

# Misallocation Measures: The Distortion That Ate the Residual\*

John Haltiwanger, University of Maryland and NBER

Robert Kulick, NERA

Chad Syverson, University of Chicago Booth School of Business and NBER

September 2018

**Abstract.** A large literature on misallocation and productivity has arisen in recent years, with Hsieh and Klenow (2009; hereafter HK) as its standard empirical framework. The framework's usefulness and theoretical founding make it a valuable starting point for analyzing misallocations. However, we show this approach is sensitive to model misspecification. The model's mapping from observed production behaviors to misallocative wedges/distortions holds in a single theoretical case, with strict assumptions required on both the demand and supply sides. We demonstrate that applying the HK methodology when there is any deviation from these assumptions will mean "distortions" recovered from the data may not be signs of inefficiency. Rather, they may simply reflect demand shifts or movements of the firm along its marginal cost curve, quite possibly in profitable directions. The framework may then not just spuriously identify inefficiencies; it might be more likely to do so precisely for businesses better in some fundamental way than their competitors. Empirical tests in our data, which allow us to separate price and quantity and as such directly test the model's assumptions, suggest the framework's necessary conditions do not hold. We then extend the HK framework to allow for more general demand and supply structures to quantify the discrepancy between the framework and the data. We find substantial deviations, particularly on the demand side. Using a decomposition derived from our extended framework, we find that much of the variation in revenue-based TFP (the measure of distortions in HK) reflects the influence of demand shifts, either directly or through distortions correlated with those shifts. We furthermore show that under general conditions, the variance of revenue-based TFP is not a sufficient statistic for efficiency losses due to misallocation.

---

\* Contact information: Haltiwanger: [haltiwan@econ.umd.edu](mailto:haltiwan@econ.umd.edu); Department of Economics, University of Maryland, College Park, MD 20742; Kulick: [robert.kulick@nera.com](mailto:robert.kulick@nera.com); NERA, 1255 23<sup>rd</sup> St. NW #600, Washington, DC 20037; Syverson: [chad.syverson@chicagobooth.edu](mailto:chad.syverson@chicagobooth.edu); University of Chicago Booth School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637. John Haltiwanger is also a Schedule A employee of the U.S. Census Bureau and Robert Kulick a Census pathways intern at the time of the writing of this paper. We thank Ron Jarmin, Pete Klenow, and many seminar and conference participants for helpful comments. This paper previously circulated as "Misallocation Measures: Glowing Like the Metal on the Edge of a Knife." Any opinions and 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.

Research has established the existence of extensive heterogeneity among producers, even within narrowly defined markets. Enormous variations in establishment and firm sizes and productivity levels are ubiquitous in the data. Researchers and policymakers who focus on productivity growth have taken a keen interest in the covariance of producers' size and productivity levels, because the extent to which the market succeeds in allocating activity across producers so that they are the "right" sizes (that is, they are as large as a social planner would want them to be given their relative productivity levels) affects market-, industry-, and economy-wide productivity.

A particular approach in this research genre attempts to measure "misallocations": the presence of wedges or distortions that cause producers to deviate from their socially efficient size. One of the seminal papers embodying this approach and introducing what has become the standard methodology for analysis of misallocations is Hsieh and Klenow (2009). The Hsieh-Klenow method combines considerable empirical power and flexibility with a straightforward measurement algorithm. From standard production microdata—revenues, along with labor and capital inputs—one can extract two producer-period-specific "wedges." One distorts the producer's input mix away from the optimal frictionless factor intensity (and through this distorts the producer's size as well), and another directly distorts the producer's size. These wedges in hand, the researcher can conduct a number of complementary empirical analyses like computing the increase in aggregate productivity if misallocations were eliminated (or brought down to some other level of interest), looking at the cross-sectional or intertemporal properties of the joint distribution of wedges, or correlating these estimated distortions with observables about the producers or the markets they operate in.

The usefulness and theoretical founding of the Hsieh and Klenow (2009) approach—hereafter HK—has driven a burgeoning and insightful literature into misallocation's productivity effects. However, we show that the empirical lynchpin of the HK approach rests on a knife's edge. The condition in the HK model that maps from observed production behaviors to misallocative wedges/distortions holds in a single theoretical case, with strict assumptions required on both the demand and supply side. Regarding the former, every producer must face an isoelastic residual demand curve. On the supply side, producers must have marginal cost curves that are both flat (invariant to quantity) and are negative unit elastic with respect to total factor productivity measured with respect to output quantity (i.e., TFPQ).

We show that applying the HK methodology to data when there is deviation from these elements will mean that the “wedges” recovered from the data may not be signs of inefficiency. They may simply reflect shifts in demand or movements of the firm along its (nonconstant) marginal cost curve. The producer may be employing the efficient input mix and be its optimal size, but the HK model would perceive this behavior as indicating inefficiencies. Researchers could infer misallocation when there is in fact none. Under several conditions the spurious wedges reflect idiosyncratic demand or cost conditions that are *good* (related to higher profits) for the business. The HK method then might not just spuriously identify inefficiencies; it might be more likely to do so precisely for businesses in some fundamental way better than their competitors.

We go into detail below about why the production-to-wedge mismapping occurs, but we summarize it briefly now. The key implication of the HK model is that an efficient market has no variation in revenue-based total factor productivity (i.e., TFPR) among producers, even if they differ greatly in their TFPQ levels. Through the model’s lens, TFPR dispersion is evidence of misallocation and the existence of distortions. This homogeneous-TFPR implication arises because in the HK model, a producer’s price has an elasticity of -1 with respect to its TFPQ level. Because TFPR is the product of a producer’s price and TFPQ, this negative unit elasticity ensures that unless there are distortions, TFPR is invariant to TFPQ differences across producers (or for that matter, differences over time for a given producer). The HK model uses this invariance implication to back out misallocation measures from the TFPR dispersion that is (inevitably) observed in the data.

We demonstrate below that this crucial negative unit elasticity only occurs under the demand and supply conditions mentioned above: every producer must face isoelastic demand, and their marginal costs must be constant in quantity and negative unit elastic with respect to TFPQ. We test whether these conditions hold using a dataset where—atypically for producer-level microdata—we can observe businesses’ quantities and prices separately. We find that, in our data spanning 11 different product markets, this condition does not hold in any market.<sup>1</sup> Applying the HK framework would therefore yield spurious measures of distortions.

---

<sup>1</sup> Because distortions are not directly observable, we actually must test the joint hypothesis of price being negative unit elastic with respect to TFPQ *and* that distortions are uncorrelated with producers’ cost and demand fundamentals (which are, respectively, TFPQ in the baseline model and log-linear demand shifts in the extended HK model as described below). If distortions were correlated with fundamentals, the elasticity of price with respect to

Moreover, the elasticities of price with respect to TFPQ are consistently and considerably smaller in magnitude than one; price does not fully respond to TFPQ differences. More technically efficient businesses in our sample do not fully pass along their cost advantages to their customers through lower prices. As a result, TFPR and TFPQ are positively correlated. This positive correlation is what researchers have typically found when the data is available to compute both TFPR and TFPQ (e.g., Eslava et. al. (2013) find this using data covering all manufacturing sectors in Colombia)) and is also implied by the extensive literature on cost pass through.<sup>2</sup> This suggests the elasticity of price with respect to TFPQ may be less than unit elastic in magnitude more generally than just in our sample.

We conduct an additional test of the HK assumptions by comparing TFPQ values measured indirectly using the HK framework to direct TFPQ measures obtained from our quantity data. We find that the indirect measures (which we denote as TFPQ\_HK) are only weakly related to the direct measures and have much higher variance. These puzzling findings are partially reconciled by using a modified HK framework with demand shocks. We find that this modified TFPQ\_HK measure is more closely related to demand shocks than TFPQ.

We next quantify the distance between the HK assumptions and the data by specifying a more general empirical model that nests, but does not impose, the HK framework. We estimate the demand and supply parameters of the model and use them to decompose TFPR into demand and supply fundamentals as well as an alternative residual measure of distortions. This measure of distortions differs from TFPR because it accounts for variations in fundamentals like TFPQ

---

TFPQ might not be -1 even if the HK assumptions about demand and technology hold. Correlated distortions, while plausible (larger businesses in some economies might be subject to greater scrutiny or regulation, for instance), pose conceptual and empirical problems. Conceptually, allowing arbitrary correlations between distortions and fundamentals grants massive degrees of freedom to the model. There would be an infinite set of combinations of distortions and fundamentals that could explain the same observed TFPR variance. A high TFPR variance could reflect large distortions or instead small distortions holding a particular correlation structure with firm-level TFPQ or demand. Empirically, this means using the HK model to identify distortions that are correlated with fundamentals necessarily requires an (untestable) assumption about the distortion-fundamental correlation structure. Further below, we generalize the HK model in a way that both allows for nonzero correlations and permits the estimation of this correlation. But we do so after more fully modeling the demand and cost sides of the model, both expanding the ability of fundamentals to explain the data and putting additional structure on the ways distortions and fundamentals could be correlated.

<sup>2</sup> The positive correlation between TFPR and TFPQ in our sample is evident in Table 1 of Foster, Haltiwanger and Syverson (2008). Kulick (2016) uses the same sample for a study of horizontal mergers in ready-mixed concrete. While it is not his focus, he also finds that there is incomplete pass through of TFPQ changes on price. The broader literature on cost pass through is quite large but some examples include Goldberg and Verboven (2001); Campa and Goldberg (2005); Nakamura and Zerom (2010); Bonnet, Dubois, Villas Boas, and Klapper (2013); and Ganapati, Shapiro, and Walker (2016).

and demand shocks that would otherwise enter into TFPR dispersion. We find that both fundamentals and the residual distortions contribute substantially to TFPR variance. We also show that our distortion measure is positively correlated with producer exit; i.e., it acts like a distortion. This is not the case for TFP, the distortion measure of HK, which is negatively correlated with exit in our sample. However, we show that once we control for TFPQ and demand shifts, TFPR does become positively correlated with exit. The sign change of the conditional correlation suggests that TFPR does contain information about factors that match the conceptualization of distortions, but this is empirically swamped by variation in fundamentals. This result, and the contrast with our residual distortion measure, suggest a general issue with misallocation measures: because they are essentially residuals, they may well indeed contain a kernel of distortions within them, but isolating this component from the effects of other (possibly efficient) sources of firm heterogeneity is empirically very difficult and can require unusually detailed data.

Our more general demand estimation implies variable markups at the producer level that are increasing in fundamentals and in turn the size of the establishment. Such variable markups underlie part of the measured variance in TFPR. We build on these empirical findings to show that, in a general setting with VES (variable elasticity of substitution) demand and heterogeneous firms that endogenously enter and exit, the variance of TFPR is not a summary welfare metric. Namely, there is not a unique mapping between the size of misallocation losses and the variance of TFPR. Zero TFPR variance does not imply zero misallocation losses, nor does dispersion in TFPR necessarily reflect inefficiencies.

The paper proceeds as follows. In section I, we review the details of the HK framework in terms of assumptions and implications. Our primary focus is to demonstrate theoretically the stringent assumptions required to use TFPR to identify distortions. Section II includes our tests of the HK assumptions and implications. Section III presents our estimates of more general demand and production function structures and quantifies their relevance for measuring distortions as well as interpreting the dispersion in TFPR. Concluding remarks are in section IV.

## **I. The Hsieh-Klenow Framework: Its Assumptions and Applications**

### *A. A Brief Overview of the Hsieh-Klenow Framework*

We first review the most critical elements of the Hsieh and Klenow (2009) framework. The HK framework posits that each industry contains a continuum of monopolistically competitive firms (indexed by  $i$ ) that differ in their TFPQ levels,  $A_i$ . Each firm combines labor and capital inputs to produce a single good. Firms in an industry face a Dixit-Stiglitz-type constant elasticity demand system, so each faces a residual demand curve with elasticity  $\eta$ . Firms choose a quantity (equivalently, price) to maximize the profit function:

$$\pi_i = (1 - \tau_{Yi})P_iQ_i - WL_i - (1 + \tau_{Ki})RK_i$$

subject to the firm's inverse residual demand curve,  $P_i = Q_i^{-1/\sigma}$ , and the production function  $Q_i = A_iL_i^\alpha K_i^{1-\alpha}$ .

The nonstandard elements here are the two wedges  $\tau_{Yi}$  and  $\tau_{Ki}$ . The former is a firm-specific scale distortion (effectively a tax or subsidy on the firm's output) and  $\tau_{Ki}$  is a firm-specific factor price wedge/distortion. Their effects in equilibrium are discussed below.

Given the isoelastic residual demand curve, Firm  $i$ 's profit-maximizing price is

$$P_i = \frac{\sigma}{\sigma - 1} MC_i$$

where  $MC_i$  is the firm's marginal cost, equal to

$$MC_i = \left(\frac{R}{\alpha}\right)^\alpha \left(\frac{W}{1-\alpha}\right)^{1-\alpha} \frac{(1 + \tau_{Ki})^\alpha}{A_i(1 - \tau_{Yi})}$$

The factor prices—assumed constant across firms—are  $R$  for capital and  $W$  for labor. Note that both wedges/distortions affect the firm's marginal cost and price, and firms with higher  $A_i$  (TFPQ) have lower marginal costs and prices.

At the optimal price and quantity, the firm's marginal products of labor and capital are proportional to the product of the factor price and functions of one or both distortions:

$$MRPL_i \propto W \frac{1}{1 - \tau_{Yi}}$$

$$MRPK_i \propto R \frac{1 + \tau_{Ki}}{1 - \tau_{Yi}}$$

Note that because of the assumption of common factor prices, in the absence of distortions, marginal revenue products of both factors would be equated across firms.

The critical result of the HK setup is that TFPR is proportional to a weighted geometric average of the marginal products of labor and capital, where the weights are the factors' output elasticities. As a result, the only firm-level variables that shift  $TFPR_i$  are the two distortions:

$$TFPR_i \propto (MRPL_i)^{1-\alpha} (MRPK_i)^\alpha \propto \frac{(1 + \tau_{Ki})^\alpha}{1 - \tau_{Yi}}$$

This key result provides the theoretical justification for empirical work seeking to infer the presence and size of misallocations from observed differences in TFPR across producers.<sup>3</sup>

### B. The Assumptions Driving HK's Result

The reason TFPR is invariant across firms in the HK model can be seen from the definition of TFPR as the product of price and TFPQ,  $TFPR_i \equiv P_i A_i$ , and by substituting the expression above for the firm's marginal cost into the HK model's optimal pricing equation:

$$P_i = \frac{\sigma}{\sigma - 1} \left(\frac{R}{\alpha}\right)^\alpha \left(\frac{W}{1 - \alpha}\right)^{1-\alpha} \frac{(1 + \tau_{Ki})^\alpha}{A_i(1 - \tau_{Yi})}$$

Notice that the elasticity of the firm's price  $P_i$  with respect to its TFPQ level  $A_i$  is -1 (as discussed in footnote 1, this implication is conditional on distortions being uncorrelated with  $A_i$ ). This means that as TFPQ levels and, therefore, prices vary across firms, the constancy of their product, TFPR, is preserved. Regardless of the characteristics of the distribution of  $A_i$  across firms, then, TFPR will not vary unless there are distortions  $\tau_{Yi}$  and  $\tau_{Ki}$ . This negative unit elasticity must hold not just on average across producers, but for every quantity that any firm might produce.

We can dig deeper into the TFPR invariance condition by using the chain rule to expand the elasticity of price with respect to TFPQ and recognizing that price is a function of marginal cost, which itself depends on TFPQ. Multiplying and dividing the resulting expression by marginal cost yields (we suppress the firm index here and below when unnecessary for clarity):

$$\begin{aligned} \varepsilon_{P,A} &= \frac{dP(MC(A))}{dA} \frac{A}{P} = \frac{dP}{dMC} \frac{MC}{P} \frac{dMC}{dA} \frac{A}{MC} = -1 \\ \varepsilon_{P,A} &= \varepsilon_{P,MC} \varepsilon_{MC,A} = -1 \end{aligned}$$

This decomposition of the key HK condition makes clear how the assumed functional forms on both sides of the market are necessary for the condition to hold. The elasticity of a firm's price with respect to marginal cost  $\varepsilon_{P,MC}$  depends on the firm's residual demand curve,

---

<sup>3</sup> This invariance of TFPR with respect to TFPQ was actually first noted by Katayama, Lu, and Tybout (2009), though they did not have distortions in their model, nor did they frame their result as informative about misallocation. Under their assumptions, TFPR does not reflect a firm's technical efficiency whatsoever, but rather only the factor prices it faces.

while the elasticity of its marginal cost to its TFPQ level  $\varepsilon_{MC,A}$  depends on its marginal cost curve (and through this, its production function).

These demand- and supply-side components of the HK condition are not completely independent, however, because they hold at the profit-maximizing price. As such the marginal cost in the expression is evaluated at the firm's optimal quantity, which depends on both the demand and cost curves. Therefore the elasticity of the firm's marginal cost with respect to TFPQ,  $\varepsilon_{MC,A}$ , depends both on direct shifts in the marginal cost curve due to TFPQ changes plus any movement along the marginal cost curve when a TFPQ change moves the intersection of the marginal cost and marginal revenue curves.

### *B.1. The Demand-Side Assumption*

We now investigate the demand- and supply-side conditions under which the HK demand and cost assumptions hold. (Recall they are connected through their evaluation of marginal cost at the firm's profit-maximizing quantity.) We begin with demand systems where the elasticity of the firm's price with respect to its marginal cost,  $\varepsilon_{P,MC}$ , equals one.<sup>4</sup>

When  $\varepsilon_{P,MC} = 1$ , the ratio of price to marginal cost is constant. That is, the price at any quantity must be a constant multiplicative markup of marginal cost,  $P = \mu \cdot MC$ . As is well known, this requires an isoelastic residual demand function,  $Q = DP^{-\sigma}$ , where  $D$  is a demand shifter and  $\sigma$  is the price elasticity of demand. Note that any  $\sigma > 1$  is consistent with the HK assumption (the  $\sigma > 1$  condition reflects the fact that profit maximization requires a firm to operate only on an elastic portion of its demand curve). As long as demand is isoelastic, it is the case that  $\varepsilon_{P,MC} = 1$  regardless of the particular value of  $\sigma$ .

Isoelastic demand is not just consistent with the HK framework, it is the *only* form of demand that is compatible with it.<sup>5</sup> If firms face any other type of residual demand curve,  $\varepsilon_{P,MC} \neq 1$ , and the necessary condition does not hold.

To see this in an example, suppose demand is linear:  $Q = a - bP$ . A firm's profit maximizing price is then  $P = (a/2b) + (MC/2)$ , where  $MC$  is the firm's marginal cost. (We

---

<sup>4</sup> While any combination of demand- and cost-side elasticities that multiply to negative one will conform to the  $\varepsilon_{P,A} = -1$  condition, the most natural case would be where  $\varepsilon_{P,MC} = 1$  and  $\varepsilon_{MC,A} = -1$ , because (as we show here) commonly assumed demand and production functions produce these results. The other cases where the product still happens to be -1 are even more arbitrary from an economic perspective than the unit elastic cases we discuss.

<sup>5</sup> Save again for the arbitrary case where a non-unitary  $\varepsilon_{P,MC}$  equals the negative reciprocal of  $\varepsilon_{MC,A}$  at all quantities.

assume MC is constant in quantity here to focus on HK's demand-side condition.) Therefore  $\varepsilon_{P,MC} = (1/2)(MC/P)$ . For any  $P \geq MC$ ,  $\varepsilon_{P,MC} \leq 1/2$ . Thus with linear demand there are no situations under which the HK assumption hold, even approximately. Another illustrative example is the constant *absolute* markup demand function  $Q = \lambda \exp(-P/M)$ , where  $M$  is the markup. Here,  $P = MC + M$  and  $\varepsilon_{P,MC} = MC/(MC + M)$ . In this case  $\varepsilon_{P,MC} = 1$  only when the market is perfectly competitive and  $M = 0$ . If there is any markup,  $\varepsilon_{P,MC} < 1$ .

Both examples have the property that the elasticity of price with respect to marginal cost is always (weakly) less than one. As noted in the prior section, the results from the empirical literature suggest this property may apply more generally in the data. Previous work has typically found TFPQ to be positively correlated with TFPR, rather than uncorrelated as implied by HK. Working from the results above, this positive correlation implies that in the data the elasticity of price with respect to TFPQ is less than one in absolute magnitude:

$$|\varepsilon_{P,MC} \varepsilon_{MC,A}| < 1$$

Or, because theory implies  $\varepsilon_{P,MC} \geq 0$  and  $\varepsilon_{MC,A} \leq 0$  under standard conditions,<sup>6</sup>

$$\varepsilon_{P,MC} < \frac{1}{|\varepsilon_{MC,A}|}$$

The intuition here is that for any given responsiveness of marginal costs to TFPQ, a sufficiently small pass through of lower costs (where costs reflect TFPQ) will ensure price stays high enough so that total revenues and TFPR rise when TFPQ does. Given the positive correlations found in empirical work, this smaller pass through appears to be the typical case in the data.

### *B.1. The Supply-Side Assumption*

We now consider the supply-side necessary condition for HK's result: the elasticity of the firm's marginal cost at its optimal quantity with respect to its TFPQ level is negative one. This holds when

---

<sup>6</sup> For smooth demand curves (those with continuous marginal revenue curves), price weakly rises with marginal cost because an increase in marginal cost reduces the firm's optimal quantity. The limit case is perfect competition, where the residual demand and marginal revenue curves are flat, and a change in the firm's marginal cost has no effect on price. The change in a firm's marginal cost resulting from a change in its TFPQ level  $A$  depends both on the direct negative effect of TFPQ on costs and any change in marginal cost resulting from the effect of TFPQ on the firm's optimal quantity. As detailed below, this total change is weakly negative, with perfect competition again representing the limit case. In that boundary case, realized marginal cost remains at the (unchanged) market price and the product of the demand- and supply-side elasticities remains less than one, although the second inequality is undefined.

$$\varepsilon_{MC,A} = \frac{\partial MC(A, Q(A))}{\partial A} \frac{A}{MC} = -1$$

where  $MC(A, Q(A))$  is the firm's marginal cost function (the derivative of its cost function with respect to quantity). We have explicitly written the firm's quantity as a function of TFPQ, but have suppressed the other arguments of the marginal cost function because they are assumed constant across firms in the HK framework.

To explore the theoretical conditions under which  $\varepsilon_{MC,A} = -1$  might hold, consider first how a change in TFPQ would qualitatively affect a firm's realized marginal cost. The total change in marginal cost depends both on the direct negative effect of TFPQ on costs—the shift in the marginal cost curve—as well as any change in marginal cost resulting from the effect of TFPQ on the firm's optimal quantity—movement along the marginal cost curve. As noted above, this total effect of a TFPQ increase is bounded from above by zero (the case under perfect competition), which requires upward-sloping marginal cost curves. The sum of these two effects—reinforcing if marginal costs decline in quantity, countervailing if they rise—must be negative unit elastic to conform to the HK model.

The simplest case where this holds is when the marginal cost curve is flat and marginal costs are negative unit elastic in TFPQ; that is, when the marginal cost curve has the form:

$$MC(A) = \frac{\Phi(\mathbf{W})}{A}$$

where  $\Phi(\mathbf{W})$  is a function of the vector of factor prices  $\mathbf{W}$ . The firm's quantity is not an argument in this function, indicating constant marginal costs in quantity. Intuitively, the negative unit elasticity holds in this case because there is no reinforcing or countervailing effect of TFPQ on the firm's optimal quantity. The only influence TFPQ has on marginal cost is its direct effect, which is negative unit elastic.

We can integrate with respect to  $Q$  to find the cost functions that satisfy the condition:

$$C(A, Q) = \int \frac{\Phi(\mathbf{W})}{A} dQ = \frac{Q}{A} \Phi(\mathbf{W}) - F$$

Where  $F$  is a (possibly zero) fixed cost. Some commonly used cost functions have this form. For example, the Cobb-Douglas production function  $Q = AL^\alpha K^\beta$  has a cost function equal to

$$C(A, Q) = \left(\frac{Q}{A}\right)^{\frac{1}{\alpha+\beta}} \left(\frac{\alpha + \beta}{\alpha^\alpha + \beta^\beta}\right)^{\frac{1}{\alpha+\beta}} W^{\frac{\alpha}{\alpha+\beta}} R^{\frac{\beta}{\alpha+\beta}}$$

As is obvious from inspection, this has the required form if  $\alpha + \beta = 1$ ; i.e., the production function exhibits constant returns to scale. This is the production function and parameterization HK assumes.<sup>7,8</sup>

The HK requirement that  $\varepsilon_{MC,A} = -1$  will not hold without constant returns to scale. This is because with nonconstant returns, the effect of TFPQ on marginal costs is not just the direct effect through shifting the marginal cost curve but also the induced movement along the curve because the firm's optimal quantity changes when TFPQ does. The size of this quantity change depends on the relative slopes of both the marginal cost and marginal revenue curves around the location of the quantity change.

### *C. A Graphical Demonstration of the Uniqueness of the HK Assumption*

In this section, we use a graphical framework to explain why the HK framework delivers the TFPR invariance result, and why any departure from either its demand- or supply-side necessary assumptions will lead TFPR to differ across firms even if there are no distortions. This will reinforce the analysis above.

There is an additional point to our exercise here, however. We introduce firm-specific demand shifts, which are not in the baseline HK model, into the framework. Under the assumptions of the HK model, demand shifts do not affect the key TFPR invariance implication. However, we show that if any of the component assumptions fail, firm-specific demand shifts will create variation in TFPR even in the absence of distortions. This creates a second channel through which applying the HK condition can yield spurious distortion measures.

---

<sup>7</sup> A similar result holds for the general CES production function  $Q = A[\alpha L^\rho + \beta K^\rho]^{\frac{v}{\rho}}$ , where  $\rho$  parameterizes the elasticity of substitution between inputs and  $v$  parameterizes the scale elasticity. In this case, the corresponding cost function is:

$$C(A, Q) = \left(\frac{Q}{A}\right)^{\frac{1}{v}} \left[ \alpha^{\frac{1}{1+\rho}} W^{\frac{\rho}{1+\rho}} + \beta^{\frac{1}{1+\rho}} R^{\frac{\rho}{1+\rho}} \right]^{\frac{1+\rho}{\rho}}$$

If the production function exhibits constant returns to scale (i.e.,  $v = 1$ ), marginal costs will be constant and negative unit elastic with respect to TFPQ.

<sup>8</sup> Note that the HK framework admits nonconstant returns to scale arising from fixed costs. However, as noted by Foster et al. (2017), in practice this will rely on the empiricist being able to measure the true, marginal  $A_i$ . If instead TFP is measured according to the common practice of taking a ratio of output to weighted inputs, this ratio will not be invariant to the firm's optimal quantity, and again the HK assumptions will be violated.

We start our analysis by imposing the HK assumptions. Residual demand is isoelastic,  $Q = DP^{-\sigma}$ . The corresponding inverse demand is  $P = D^{\frac{1}{\sigma}}Q^{-\frac{1}{\sigma}}$  and the inverse marginal revenue curve is  $MR = \left(1 - \frac{1}{\sigma}\right)D^{\frac{1}{\sigma}}Q^{-\frac{1}{\sigma}}$ . Both curves are log-linear:

$$p = \frac{1}{\sigma}d - \frac{1}{\sigma}q$$

$$mr = \ln\left(1 - \frac{1}{\sigma}\right) + \frac{1}{\sigma}d - \frac{1}{\sigma}q = \ln\left(1 - \frac{1}{\sigma}\right) + p$$

where lowercase letters are logged values. (Neither function is defined at its vertical or horizontal intercepts.)

Because  $\sigma > 1$ , the first term in the logged marginal revenue curve is negative. Thus in logged-quantity-logged-price space, the marginal revenue curve runs parallel to demand at a distance  $\ln(1 - 1/\sigma)$  below it. As we show below, this parallelism is important to the HK result.

We also impose constant returns to scale with a cost function of

$$C(A, Q) = \frac{Q}{A}\Phi(W)$$

Marginal costs, of course, do not depend on output, and their elasticity with respect to TFPQ is -1. The log of marginal cost is:

$$mc = \phi(w) - a$$

These elements—the demand curve, the marginal revenue curve, and the marginal cost curve—are combined in the solution to the standard monopolist's price/quantity problem in Figure 1. The firm's optimal (logged) quantity is where  $mr = mc$ ,  $q^*$ , and its optimal price is  $p^*$ .

The figure also demonstrates how a change in (logged) TFPQ,  $a$ , affects the optimal quantity and price. The HK condition requires that TFPR, which is the product of  $P$  and  $A$ , be invariant to changes in  $A$ . In the logged space shown in the figure, it means that any change in TFPQ,  $\Delta a$ , must induce a price change  $\Delta p = -\Delta a$ .

Figure 1 makes clear why this result always holds in the HK setting. Suppose TFPQ rises from  $a$  to  $a'$ , so  $\Delta a = a' - a$ . HK's assumed  $\varepsilon_{MC,A} = -1$  implies that  $\Delta mc = -\Delta a$ . This drop in marginal cost raises the firm's optimal quantity to  $q'^*$  with a corresponding price change from  $p^*$  to  $p'^*$ , as shown in the figure. Here is the key result: because the marginal revenue and demand curves  $mr(q)$  and  $p(q)$  are parallel and the marginal cost curve horizontal, it must be that

the drop in logged marginal revenue at the optimum quantity must exactly equal the drop in logged price. Thus  $\Delta p^* = \Delta mr^* = \Delta mc^* = -\Delta a$ , the HK result.

Note that both elements of the HK framework are necessary for this result. Only isoelastic demand creates parallel demand and marginal revenue curves. This ensures a given change in logged marginal revenue at the optimal quantity translates into the same-sized change in logged price. In other words, the ratio of (the level of) price to (the level of) marginal cost stays the same, so the elasticity of price with respect to marginal cost is one. The constant returns assumption creates the horizontal marginal cost curve. This ensures that the total effect on the firm's marginal cost at its optimal quantity,  $\Delta mc^*$ , is only the direct effect of the shift in the curve  $\Delta a$ . There is no reinforcing (if the marginal cost curve is downward sloping) or countervailing (upward sloping) effect on marginal costs through induced shifts along the marginal cost curve when the firm's optimal quantity changes.

Violating either of these conditions ensures that  $\Delta p \neq -\Delta a$  and failure of TFPR invariance with respect to TFPQ.

It is obvious from inspection of Figure 1 that any other demand curve, because it does not have a parallel marginal revenue curve, will cause any change in logged marginal cost—even in the presence of a horizontal marginal cost curve—to lead to a disproportionate change in the firm's optimal price. (Recall that proportionalism in levels is graphically reflected in parallelism in logged values.)

Regarding the HK assumption about the marginal cost curve, Figure 2 preserves CES demand but shows the effect of an increase in TFPQ when marginal costs rise with output. As in Figure 1, an increase in logged TFPQ from  $a$  to  $a'$  shifts down the marginal cost curve by  $\Delta a$ . Here, however, because the marginal cost curve is not horizontal, the effect of this TFPQ change on the firm's marginal cost is not just the drop in the  $mc$  curve. It is also the effect of moving along the new  $mc$  curve from the old optimal quantity  $q^*$  to the new one  $q'^*$ . This total effect is necessarily less than  $\Delta a$  because  $mc$  is upward sloping. As a result, price doesn't fall as much as the marginal cost curve shifts down, and  $\Delta p \neq -\Delta a$ . Similarly, a downward-sloping marginal cost curve would create a movement along the  $mc$  curve that would make the total effect of a change in TFPQ on marginal costs *greater* than  $\Delta a$ . Again, it is the case that  $\Delta p \neq -\Delta a$ .

## II. Testing the Assumptions of the Hsieh-Klenow Framework

### *A. Elasticity of Prices with Respect to TFPQ*

We first test the core implication of the HK setup: producer prices are negative unit elastic with respect to TFPQ levels (under the assumption, as described above, that distortions are uncorrelated with TFPQ).

One needs to observe prices and TFPQ levels to conduct this test. While techniques have been developed to back out otherwise unobservable price and quantity information from revenue data (see, e.g., Klette and Griliches, 1996; Katayama, Lu, and Tybout, 2009; De Loecker and Warzynski, 2012), these require assumptions, making any test a joint test not only of the assumptions of the HK model but these techniques as well.

Fortunately, we collected a dataset in earlier work (Foster, Haltiwanger, and Syverson, 2008, 2016) that includes separate quantity and price information at the individual producer level. Those papers extensively detail this data, so we only very briefly review its contents here.

Our microlevel production data is a subset of the 1977, 1982, 1987, 1992, and 1997 U.S. Census of Manufactures (CM). The CM collects information on plants' shipments not just in the standard revenue sense (i.e., dollar values), but physical units as well. The sample includes producers of one of eleven products: corrugated and solid fiber boxes (which we will refer to as "boxes" from now on), white pan bread (bread), carbon black, roasted coffee beans (coffee), ready-mixed concrete (concrete), oak flooring (flooring), gasoline, block ice, processed ice, hardwood plywood (plywood), and raw cane sugar (sugar).<sup>9</sup> We chose these products based in part on their physical homogeneity, which allows plants' output quantities and unit prices to be more meaningfully compared.

From these product-level revenue and physical quantity data, we can construct important inputs to our analyses here. (The details of construction can be found in our earlier work.) First, we can compute producers' average unit prices. Second, we can measure TFPQ directly, using physical quantity as the output measure in the productivity numerator. Third, we can back out idiosyncratic demand shifts (alternately referred to as "shifts" and "shocks" below) for every producer. We describe this process in Foster, Haltiwanger, and Syverson (2008), but in brief, we

---

<sup>9</sup> We exclude observations with imputed physical quantity data. For this purpose, we take advantage of newly recovered item impute flags developed and described in White, Reiter and Petrin (2014). We use inverse propensity score weights in our analysis to deal with possible non-randomness in the likelihood of observations being imputed. We find that results are largely robust to not using such weights. We use the same approach as in Foster, Haltiwanger and Syverson (2016) for this purpose. See the latter paper for details.

impose a CES demand system for each industry—using TFPQ as a cost-shifting instrumental variable—and take the residual as a measure of the producer-specific demand shift.

In our basic specification, we regress a producer’s logged price on its contemporaneous logged TFPQ for each product separately:

$$p_{it} = \alpha_0 + \alpha_1 \text{tfpq}_{it} + \eta_t + \varepsilon_{it}$$

where  $\eta_t$  is a fixed effect corresponding to the CM year, which removes any shifts in prices across time that are common across all producers. Under the HK assumptions,  $\alpha_1 = -1$ . We, therefore, test industry-by-industry the null hypothesis that  $\alpha_1 = -1$ .

We also estimate a pooled specification on the combined dataset. Here the specification is the same, except rather than just having CM year fixed effects we include industry-CM year fixed effects, so identification of the relationship between price and TFPQ comes from within-industry-year variation. In this case we are imposing a common value of  $\alpha_1$  across all industries.

The results are shown in Table 1. The magnitudes of the estimated elasticities  $\alpha_1$  are considerably less than one for every industry. The null hypothesis of the HK conditions is clearly rejected; the smallest  $t$ -statistic rejecting the null is 4.4, for carbon black. In the pooled specification, we estimate an average elasticity of price with respect to TFPQ of -0.450, and reject the null with a  $t$ -statistic of 86.4. Thus the average elasticity of a producer’s price with respect to its TFPQ level is less than half the magnitude of that implied by the HK assumptions. Given that price is less than negative unit elastic with respect to TFPQ, TFPR in our data is positively correlated with TFPQ. Producers with low costs (high TFPQ) do not fully pass onto consumers their cost advantages.

We estimate an alternative pooled specification as a check because, as noted in Foster, Haltiwanger, and Syverson (2008), the fact that we measure unit prices as the quotient of reported revenues and physical quantities means that measurement error in quantities might create division-bias-based measurement error in a regression of price on TFPQ. We therefore instrument using the producer’s TFPQ level in the previous CM. The first stage results indicate this instrument has considerable explanatory power with respect to current TFPQ. The results of the second stage, shown in Table 1, are consistent with the OLS results. The point estimate of the elasticity  $\varepsilon_{P,A}$  is well below one, economically and statistically.

In sum, we find consistent evidence that the elasticity of price with respect to TFPQ is well below one in magnitude. Of course, this result applies to our particular sample, which is by

no means representative of all production settings. It cannot elucidate whether the less than complete response of prices to TFPQ-driven cost changes holds more generally. On this point, however, we can also appeal to a separate and very large empirical literature on pass through rates that indicates our result is indeed typical. Some examples of this literature include Goldberg and Verboven (2001); Campa and Goldberg (2005); Nakamura and Zerom (2010); Bonnet, Dubois, Villas Boas, and Klapper (2013); and Ganapati, Shapiro, and Walker (2016). These studies reflect the results found throughout the literature: across diverse market settings, pass through of costs into prices is less than one-for-one.<sup>10</sup>

### *B. Relationship between Direct TFPQ Measures and TFPQ from HK Framework*

We conduct a second test of the HK framework using our sample of homogenous-product manufacturers. Namely, we back out the TFPQ implied by the HK model from our data and compare it to the TFPQ that we can measure directly. This gives us the ability to gauge how closely a key unobservable derived from the HK framework resembles its direct measure.

As HK show, one can recover a producer’s implied TFPQ as follows:

$$TFPQ_{HK_i} = \kappa \frac{(P_i Q_i)^{\frac{\sigma}{\sigma-1}}}{K_i^\alpha L_i^{1-\alpha}}$$

Intuitively, the numerator is output as backed out from observed revenue via the demand elasticity  $\sigma$ . We allow the elasticity to vary by industry, using our industry-specific demand estimates (described in more detail below). The denominator is the standard composite TFP input. The constant  $\kappa$  is the same across all producers, so it can be ignored in all comparisons we make below.

We compare this to our directly measured TFPQ:<sup>11</sup>

$$TFPQ_i = \frac{Q_i}{K_i^\alpha L_i^{1-\alpha}}$$

---

<sup>10</sup> Note that constant-elasticity demand implies complete pass through of *logged* costs into *logged* prices. With a markup, therefore, the pass through of cost *levels* into price *levels* will be greater than one-to-one. Some of the cited studies measure pass through in levels rather than logs. Given that they find less than one-to-one pass through in levels, this also implies less than one-to-one pass through in logs. Indeed, one notable “exception” paper in the literature known for finding close to complete pass through in its empirical setting is Fabra and Reguant (2014). However, their result of near-complete pass through is in levels, indicating incomplete pass through in logs.

<sup>11</sup> In practice, we use a gross output production function to obtain TFPQ. We present a version based on a value added production function to match the notation in HK (2009). Note also that we assume the production function has constant returns to scale, just as the HK framework.

It is apparent that the model-driven transformation from revenue to implied output is what separates TFPQ\_HK from TFPQ.

Our data reveal that this indirect-versus-direct distinction in quantity measurement makes a big difference. The correlation between TFPQ\_HK and TFPQ across our entire sample is only 0.09. That is, the physical efficiency of producers in our data as derived from the HK model is weakly correlated with its directly measured value. Part of this poor fit reflects the fact that there is much more variability in TFPQ\_HK than TFPQ. The standard deviation of TFPQ is 0.28, while for TFPQ\_HK it is an enormous 3.29.<sup>12</sup> The source of this large variance can be observed in the expression for TFPQ\_HK above: a demand elasticity  $\sigma$  near one requires huge variation in implied quantity to explain observed revenue variation. Two of our sample industries, carbon black and gasoline, have estimated demand elasticities that are relatively close to unity and as such have highly variable implied output quantities. If we remove these from the sample, the standard deviation of TFPQ\_HK falls considerably, to 1.03. However, this is still much larger than the TFPQ standard deviation for this restricted sample of 0.28, and in any case the main message stands: TFPQ\_HK and TFPQ are only weakly correlated, with a correlation coefficient of 0.29 in this restricted sample.

At the same time, TFPQ\_HK is uncorrelated with producers' prices (correlation coefficients of 0.01 in the whole sample and 0.01 in the sample excluding carbon black and gasoline). This contrasts with a correlation between directly measured TFPQ and prices of -0.59 in both the whole and restricted samples. As Foster, Haltiwanger, and Syverson (2008) point out, this negative correlation is consistent with the notion that TFPQ differences are cost differences: higher TFPQ implies lower costs, and these costs are then (partially) passed through in the form of lower prices. The fact that increases in TFPQ\_HK do not correspond to lower prices raises questions about the extent to which TFPQ\_HK captures firms' cost efficiencies.

We next compare TFPR to alternative measures of TFPQ and our measured estimated producer-level shifts. TFPR has slightly lower dispersion (standard deviation of 0.23) than directly measured TFPQ, but it is much less dispersed than TFPQ\_HK. TFPR is highly correlated with directly measured TFPQ (about 0.66) and positively correlated with demand (0.29). On the other hand, it is less correlated with TFPQ\_HK (0.11). These patterns are for the

---

<sup>12</sup> These calculations use values where we have removed industry-year means from the sample. Bear in mind that these TFPQ values are in logged units of output, so a log difference of 3.29 implies a 27-fold ratio in levels.

full sample but are quite similar for the restricted sample.<sup>13</sup> This further emphasizes that the measures of TFPQ derived from the HK model do not behave the way directly measured TFPQ does in our sample.

One potential source of the unusual relationships between TFPQ\_HK and directly measured TFPQ and prices is that we apply the baseline HK model to derive TFPQ\_HK (except that, unlike HK, we use industry-specific demand elasticities). The baseline model has no demand shifts across producers; all heterogeneity comes through TFPQ and distortions. However, it is possible—and indeed a burgeoning literature suggests it is likely—that producers face idiosyncratic demand shocks along with having different productivity levels. Hsieh and Klenow (2009) show in an appendix that their model can be augmented to include demand shocks (horizontal shifters in firms’ CES demand curves) while still preserving the basic logic of the model. To see how allowing demand variations might improve the fit of TFPQ\_HK to TFPQ, we apply this augmented version of their model to our data.

In the demand-augmented HK framework, TFPQ\_HK is now<sup>14</sup>

$$TFPQ\_HK\_WD_i = \kappa \frac{(P_i Q_i)^{\frac{\sigma}{\sigma-1}}}{K_i^\alpha L_i^{1-\alpha}} = \kappa \frac{Q_i}{K_i^\alpha L_i^{1-\alpha}} D_i^{\frac{1}{\sigma-1}}$$

where  $D_i$  is firm  $i$ ’s idiosyncratic demand. We mnemonically name the object TFPQ\_HK\_WD to denote “with demand.” Intuitively, this is a composite measure reflecting  $TFPQ_i = \frac{Q_i}{K_i^\alpha L_i^{1-\alpha}}$  and idiosyncratic demand shifts. Decomposing this composite into its demand and TFPQ components is not feasible with standard production data with revenue and inputs, but we can in our data. Under the HK assumption of a CES demand system, the estimated  $D_i$  by design satisfies the above composite relationship. A critical point to emphasize is that the equivalence between TFPQ\_HK and the composite shock requires estimating the demand elasticities and demand shocks in an internally consistent manner. In practice, HK and others who implement the TFPQ\_HK methodology typically impose the same elasticities across industries (and countries).

---

<sup>13</sup> The correlation between TFPR and TFPQ\_HK rises to 0.36 in the restricted sample. The other correlations are very similar to the full sample. Both Hsieh and Klenow (2009) and Bils, Klenow, and Ruane (2017) report a low correlation between TFPR and TFPQ\_HK, about 0.10.

<sup>14</sup> The  $D_i$  shifter we are now including is from the specification  $Q_i = D_i P_i^{-\sigma}$ , so that  $P_i = D_i^{\frac{1}{\sigma}} Q_i^{-\frac{1}{\sigma}}$ . That is, it is the shift in quantity demanded  $Q_i$  holding price constant. In this specification  $(P_i Q_i)^{\frac{\sigma}{\sigma-1}} = D_i^{\frac{1}{\sigma-1}} Q_i$

We find that TFPQ\_HK has a stronger correlation with our measure of demand (about 0.28 in the full sample and 0.55 in the restricted sample) than with TFPQ. This suggests that a considerable amount of the variation in TFPQ\_HK is actually driven by demand shifts rather than TFPQ differences.

To add further insights, our final investigation of the properties of TFPQ\_HK explores its relationship with survival and compares this to other producer-level metrics. Table 2 shows the results. High TFPR, high directly measured TFPQ, and high demand  $D_i$  are each negatively associated with exit. Demand shocks play the dominant quantitative role, with a one standard deviation increase in demand associated with a 9-percentage-point drop in the probability of exit. In contrast, a one standard increase in TFPQ is tied to a decline in the probability of exit of 1 percentage point. High TFPQ\_HK plants are also more likely to survive. A one standard deviation increase in TFPQ\_HK corresponds to a 1.6-percentage-point decline in the exit probability. Thus, the most important predictor of exit is the demand shift, with a one standard deviation increase yielding a drop in the exit rate that is nine times larger than that of a similar sized shift in TFPQ, and more than five times that of TFPQ\_HK.

Our evidence suggests TFPQ\_HK is best thought of as a composite measure that reflects both TFPQ and demand shocks. While it is interpretable as a composite, it has less predictive value in accounting for key outcomes like survival than its underlying components. Moreover, this composite interpretation requires that demand be estimated in a manner that is internally consistent with the micro data.

### *C. Demand Variations and the Hsieh-Klenow Framework*

The previous section's analysis makes it clear that demand variations across producers are important. We explore the empirical relationship between TFPR and demand here, but we first discuss how demand variations fit into the HK framework more generally.

Under the joint assumptions of isoelastic demand and constant marginal costs, shifts in a firm's residual demand curve will not change its TFPR level in the absence of distortions. The inverse is also true: if either or both of these assumptions do not hold, variation in demand will create variation in TFPR.

The invariance of a firm's TFPR to demand shifts under the HK conditions is shown in Figure 3. The firm's initial demand and marginal revenue curves are  $p(q)$  and  $mr(q)$ , and the

firm's optimal price and quantity are  $p^*$  and  $q^*$ . The inverse demand curve then shifts by  $\Delta d$ . This shifts out marginal revenue by  $\Delta d$  as well. As a result, the firm's profit-maximizing quantity rises to  $q'^*$ . However, because isoelastic demand implies a constant multiplicative markup, the profit-maximizing price remains  $p^*$ . Because the firm's price does not change and TFPQ is unaffected by the demand shift, TFPR does not change.

To see how departures from the HK assumptions cause TFPR to be correlated with demand even in the absence of distortions, consider the cases in Figure 4. Panel A shows an example of a non-isoelastic residual demand curve but constant marginal costs. A shift in the firm's residual demand by  $\Delta d$  no longer creates a parallel shift in the marginal revenue curve because the markup varies with quantity. As a result, even though marginal costs are constant, the markup, and hence price, is not. The change in price changes TFPR. Thus, demand shifts TFPR if demand is not isoelastic.

In Panel B, demand is again isoelastic, but marginal costs are no longer constant. Instead the firm's marginal cost rises with its quantity. As opposed to the HK case in Figure 3, a demand shift changes not just the firm's optimal quantity but its price too. The multiplicative markup has not changed, but the firm's marginal cost has because of nonconstant returns. As a result, the demand shift changes TFPR. Here TFPR increases with a positive shift in demand; TFPR would fall if the marginal cost curve were downward sloping.

The comparison of Figures 3 and 4 suggests a test. If one can measure demand shifts (either across firms or within firms over time) that are orthogonal to TFPQ variations, one can see if these demand changes are correlated with TFPR levels. Rejecting the null hypothesis of no correlation would indicate that either the HK assumptions do not hold or that the distortions are correlated with demand. Because the invariance of TFPR to demand changes depends on prices being invariant to demand, a corollary test that we conduct is to see if demand changes are correlated with plant-level prices.

We begin with the demand shifts we used in the prior section to explore the properties of TFPQ\_HK\_WD. For each product, we estimate the simple specification

$$tfpr_{it} = \beta_0 + \beta_1 demand_{it} + \eta_t + \varepsilon_{it}$$

where  $tfpr_{it}$  is (log) TFPR for plant at time  $t$ ,  $demand_{it}$  is the idiosyncratic demand shift identified as described above,  $\eta_t$  is a CM year fixed effect, and  $\varepsilon_{it}$  is the residual. We also estimate a pooled specification where we include a full set of product-by-year effects, and a first-

difference version.<sup>15</sup> We also estimate an analogous specification using the producer's (log) price in period  $t$  as the dependent variable.

The results of the TFPR level specifications are in panel A of Table 3a. Demand is positively correlated with TFPR. The estimated elasticities  $\beta_1$  are positive for every product and statistically significant at the five percent level for all but two products. In the pooled specification, we estimate an average elasticity of TFPR with respect to demand of 0.064, and reject the null with a  $t$ -statistic of 29.9. This elasticity implies that a one standard deviation increase in plant-specific demand corresponds to an increase in TFPR of one-third of a standard deviation. The first difference specification results in panel B also reject the hypothesis of zero covariance between TFPR and demand. The pooled estimates imply that a one standard deviation increase in plant-specific demand yields an increase in TFPR of about 40 percent of a standard deviation in TFPR. (For the sake of comparison to results we describe immediately below, we also run the specification separately on the subset of our sample composed of ready-mixed concrete producers. We find similar results.)

The results using the (log) of plant-level price as the dependent variable are reported in Table 3b. The results closely mimic those for TFPR. The magnitudes of the estimated elasticities are positive and significant at the 5 percent level for seven of the eleven individual products. The pooled average elasticity of price with respect to demand is 0.059, and we reject the null with a  $t$ -statistic of 29.9. This elasticity implies that a one standard deviation increase in plant-specific demand corresponds to a price increase of about 30 percent of a standard deviation. First difference results also yield a large positive and statistically significant elasticity of price with respect to plant-specific demand in both the pooled sample and when restricted to ready-mixed concrete producers.

The results of these tests are especially interesting because in the prior section we highlighted that TFPQ\_HK, both theoretically and empirically, depends strongly on demand. Under the assumptions of the HK framework, TFPR and prices should be invariant to demand, and by implication, TFPQ\_HK as well. Yet when we measure TFPQ\_HK imposing the assumptions of the HK framework, we find that it is—contrary to the implications of the framework—correlated with price and TFPR variation. This internal inconsistency is another

---

<sup>15</sup> We have also estimated the first difference specification industry-by-industry with year effects and pooled sample first differences with product-by-year effects and obtained very similar results. We use the same inverse propensity score weight for the first differences as for the levels.

sign that the conditions necessary to interpret TFPR variation as reflecting distortions do not hold in the data.

We consider a second approach to testing for a relationship between TFPR and demand. This has the advantage that, in principle, one can apply it to a much wider range of data without having direct measures of prices and quantities. As such, it may be of broader applicability for researchers. It uses geographic and vertical distance measures to identify shifts in local downstream demand. We apply this methodology for the products in our dataset that are primarily sold near to where they are produced (Boxes, Bread, Concrete, and Ice).

For each of the local products, we use the detailed U.S. input/output matrix to identify the top ten downstream industries. We combine this with the Longitudinal Business Database to measure employment at the BEA Economic Area level in each downstream industry. Our downstream demand metric for each producer is the weighted average of local employment in each of the downstream demand industries (where the weights are computed using the input/output matrix). We use the log of this value in our tests.

To motivate this approach, consider ready mixed concrete. Demand for concrete is very local; almost all of it is shipped short distances. Further, as emphasized by Syverson (2004), the construction sector accounts for 95% of the ready mixed concrete industry's revenues, but ready mixed accounts for less than 5% of construction sector's intermediate input costs. Thus (local) construction demand drives (local) ready mixed concrete outcomes and not vice versa. We extend this same logic to our other local products.<sup>16</sup>

As before, we consider level and first difference specifications. The former is

$$tfpr_{it} = \beta_0 + \beta_1 \text{downdemand}_{mt} + \beta_m + \eta_t + \varepsilon_{it}$$

where  $\text{downdemand}_{mt}$  is the downstream demand measure in market  $m$  at time  $t$ ,  $\eta_t$  is a period fixed effect, and  $\beta_m$  is a BEA Economic Area (market) fixed effect. We estimate this specification for ready mixed concrete and a pooled estimate for all local market products. The pooled estimates include year-by-Economic Area and product-by-Economic-Area fixed effects. Standard errors are clustered by Economic Areas. Under the HK assumptions,  $\beta_1 = 0$ . The first

---

<sup>16</sup> For ice and bread the top downstream industry is grocery stores. The top downstream industries for boxes are in the wholesale and retail trade sectors. In all of these industries, the share of downstream costs accounted for by the upstream industry is small, just as with concrete.

difference specification uses plants that continue operations across at least two consecutive Economic Censuses. This specification is

$$\Delta tfpr_{it} = \delta_0 + \delta_1 \Delta \text{downdemand}_{mt} + \varepsilon_{it}$$

where under the HK assumptions,  $\delta_1 = 0$ . SEs are again clustered by Economic Areas.

The results of this second test using downstream demand indicators are reported in Tables 4a for TFPR and 4b for price. The magnitudes of the estimated elasticities  $\beta_1$  are positive and statistically significant for price but only marginally statistically significant for the ready mixed concrete and pooled results for TFPR using the level specifications. However, the first difference specifications reject the null hypothesis of zero covariance between TFPR and demand and as well as price and demand at a five percent level. To benchmark the magnitudes of these relationships, the first difference estimates for the pooled specification imply that a one standard deviation increase in downstream demand raises TFPR by about 35 percent of its standard deviation, the same order of magnitude we found with our other demand measures.

It seems from these tests that either that the HK assumptions are violated or, alternatively, that distortions are positively correlated with demand. This raises an obvious question: Can one separately measure distortions and demand in the HK framework if they are in fact correlated? It would be important to do so because from a positive standpoint these firm-level primitives could have very different statistical properties (persistence, variance, etc.) as well as from a normative standpoint because there are different policy implications depending on whether firm outcomes reflect distortions or demand variation.

Distortions and demand cannot be separately identified in standard production data. As we emphasize above, only with CES demand and constant marginal costs is TFPR invariant to TFPQ and demand shifts in the absence of distortions. Departures from these assumptions result in TFPR being a function of fundamentals, creating an identification problem. However, with price and quantity data and more flexible demand and technology structures, further progress can be made distinguishing between fundamentals and distortions. This also permits decomposing TFPR into its fundamental- and distortion-based components. We explore these issues in the next section.

### **III. Quantifying Departures from HK and Effects on Measuring Misallocations**

Given the empirical findings that the HK assumptions do not hold in our data, we now quantify, at least partially, the effects of departures from HK's assumptions on misallocation measurement. To implement this analysis, we need additional structure on both the demand and supply sides of the market. We use a framework that departs from CES demand and constant marginal costs, but nests for both, allowing us to quantify the distance between the data generating process and the assumptions of the HK model.

### A. Generalizing the Model

We assume that the utility function has the hyperbolic absolute risk aversion (HARA) form. HARA preferences are fairly general; they include both CARA and CRRA as special cases.<sup>17</sup> The particular form that we assume is that for the good produced for firm  $i$ :

$$u(Q_i) = \frac{D_i \left[ \left( \frac{Q_i}{1-\rho} + \alpha \right)^\rho - \alpha^\rho \right]}{\frac{\rho}{1-\rho}}$$

where  $Q_i$  is the quantity consumed of firm  $i$ 's product variety,  $D_i$  is a firm-specific demand shifter, and  $\alpha$  and  $\rho$  are parameters common across all varieties. When  $\alpha = 0$ , this simplifies to the CES utility function, so we can quantify demand-side departures from CES in a single parameter.

The utility function for the product of firm  $i$  implies the following inverse demand function:

$$P_i = u'(Q_i) = D_i \left( \frac{Q_i}{1-\rho} + \alpha \right)^{\rho-1}$$

Profit maximization under monopolistic competition implies firms' markups are

$$\frac{P_i}{MC_i} = \frac{u'(Q_i)}{u'(Q_i) + u''(Q_i)Q_i}$$

which, given the utility function, simplifies to

$$\frac{P_i}{MC_i} = \frac{Q_i + \alpha(1-\rho)}{\rho Q_i + \alpha(1-\rho)}$$

If  $\alpha = 0$ , this collapses to the standard CES markup of  $1/\rho$ .

---

<sup>17</sup> Perets and Yashiv (2015) argue HARA utility is not just useful but also an economically "essential restriction."

On the production side, we consider a generalized cost function of the form

$$C(A_i, Q_i) = \left(\frac{Q_i}{A_i}\right)^{\frac{1}{\nu}} \Phi(W)$$

where  $\nu$  is a scale parameter;  $\nu > 1$  ( $\nu < 1$ ) reflects economies (diseconomies) of scale.

To introduce distortions into the model we apply the logic of Hsieh and Klenow (2009) who demonstrate that revenue distortions (the  $\tau_{Yi}$  in their model) are effectively shifters of the marginal cost curve.<sup>18</sup> For expositional convenience it is useful to specify the distortion as proportional to revenue, so we work with a distortion  $T_i$  that is equivalent to  $(1 - \tau_{Yi})^{-1}$  in the HK model. In this proportional form, all distortions  $T_i$  are positive; a “tax” involves  $T_i > 1$  and a “subsidy” is  $T_i < 1$ . Adding this distortion means marginal costs are

$$MC(A_i, Q_i) = \frac{1}{\nu} Q_i^{\frac{1}{\nu}-1} A_i^{-\frac{1}{\nu}} \Phi(W) T_i$$

It will be useful later to be able to measure  $T_i$  in the data. Defining the markup as  $\Psi_i \equiv \frac{P_i}{MC_i}$ , one can show that profit maximization implies

$$T_i = \frac{\nu R_i}{\Psi_i C_i}$$

where the firm’s total revenue and total costs are  $R_i$  and  $C_i$ , respectively. We measure both directly in the data and will estimate firms’ demand and cost functions to obtain  $\nu$  and  $\Psi_i$ .

### B. Decomposing the Variance of TFPR

The next step in our analysis is to decompose the variance of TFPR under our more general demand and cost structures.

We can write TFPR for a producer  $i$  as:

$$TFPR_i \equiv P_i \cdot A_i = \frac{P_i}{MC_i} MC_i \cdot A_i = \Psi_i S_i$$

Where  $\Psi_i \equiv \frac{P_i}{MC_i}$  as above and  $S_i \equiv MC_i \cdot A_i$ . This lets us write the variance of logged TFPR as

(lowercase denotes logged values)

$$V(tfpr_i) = V(\psi_i) + V(s_i) + 2cov(\psi_i, s_i)$$

---

<sup>18</sup> See their equation (6). Note that because they assume constant returns, TFPQ and the distortion term share a common unit exponent in their marginal cost expression. Here, however, because we allow non-constant returns, the exponents will differ.

Using the definition  $S_i \equiv MC_i \cdot A_i$  and taking logs gives

$$s_i = \ln \frac{1}{\nu} + \ln \Phi(W) + \left( \frac{1}{\nu} - 1 \right) (q_i - a_i) + t_i$$

where  $t_i \equiv \ln T_i$ . The first and second terms are constants. It is convenient to define:

$$f_i \equiv \left( \frac{1}{\nu} - 1 \right) (q_i - a_i)$$

Where  $f_i$  is the firm-specific component of  $s_i$ . This only varies across producers if there are non-constant scale economies. Using this definition, we can write the variance of  $s_i$  as

$$V(s_i) = V(f_i) + V(t_i) + 2cov(f_i, t_i)$$

Substituting this into the above expression for the variance of logged TFPR, we have the following decomposition of logged TFPR in terms of markups, non-constant returns to scale effects, and distortions:

$$V(tfpr_i) = V(\psi_i) + V(f_i) + V(t_i) + 2cov(\psi_i, f_i) + 2cov(\psi_i, t_i) + 2cov(f_i, t_i)$$

Under the HK assumptions ( $\alpha = 0$  and  $\nu = 1$ ),  $\psi_i$  and  $f_i$  do not vary across producers, so the first five terms of the decomposition are zero and  $V(tfpr_i) = 0$  absent any distortions. Our more general model lets us measure how deviations from the HK assumptions quantitatively map into TFPR variation. This is decomposed into variations due to firm specific variation in markups and the effects of non-constant returns to scale. In turn, both of the latter are functions of the underlying technology and demand shocks. For example, with  $\alpha > 0$ , HARA yields firm-specific markups that are increasing in firm-specific technology and demand shocks.

### C. Empirical Implementation of the Variance Decomposition

We need estimates of the demand function and the scale elasticity to implement the decomposition. For the demand parameters, we first take the log of the inverse demand function above and do a first-order Taylor expansion of the first term around  $\frac{Q_i}{1-\rho}$ . This gives, after simplifying (lower case denotes logged variables):

$$p_i \approx (1 - \rho) \ln(1 - \rho) - (1 - \rho)q_i - \alpha(1 - \rho)^2 \frac{1}{Q_i} + d_i = b_0 + b_1 q_i + b_2 \frac{1}{Q_i} + d_i$$

As seen in this expression, we can estimate  $\rho$  and  $\alpha$  as well as producers' demand shifters  $D_i$  using a linear regression of logged price on logged quantity and the inverse of the quantity.<sup>19</sup> To obtain consistent estimates, we follow the logic of Foster et al. (2008, 2016) and use functions of the firm's TFPQ as cost-shifting instruments. For the pooled estimates below, we include a full set of product-specific fixed effects in the estimation and also interact them with the third term in the regression to allow  $\alpha$  to vary at the product level.<sup>20</sup>

To obtain the scale elasticity, we estimate the production function applying Wooldridge (2009) through two approaches. In one, we estimate the scale elasticity directly by measuring the elasticity of output to a composite input that is a cost-share-weighted sum of the individual logged inputs (labor, capital, materials, and energy). In the second approach, we estimate each factor elasticity separately and sum them. We find very similar returns to scale estimates either way and report the composite input results here.

We estimate these specifications using the pooled data while controlling for product-by-year effects. This yields estimates of the pooled demand and return to scale parameters. We also estimate these parameters for concrete producers specifically, the product with the largest sample size. We are limited in our ability to obtain precise product-specific estimates because the Wooldridge (2009) method requires lagged instruments and such proxy methods use high order polynomials (we use cubics) that are more reliable with larger samples (see Foster et al. (2017)). Concrete has a sufficient number of observations to allow product-specific implementation.

Panel A of Table 5 reports the parameter estimates. In the pooled sample, the average implied markup is 2.08 (corresponding to an elasticity of demand of -1.93), and the average value of  $\alpha$  is 26.<sup>21</sup> We obtain an estimate of returns to scale statistically equal to one. For

---

<sup>19</sup> Because  $Q$  enters the demand estimation not just in logs but also in levels, there is a practical matter to be dealt with in terms of the units of  $Q$ . We have shown in unreported results that there is a proportionality relationship between  $Q$  and  $\alpha$ . That is, choose any factor of proportionality for  $Q$  (e.g., multiply  $Q$  by 10), and  $\alpha$  will change by exactly that factor. This implies, importantly, that the estimated markup is independent of the choice of units for  $Q$ . For the sake of comparability, where we want to report an average value of  $\alpha$  across products, we renormalize the units of quantity for each product (which we observe directly in physical units) so that the median producer of each product has  $Q = 100$ . This normalization has no influence on the dispersion of log markups or log(TFPR) and thus no influence on the TFPR decomposition below.

<sup>20</sup> The instruments in the pooled specification are TFPQ,  $\log(\text{TFPQ})$ , and  $(1/\text{TFPQ})$ , where the last is interacted with product fixed effects. For the concrete-only sample, the interaction terms are of course not present.

<sup>21</sup> The parameter  $\alpha$  is product specific because  $\alpha = b_2/b_1^2$ , and  $b_2$  is product specific. We cannot report these specific estimates for disclosure reasons. However, we find that  $b_2$  is positive for all products and statistically significantly so at the 10% level for 7 of the 11 products.

concrete, consistent with our prior work we find a smaller implied markup, 1.52 (corresponding to a demand elasticity of -2.92). We also cannot reject constant returns to scale.<sup>22</sup>

Before turning to the decomposition, panel B of Table 5 reports key summary statistics of objects in the data, whether measured directly or estimated using the model. Considerable heterogeneity exists among producers of a given product. In the whole sample, the (within-product) standard deviation of log markups ( $\psi_i$ ), logged demand ( $d_i$ ), and logged distortions ( $t_i$ ) are 0.195, 0.478, and 0.284 respectively. Each variance is smaller among concrete producers but still considerable. As for correlations, TFPR and our distortion measure are positively related, but the correlation is far from one. Thus in our more flexible model, TFPR variation does not summarize the extent of distortions. TFPR is also positively correlated with the demand shock—indicating that, as discussed above, demand primitives end up reflected in TFPR in settings outside the HK framework. The correlations between  $t_i$  and fundamentals vary in sign; distortions covary positively with TFPQ but negatively with demand shocks.

The variance decomposition results, shown in panel C of Table 5, indicate that fundamentals account for an important fraction of TFPR variation that is independent of distortions. In the pooled results, variance in fundamentals accounts for 80 percent of TFPR variation.<sup>23</sup> It is roughly 20 percent for concrete. The variance of our distortion measure is also an important contributor to TFPR, equaling 169% of the TFPR variance in the pooled results and 111% in the concrete sample. The balance of the TFPR variance is accounted for by the negative covariance between fundamentals and distortions, which, given that returns to scale are constant, reflects a negative correlation between firms' markups (driven by TFPQ or demand variation) and the measured distortion. Given  $\alpha > 0$ , markups are increasing in output and in turn TFPQ and demand shocks. As noted by Dhingra and Morrow (forthcoming), this also implies that firms with low fundamentals (e.g., low demand or TFPQ) will be larger than they should be relative to the allocation of the social planner. In contrast, residual distortions are inversely correlated with output and demand shocks (unlike TFPR which is positively correlated with both).

---

<sup>22</sup> Using a very different (Klette and Griliches (1996)) approach, Foster et al. (2017) find evidence of mild increasing returns—with estimates of returns to scale that average 1.09—on a much larger sample of industries.

<sup>23</sup> The fact that our estimated scale elasticities are close to one implies that for both the pooled and concrete samples, the second, fourth and fifth terms in the TFPR variance decomposition are close to zero. Because we cannot statistically reject constant returns, we assume these terms are zero in what follows.

We can therefore explain part of the variation of TFPR as purely reflecting variation in demand and cost fundamentals once we allow for more general demand and cost structures that nest for, but reject, the HK model. In these particular samples, the variation in the firm specific markup is the key source of differences between the data and the HK assumptions. Unlike in the HK model, markups are increasing in both demand shocks and TFPQ and, in turn, output. Distortions do still have a considerable explanatory role, both directly and through their covariance. However, it is important to note that just as with the HK framework, our distortion measure is also a residual. Now, rather than being TFPR itself, it is the part of TFPR that we cannot account for with our more flexible demand and cost structures. Remaining departures of the data generating process from *our* augmented framework would be labeled distortions even if they were not.

#### D. *The Relationship between Measured Distortions and Survival*

Table 6 reports estimates of the marginal effect of various measures and combinations of measures on the probability of exit. For the sake of comparison, columns 1-3 reprint the results from Table 2. These show that in our sample, as in much of the literature, businesses with higher TFPR, higher TFPQ, and higher demand are less likely to exit.

We find that businesses with higher measured distortions in our framework are more likely to exit, as seen in columns 5 and 6. This is an interesting contrast to the negative correlation between TFPR and exit rates in column 1. The HK framework interprets TFPR variation as reflecting distortions, yet column 1 would imply more distorted producers (those facing a higher “tax”) are *more* likely to survive. This seems an odd empirical property of distortions. On the other hand, the positive correlation between exit and our distortion suggests that, by being less restrictive about supply and demand—thereby allowing the model to explain more of the data in terms of supply and demand fundamentals rather than distortions—we obtain a residual distortion measure that empirically behaves more like one might expect.

The result in column 7 bolsters this logic. Once we control for supply and demand fundamentals (TFPQ and the demand shock, the latter obtained imposing the CES demand of the HK framework for the sake of internal consistency with that model), TFPR and exit are now positively correlated. In other words, the part of TFPR that is independent of fundamentals does look more like a true distortion, but this component is empirically swamped by positive

fundamentals about producer profitability. Our more flexible model captures those variations within the demand and supply specifications themselves, leaving the distorting influences to the distortion metric.

These results are consistent with the identification problem we mentioned above. In the end, empirical distortion measures are a residual. Their separate identification from fundamentals exists only to the extent that one believes the modeled structure of producer demand and costs. Measured “distortions” may still embody elements of producers’ idiosyncratic demand or costs that are—contrary to the concept of a distortion that acts as an implicit tax—“good news” about the producer’s survival prospects. Unmodeled idiosyncratic demand and cost conditions would be misinterpreted as misallocation. We show that one can make progress on reducing the extent to which this confound occurs by using more flexible modeling structures. We were able to leverage price and quantity data that most researchers in this literature do not have access to, however. Accounting for model misspecification without such data is a considerably more difficult task, raising the likelihood in more general settings that misallocation measures will confound distortions with other components of idiosyncratic profitability.

#### *E. Variable-Elasticity Demand and Distortions*

The results above highlight the role of variable markups, resulting from patterns of non-CES demand in the data, in explaining TFPR variation. Even in the complete absence of distortions, variation in firms’ demand and cost fundamentals will shift their optimal outputs to different quantities on their residual demand curves, resulting in markup (and therefore TFPR) dispersion.

This variable-markup channel for TFPR variation raises an interesting and more general question about its welfare effects. One could argue that markup variation is itself a type of distortion; it means products with the same marginal cost could sell at different prices. Allocative efficiency in many models would require a common markup. While outside the scope of the HK framework, one might argue that under variable-elasticity demand, TFPR variation might still reflect distortions. These distortions are not the free-floating “tax” of HK, but rather would be induced by, and necessarily correlated with, heterogeneity in producers’ demand and supply primitives.

However, in this section we build on our results above to show that in a general setting with non-CES demand and heterogeneous firms that enter endogenously into a market, the variance of TFPQ is still not a summary welfare metric. Namely, there is not a unique mapping between the size of misallocation losses and the variance of TFPQ, and zero TFPQ variance does not imply zero misallocation losses.

The theoretical foundation for our exploration here is laid out in Dhingra and Morrow (forthcoming). Their work sums up the logic of efficiency in markets with heterogeneous producers facing variable-elasticity demand and engaging in monopolistic competition. They show in general that the free-entry market equilibrium is not allocatively efficient (indeed, the only exception to this is when demand is CES—essentially extending the Dixit and Stiglitz (1977) result to heterogeneous producers).<sup>24</sup>

Dhingra and Morrow characterize how allocative inefficiency relates to the shape of the demand curve. When demand is such that markups grow in quantity—which is the case for HARA utility when  $\alpha > 0$ —there are three departures from efficiency. First, the largest producers in the market (those with the highest combinations of TFPQ and demand) will be smaller in a market equilibrium than their socially optimal size. These firms under-produce because the social surplus created by a marginal unit of their output exceeds its marginal private profit. Second, below some threshold TFPQ-demand combination, firms will be larger than their socially optimal size.<sup>25</sup> Here, the marginal profit of their output exceeds its social surplus. Third, selection will be biased to be too low with a set of producers with sufficiently low realized combination of TFPQ and demand who are profitable enough to stay in the market in equilibrium, but whose social surplus is not large enough to justify their operations under the social planner. There is not an unambiguous prediction regarding the bias in the pace of entry in the private market vs. the social planner. However, while the sign of the bias is ambiguous, entry is typically inefficient.

The selection and entry components of allocative inefficiency have important implications for the relationship between the variance of markups, TFPQ, and welfare. Even if

---

<sup>24</sup> The Dhingra and Morrow (forthcoming) model is a standard ex ante homogeneous, ex post heterogeneous entry, exit, produce model. Potential entrants pay a fixed cost to learn their combined TFPQ and demand shock. After the latter are revealed, the firm decides whether to produce given there are fixed costs of operation. The novelty of the model and insights is the consideration of VES demand structure. We follow this model very closely in our calibrated analysis allowing for ex post heterogeneity both through cost (TFPQ) and demand shocks.

<sup>25</sup> Dhingra and Morrow (forthcoming) note that this threshold may not be interior.

one were to force reallocations among producers in a market in order to eliminate markup dispersion (and TFPR dispersion in absence of HK-type distortions), there would still be misallocative losses because some firms shouldn't even be operating. Additionally, the same observed TFPR variance will be consistent with very different levels of efficiency loss depending on the particulars of the demand system and its effect on relative sizes, entry and exit.

We quantify this logic in our setting by solving a version of the Dhingra and Morrow framework using HARA demand with parameter settings guided by our data. We model a product market as comprising a set of ex-ante identical producers who pay a fixed cost  $f$  to learn their TFPQ and demand shift draws from known distributions. Those receiving a combination of draws such that production is profitable enter the market and produce the quantity that maximizes their profits given their draws as well as the number and types of other entrants.<sup>26</sup> We parameterize demand using our (pooled) estimates of  $\rho$  and  $\alpha$  from above. We assume the standard deviation of the sum of (logged) TFPQ  $a_i$  and (logged) demand shock  $d_i$  is 0.54, as in our pooled sample. Using the three remaining free parameters of the model—the mean of the TFPQ-demand composite (the distribution is assumed to be lognormal), a fixed cost of operating conditional on entry, and fixed cost of entry  $f$ —we seek to match the following aspects of the data. First, we set the median normalized quantity level for products (derived as above) to exactly 100. We seek an exact match here in order to replicate aggregate output by product. Second, we try to match the observed dispersion in markups across producers of a given product.

We find parameters that best fit the data (matching the median normalized quantity exactly while implying a standard deviation of the logged markup of 0.12) are a fixed cost of entry  $f = 3.75$ , a fixed cost of operating each period equal to 3.5, and a mean of the composite (log) shock of 2.3. Given these estimates, the calculation in Dhingra and Morrow implies that the market equilibrium has 25% lower social surplus than the efficiency-maximizing allocation. Thus, the largest producers in equilibrium are too small, the smallest are too large, and a social planner would not have a large set of the remaining producers operating at all. For our parameter estimates, the social planner would reduce the normalized output of the median-sized producer from 100 to about 70. The social planner's and market equilibrium normalized quantities cross at about 1250; above this size, the social planner would make producers larger than they are market

---

<sup>26</sup> This “ex-ante identical, heterogeneous type, endogenous entry” theoretical framework is used in a large class of models, including for example Melitz (2003), Asplund and Nocke (2006), and Foster et al. (2008).

equilibrium. The fraction of entrants that exit upon learning their composite TFPQ and demand shock is much lower in the market equilibrium (about 5%) than in the efficient outcome (about 50%). The mass of entrants per capita (i.e., the ratio of the mass of entrants to the mass of workers) is about the same in equilibrium as it is for the social planner (about 6%).

Importantly, the variance of TFPR (which in this model, given that we have not added any distortions and have constant returns to scale, comes completely from variable markups) is not a sufficient statistic for misallocative welfare losses. Other combinations of our three chosen parameters yield the same in TFPR but have very different implied welfare losses. For example, it is straightforward to keep the variance of the TFPR by simultaneously lowering (increasing) the entry cost and the fixed operating cost while keeping the mean of the composite shock the same. Even though TFPR variance remains invariant, this results in a greater (lower) welfare loss, as more (fewer) producers become profitable enough to enter the market despite not justifying, from a social planner's perspective, their social entry cost. The implied welfare differences among the various scenarios are large. We found parameter combinations that could double or halve the estimated 25% welfare loss, all while holding the observed dispersion in TFPR fixed.

For example, with the fixed cost of entry at 0.75 and the fixed cost of operating at 3.45, the variance of TFPR remains at 0.12 but the welfare loss is 50%. In contrast, with the fixed cost of entry at 6.25 and the fixed cost of operating at 3.55, the variance of TFPR remains at 0.12 but the welfare loss is 12.5%. The substantial changes in the welfare loss with invariant variance of TFPR reflects substantially larger (smaller) entry and exit with lower (higher) fixed costs of entry and operating. Even though the range of producers changes substantially across this range of fixed costs, the variance of TFPR remains invariant in part because the elasticity of the markup with respect to the composite shock changes substantially as well. At fixed cost of entry of 0.75, this elasticity is 0.22 while it is 0.18 at 6.25.

#### **IV. Concluding Remarks**

Measuring misallocation—identifying idiosyncratic distortions that adversely impact the allocation of resources—is a first order issue. Our analysis highlights difficult identification challenges for measuring distortions. In particular, we view our paper as sounding a note of caution about using differences across producers' measured revenue productivity (TFPR) levels

to measure distortions. The stringent assumptions of the Hsieh and Klenow (2009) framework that enables such identification typically do not hold in the US data where price and quantity data are available, and other evidence suggests this may be a more general issue.

We find that there is incomplete pass-through of TFPQ in plant-level prices, one of the implications of the stringent assumptions of the HK framework. Perhaps as a result of this departure from the framework's assumptions, TFPQ measures derived indirectly using the framework are only weakly correlated with and have much more dispersion than directly measured TFPQ. Moreover, the indirect measures of TFPQ are inversely related to firm survival (in contrast to the direct measures), inconsistent with economic theory.

To quantitatively account for these patterns, we augment the HK framework to allow for departures from CES demand and constant marginal costs. We find evidence of such departures in our data and find that (non-distortionary) demand and cost fundamentals explain a considerable portion of the observed TFPR variation. Interestingly, the residual measure of distortions that emerges from our model is positively related to exit, as one might expect from a distortion, while TFPR (the measure of distortions in the HK framework) is negatively related. Finally, we show that in general settings with heterogeneous producers, the variance of TFPR is not a sufficient statistic for misallocative welfare losses. Taken together, these findings raise questions about using TFPR as a measure of distortions.

While much of the message of our paper is to sound a note of caution, one of our findings suggests there is interesting information captured by TFPR (and our residual measure) once one controls for fundamentals. Specifically, we find that after controlling for TFPQ and demand shocks, plants with high TFPR and high residual measures of distortion are more likely to exit. Thus, if independent information on fundamentals can be measured (feasible with price and quantity data but a challenge in their absence) then one might be able to make progress at isolating a sharper measure of true distortions.

We close by noting that we have focused here on the role of model misspecification in accounting for reasons why misallocation measures may not simply reflect wedge-like distortions. However, there are several additional reasons why TFPR might vary across firms in the absence of distortions. These include differences in factor prices (Katayama, Lu, and Tybout (2009)), factor quality, heterogeneity in factor demand and elasticities, adjustment costs (Asker,

Collard-Wexler, and De Loecker (2014) and Decker et. al. (2018) offer extensive analysis of the role of adjustment costs) and measurement error (Bils, Klenow and Ruane (2017)).

It is beyond the scope of this paper to consider all of these alternatives, but they too provide reasons for applying considerable caution when measuring misallocation using revenue productivity dispersion. We also mention these because we think our analysis provides guidance that can be potentially used to help differentiate amongst them. Our findings highlight the following properties observed when price and quantity data are available. First, TFPQ and TFPR are strongly positively correlated. Second, TFPR is strongly positively correlated with producers' idiosyncratic demand levels. Third, the elasticity of prices with respect to TFPQ is less than one. Fourth, survival is greater for plants with higher TFPR, higher TFPQ, and higher demand. Fifth, interestingly, when all three of these measures are considered jointly, plants with higher TFPQ and demand are more likely to survive holding the other factors constant, but plants with higher TFPR are less likely to survive holding TFPQ and demand constant. Thus, researchers should take several moments into account when evaluating models that account for TFPR dispersion.<sup>27</sup>

Of the alternative explanations for TFPR variation above, one that we regard as especially relevant and promising for being consistent with our evidence is factor adjustment costs. A firm with a positive realization of TFPQ wants to become larger. In a frictionless environment the firm increases factors to the point where marginal revenue products equal the input factor costs. Output rises and price falls. Under the HK assumptions, price falls just enough to counteract the increase in TFPQ. If there are adjustment frictions, however, the increase in inputs and output will be smaller, making the decline in prices smaller too. Accordingly, the positive TFPQ realization will result in an increase in TFPR. Putting the pieces together, TFPR will be positively correlated with TFPQ, prices will have a less than unit elastic response to TFPQ, and—given the positive correlation between TFPR and fundamentals—higher TFPR firms will be more likely to survive. In short, adjustment frictions have implications that match many of the core findings of our analysis.

---

<sup>27</sup> Foster et al. (2017) include a complementary analysis with related but distinct findings. They contrast and compare TFPR to residuals from revenue function estimation, highlighting that the latter are conceptually different from TFPR. Under CES demand, the revenue function residuals are a function of fundamentals, TFPQ and demand shocks. The reason is that the estimated parameters of the revenue function are revenue elasticities reflecting both output elasticities and the demand elasticity. They find that TFPR and revenue function residuals are highly correlated, exhibit similar dispersion and are both positively related to survival. This provides a distinct set of moments that should be taken into account in modeling TFPR dispersion.

## References

- Asker, John, Allan Collard-Wexler and Jan De Loecker. 2014. “Dynamic Inputs and Resource (Mis)Allocation,” *Journal of Political Economy*, 122(5), 1013-1065.
- Asplund, Marcus and Volker Nocke. 2006. “Firm Turnover in Imperfectly Competitive Markets.” *Review of Economic Studies*, 73(2): 295-327.
- Bils, Mark, Peter J. Klenow, and Cian Ruane. 2017. “Misallocation or Mismeasurement?” Working Paper.
- Bonnet, Céline, Pierre Dubois, Sofia B. Villas Boas, and Daniel Klapper. 2013. “Empirical Evidence on the Role of Nonlinear Wholesale Pricing and Vertical Restraints on Cost Pass-Through.” *Review of Economics and Statistics*, 95(2): 500–15.
- Campa, José Manuel and Linda S. Goldberg. 2005. “Exchange Rate Pass-Through into Import Prices.” *Review of Economics and Statistics*, 87(4): 679–90.
- Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2018. “Changing Business Dynamics and Productivity: Shocks vs. Responsiveness.” NBER Working Paper No. 24236.
- De Loecker, Jan and Frederic Warzynski. 2012. “Markups and Firm-level Export Status.” *American Economic Review*, 102(6), 2437-71.
- Dhingra, Swati and John Morrow. Forthcoming. “Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity.” *Journal of Political Economy*.
- Dixit, Avinash K. and Joseph E. Stiglitz. 1977. “Monopolistic Competition and Optimum Product Diversity.” *American Economic Review*, 67(3): 297-308.
- Fabra, Natalia and Mar Reguant. 2014. “Pass-Through of Emissions Costs in Electricity Markets.” *American Economic Review*, 104(9): 2872-99.
- Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?” *American Economic Review*, 98(1): 394-425.
- Foster, Lucia, John Haltiwanger, and Chad Syverson. 2016. “The Slow Growth of New Plants: Learning about Demand.” *Economica*, 83(329): 91-129.
- Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2017. “Macro and Micro Dynamics of Productivity: From Devilish Details to Insights.” NBER Working Paper No. 23666.

- Ganapati, Sharat, Joseph S. Shapiro, and Reed Walker. 2016. "Energy Prices, Pass-Through, and Incidence in U.S. Manufacturing." Working paper.
- Goldberg, Penelopi K. and Frank Verboven. 2001. "The Evolution of Price Dispersion in the European Car Market." *Review of Economic Studies*, 68(4): 811-48.
- Hsieh, Chang-Tai and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403-48.
- Katayama, Hajime, Shihua Lu, and James R. Tybout. 2009. "Firm-Level Productivity Studies: Illusions and a Solution." *International Journal of Industrial Organization*, 27(3): 403–13.
- Klette, Tor Jakob and Zvi Griliches. 1996. "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous," *Journal of Applied Econometrics*, 11(4): 343–361.
- Kulick, Robert. 2016. "Horizontal Mergers, Prices and Productivity," Mimeo.
- Melitz, Marc. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695-1725.
- Nakamura, Emi and Dawit Zerom. 2010. "Accounting for Incomplete Pass-Through." *Review of Economic Studies*, 77(3): 1192-1230.
- Perets Gadi S. and Eran Yashiv. 2015. "The Fundamental Nature of HARA Utility." Discussion Papers 1522, Centre for Macroeconomics (CFM).
- White, T. Kirk, Jerome P. Reiter, and Amil Petrin. 2014. "Plant-level Productivity and Imputation of Missing Data in U.S. Census Manufacturing Data." Mimeo, Center for Economic Studies.
- Wooldridge, Jeffrey. 2009. "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables." *Economics Letters*, 104(3): 112-114.

Figure 1. Effect of a Change in TFPQ in the Hsieh-Klenow Framework

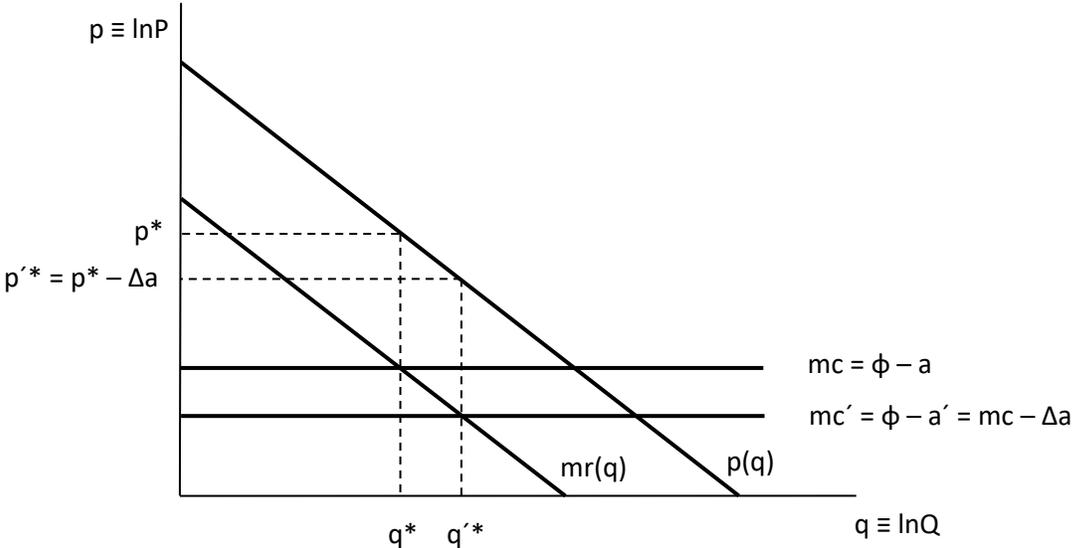


Figure 2. Effect of a Change in TFPQ when the Marginal Cost Curve Is Not Horizontal

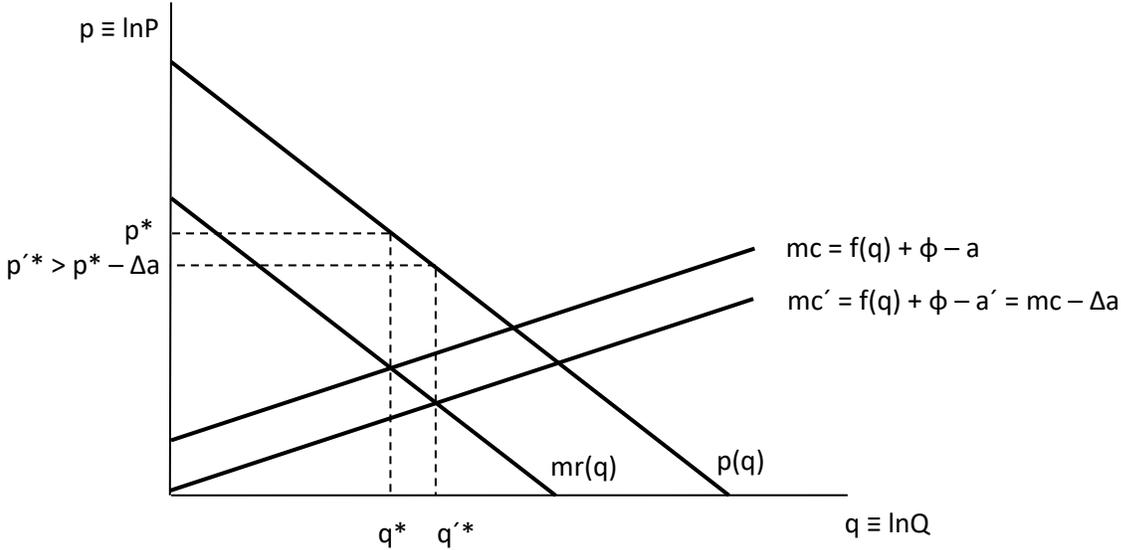


Figure 3. Demand Shifts Do Not Change TFPR in HK

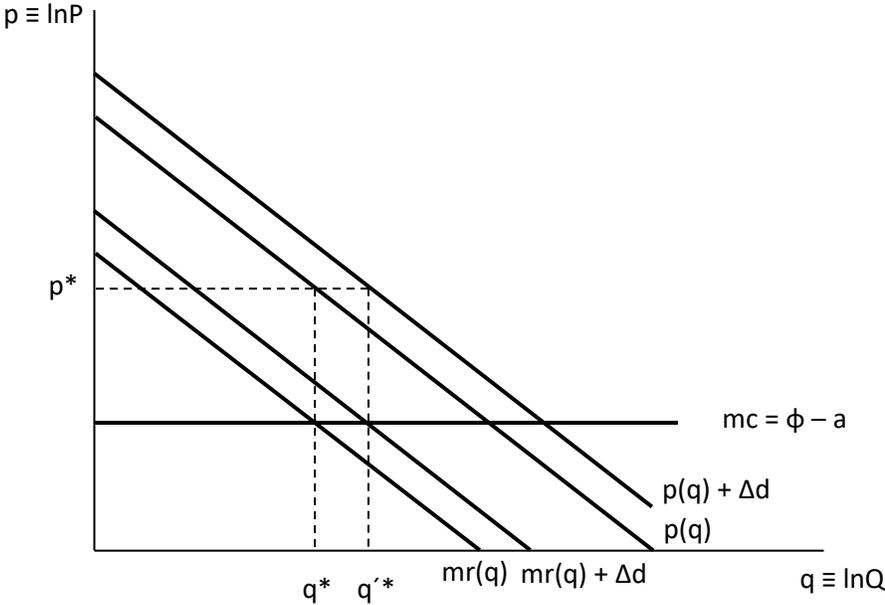
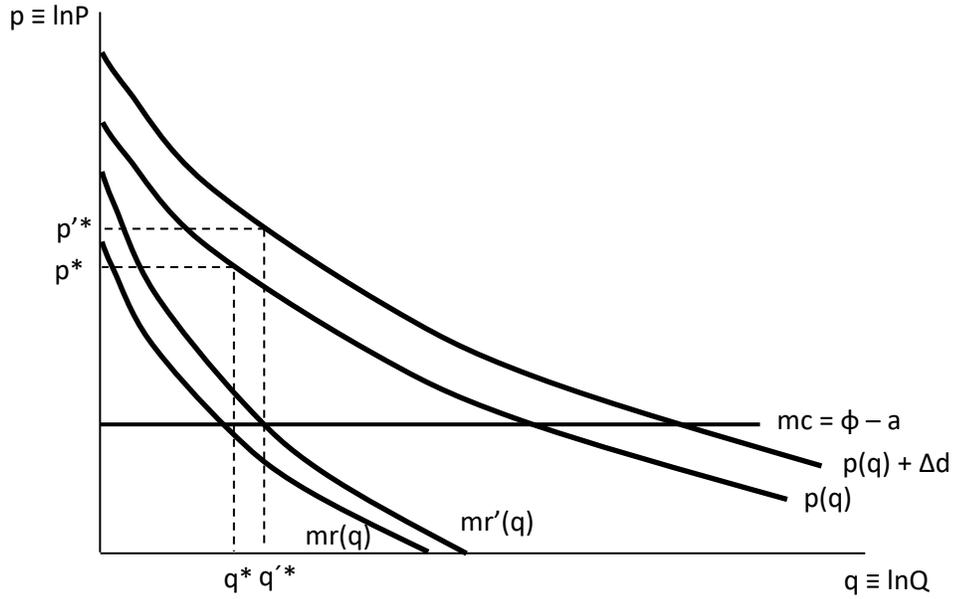


Figure 4. Demand Shifts Change TFPR If HK Assumptions Do Not Hold

A. Non-Isoelastic Demand but Constant Marginal Costs



B. Isoelastic Demand but Non-Constant Marginal Costs

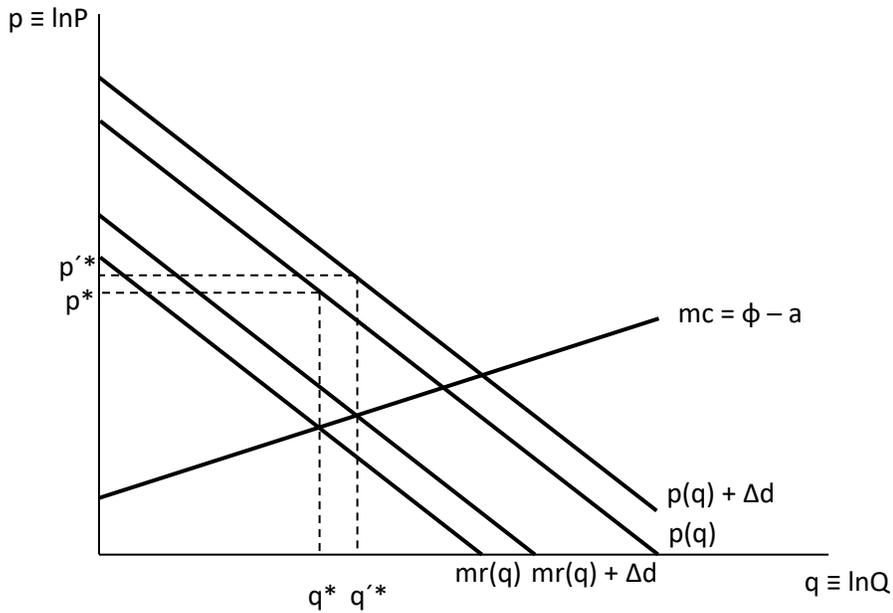


Table 1. Elasticity of Plant-Level  $\log(\text{Price})$  to  $\log(\text{TFPQ})$

Product	Point Estimate	Std. Error	t-stat for $H_0: \alpha_1 = -1$
Boxes	-0.825	0.013	-13.4
Bread	-0.521	0.031	-15.6
Carbon Black	-0.691	0.071	-4.4
Coffee	-0.527	0.038	-12.5
Concrete	-0.265	0.008	-91.9
Flooring	-0.724	0.064	-4.3
Gasoline	-0.251	0.024	-31.3
Block Ice	-0.569	0.067	-6.4
Processed Ice	-0.521	0.041	-11.8
Plywood	-0.862	0.020	-6.9
Sugar	-0.177	0.035	-23.5
Pooled, OLS	-0.450	0.006	-86.4
Pooled, IV (Lagged TFPQ)	-0.537	0.043	-10.7

Notes: The total sample (pooled) is approximately 9600 observations. All specifications use inverse propensity score weights to account for selection in using only non-imputed physical product data. By product estimates include year effects. Pooled specifications include product by year effects. Differences in reported t-statistics and ratio of reported point estimates and standard errors subject to rounding error.

Table 2. Selection on Alternative TFP Measures and Demand

Specification:	[1]	[2]	[3]	[4]
TFPR	-0.039 (0.019)			
TFPQ		-0.035 (0.017)		
Demand Shock			-0.055 (0.005)	
TFPQ_HK				-0.005 (0.001)

Note: These results show marginal effects from of plant exit by the next census (shown by column) on plant-level (logged) measures as well as a full set of product-year fixed effects. The sample is the pooled sample of approximately 9600 observations. Standard errors, clustered by plant, are in parentheses.

Table 3a. Elasticity of Plant-Level  $\ln(\text{TFPR})$  to Plant-Level  $\ln(\text{Demand})$

A. Levels:

Product	Point Estimate	Std. Error	t-stat for $H_0: \beta_1 = 0$
Boxes	0.029	0.003	8.8
Bread	0.118	0.010	12.4
Carbon Black	0.087	0.045	1.9
Coffee	0.074	0.008	9.3
Concrete	0.068	0.003	24.2
Flooring	0.069	0.028	2.4
Gasoline	0.004	0.005	0.7
Block Ice	0.195	0.060	3.3
Processed Ice	0.098	0.030	3.2
Plywood	0.008	0.015	0.5
Sugar	0.085	0.031	2.8
Pooled, All Products	0.064	0.002	29.9

B. First Difference Specification for Continuing Plants:

Product	Point Estimate	Std. Error	t-stat for $H_0: \delta_1 = 0$
Concrete	0.135	0.007	20.7
Pooled, All Products	0.133	0.005	25.8

Notes: The total sample (pooled) is approximately 9600 observations. All specifications use inverse propensity score weights to account for selection in using only non-imputed physical product data. For the level specifications, by product estimates include year effects and pooled specification includes product by year effects. For the first difference specification, pooled specification includes product effects. Differences in reported  $t$ -statistics and ratio of reported point estimates and standard errors subject to rounding error.

Table 3b. Elasticity of Plant-Level ln(Price) to Plant-Level ln(Demand)

A. Levels:

Product	Point Estimate	Std. Error	t-stat for $H_0: \beta_1 = 0$
Boxes	0.028	0.006	4.9
Bread	0.118	0.010	11.8
Carbon Black	0.054	0.059	0.9
Coffee	0.074	0.008	8.8
Concrete	0.061	0.002	32.4
Flooring	0.068	0.044	1.6
Gasoline	0.004	0.003	1.1
Block Ice	0.192	0.069	2.8
Processed Ice	0.113	0.031	3.6
Plywood	-0.001	0.043	0.0
Sugar	0.071	0.015	4.9
Pooled, All Products	0.059	0.002	29.9

B. First Difference Specification for Continuing Plants:

Product	Point Estimate	Std. Error	t-stat for $H_0: \delta_1 = 0$
Concrete	0.133	0.003	40.8
Pooled, All Products	0.159	0.003	46.2

Notes: The total sample (pooled) is approximately 9600 observations. All specifications use inverse propensity score weights to account for selection in using only non-imputed physical product data. For the level specifications, by product estimates include year effects and pooled specification includes product by year effects. For the first difference specification, pooled specification includes product effects. Differences in reported  $t$ -statistics and ratio of reported point estimates and standard errors subject to rounding error.

Table 4a. Elasticity of Plant-Level  $\ln(\text{TFPR})$  to Downstream Demand

A. Level:

Product	Point Estimate	Std. Error	$t$ -stat for $H_0: \beta_1 = 0$
Concrete	0.046	0.025	1.82
Pooled, Local Products	0.042	0.024	1.74

B. First Difference Specification for Continuing Plants:

Product	Point Estimate	Std. Error	$t$ -stat for $H_0: \delta_1 = 0$
Concrete	0.127	0.052	2.42
Pooled, Local Products	0.115	0.050	2.33

Table 4b. Elasticity of Plant-Level  $\ln(\text{Price})$  to Downstream Demand

C. Level:

Product	Point Estimate	Std. Error	$t$ -stat for $H_0: \beta_1 = 0$
Concrete	0.075	0.022	3.42
Pooled, Local Products	0.076	0.022	3.51

D. First Difference Specification for Continuing Plants:

Product	Point Estimate	Std. Error	$t$ -stat for $H_0: \delta_1 = 0$
Concrete	0.108	0.032	3.42
Pooled, Local Products	0.107	0.029	3.72

Notes: The total sample (pooled for local products) is approximately 8000 observations. All specifications use inverse propensity score weights to account for selection in using only non-imputed physical product data. For the level specifications, by product estimates include year effects and economic area effects, and pooled specification includes product, year and economic area effects. For the first difference specification, pooled specification includes product effects. Differences in reported  $t$ -statistics and ratio of reported point estimates and standard errors subject to rounding error.

Table 5: Distortion Estimates with HARA Demand and Non-constant Marginal Costs

Panel A: Parameter Estimates

Statistic	Pooled	Concrete
Coefficient on $q$ (demand parameter)	-0.643 (0.041)	-0.426 (0.114)
Coefficient on $1/Q$ (demand parameter)	-5.52 (3.619)	-6.78 (3.870)
Implied average value of $\alpha$ across products	26	
Average implied markup $\Psi_i$	2.08	1.52
Estimated scale elasticity, $\nu$	1.00 (0.02)	1.02 (0.02)

Panel B: Dispersion and Correlations

Statistic	Pooled	Concrete
$SD(\psi_i)$	0.195	0.095
$SD(d_i)$	0.478	0.325
$SD(\tau_i)$	0.284	0.233
$\text{Corr}(p_i, t_i)$	0.33	0.32
$\text{Corr}(tfpr_i, t_i)$	0.57	0.75
$\text{Corr}(a_i, t_i)$	0.22	0.50
$\text{Corr}(d_i, t_i)$	-0.27	-0.10
$\text{Corr}(tfpr_i, d_i)$	0.15	0.24
$\text{Corr}(tfpr_i, a_i)$	0.66	0.78

Panel C: Variance Decomposition

Fraction of Variance of TFPR from:	Pooled	Concrete
Fundamentals	0.80	0.19
Distortions	1.69	1.11
Covariance of fundamentals and distortions	-1.48	-0.30

Notes: Approximate sample sizes are 9600 in the pooled sample, 5800 for concrete. Standard errors in parentheses in panel A. Pooled results control for product by year effects. In panel C, “Fundamentals” include the first and second terms of the decomposition. “Distortions” reflect the variance of the third term, and the “Covariance” terms are the fourth, fifth, and sixth terms of the decomposition.

Table 6. Selection on TFP Measures, Demand, and Distortions

Specification:	[1]	[2]	[3]	[4]	[5]	[6]	[7]
TFPR	-0.039 (0.019)						0.120 (0.028)
TFPQ		-0.035 (0.017)				-0.060 (0.017)	-0.101 (0.023)
Demand shock, CES			-0.055 (0.005)				-0.062 (0.005)
Demand shock, HARA				-0.064 (0.013)		-0.047 (0.013)	
Distortions ( $t$ )					0.113 (0.018)	0.107 (0.019)	

Note: These results show marginal effect from probits of plant exit by the next census (shown by column) on plant-level measures as well as a full set of product-year fixed effects. The sample is the pooled sample of approximately 9600 observations. Standard errors, clustered by plant, are in parentheses.