JOHN HALTIWANGER
University of Maryland

Firm Dynamics and Productivity: TFPQ, TFPR, and Demand-Side Factors

ABSTRACT Two common findings in the firm dynamics literature are that there is large dispersion across firms in productivity within narrowly defined industries and that firms that are high in the within-industry distribution are more likely to survive and grow. These findings underlie a rich class of models relating the level and growth of aggregate (industry-level) productivity to the reallocation of resources away from less productive to more productive firms. While these findings are common, there are a variety of empirical measures of firm-level total factor productivity that have been used in the literature to generate these findings. These include measures that are closer to the concepts of technical efficiency common in many models to measures that encompass demand-side factors as well. In addition, the recent literature has developed methods to extract measures of distortions from specific measures of dispersion in productivity given assumptions about the production and demand functions in the economy. In this paper, I discuss the relationship between the alternative measures that have been proposed and used in the literature and, in turn, the implications of these relationships for our understanding of observed firm dynamics.

JEL classifications: D61, E23, O47

Keywords: productivity; growth; allocative efficiency

The development of rich longitudinal business databases for many countries in the last few decades has generated new core facts about the joint distributions of firm-level productivity, firm size, and the pace of the reallocation of outputs and inputs across firms.[1] First, there is tremendous dispersion in productivity across firms in the same industry. Second, there is tremendous dispersion and skewness in the size distribution of firms across firms in the same industry. Third, there is a high pace of the reallocation of outputs and inputs across firms within industries. Fourth, firm entry and exit contribute

ACKNOWLEDGMENTS This paper draws heavily on joint work with Ryan Decker, Marcela Eslava, Lucia Foster, Cheryl Grim, Ron Jarmin, Javier Miranda, Chad Syverson, and Zoltan Wolf. The author thanks Marcela Eslava, Lucia Foster, and Chad Syverson for providing comments on an early draft.

1. See Syverson (2011) for a recent survey.

substantially to the pace of reallocation. The recent evidence suggests that these patterns hold widely in both advanced and emerging economies.

Economic theory and accumulating empirical evidence show that these facts are related in ways that vary systematically across countries. Allocative efficiency implies that the more productive firms should be larger or becoming larger, while less productive firms should be smaller or becoming smaller. These patterns relating size and growth to productivity seem to hold to a greater extent in advanced economies than in emerging economies. This has led to the working hypothesis that the large observed differences in gross domestic product (GDP) per capita across countries are driven by distortions to these allocation dynamics within countries. This has, in turn, yielded a growing literature that seeks to identify the distortions generating misallocation.[2] Progress has been made, but there is still much to be done to identify the most relevant distortions in specific countries, industries, and time periods.

One complicating factor in terms of both theory and evidence is that the observed dispersion in productivity across firms is not time invariant. Idiosyncratic productivity differences are persistent, but not permanent.[3] This has led to a richer characterization of firm dynamics: in well-functioning economies, firms with positive idiosyncratic shocks will endogenously innovate successfully or successfully adapt to a changing economic environment, resulting in the growth of the firm, while firms with adverse shocks will fail to innovate or adapt and thus will shrink and exit. This implies that the nature and consequences of distortions may be affecting the dynamic aspects of the reallocation process. Differences across countries may show up in the extent to which businesses with positive innovations in productivity grow, while businesses with negative innovations in productivity contract.

While there are many open questions in this expanding literature, I focus on two related areas of inquiry that reflect the interaction between economic theory and economic measurement. The first area of inquiry aims to identify the most appropriate measure of firm performance when studying the evolution of firms. The second explores the role of demand-side factors in firm dynamics. An understanding of these two issues is critical to keep important concepts from becoming muddled.

The first area of inquiry stems from the emphasis in the literature on firm-level productivity as the measure of interest. However, most firm-level

2. See, for example, Restuccia and Rogerson (2008); Hsieh and Klenow (2009, 2014); Bartelsman, Haltiwanger, and Scarpetta (2013).

3. See Foster, Haltiwanger, and Syverson (2008); Decker and others (2015).

databases do not permit measuring firm-level prices, so the most commonly used empirical measures of productivity are revenue-based measures.[4] The key measurement issue is that firm-level real output is measured as firm nominal revenue divided by an industry-level price deflator. This implies that firm-specific price variation is captured in measures of firm-level productivity. Thus, the typical empirical measure used differs from the theoretical concept of productivity, which reflects output per unit of composite input, taking into account the production technology relating inputs to output. This measurement issue has potentially important theoretical implications: since firm-level prices are likely endogenous, these alternative measures may have quite different properties.

Foster, Haltiwanger, and Syverson provide a structure to think about these issues by distinguishing between what they denote as TFPQ and TFPR.[5] The former is the measure of technical efficiency that emerges from the production function; the latter they define as the firm-level price times TFPQ. The TFPR measure has, as they show, natural appeal conceptually since it is equivalent to revenue per unit of composite input when there are constant returns to scale. Using this distinction, Foster, Haltiwanger, and Syverson show that if firms face linear demand schedules, then TFPR will reflect a combination of demand, cost, and productivity (TFPQ) factors.[6] They use this structure (along with direct information on firm-level prices and quantities) to draw out the respective role of these factors for market selection.

Hsieh and Klenow highlight this distinction between TFPQ and TFPR so defined.[7] Under specific assumptions about the functional forms of the demand and production functions, they show there will be zero dispersion in TFPR in an economy without distortions even if there is dispersion in TFPQ.[8] They argue accordingly that observed TFPR dispersion must reflect distortions. Hsieh and Klenow use this powerful identifying assumption to explore

4. In practice, the issue is often that establishment-level prices are not observed. In much of the discussion in the paper, the distinction between firms and establishments is ignored for expositional convenience. In empirical analysis, this distinction is important. There are establishment-level studies where establishment-level prices are observed, including the analyses in Foster, Haltiwanger, and Syverson (2008), Foster and others (2015), and Eslava and others (2004, 2013). This evidence is used to guide the discussion of the TFPR versus TFPQ debate in the literature.

5. Foster, Haltiwanger, and Syverson (2008).

6. Foster, Haltiwanger, and Syverson (2008).

7. Hsieh and Klenow (2009, 2014).

8. Their assumptions are quite different from the assumptions made by Foster, Haltiwanger, and Syverson (2008).

cross-country differences in observed differences in the dispersion of TFPR, as well as cross-country differences in the relationship between TFPR, firm size, and firm age. Their provocative work highlights one potentially important complication of distinguishing between TFPQ and TFPR. This paper explores this distinction further, drawing out how this implication is driven by specific assumptions about the production and demand functions.

Many revenue productivity measures considered in the literature are related to the conceptual TFPR measure (price times TFPQ), but they are not equivalent to it. Specifically, given consistent estimation, which itself is a challenge, estimation methods for the revenue function yield a residual that typically reflects both demand and TFPQ. This sounds like Foster, Haltiwanger, and Syverson's TFPR as described above, but in fact the revenue residual is typically not equivalent to price times TFPQ.[9] Most studies that estimate the revenue function do not explicitly recognize that the estimated coefficients on the inputs in the revenue function are not estimates of factor elasticities, but rather estimates of some combination of factor elasticities and parameters of the demand structure. Additional structure and data are needed to disentangle the estimates into factor elasticities and demand parameters. This subtle issue is important since it implies that even if Hsieh and Klenow's strong assumptions hold, it is not the case that revenue function residuals should exhibit no dispersion in the absence of distortions.[10] Put differently, revenue function residuals should, in general, only reflect market fundamentals such as demand shocks and TFPQ.

The second area of inquiry is a consequence of attempts to address the challenges of the distinction between TFPR and TFPQ or, more generally, between technical efficiency and revenue productivity measures. In exploring the empirical relevance and implications of using TFPR measures of productivity, a number of studies take advantage of the limited number of firm-level databases in which firm-level prices are available.[11] Firm-level prices make it possible to decompose revenue into the price and quantity components, which in turn makes it possible to construct TFPQ measures of productivity. Moreover, having price and quantity variation supports the investigation of the role of demand-side factors. Evidence is emerging that the differences in firm size and firm growth dynamics are much more driven by demand-side differences

9. Foster, Haltiwanger, and Syverson (2008).

10. Hsieh and Klenow (2009, 2014).

11. See Foster, Haltiwanger, and Syverson (2008); Foster and others (2015); Eslava and others (2004, 2013).

than by TFPQ differences. The latter are important, but demand differences are more important. This has important implications since the frictions and distortions involving demand-side factors are potentially different from those involving cost and supply-side factors, such as TFPQ.

This paper proceeds as follows. The next section discusses the distinction between TFPQ, TFPR, and revenue function residuals using a simple illustrative framework. Following this discussion, the basic facts from the firm-level productivity literature are reviewed in light of this perspective. The potential role of demand shocks is then discussed, and the final section presents concluding remarks.

Understanding Alternative Measures of Productivity: TFPQ versus TFPR

To establish the conceptual distinction between TFPR and TFPQ, I use a very simple model that is motivated by the insights of Foster, Haltiwanger, and Syverson; Hsieh and Klenow; and the related literature.[12] Consider a firm that faces the following production and demand functions:

(1)
$$Y_{it} = A_{it} x_{it}^{\gamma};$$
$$P_{it} = D_{it} Y_{it}^{-\epsilon};$$

where Y_{it} is output for firm i at time t, x_{it} is the composite input for firm i at time t, P_{it} is the price of the output of firm i at time t, A_{it} is the productivity of firm i at time t, D_{it} is the demand or quality of firm i at time t, γ is the returns to scale, and ϵ is the inverse of the elasticity of demand (which is assumed for now to be the same for all firms). In this simple model, conceptually TFPQ = A_{it} and TFPR = $P_{it}A_{it}$.[13] Given the model, TFPR = $D_{it}A_{it}^{1-\epsilon}x_{it}^{-\gamma\epsilon}$. How close empirical measures come to these conceptual measures depends critically on assumptions. In this simple framework, TFPQ reflects technology factors only, while TFPR likely reflects both technology and demand factors. It may reflect those factors in complex ways through endogenous prices, which raises the question as to which measure is preferable as a measure of firm performance. For the most part, I do not take a stand on

12. Foster, Haltiwanger, and Syverson (2008); Hsieh and Klenow (2009, 2014).

13. This is the definition of TFPR from Foster, Haltiwanger, and Syverson (2008), which explicitly defines this as price times TFPQ.

which measure is preferable, but rather discuss the properties of the alternative conceptual measures and how empirical measures correspond to these conceptual measures.

A few background remarks are useful here. First, the discussion proceeds initially under the assumption that firms in the same industry face the same production and demand parameters. The implications of firm heterogeneity on these dimensions are discussed below. Second, the simple illustrative framework highlights the key parameters affecting the curvature of the profit function (that is, the demand elasticity and the returns to scale). Underlying the returns to scale are the factor elasticities for specific factors in a multi-factor environment. The discussion here largely abstracts from multi-factor specifications to make the discussion more transparent, but these issues are important in practice.[14] Third, many of the issues discussed here are moot if the law of one price holds within an industry, since then TFPQ and TFPR are identical. However, even for commodity-like products, there is price heterogeneity and evidence of market power within the industry. This motivates much of the recent literature and the discussion in this paper.

A key issue is what is observed in the firm-level data. It is relatively rare to measure output (Y_{it}) directly, so most empirical researchers start with the firm's revenue ($P_{it}Y_{it}$). Firm-level prices are typically not available, so in practice this is nominal revenue for the firm deflated by an industry deflator.[15] In this respect, P_{it} is the relative price of firm i vis-à-vis the average price in the industry. Using the above model, revenue is a measure of $D_{it}A_{it}^{1-\epsilon}\, x_{it}^{\gamma(1-\epsilon)}$.

It is also critical to consider what methods researchers use to measure productivity empirically. In the current framework, it is useful to define the concept of revenue per unit of composite input (that is, the ratio of revenue to the composite input as given by $P_{it}Y_{it}/x_{it}$). This measure is denoted here as

14. Foster and others (2015, 2016) provide an extensive discussion of these issues. In the latter paper, some of the same issues are discussed in terms of alternative empirical estimates of revenue productivity.

15. There is an analogous issue of not measuring firm-specific prices for inputs. The difference is that it is more reasonable to assume that firms do not have market power in input markets. A more reasonable assumption is some of the differences in input prices that firms face reflect differences in input quality. In that respect, this is seemingly less of a problem, since empirical measures have built-in controls for input quality. However, differences in input quality are likely related to differences in output quality, so it is very important to take the covariance between output and input quality into account. Recent work by de Loecker and others (2016) explores these issues in the context of having data on both firm-level output and input prices.

RPI. Using the above model, RPI = $D_{it}A_{it}^{1-\epsilon}x_{it}^{\gamma(1-\epsilon)-1}$. Under constant returns to scale (CRS), RPI is equivalent to the conceptual measure of TFPR. It is this property that has led many researchers to use growth-accounting methods to directly measure revenue per unit of input. For example, under the assumption of cost minimization and CRS, the composite input can be constructed using individual inputs and factor elasticities, estimated as factor costs as a share of total factor costs.

Alternatively, researchers may be estimating the revenue function and recovering revenue function residuals. The revenue function residual is denoted as REVRES. If consistent estimates of the parameters of the revenue function are obtained, then REVRES = $D_{it}A_{it}^{1-\epsilon}$. REVRES will typically not be equal to RPI and will reflect both demand shocks and TFPQ. REVRES and RPI are equivalent if $\gamma = 1$ (CRS) and $\epsilon = 0$ (firms have no market power).

Evaluating the measures that are recovered empirically depends on the assumptions about the endogenous determination of prices and inputs. A useful starting (but not ending) point is to assume that there are no frictions or distortions, so that prices and inputs are determined by simple static profit maximization. Let static profits be given by

$$\text{(2)} \qquad \pi_{it} = P_{it}Y_{it} - c_{it}x_{it},$$

where c_{it} is the cost of the composite input for firm i at time t. Idiosyncratic differences in input costs are considered as they are potentially important for the distinction between TFPR, TFPQ, RPI, and REVRES.

The optimal composite input is given by

$$\text{(3)} \qquad x_{it} = \left[\frac{\gamma(1-\epsilon)D_{it}A_{it}^{1-\epsilon}}{c_{it}}\right]^{1/[1-\gamma(1-\epsilon)]}.$$

Thus, the composite input is increasing in productivity and demand and decreasing in costs. Under these assumptions, TFPR is given by

$$\text{(4)} \qquad \text{TFPR} = P_{it}A_{it} = \frac{c_{it}}{\gamma(1-\epsilon)}x_{it}^{1-\gamma}$$

$$= \frac{c_{it}}{\gamma(1-\epsilon)}\left[\frac{\gamma(1-\epsilon)D_{it}A_{it}^{1-\epsilon}}{c_{it}}\right]^{(1-\gamma)/[1-\gamma(1-\varepsilon)]}.$$

Revenue per unit of composite input will, in this case, be given by

$$\text{(4}'\text{)} \qquad \text{RPI} = \frac{P_{it}Y_{it}}{x_{it}} = \frac{c_{it}}{\gamma(1-\epsilon)}.$$

In the absence of CRS, equations 4 and 4′ show that RPI is not equal to TFPR. However, with constant returns to scale, TFPR and RPI become identical. With the additional assumption that all firms face the same input costs ($c_{it} = c_t$), TFPR and revenue per unit are equalized and given by

$$\text{(5)} \qquad \text{TFPR} = P_{it}A_{it} = \text{RPI} = \frac{P_{it}Y_{it}}{x_{it}} = \frac{c_t}{(1-\epsilon)}.$$

Equation 5 is the starting point of Hsieh and Klenow.[16] Under these assumptions, TFPR and RPI is the same across all firms in the same industry. Hsieh and Klenow account for any observed dispersion in TFPR by adding idiosyncratic distortions to the above profit maximization. Specifically, consider a profit function given by

$$\text{(6)} \qquad \pi_{it} = (1-\tau_{it})P_{it}Y_{it} - c_{it}x_{it},$$

where τ_{it} is the idiosyncratic distortion for firm i at time t.[17] If the same assumptions (CRS and all firms face the same input costs) hold, then equation 5 becomes

$$\text{(5}'\text{)} \qquad \text{TFPR} = P_{it}A_{it} = \text{RPI} = \frac{P_{it}Y_{it}}{x_{it}} = \frac{c_t(1-\tau_{it})}{(1-\epsilon)}.$$

In this case, dispersion in TFPR and RPI reflects dispersion in idiosyncratic distortions. In addition, the inference that emerges from these assumptions

16. Hsieh and Klenow (2009, 2014).

17. Consistent with Hsieh and Klenow (2009), I take the view here that distortions do not enter the production and demand functions directly, but rather affect the profit function. There could be distortions that affect the innovation dynamics of firms (and thus the production function), as well as distortions that affect the accumulation of the customer base (and thus the demand function).

is that plants with a high TFPR and RPI are not inherently high-productivity plants, but rather high-distortion plants.

Hsieh and Klenow's interpretation of TFPR reflecting distortions requires strong assumptions that may not be warranted. First, firms may face different input costs. Second, returns to scale may not be equal to one, so there is a wedge between the conceptual TFPR and RPI. Third, it may be more appropriate to interpret distortions as reflecting frictions that are present even in well-functioning economies. With factor-adjustment frictions, firms with positive realizations of fundamentals take time to adjust factors, implying that they will have high measured RPI.[18] Fourth, heterogeneous demand elasticities, idiosyncratic costs, and idiosyncratic returns to scale (or, more generally, factor elasticities) all drive wedges in these measures across firms even without distortions.

Alternative functional forms for production and demand functions also yield dispersion in TFPR and RPI without distortions. Consider, for example, a modified production function that incorporates the potential role of overhead input requirements.[19] Specifically, suppose the production function is given by

(7) $$Y_{it} = A_{it}\left(x_{it} - f\right)^{\gamma},$$

where f is the fixed input requirements (which could be fixed overhead capital or overhead labor in a more general model). In this case, the optimal choice of inputs is given by

(8) $$x_{it} = f + \left[\frac{\gamma(1-\epsilon)D_{it}A_{it}^{1-\epsilon}}{c_{it}}\right]^{1/(1-\gamma(1-\epsilon))}.$$

RPI, under CRS and common input costs, is then given by

(9) $$\text{RPI} = \frac{P_{it}Y_{it}}{x_{it}} = \frac{c_t}{(1-\epsilon)}\left(1 - \frac{f}{x_{it}}\right).$$

18. Asker, Collard-Wexler, and de Loecker (2014) develop a rich framework to show the importance of adjustment frictions in accounting for the observed patterns of TFPR.

19. This is the case considered by Bartelsman, Scarpetta, and Haltiwanger (2013).

Since x_{it} is increasing in A_{it} and D_{it}, this implies that RPI will be increasing in both demand and productivity shocks. TFPR, under CRS and common input costs, is given by

$$\text{TFPR} = \frac{c_t}{(1 - \epsilon)}. \tag{10}$$

TFPR is equalized across firms in this case because marginal revenue products are still equalized with this production function and the assumption of common input costs and common demand elasticities. However, there is a wedge between average and marginal revenue products, so RPI is not equal to TFPR (even with CRS) and is not equalized across firms. TFPR is not straightforward to estimate in this case, given the presence of the fixed overhead costs.

Another alternative model departs from the iso-elastic demand model. Melitz and Ottaviano develop a framework in which quadratic preferences yield linear demand schedules for firms.[20] Foster, Haltiwanger, and Syverson use a related model to develop a framework to think about the distinction between TFPQ and TFPR.[21] Consider a simplified version where the inverse demand curve is given by

$$P_{it} = \alpha - \beta Y_{it} + D_{it}. \tag{11}$$

Assuming CRS so that output is given by $Y_{it} = A_{it} x_{it}$, optimal x_{it} is given by

$$x_{it} = \frac{(\alpha + D_{it})A_{it} - c_{it}}{2\beta A_{it}^2}. \tag{12}$$

This implies that both TFPR and RPI will be given by

$$\text{TFPR} = P_{it}A_{it} = \text{RPI} = P_{it}Y_{it}/x_{it} = \frac{(\alpha + D_{it})A_{it} - c_{it}}{2}. \tag{13}$$

With this alternative demand structure and constant returns to scale, TFPR and revenue per unit of input will be increasing in demand and productivity

20. Melitz and Ottaviano (2008).
21. Foster, Haltiwanger, and Syverson (2008).

and decreasing in costs. Each firm is equalizing marginal revenue to marginal cost, but the demand function implies endogenously different markups across firms. It is this feature that yields dispersion in TFPR and RPI across firms, in relation to demand shocks and TFPQ.

The discussion thus far highlights why TFPR and RPI likely exhibit dispersion and are correlated with TFPQ even in the absence of distortions. A related and important issue for interpreting the empirical findings in the literature is that estimation of the revenue function residuals does not yield a measure that is equal to either TFPR or RPI. Recall that using the simple framework set out at the beginning of this section and making the Hsieh-Klenow assumptions, the estimation of the revenue function yields a residual equal to REVRES $= D_{it}A_{it}^{1-\epsilon}$. Here, REVRES is not equalized across firms, but reflects demand and TFPQ variation. This point holds more generally for any revenue function estimation. If the parameters are estimated consistently, then the revenue residual will generally be a function of demand and technology shocks. Thus, researchers who estimate revenue productivity residuals should not presume that such residuals reflect distortions even if they make the same assumptions as Hsieh and Klenow.[22] Instead, such residuals will reflect only market fundamentals.

A key point here is that in the estimation of the revenue function, the estimated coefficients reflect both factor elasticities and demand parameters. As such, the residual is not equal to $P_{it}A_{it}$.[23] It is helpful to return to the setting underlying equation 5 and compare it to the residual. Under CRS, the residual from the revenue function estimation is given by

$$\text{REVRES} = \frac{P_{it}Y_{it}}{x_{it}^{(1-\epsilon)}} = D_{it}A_{it}^{1-\epsilon}. \tag{14}$$

The difference between equations 5 and 14 is apparent. TFPR is equal to RPI under CRS, and that is equalized with common costs. The reason it is equalized is that price decreases one-for-one with TFPQ. The revenue residual does not have this property, however, and it will not be equalized across firms. The equivalent of REVRES can be constructed directly using the cost share approach underlying RPI, along with estimates of γ and ϵ.

22. Hsieh and Klenow (2009, 2014).

23. Given the simple representation with a composite input, this discussion neglects the estimation of the factor elasticities and the demand parameters. But these arguments readily generalize to the multiple-input case.

Taking stock, under some strong assumptions (namely, CRS, iso-elastic demand, no overhead input costs, common input costs, common demand, and production parameters), TFPR is equal to revenue per unit of input. Moreover, in the absence of frictions and distortions, TFPR and RPI should be equalized across firms. This implies that under these conditions, measured dispersion in TFPR and RPI across firms in the same industry reflects frictions and distortions. Departures from any of these assumptions yields TFPR dispersion in equilibrium, which sometimes will and sometimes will not be well captured by revenue per unit of input. Regardless of these assumptions, estimated revenue function residuals will typically reflect both demand and productivity factors and exhibit dispersion across firms. Only in special cases will the estimated residual of the revenue function be equivalent to revenue per unit of input.

Many, but not all, of these complications are eliminated if firm-level prices of outputs and inputs are directly observed, as well as revenue and expenditures on inputs. With direct measures of Y_{it}, x_{it}, and P_{it}, TFPQ and TFPR are measured directly with accounting under CRS. One first measures $\text{TFPQ} = A_{it}$ and then constructs $\text{TFPR} = P_{it}A_{it}$. Estimation of the production function residual should yield the same TFPQ. This estimation might be needed to estimate the composite input (that is, to estimate the factor elasticities needed to construct the composite input). Of course, growth accounting under the assumptions that first-order conditions with cost minimization hold at least in the long run and on average across firms can yield factor elasticities using cost shares of total costs. In this respect, having firm-level prices and quantities avoids many of the issues discussed above. Even with departures from CRS, with some method to estimate returns to scale, measurement of TFPQ and TFPR is straightforward when prices and quantities are measured directly.

A Brief Review of Basic Facts

The previous section is full of subtleties, including a knife-edge case where RPI is equivalent to TFPR, which is different from REVRES, which in turn is different from TFPQ. However, in practice, these alternative measures (that is, RPI, REVRES, TFPR, and TFPQ) have been shown to be closely related when some or all of these measures can be constructed. These basic facts are reviewed here, followed by a discussion of how to think about the subtleties of the previous section in light of the evidence. In reviewing the basic facts, it is useful to consider two different types of empirical studies. First, in studies where firm-level prices are not observed, measures of RPI and REVRES can be readily estimated. Second, in studies where firm-level prices are observed, measures of TFPQ can be directly constructed and then combined with prices to construct TFPR. Moreover, when firm-level prices are observed, it is possible to construct measures of RPI and REVRES as well.

Foster and others provide a useful starting point when prices are not observed.[24] They show that RPI and REVRES measures are highly correlated empirically. Using U.S. manufacturing plant-level data, they find that RPI measures (assuming CRS and using cost shares to estimate the composite input) have a correlation of about 0.7 with REVRES estimated via the Levinsohn-Petrin method.[25] Moreover, the dispersion in these alternative revenue productivity measures is large and quite similar in magnitude. They also find that the relationship between firm-level growth and survival and these measures of firm performance is quite similar both qualitatively and even quantitatively using these alternative measures. Thus, Foster and others produce results consistent with Syverson's survey.[26] The latter emphasizes two key findings in the firm-level productivity literature. First, many studies find large dispersion in measured productivity within industries. Second, many studies find that firms with high measured productivity are more likely to grow and less likely to exit. These findings hold across a wide variety of methods and data sets. They hold across methods that measure productivity as revenue per unit of input and as the revenue function residual.

The basic facts discussed above are all about revenue productivity measures and not about where TFPQ fits in. For the latter, it is useful to turn to studies that measure TFPQ directly and thus also measure TFPR more directly. Foster, Haltiwanger, and Syverson show that for selected products in the United States, TFPQ measures of productivity exhibit a high level of within-industry dispersion that exceeds the within-industry dispersion in TFPR.[27] Using Colombian data, Eslava and others similarly find high levels of within-industry dispersion in both TFPQ and TFPR, with the dispersion in TFPQ exceeding that of TFPR.[28] These studies that include both TFPR and TFPQ also find a high correlation between the two. Foster, Haltiwanger, and

24. Foster and others (2015).

25. Levinsohn and Petrin (2003). Eslava and Haltiwanger (2016) report that correlations between TFPQ, TFPR, and a revenue residual are quite high in Colombia.

26. Foster and others (2015); Syverson (2011).

27. Foster, Haltiwanger, and Syverson (2008, 2016). Recall that when TFPQ is measured directly, measuring TFPR is straightforward.

28. Eslava and others (2004, 2013).

Syverson and Foster and others find a correlation of 0.75, while Eslava and others find a correlation of 0.69.[29]

Taken together, these findings raise a variety of questions and issues about the knife-edge case considered by Hsieh and Klenow.[30] Under the assumptions in the latter paper, RPI should equal TFPR and should reflect only distortions. In the United States, however, RPI is highly correlated with empirical measures of REVRES, which will capture both demand and TFPQ variation. Moreover, TFPR is highly correlated with TFPQ when the latter can be directly measured (so that TFPR can also be directly measured). Putting these pieces together, to treat RPI as a proxy for TFPR and a measure of distortions requires making some assumptions about correlations across these measures. In particular, to reconcile the empirical patterns described above, distortions must be assumed to be highly correlated with technical efficiency and demand shocks, and dispersion would have to be about the same as technical efficiency and demand shocks combined. Rather than make these strong assumptions about correlated distortions, a more reasonable interpretation of these patterns might be that adjustment frictions or other factors are causing TPFR and RPI to exhibit dispersion and correlation with fundamentals without distortions.

Further comparisons of the properties of TFPQ and TFPR when prices and quantities are directly measured raise further doubts about interpreting empirical estimates of TFPR dispersion as reflecting only distortions. In the United States, the standard deviation of TFPR is about 0.2 for manufacturing plants within industries. In Colombia, the standard deviation is about 0.7 for manufacturing plants within industries. At first glance this might be interpreted as consistent with the distortions view of TFPR, since Colombia presumably has a much higher dispersion of idiosyncratic distortions than the

29. Foster, Haltiwanger, and Syverson (2008); Foster and others (2015); Eslava and others (2004, 2013). Foster, Haltiwanger, and Syverson (2008) construct their TFPR measure so that TFPR = P*TFPQ. They also consider a traditional revenue productivity measure using output measured as total revenue deflated by an industry-level deflator. For their TFPT measure, they use the same cost-share-based factor elasticities as for TFPQ and TFPR. They find that TFPT has a high correlation with TFPR (about 0.86). For TFPQ, Eslava and others (2004, 2013) use total revenue of the plant deflated by a plant-level deflator. Their TFPR measure uses the same revenue but deflates by an industry-level deflator. Their TFPQ and TFPR measures use the same estimated factor elasticities from estimating the production function (using the measure of output with plant-level prices). A small point is that their TFPQ measure uses materials inputs as materials expenditures deflated with plant-level materials prices, while their TFPR measure uses materials expenditures deflated with an industry-level deflator.

30. Hsieh and Klenow (2009).

United States. As discussed earlier, however, the evidence shows that TFPQ dispersion in the United States is larger but about the same as TFPR dispersion, with both being highly correlated (the standard deviation of TFPQ is 0.26, versus 0.22 for TFPR). Similarly, TFPQ dispersion is slightly higher than TFPR dispersion in Colombia (both about 0.7). This is consistent with the view that many factors make TFPQ and TFPR closely related. Moreover, if distortions are much higher in Colombia than in the United States (as might be presumed), then it is an open question how to reconcile the small gaps between TFPQ and TFPR within both Colombia and the United States.

Another reason to be skeptical is that the rank ordering of countries in terms of TFPR dispersion seems inconsistent with likely rank ordering of countries by distortions. Hsieh and Klenow find that China and India have higher TFPR dispersion than the United States, and the *World Development Report* indicates that for selected countries, TFPR dispersion in emerging economies is higher than in the United States.[31] However, studies that look at a broader range of countries raise doubts. Bartelsman, Haltiwanger, and Scarpetta report TFPR dispersion for eight countries including the United States and countries in western and eastern Europe.[32] The United States is about in the middle of the range, with countries like Slovenia having much lower TFPR dispersion. The authors also document the change in TFPR over the course of the 1990s. The transition economies exhibit only a very modest decline in TFPR dispersion during a decade with substantial market reforms and presumably lower distortions.

I am not implying here that idiosyncratic distortions are not present or are not important for accounting for variation in economic performance across and within countries. Such distortions are likely important for both the dispersion of TFPR and the covariance between size and productivity in countries. Distortions will affect the relationship between TFPQ and input and output size. In many of the cases considered above, they will also affect the relationships between TFPR, its proxies like RPI, and input and output growth. Put more broadly, both distortions and frictions will influence the observed distribution of TFPR and RPI. Frictions and distortions will have an impact not only on the distribution of TFPR and RPI, but also on the observed distribution of TFPQ if entry and exit are important, to the extent that they influence

31. Hsieh and Klenow (2009); World Bank (2013).

32. Bartelsman, Haltiwanger, and Scarpetta (2013). Because they use cost shares for factor elasticities, they are, in principle, using factor elasticities of the production function and not revenue elasticities.

the pace and nature of entry and exit. In turn, such selection will affect the observed distribution of TFPQ (if it can be directly measured).[33]

One inference from this discussion is that insights and progress on these issues will be more straightforward when both prices and quantities are observed. However, this also does not mean that researchers should inherently prefer using measures of TFPQ over TFPR. TFPR measures have the virtue that they will reflect idiosyncratic profitability factors beyond TFPQ. That is, in general TFPR measures will reflect both demand and productivity factors, as will revenue residuals. Since both are likely important for firm dynamics, it is useful to capture both, but the preferred approach is to measure TFPQ and demand factors (and cost factors) separately. One reason is that TFPR will reflect endogenous prices in potentially complex ways. A second reason is that frictions and distortions have a differential impact on demand-side versus cost-side factors. The importance of capturing both demand and productivity factors is discussed in more detail in the next section.

Another inference that emerges from this discussion is that if the objective is to obtain a revenue productivity measure that is only a function of market fundamentals, then REVRES measures potentially have great appeal.[34] The challenge here is to estimate the parameters of the revenue function consistently.[35] Much of the recent literature focuses on proxy methods, but other approaches (such as panel estimation methods) may be appropriate in this context as well.[36]

Distinguishing between TFPQ and Demand-Side Factors

If a conclusion from the prior section is that empirical measures of revenue productivity (either RPI or REVRES) are likely to reflect both TFPQ and demand-side factors, the question is whether it is critical to disentangle the different components. As noted, many of the empirical regularities found in the literature on the relationship between productivity, growth, and survival are based on relating revenue productivity measures to firm growth

33. Eslava and others (2013, 2015) provide direct evidence of the impact of changes in distortions on all of the moments discussed in this paragraph.

34. This inference relies on distortions not directly entering the production and demand functions.

35. See the discussion in Wooldridge (2009).

36. One recent example is Cooper and Haltiwanger (2006), who estimate the short-term profit function using a quasi-differenced approach. That same method can be used for the revenue function as well.

and survival. More generally, the researcher's interest often lies in the joint evolution of the determinants of firm profitability and economic outcomes, so revenue productivity measures may be preferred since they capture more of the determinants of firm profitability.

However, the TFPR and RPI measures arguably also reflect endogenous components of profitability through endogenous prices.[37] They may be a proxy for the many factors discussed above, but the nature of the endogeneity of those factors may differ across components. Likewise, the frictions associated with these endogenous factors may differ substantially. This should be a rich area for future research, and this section discusses some possibilities based on recent research. The discussion is not intended to be exhaustive, but rather suggestive of the open questions in this area.

Foster, Haltiwanger, and Syverson find that much of the evolution of the size distribution by plant age reflects demand-side factors rather than TFPQ.[38] A well-known finding in the literature is that young businesses are small, and it takes time for young businesses to achieve their steady-state size.[39] Foster, Haltiwanger, and Syverson show that the slow growth of young plants is driven by demand-side factors rather than TFPQ.[40] Figure 1 provides a simple, transparent summary of that finding, which is based on direct estimates of TFPQ and demand-side factors from data that have both price and quantity variation. In the notation of the prior section, figure 1 shows the evolution of A_{it} and D_{it}. The differences between young and mature plants are overwhelmingly driven by D_{it} rather than A_{it}.

A growing theoretical literature finds that developing a customer base is a slow process.[41] Foster, Haltiwanger, and Syverson develop a simple model to help account for this slow growth of demand.[42] They posit two distinct processes for accumulating a customer base over time. The first they call accumulating by being: the longer a firm is in existence, the more customers learn about the business and its attributes (that is, its product characteristics, the quality and quantity of its bundled services, the consistency of its operations,

37. TFPQ should probably be considered endogenous as well. Firms engage actively in investment in research and development and other forms of innovation.

38. Foster, Haltiwanger, and Syverson (2008, 2016).

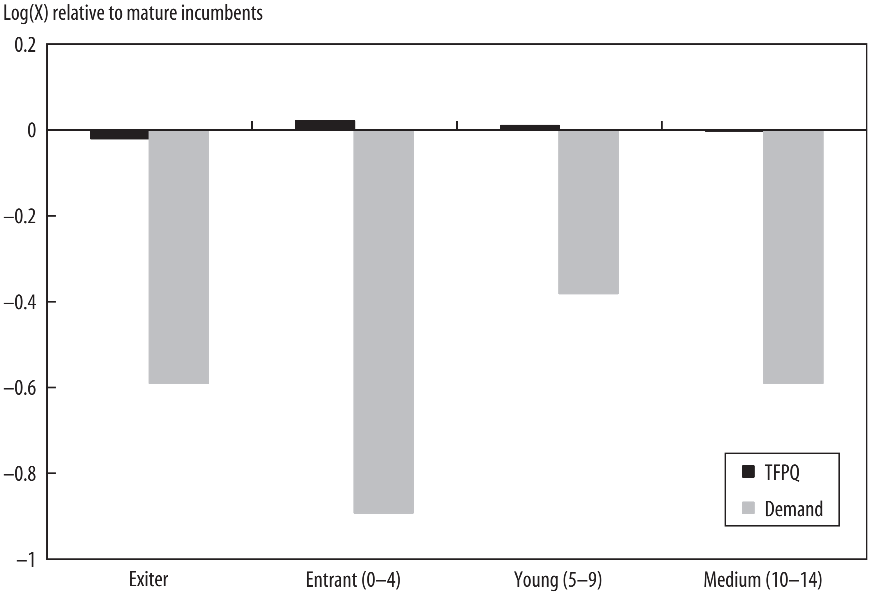
39. See, for example, Dunne, Roberts, and Samuelson (1989). The literature suggests that this is especially true in manufacturing compared with retail trade; see Ericson and Pakes (1995).

40. Foster, Haltiwanger, and Syverson (2008, 2016).

41. See, for example, Droszd and Nosal (2012); Gourio and Rudanko (2014).

42. Foster, Haltiwanger, and Syverson (2016).

FIGURE 1. The Evolution of TFPQ versus Demand



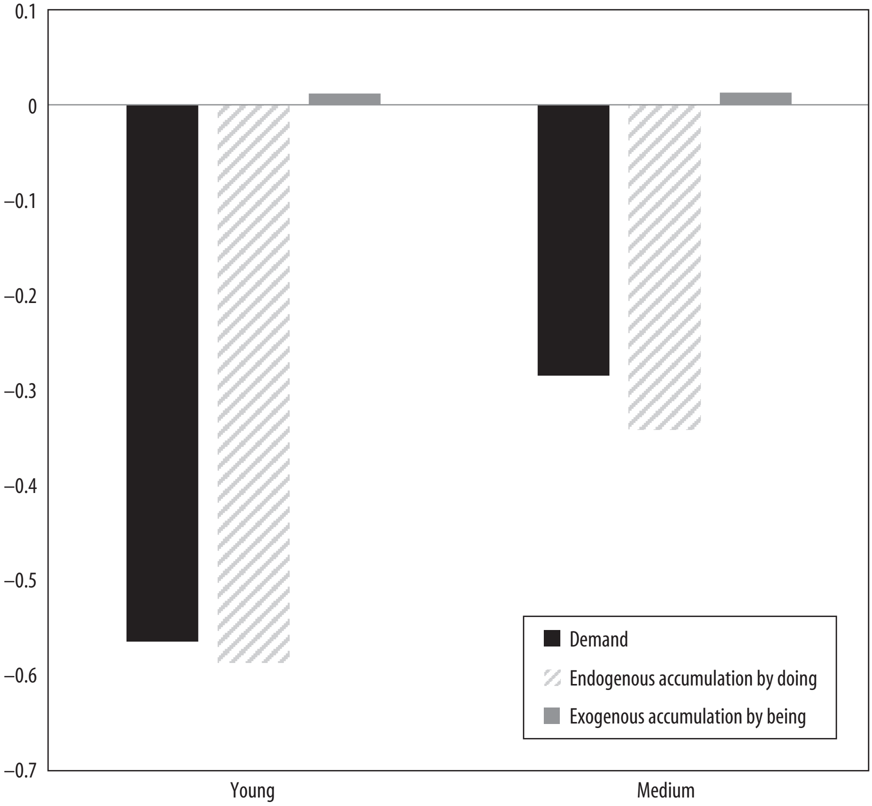
Source: Tabulations from Foster, Haltiwanger, and Syverson (2016, table 1).

the expected longevity of operations, and so forth). This suggests that a customer base would exogenously accumulate as a business ages, but it would perhaps be a slow process given the many different dimensions of learning that might be important. The second they call accumulating by doing, which requires some form of active investment in the customer base. The authors use a reduced-form approach consistent with the learning-by-doing literature on the supply side. Specifically, they assume that firms can accumulate by doing if they lower prices to increase customer demand. That is, the need to build up a customer base gives young firms a greater incentive to charge lower prices than would be predicted by the optimal static markup.

Foster, Haltiwanger, and Syverson find overwhelming evidence that the endogenous component of demand is the source of the slow growth of demand of young businesses.[43] This finding is illustrated in figure 2. The demand for young businesses is more than 50 log points lower than for mature businesses.

43. Foster, Haltiwanger, and Syverson (2016).

FIGURE 2. Endogenous versus Exogenous Evolution of Demand



Source: Tabulations from Foster, Haltiwanger, and Syverson (2016, table 6).

Virtually all of that is accounted for by the endogenous component of building up demand through charging lower prices when young.

These results imply that post-entry growth dynamics are a time-intensive process. It takes a long time for new businesses to build enough relationship-specific capital with their potential customers that they can expect to sell the same amount of output as do their more established competitors (at the same price). This buttresses the recent literature pointing toward the importance of idiosyncratic demand factors in explaining the fortunes of businesses, and it has implications for the nature of competition in markets, firm valuations, the evolution of industries, and the prospects for exporting in new markets.

A clear next step that researchers can take based on these results is to explore the particular mechanisms that underlie the endogenous and exogenous demand accumulation processes in this model. Several basic questions present themselves. How much of this reflects brand effects, reputation, or other aspects of buyer-supplier relationships? Does the specific mechanism at work differ across markets, and, if so, how? Will active accumulation processes always quantitatively dominate passive processes? What affects the extent to which a firm's demand capital spills over to its newly built or newly acquired plants?

The important role of demand-side factors is potentially quite significant for the firm dynamics literature, since it suggests that the frictions associated with breaking into a market may be critical for young businesses. The literature focuses overwhelmingly on productivity differences, as opposed to demand differences, as the source of firm heterogeneity, with much less attention on how these frictions might vary across countries and sectors. Nevertheless, distortions in demand dynamics may be especially important for the misallocation literature. Similarly, distortions from the demand structure may be preventing young businesses from growing in emerging economies.

Concluding Remarks

The literature on firm heterogeneity and firm dynamics has had a major impact on a wide range of questions ranging from macroeconomics, labor economics, development economics, and international trade. It is increasingly recognized that cross-country differences in economic performance are likely related to the underlying efficiency of firm dynamics. The impact of international trade on an economy is now typically understood through its impact on the reallocation of resources across firms and on the market structure of economies. Labor market dynamics are increasingly understood in relation to how heterogeneous firms interact with heterogeneous workers.

Underlying the enormous progress in understanding firm heterogeneity and dynamics is the finding that measured productivity varies widely across firms within narrowly defined sectors. Moreover, this dispersion in measured productivity is closely tied to the reallocation dynamics discussed above. In the midst of this burgeoning literature, two interesting and related issues have arisen that overlap theoretical and measurement concerns in this literature.

The first is that most of the empirical literature captures various measures of revenue productivity rather than solely measures of technical efficiency. The second is that in attempting to distinguish between revenue productivity measures and TFPQ, one can distinguish between and understand the dynamics of demand versus supply factors for the evolution of firms.

Revenue productivity measures almost inherently reflect demand-side factors as well as supply-side factors like TFPQ. But added to the mix is the recognition that firm-level prices are likely endogenous. Moreover, since it seems increasingly likely that even for commodity-like products, there is product differentiation, and the typical firm faces a downward-sloping demand curve. This is critical since it implies that TFPQ will be inversely related to firm-level prices.

One particular and frequently used measure of revenue productivity is TFPR, which is defined as the product of the firm-level price and TFPQ. There is an interesting knife-edge case in which the inverse relationship between TFPQ and prices at the firm level implies, in the absence of frictions or distortions, that TFPR will exhibit no dispersion in equilibrium even with substantial dispersion in TFPQ (and an accompanying substantial dispersion in the size distribution of activity). High-productivity firms will be larger in size and move down their demand curves charging lower prices. In the knife-edge case (Cobb-Douglas production functions with CRS and iso-elastic demand), the reduction in prices will exactly offset the higher productivity, so that TFPR is the same across firms. This insight has led some analysts to interpret all of the dispersion in empirical measures of TFPR observed in the literature as reflecting distortions.

This paper emphasizes that empirical measures of revenue productivity may not correspond to the conceptual TFPR measure, which is price times TFPQ. Revenue per unit of composite input yields an empirical estimate of TFPR under CRS and Cobb-Douglas production functions, but revenue function residuals do not typically yield a measure that is equivalent to TFPR or revenue per unit of input. Revenue function residuals have the attractive property that they are, with consistent estimation, only a function of market fundamentals. Despite these conceptual differences, empirical estimates of revenue per unit of input and revenue residuals are highly correlated and exhibit similar dispersion. This evidence implies that distortions are highly correlated with distortions or that, as argued here, frictions or other factors cause departures from the assumptions that imply that dispersion in empirical measures of TFPR reflects distortions.

Rather than view these complications as distractions, this paper takes the view that resolving them will lead to further progress on illuminating the role of firm heterogeneity and firm dynamics. That is, distinguishing between TFPR (and empirical measures of revenue productivity that do not necessarily correspond to TFPR) and TFPQ is not a bug but a feature of the literature. Similarly, distinguishing between TFPR and TFPQ pushes the literature to a much richer consideration of demand-side factors in understanding observed firm dynamics. The idea that the endogenous process of breaking into a market and building up a customer base is a potentially important dynamic deserves further attention.

References

Asker, John, Allan Collard-Wexler, and Jan de Loecker. 2014. "Dynamic Inputs and Resource (Mis)Allocation." *Journal of Political Economy* 122(5): 1013–65.

———. 2013. "Cross-Country Differences in Productivity: The Role of Allocative Efficiency." *American Economic Review* 103(1): 305–34.

Cooper, Russell, and John Haltiwanger. 2006. "On the Nature of Capital Adjustment Costs." *Review of Economic Studies* 73(3): 611–33.

Decker, Ryan, and others. 2015. "Changing Business Dynamism: Volatility of Shocks versus Responsiveness to Shocks." Washington: Board of Governors of the Federal Reserve System.

De Loecker, Jan, and others. 2016. "Prices, Markups, and Trade Reform." *Econometrica* 84(2): 445–510.

Drozd, Lukasz A., and Jaromir B. Nosal. 2012. "Understanding International Prices: Customers as Capital." *American Economic Review* 102(1): 364–95.

Dunne, Timothy, Mark Roberts, and Larry Samuelson. 1989. "Plant Turnover and Gross Employment Flows in the U.S. Manufacturing Sector." *Journal of Labor Economics* 7(1): 48–71.

Ericson, Richard, and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62(1): 53–82.

Eslava, Marcela, and John Haltiwanger. 2016. "The Drivers of Post-Entry Growth: TFPQ versus Demand and Input Costs." University of Maryland.

Eslava, Marcela, and others. 2004. "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia." *Journal of Development Economics* 75(2): 333–72.

———. 2013. "Trade Reforms and Market Selection: Evidence from Manufacturing Plants in Colombia." *Review of Economic Dynamics* 135–58.

———. 2015. "Scarring Recessions and Credit Constraints: Evidence from Colombian Firm Dynamics." CEDE Working Paper 2010-07. Bogotá: Universidad de los Andes.

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.

———, 2016. "The Slow Growth of Young Plants: Learning about Demand." *Economica* 83(329): 91–129.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2015. "From Devilish Details to Insights: Micro and Macro Dynamics of Productivity." Washington: U.S. Census Bureau, Center for Economic Studies.

———. 2016. "Firm-Level Dispersion in Productivity: Is the Devil in the Details?" *American Economic Review* 106(5): 95–98.

Gourio, François, and Leena Rudanko. 2014. "Customer Capital." *Review of Economic Studies* 81(3): 1102–36.

Hsieh, Chang-Tai, and Peter Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124(4): 1403–48.

———. 2014. "The Life Cycle of Plants in India and Mexico." *Quarterly Journal of Economics* 129(3): 1035–84.

Levinsohn, James A., and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies* 70(2): 317–41.

Melitz, Marc J., and Gianmarco Ottaviano. 2008. "Market Size, Trade, and Productivity." *Review of Economic Studies* 75(1): 295–316.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Plants." *Review of Economic Dynamics* 11(4): 707–20.

———. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49(2): 326–65.

Wooldridge, Jeffrey. 2009. "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables." *Economics Letters* 104(3): 112–14.

World Bank. 2013. *World Development Report 2013: Jobs.* Washington.