alpha_{Ms}}{\gamma_s}}{\sum_s^s \theta_s \left(1-\frac{\alpha_{Ms}}{\gamma_s}\right)}$, $\delta_{s3} = \frac{\sum_s^S \theta_s \frac{\alpha_{js}}{\gamma_s}\left(1-\frac{\alpha_{Ms}}{\gamma_s}\right)}{\sum_s^s \theta_s \left(1-\frac{\alpha_{Ms}}{\gamma_s}\right)}$. Defining $\widetilde{\alpha}_j = \frac{\sum_s^S \theta_s \left(\frac{\alpha_{Ms}}{\gamma_s}\right)\alpha_{js}}{\sum_s^S \theta_s \left(\frac{\alpha_{Ms}}{\gamma_s}\right)}$ and aggregate consumption or value added as output less intermediate input $C = Q - M$, the expression for aggregate TFP is given by TFP$=C/\prod_{j\neq M} X_j^{\widetilde{\alpha}_j}$. Let $\overline{T}$ denote the part of the expression that depends only on sectoral distortions and parameters. In addition, adjust equation (26) for intermediate input use. Then the following expression can be obtained for aggregate TFP:

$$TFP = \overline{T} \times \prod_s^S TFP_s^{\frac{\theta_s}{\sum_s \theta_s \left(1-\frac{\alpha_{Ms}}{\gamma_s}\right)}}. \tag{27}$$

A.2.6 Accounting for input demand

It can be shown that if sectoral level inputs with inelastic aggregate supply, so long as *average* sectoral distortions are the same under a new distribution of distortions, sectoral capital and labor are unchanged in the "undistorted" counterfactual. For our analysis, we ignore between sector distortions and assume that average sectoral distortions are the same. Since aggregate labor and capital supply is assumed to be inelastic, AE$_s$ is given by

$$AE_s = \frac{A_s}{A_s^*} = \left(\sum_i \left(\frac{A_{is}}{\tilde{A}_s} \right)^{\frac{\rho_s}{1-\rho_s\gamma_s}} \left(\frac{\tau_{is}}{\tau_s} \right)^{\frac{\rho_s\gamma_s}{1-\rho_s\gamma_s}} \right)^{\frac{1-\rho_s\gamma_s}{\rho_s}} \left(\frac{M_s^*}{M_s} \right)^{\frac{\alpha_{Ms}(1-\gamma_s)}{\gamma_s}}. \tag{28}$$

In addition, sectoral intermediate input demand is proportional to total output: $M_s = \theta_s Q \rho_s \gamma_s (1-\gamma_s)\frac{1}{1+\tau_s^M}$. So long as sectoral average intermediate distortions are held constant then, the ratio of undistorted inputs to distorted inputs in industry s can be expressed as:

$$M_s^*/M_s = Q_s^*/Q, \tag{29}$$

where Q_s^* is the aggregate output under the regime where distortions in sector s are equalized, holding average distortions constant. Thus, the ratio of aggregate "s-undistorted" output to realized output is equivalent to the ratio of intermediates. Using equation (26), we can obtain an expression for Q_s^*. On condition that average distortions are held constant between the actual and counterfactual cases, the only change relative to (26) is that the leading term of Q_s^* is given by $(A_s^*)^{\frac{\theta_s}{\sum_k \theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)}} \prod_{k\neq s} A_k^{\frac{\theta_k}{\sum_k \theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)}}$, i.e. only the sectoral productivity of the sth industry changes. It follows that the ratio in equation (29) can be written as a function of allocative efficiency:

$$Q_s^*/Q = (A_s^*/A_s)^{\frac{\theta_s}{\sum_k \theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)}} = (AE_s)^{\frac{-\theta_s}{\sum_k \theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)}}. \tag{30}$$

Substituting (30) into (28), we see that allocative efficiency is a function of AE_k^{COV} and itself:

$$AE_s = \left(\sum_i \left(\frac{A_{is}}{\widetilde{A}_s}\right)^{\frac{\rho_s}{1-\rho_s\gamma_s}} \left(\frac{\tau_{is}}{\widetilde{\tau}_s}\right)^{\frac{\rho_s\gamma_s}{1-\rho_s\gamma_s}}\right)^{\frac{1-\rho_s\gamma_s}{\rho_s}} \times AE_s^{\frac{-\theta_s\alpha_{Ms}(1-\gamma_s)/\gamma_s}{\sum_k \theta_k(1-\alpha_{Mk}/\gamma_k)}},$$

which means we can solve for AE$_s$:

$$AE_s = \left[\left(\sum_i \left(\frac{A_{is}}{\widetilde{A}_s}\right)^{\frac{\rho_s}{1-\rho_s\gamma_s}} \left(\frac{\tau_{is}}{\widetilde{\tau}_s}\right)^{\frac{\rho_s\gamma_s}{1-\rho_s\gamma_s}}\right)^{\frac{1-\rho_s\gamma_s}{\rho_s}}\right]^{\frac{\sum_k \theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)}{\sum_k \theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)+\theta_s\alpha_{Ms}\frac{1-\gamma_s}{\gamma_s}}}. \tag{31}$$

As in BKR, the contribution of sectoral TFP to aggregate TFP can be written as

$$TFP \propto \prod_s AE_s^{\frac{\theta_s}{\sum_s \theta_s\left(1-\frac{\alpha_{Ms}}{\gamma_s}\right)}} \prod_s (A_s^*)^{\frac{\theta_s}{\sum_s \theta_s\left(1-\frac{\alpha_{Ms}}{\gamma_s}\right)}}$$

Thus, we can separate the "undistorted" effect of sectoral TFP (which incorporates the influence of returns to scale) on aggregate TFP from the allocative efficiency effect on TFP. To relate to the variance-covariance term, we would need to aggregate expand the exponent, yielding:

$$TFP \propto \prod_s \left(AE_s^{COV}\right)^{\frac{\theta_s}{\sum_k \theta_k\left(1-\frac{\alpha_{Mk}}{\gamma_k}\right)+\theta_s\frac{\alpha_{Ms}}{\gamma_s}(1-\gamma_s)}} \tag{32}$$

A.2.7 Impacts of ρ_s and γ_s

To understand why ρ_s and γ_s enter asymmetrically into sectoral TFP given the following CES aggregator, note we can write plant level output as follows:

$$Q_{is} = (\rho_s\gamma_s P_s Q_s^{1-\rho_s})^{\frac{\gamma_s}{1-\rho_s\gamma_s}} \left(\frac{A_{is}}{\tau_{is}^{\gamma_s}}\right)^{\frac{1}{1-\rho_s\gamma_s}} \left[\prod_j \left(\frac{\alpha_{js}}{w_{js}}\right)^{\alpha_{js}}\right]^{\frac{\gamma_s}{1-\rho_s\gamma_s}} \tag{33}$$

Note here that if the plant does not account for the impact of its decisions on aggregate output and prices (which by assumption it does not), then the elasticity of production with respect to a change in A_{is} is $1/(1-\rho_s\gamma_s)$. Here we see that both returns to scale γ_s and downward-sloping demand ρ_s play a role in the impact of shocks on output. Demand parameter ρ_s impacts output through prices. As TFP increases, the plant can produce more output, but prices fall in response, dampening the effect of TFP shocks on output. In the CRS case, with a lower ρ_s (higher markups), the lower the elasticity. Now consider returns to scale: if $\gamma_s < 1$, i.e. DRS, then the elasticity of output with respect to changes in technical efficiency is smaller than the CRS case, and vice versa for increasing returns.

Now consider the aggregator for output in the sector again:

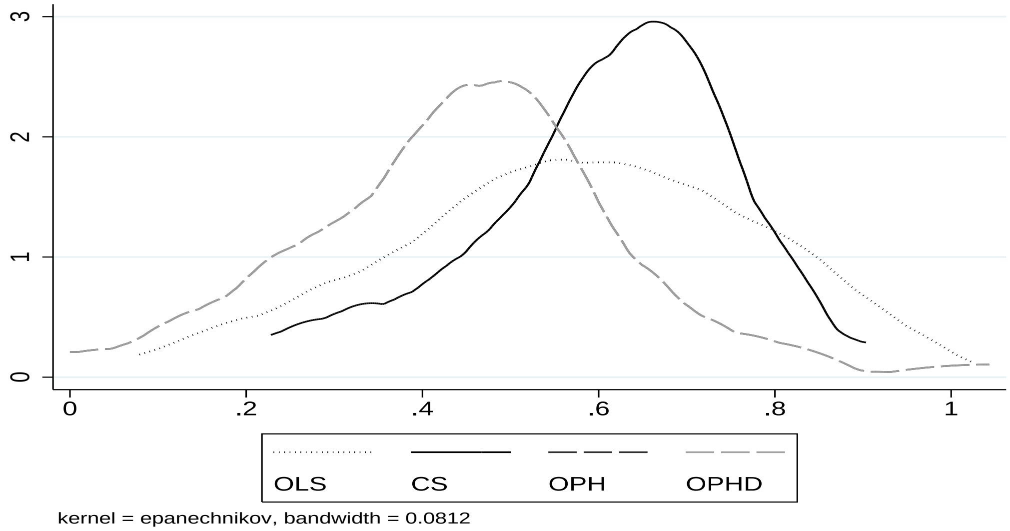
$$Q_s = \left(\sum_{i=1}^{N_s} Q_{is}^{\rho_s}\right)^{\frac{1}{\rho_s}} \tag{34}$$

Here note that the elasticity of total Q_s with respect to plant-level output is the following:

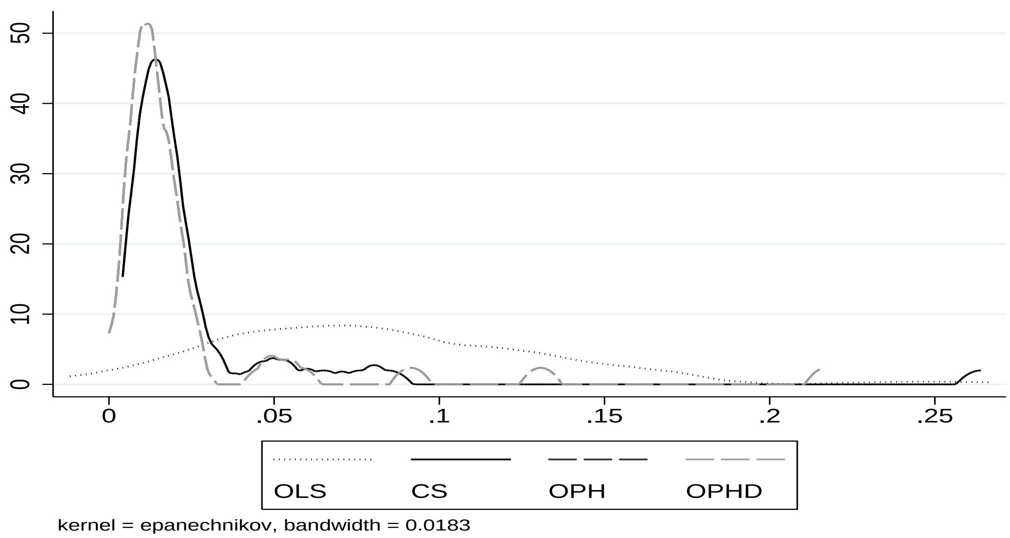
$$\epsilon_{Q_s,Q_{is}} = \left(\frac{Q_s}{Q_{is}}\right)^{1-\rho_s} \tag{35}$$

Note that this is independent of any returns to scale at the plant-level. So, we see the following: ρ_s impacts output in two ways: first through the impact on firm-level decisions, as firms take into account the impact of their choice of output on prices. This effect can be amplified or mitigated by returns to scale $\neq 1$. Second, given an increase in output of a plant, ρ_s dampens the effect of a single plant on sectoral output. This effect is the same no matter what the returns to scale at the plant-level are.

A.3 Additional Figures and Tables



(a) $\hat{\beta}_m$



(b) $\hat{\beta}_e$

Figure A1: Cross-industry distributions of elasticities of $\hat{\beta}_m$ and $\hat{\beta}_e$.

Table A1: Sample size in the specifications shown in table 3. Sample size is measured as the total number of plant-year observations used in a regression (in thousands).

	OLS	CS	OPH	OPHD
overall growth	424	424	405	405
exit	424	424	405	405
conditional growth	407	407	388	388