Very preliminary; please do not cite or circulate

Misallocation Measures:

Glowing Like the Metal on the Edge of a Knife [Placeholder Title][*]

John Haltiwanger, University of Maryland and NBER

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

March 2016

Abstract. A large literature on misallocation and productivity has arisen in recent years. The standard empirical framework in this line of work is Hsieh and Klenow (2009; hereafter HK). The framework's usefulness and theoretical founding make it a valuable starting point for analyzing misallocations. However, we show that the empirical lynchpin of this approach rests very much on a knife's edge. The condition in the HK model that maps from observed production behaviors to the misallocative wedges/distortions holds in a single theoretical case, with strict assumptions required on both the demand and supply side. We demonstrate that applying the HK methodology to data when there is any deviation from these assumptions will mean that the "wedges" recovered from the data may not be signs of inefficiency at all. Rather, they may simply reflect demand shifts or movements of the firm along its marginal cost curve. Moreover, there are several conditions under which the spurious wedges actually reflect idiosyncratic demand or cost conditions that are good (related to higher profits) for the business. The framework may then not just spuriously identify inefficiencies; it might be more likely to do so precisely for businesses that are in some fundamental way better than their competitors. Preliminary 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 also suggest a more general empirical pre-test of the framework's required assumptions that researchers can apply in the more typical case where only producer revenues are observed rather than prices and quantities separately.

[*] Contact information: Haltiwanger: haltiwan@econ.umd.edu; Department of Economics, University of Maryland, College Park, MD 20742; Syverson: chad.syverson@chicagobooth.edu; University of Chicago Booth School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637. 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 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 be either too large or too small relative to their socially efficient size. One of the seminal papers espousing 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 misallocation's productivity effects. However, we show that the empirical lynchpin of the HK approach rests very much on a knife's edge. The condition in the HK model that maps from observed production behaviors to the 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 that have a negative unit elasticity with respect to TFPQ. We show that applying the HK methodology to data when there is any deviation from

these elements will mean that the "wedges" recovered from the data may not be signs of inefficiency at all. 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. What is more, there are several conditions under which the spurious wedges actually 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 in fact be more likely to do so precisely for businesses that are in some fundamental way better than their competitors.

We go into detail below 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 physical (quantity-based) efficiency (aka TFPQ). This means that through the lens of the model, any observed 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 TFPR is invariant to TFPQ differences across producers (or for that matter, differences over time for a given producer). For every 1% increase (decrease) in TFPQ, price falls (rises) by 1%. These two changes cancel each other out in their product, so TFPR doesn't change. The HK model uses this invariance implication to back out misallocation measures from the TFPR dispersion that is (inevitably) observed in the data. The model reads TFPR differences as inefficiencies.

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 demonstrate this in detail below, but the intuition is straightforward. The conditions on cost imply for each percentage increase in TFPQ, the marginal cost curve falls by that same amount. Because marginal costs are constant in quantity, this percentage change is the same at any quantity the firm might produce (and a producer's optimal quantity does change as TFPQ does in most models). On the demand side, isoelastic demand ensures a constant multiplicative markup—that is, the ratio of a producer's price to its marginal cost does not change. As a result, the producer's price falls in

exact proportion to its marginal cost. Putting these cost and demand implications together implies that any change in TFPQ creates an equal-sized and opposite-signed proportional change in price. This is the HK result: TFPR, which is the product of price and TFPQ, is invariant.

After demonstrating the specialness of the HK assumptions, we test whether these conditions hold in the data. We do so using a dataset where we—atypically for producer-level microdata—can observe businesses' quantities and prices separately. This allows us to directly test the model's key implication of price having an elasticity of -1 with respect to TFPQ. We find that, at least in our data which spans 11 different product markets, this condition does not hold in any market. Applying the HK framework to our data would therefore yield spurious measures of distortions.

Moreover, the elasticities of price with respect to TFPQ are consistently and considerably smaller in magnitude than one. This implies that 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 in our sample. This positive correlation is what researchers have typically found in other samples when the data is available to compute both TFPR and TFPQ. This suggests the elasticity of price with respect to TFPQ may be less than unit elastic in magnitude not just in our sample but also more generally.

We recognize that separate price and quantity data are typically unavailable in producer microdata, which precludes direct testing of the HK model's key implication. However, we present below another test that requires only producer-level revenue data as an output measure. This test works off another implication of the HK model: producers' TFPR levels should be invariant not just to TFPQ, but to any shifts in demand as well. Again, the intuition behind this condition is straightforward. If, as the HK framework assumes, the price-to-marginal-cost ratio is constant and marginal costs don't vary with quantity, any quantity change resulting from a shift in the firm's residual demand curve should leave price, and therefore TFPR, unchanged. We conduct this test in an even more expansive set of producer-level data on U.S. manufacturers. We reject the null of TFPR invariance to demand shocks, again indicating the HK conditions don't hold.

This second test is convenient. It uses the same revenue and inputs data that the HK methodology uses, so researchers can use it as a pretest. Before applying HK and conducting

inference on the implied misallocations, researchers can see if the most basic assumptions of the HK model hold. If they can reject the null hypothesis of invariance, false inference on misallocation patterns are likely if they apply the HK methodology. The largest measurement hurdle for conducting this test is that the researcher be able to measure producer-specific demand shifts that are likely to be orthogonal to producers' TFPQ differences. We discuss flexible and practical approaches for doing so below.

We explore what factors affect the quantitative extent of mismeasurement if the HK framework were applied in settings where its assumptions do not hold. We show it depends on multiple factors. They aren't always directly measurable in every dataset, but researcher introspection as to the specifics of the market they are studying will often offer guidance as to how extensive the spurious misallocation problem might be.

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. Readers seeking more detail are of course referred to the article.

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 η. 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, $P_i = Q_i^{-\sigma}$, and the production function $Q_i = A_i L_i^{\alpha} K_i^{1-\alpha}$.

The nonstandard elements here are the two wedges τ_{Yi} and τ_{Ki}. The former is a firm-specific scale distortion (effectively a tax on the firm's output) and τ_{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 then

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

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 TFPQ A_i 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, under its assumptions, 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 is what lets those who impose the HK framework to infer the presence and size of misallocations from observed differences in TFPR across producers.[1]

B. The Assumptions Driving HK's Result

The reason TFPR is invariant across firms in the HK model can be seen by recalling 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:

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

[1] 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 were they framing their result as being informative about misallocation. Their work points out that under their assumptions, TFPR doesn't reflect a firm's technical efficiency whatsoever, but rather only the factor prices it faces. This result, when combined with the homogeneous factor prices assumption of HK, yields the TFPR invariance condition.

Notice that the elasticity of the firm's price P_i with respect to its TFPQ level A_i is -1. 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 τ_{Yi} and τ_{Ki}.

We can dig deeper into the TFPR invariance condition by using the chain rule to expand the elasticity of price with respect to TFPQ, accounting for the fact that the firm's 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 it is not necessary for clarity):

$$\varepsilon_{P,A} = \frac{dP\big(MC(A)\big)}{dA}\frac{A}{P} = \left(\frac{dP}{dMC}\frac{dMC}{dA}\right)\frac{A}{P} = -1$$

$$\frac{dP}{dMC}\frac{MC}{P}\frac{dMC}{dA}\frac{A}{MC} = -1$$

$$\varepsilon_{P,MC}\varepsilon_{MC,A} = -1$$

Equivalently,

$$\varepsilon_{P,MC} = \frac{1}{-\varepsilon_{MC,A}}$$

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, 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. This quantity depends on both the demand and cost curves. The elasticity of the firm's marginal cost with respect to TFPQ, $\varepsilon_{MC,A}$, depends both on the direct effect that TFPQ changes have on the marginal cost curve plus any movement along the marginal cost curve that a TFPQ change would induce due to a shift in the intersection between the marginal cost and marginal demand curves.

Further to this point, the $\varepsilon_{P,A}$ = -1 condition must hold at all quantities firms might produce to obtain the HK invariance condition. For example, while $\varepsilon_{P,MC}$ = 1 may hold at a particular quantity for a variable-elasticity demand system, only firms producing this exact

quantity would conform to the assumptions of the HK model. All other industry firms would not, and the invariance of their TFPR levels to their TFPQ levels would not hold.

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 below) commonly assumed demand and production functions produce these results. The other cases where the product still happens to be -1 are even more "just-so" conditions than the unit elastic cases discussed here.

B.1. The Demand-Side Assumption

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

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 σ 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 σ.

Isoelastic demand is not just consistent with the HK framework, it is the *only* form of demand that is compatible with it.[2] 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 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 $P \geq MC$, $\varepsilon_{P,MC} \leq ½$. Thus with linear demand there are no situations under which the HK assumption hold, even approximately. Another illustrative example is the

[2] Save again for the coincidental case where a non-unitary $\varepsilon_{P,MC}$ is equal to the negative of the reciprocal of $\varepsilon_{MC,A}$ at all quantities.

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, $\varepsilon_{P,MC} \leq 1$.

These two examples both 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:

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

Or, because theory implies $\varepsilon_{P,MC} \geq 0$ and $\varepsilon_{MC,A} \leq 0$ under standard demand and cost conditions,[3]

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

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 on the issue, 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. By definition, this holds when

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

[3] 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, running up the marginal revenue and demand curves. 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 again the limit case being perfect competition. 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, though of course the second inequality is undefined in this case.

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 such as factor prices 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 in the prior section, 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 would hold 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:

$$MC(A) = \frac{\Phi(\boldsymbol{W};\theta)}{A}$$

where $\Phi(\boldsymbol{W};\theta)$ is a function of the vector of factor prices $\boldsymbol{W}$ and parameters θ. 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:

$$C(A,Q) = \int_0^Q \frac{\Phi(\boldsymbol{W};\theta)}{A} dQ = \frac{Q}{A}\Phi(\boldsymbol{W};\theta) - F$$

where F is a 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 that HK assume.[4]

However, the HK requirement that $\varepsilon_{MC,A} = -1$ will not hold without constant returns to scale. 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, but in any case the induced shift will not generally lead to a negative unit elastic response of marginal cost.

To see this in an example, suppose that the firm faces an isoelastic residual demand curve $Q = DP^{-\sigma}$ as the HK methodology assumes. Therefore its inverse marginal revenue curve is $MR = D_1 Q^{-\frac{1}{\sigma}}$, where D_1 is a constant.

Now suppose the firm's cost function has the generalized form

$$C(A, Q) = \left(\frac{Q}{A}\right)^{\frac{1}{\nu}} \Phi(\boldsymbol{W}; \theta)$$

where ν parameterizes the scale elasticity ($\nu = 1$ implies constant returns to scale in the production function). Marginal costs are then $MC = \frac{1}{\nu} Q^{\frac{1}{\nu}-1} A^{-\frac{1}{\nu}} \Phi(\boldsymbol{W}; \theta)$.

The firm's optimal quantity equates MR and MC. Equating the logs of these two values and solving for the firm's logged optimal quantity, we have:

$$q = \left(\frac{\nu\sigma}{\sigma - \nu - \nu\sigma}\right)\left[d_1 - c_1 + \frac{1}{\nu} a\right]$$

where lowercase q and a denote logs of quantity and TFPQ and d_1 and c_1 are respectively demand- and cost-side constants. Substituting this back into the expression for logged marginal costs, we have

---

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

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

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

$$mc = c_1 + \left(\frac{1}{\nu} - 1\right)\left(\frac{\nu\sigma}{\sigma - \nu - \nu\sigma}\right)\left[d_1 - c_1 + \frac{1}{\nu}a\right] - \frac{1}{\nu}a$$

Therefore the elasticity of marginal cost with respect to TFPQ is

$$\varepsilon_{MC,A} = \left(\frac{1}{\nu} - 1\right)\left(\frac{\nu\sigma}{\sigma - \nu - \nu\sigma}\right)\frac{1}{\nu} - \frac{1}{\nu} = -\frac{1}{\nu - \sigma + \nu\sigma}$$

This has the HK-required value of -1 when $\nu - \sigma + \nu\sigma = 1$. Solving for *ν*:

$$\nu = \frac{1 + \sigma}{1 + \sigma} = 1$$

Thus regardless of the slope of the residual demand and marginal revenue curves -1/*σ*, the only scale parameter that is consistent with the HK condition is *ν* = 1. Any departure from constant marginal costs, whether downward- or upward-sloping, will violate the necessary condition because the reinforcing or countervailing effect of TFPQ moving a firm along its marginal cost curve will make the elasticity of the firm's realized marginal cost with respect to its TFPQ different from -1.

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.

However, there is an additional point to our exercise here. We introduce firm-specific demand shifts, which are not in the HK model, into the framework. We show that under the assumptions of the HK model, the absence of demand shocks does not affect the key TFPR invariance implication. However, we also show that if any of the component assumptions fail, firm-specific demand shocks will create variation in TFPR even in the absence of distortions. This creates a second channel through which applying the HK condition will yield spurious distortion measures. It also suggests to us an empirical test for the HK conditions that researchers can apply to their data before applying the methodology. As we explain below, the nice thing about this pretest is that it does not require separate firm-level output quantity and price information; the same revenue data that the HK methodology uses can be used in the pretest.

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 inverse marginal revenue curve is $MR = \left(1 - \frac{1}{\sigma}\right) D^{\frac{1}{\sigma}}Q^{-\frac{1}{\sigma}}$.

Both of these 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 the demand curve at a distance $\ln(1 - 1/\sigma)$ below it. As we will see below, this parallelism is important to the HK result.

We impose the HK assumption of constant returns to scale in the production function. The corresponding cost function is

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

Marginal costs of course do not depend on output here, 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, Δa, 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, 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. Isoelastic demand, and 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. Constant returns creates the horizontal marginal cost curve. This ensures that the total effect on the firm's marginal cost at its optimal quantity, Δmc^*, is only the direct effect of the shift in the curve Δa; 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 will ensure 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 Δa. 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 necessary less than Δa 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 Δa. Again, it is the case that $\Delta p \neq -\Delta a$.

C.1. A Suggested Test for the HK Assumptions

This graphical analysis also makes transparent another implication of the HK assumptions—one which we believe suggests a feasible test for those assumptions that researchers can use before deciding to apply the framework to their data.

The core logic of the test rests on the implication that, under the joint assumptions of isoelastic demand and constant marginal costs, shifts in the firm's residual demand curve will not change its TFPR level. 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. Initially, the firm's demand and marginal revenue curves are $p(q)$ and $mr(q)$, respectively. The firm's optimal price and quantity are p^* and q^*. The inverse demand curve then shifts by Δd. This shifts out marginal revenue by Δd as well. As a result, the firm's profit-maximizing quantity rises to q'^*. However, the profit-maximizing price remains p^*. Because the firm's price doesn't change and TFPQ is not affected by the demand shift, TFPR doesn't change.

The intuition for this result is straightforward. Because isoelastic demand implies a constant multiplicative markup, if marginal cost doesn't change, price won't either. When the marginal cost curve is flat, as it is in the HK model, shifts in a firm's demand that are uncorrelated with shifts in its marginal cost curve will not change the firm's optimal price. As a result, TFPR does not move with demand. The same implication holds across firms: differences in demand that are not correlated with TFPQ differences will not create price, and therefore TFPR, variation.

To see how departures from the HK assumptions cause TFPR to be correlated with demand, consider the cases in Figure 4. Panel A shows an example of a non-isoelatic residual demand curve but constant marginal costs. A shift in the firm's residual demand by Δd 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 isoelatic.

In Panel B, demand is again isoelastic but the assumption of constant marginal cost no longer holds. Instead, the firm's marginal cost rises with its quantity. As opposed to the HK case in Figure 3, here 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. In this case, 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 suggest a test for whether the HK conditions hold in the data. If one can measure demand shifts (either across firms or within firms over time) that are orthogonal to TFPQ variations, one can test if these demand changes are correlated with TFPR levels. Rejecting the null hypothesis of no correlation would indicate that the HK assumptions do not hold.

A nice feature of this test is that, while direct tests of the HK implication that producers' prices are negative unit elastic with respect to TFPQ require separate data on prices or quantities (or assumption-laden models used to infer these values), all that it requires as an output measure are revenues. The very same TFPR measures that one would apply the HK methodology to are the object of our suggested test. In other words, if one has enough production data to be considering applying HK, one has enough to conduct our suggested test.

If there is any practical challenge in carrying out the test, it is in measuring demand shifts that are uncorrelated with TFPQ. Sometimes exogenous demand variations are readily available due to institutional or historical reasons related to the market being studied. Below we discuss and apply in our data a more general approach suggested by Shea (1993a,b): using changes in the level of activity in downstream industries as measures of demand for firms in an upstream industry of interest.

II. Testing the Assumptions of the Hsieh-Klenow Framework

We first test the core implication of the HK setup, that producer prices are negative unit elastic with respect to TFPQ levels.

To conduct this test, one needs to observe prices and TFPQ levels. 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 have collected a dataset in our earlier work (Foster, Haltiwanger, and Syverson, 2008, 2016) that includes separate quantity and price information at the level of the

individual producer. Those papers extensively detail this data, so we only very briefly review its contents here.

Our microlevel production data is an extract 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. From these product-level revenue and physical quantity data, we can also back out producers' average unit prices. 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).[5] We chose these products based in part on their physical homogeneity, which allows plants' output quantities and unit prices to be more meaningfully compared.

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

$$p_{it} = \alpha_0 + \alpha_1 tfpq_{it} + \eta_t + \varepsilon_{it}$$

where η_t is a fixed effect corresponding to the CM, which removes any shifts in prices across time that are common across all producers. Under the HK assumptions, α_1 = -1. We therefore test industry-by-industry the null hypothesis that α_1 = -1.

In addition to these industry-specific tests, we estimate a pooled specification on the combined dataset. Here the specification is the same, except rather than just having CM fixed effects we include industry-CM fixed effects, so all identification of the relationship between price and TFPQ still comes from within-industry-year variation. Of course in this case we are imposing a common value of α_1 across all industries.

RESULTS PENDING CENSUS REVIEW.

III. Discussion of Results

TBD

[5] 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).

IV. Conclusion

TBD

References

De Loecker, Jan and Frederic Warzynski. 2012. "Markups and Firm-level Export Status." *American Economic Review*, 102(6), 2437-71.

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.

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.

Shea, John. 1993a. "The Input-Output Approach to Instrument Selection." *Journal of Business & Economic Statistics*, 11(2): 145-55.

Shea, John. 1993b. "Do Supply Curves Slope Up?" *Quarterly Journal of Economics*, 108(1): 1-32.

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.

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

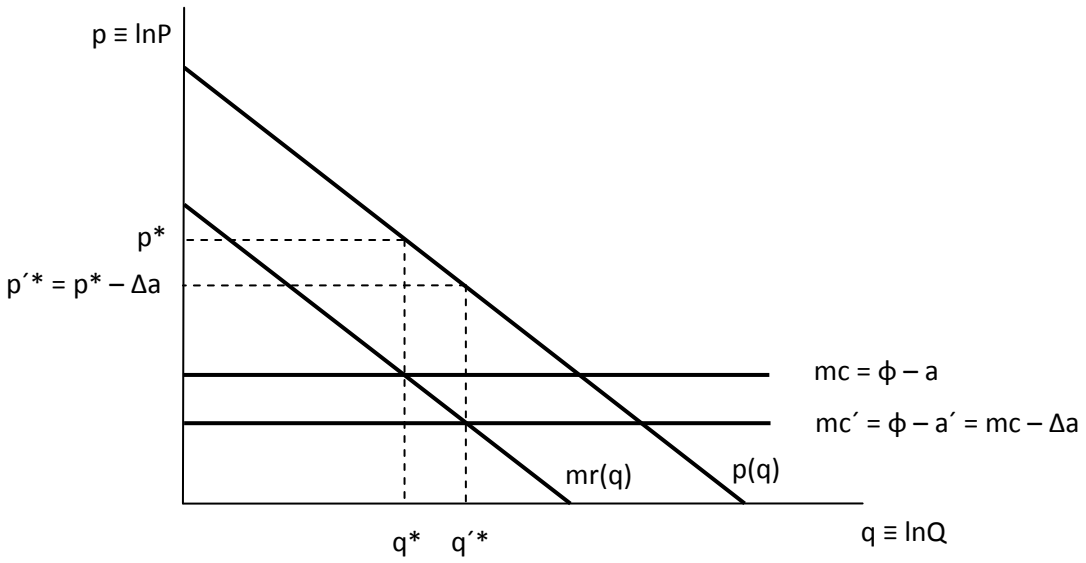


Figure 2. Effect of a Change in TFPQ when the Marginal Cost Curve Is Not Horizontal

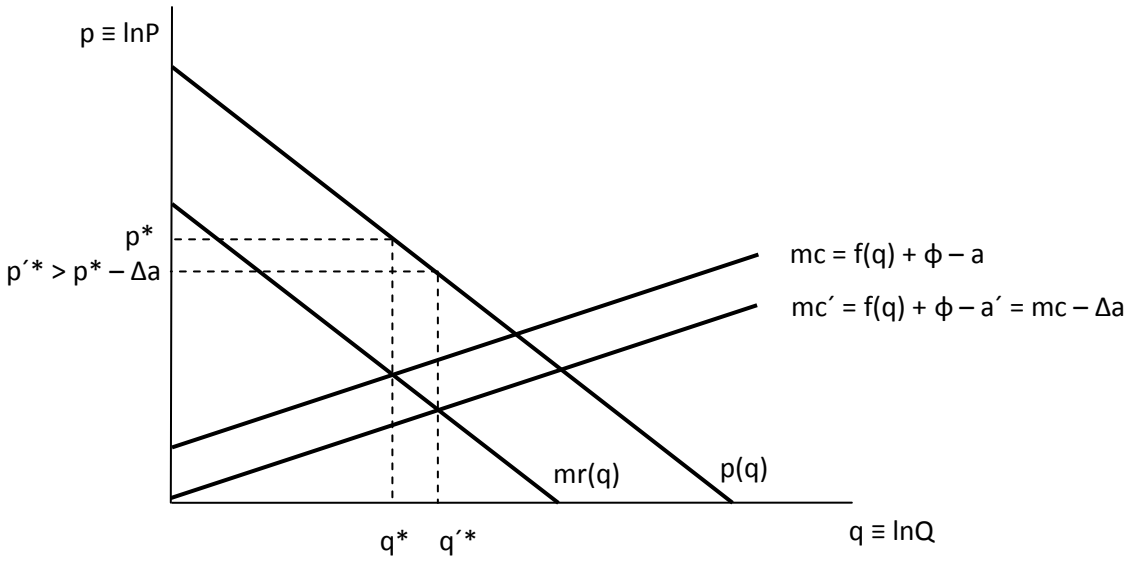


Figure 3. Demand Shifts Do Not Change TFPR in HK

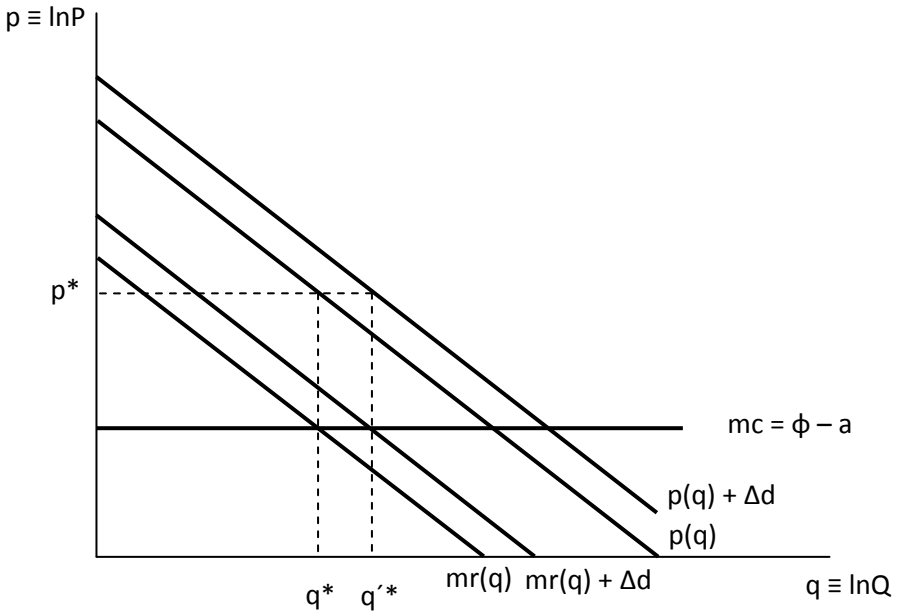
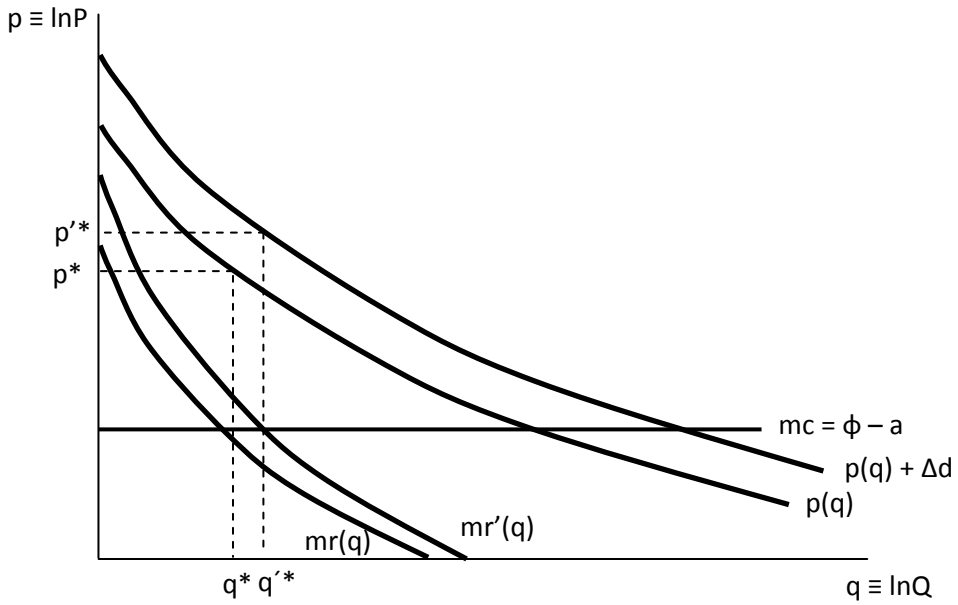


Figure 4. Demand Shifts Change TFPR If HK Assumptions Do Not Hold

A. Non-Isoelastic Demand but Constant Marginal Costs



B. Isoelastic Demand but Non-Constant Marginal Costs

