The Life-cycle Growth of Plants: The Role of Productivity, Demand and Wedges.*

Marcela Eslava† and John Haltiwanger‡

March 12, 2019

Abstract

Using rich product level data on prices and quantities of both inputs and outputs at the establishment level for Colombia, we develop a methodological approach to decomposing sales and output growth over an establishment's life cycle. Our approach brings together strands of the literature that have either focused on the relative roles of broadly defined productivity vs. wedges using data on revenue and input expenditure, or on the roles of cost vs. demand-side components of productivity using data on prices and quantities of output. Our findings show that the literature using just price and quantity data on output understates the role of cost and productivity factors in accounting for sales volatility especially at young ages. The reason is that such approaches implicitly combine the role of wedges that dampen volatility

*We thank Alvaro Pinzón for superb research assistance, and Innovations for Poverty Action, CAF and the World Bank for financial support for this project. We also thank DANE for permitting access to the microdata of the Annual Manufacturing Survey, as well as DANE's staff for advice in the use of these data. The use and interpretation of the data are the authors' responsibility. We gratefully acknowledge the comments of David Atkin, Steven Davis, Chad Syverson, Robert Shimer, Diego Restuccia, Peter Klenow, Stephen Redding, Gabriel Ulyssea, Irene Brambilla, Manuel García-Santana and those from participants at the 2017 Society for Economic Dynamics Conference; 2017 meeting of the European chapter of the Econometric Society; 2017 NBER Productivity, Development, and Entrepreneurship workshop; the 2016 Trade and Integration Network of LACEA; and seminars at Universitat Pompeu Fabra, Universitat Autónoma de Barcelona, the University of Chicago and Universidad de Los Andes.

†Universidad de Los Andes, Bogotá. meslava@uniandes.edu.co

‡University of Maryland at College Park. haltiwan@econ.umd.edu

with cost and technology factors that exacerbate volatility. At the same time, our inferences about the respective role of fundamentals and wedges are quite different from those drawn from revenue and input data alone. Use of the latter (which is the common approach in the literature) yields a substantial overstatement of the contribution of role of wedges in accounting for sales growth volatility. We find that at young ages, technology and demand shocks are about equally important in accounting for variability in sales growth volatility but the latter is dampened by wedges that are positively correlated with these fundamentals. As plants age, wedges have a less dampening effect. Moreover, demand differences increasingly dominate sales volatility for older plants. The dominance of demand is driven by superstar plants that are in the top quartile of life cycle growth.

Keywords: post-entry growth; TFPQ; demand; distortions.

JEL codes: O47; O14; O39

1 Introduction

Robust firm size and firm growth distributions are key features of development: within narrowly defined sectors, firms in richer countries are larger and grow faster than their counterparts in less developed economies, with differences concentrated in the tails of the respective distributions.[1] Having businesses that grow such as those in India or China rather than as those in the US is associated with an aggregate TFP differential as large as 25 percent (Hsieh and Klenow, 2014). Understanding the determinants of firm growth and size is thus crucial to our understanding of development.

Research on this question has been constrained by data availability. One strand of the literature uses revenue and input data to decompose firm growth into the contributions of productivity vs. residual wedges, or alternatively, productivity vs. markups (Hsieh and Klenow, 2014; De Loecker et al, 2015). Another branch uses detailed information on prices and quantities of the goods produced by the firm to separate growth into that attributable to cost shocks and demand shocks (Hottman, Redding and Weinstein, 2016). Wedges in the latter approach are subsumed into residual cost shocks, while the former approach lumps demand shocks and cost shocks from technical efficiency into "productivity", and cost shocks from input prices into wedges.

[1] See, e.g. Bento and Restuccia (2017, 2018) ; Hsieh and Klenow (2014), Eslava, Haltiwanger and Pinzón (2018).

Bringing both approaches together to separate the role of these different components requires simultaneous access to data on prices, quantities and values for both inputs and outputs, and a structure suitable to take advantage of this data. In the context of explaining life cycle growth, further availability of these detailed data for long periods of time for each business unit is required.

Put differently, the paucity of firm-level price and quantitity data on both outputs and inputs has implied that the inferences on the role of productivity, demand and wedges in the literature has relied on a high ratio of assumptions to data. One of the common simplifying assumptions is constant returns to scale in production. An alternative is that estimation methods that in fact yield elasticities of the revenue function are assumed to be proxies for the needed elasticities of the production function. In a related way, it is often assumed that the production technology can be inferred without jointly specifying and inferring the structure of demand. However, particularly for multi-product firms, defining and measuring real output and inputs at the firm-level requires computing a firm-level price index for both outputs and inputs. Following the insights of Hottman, Redding and Weinstein (2016) and Redding and Weinsten (2018), this implies that a nested demand structure must be specified at the product-level within firms to enable construction of such firm-level price indices.

We develop a conceptual, measurement and estimation structure that overcomes these limitations by taking advantage of uniquely rich data that tracks product-level outputs and input prices and quantities within plants over time. Our novel approach shows that integrating these different dimensions of data is crucial to inferences regarding the role of wedges and fundamentals in explaining the life cycle growth of output and sales as well as the relative importance of demand vs. cost factors in sales and output volatlity. For this purpose, we use the Colombian Annual Manufacturing Survey which is a census of non-micro Colombian manufacturing plants with data on quantities and prices, at the detailed product class for outputs and inputs within plants. Individual plants can be followed for up to thirty years (1982-2012). The availability of price and quantity data for both outputs and inputs at the product level permits separate measurement of fundamental attributes of plants on the technology, the demand, and the cost sides, as well as idiosyncratic markups. By technology or technical efficiency we refer to a production function residual, where production is plant-level revenue deflated with a quality adjusted plant-level deflator. We will refer to this technical efficiency dimension as $TFPQ$, as in Foster, Haltiwanger and Syverson (2008)[2] On the demand side, we estimate plant-specific demand function residuals, that identify greater appeal/quality as the ability to charge higher prices for one unit of the same product. Input costs are directly measured from input price data. Our specification of demand and competition allows for idiosyncratic markups that vary with the plant's market share and with the elasticity of substitution in the plant's sector. With all of these elements at hand, we measure the contribution of each to the variability of sales growth. Wedges, defined by the gap between actual size at any point of the life cycle and size implied by the different fundamentals in a frictionless benchmark at that point, can also be identified once fundamentals are measured.[3]

Key to our approach is the construction of plant-level price indices, which allows measuring output as deflated revenue. Our demand function and exact price index are derived following Hottman et al's (2016) modelling of preferences. To construct the theoretical price index relying solely on observable price and quantity data, we follow Redding and Weinstein's (2018) recent Unified Price Index approach.

Our approach requires, and the richness of the data permits, estimating the parameters of the production and demand functions for each sector both to obtain $TFPQ$ and appeal as residuals of these functions. We introduce an estimation technique that jointly estimates the production factor elasticities and the elasticity of demand for plants, bringing together insights from recent literature on estimating production functions using output and input use data,[4] and literature on estimating demand functions using P and Q data.[5] As in the former, we rely on assumptions regarding the dynamics of

[2]In contrast to Foster et al's application to producers of one homogeneous good, we use the term $TFPQ$ in the context of multiproduct plants and potentially heterogeneous products, where product is a quality-adjusted bundle of quantities of differentiated goods, operationalized as deflated revenue. Hsieh and Klenow (2009, 2014) also use the term $TFPQ$, but they use it to refer to a composite productivity measure that lumps together technical efficiency and demand shocks. We refer to this composite concept further below as $TFPQ\_HK$, as a reference to Hsieh and Klenow. Haltiwanger, Kulick and Syverson (2018) explore properties of $TFPQ\_HK$ using U.S. data.

[3]These wedges are also frequently termed "distortions", but we prefer the former term since the idiosyncratic gaps we identify may represent sources of productivity or welfare loss that even the social planner would incur, as they may stem from constraints more technological in nature, such as adjustment costs.

[4]e.g. Ackerberg, Caves and Frazer (2015); De Loecker et al. (2016)

[5]E.g. Hottman, Redding and Weinstein (2016); Foster, Haltiwanger and Syverson

input use and input prices to form moments that identify production function coefficients. As in the latter, we rely on supply shocks to identify the slope of the demand function. But, in contrast to much of that literature, we identify the slope of the demand function by assuming that current period innovations to technology are orthogonal to lagged demand shocks. In this way, we allow $TFPQ$ and demand to be correlated if for instance, as plausible, higher quality is more difficult to produce, or investments in improving fundamentals depend on previous profitability. Estimating production and demand jointly ensures consistency and thus proper separate identification of revenue vs. production parameters.

We find that exploiting price and quantity data for both outputs and inputs yields distinct inferences relative to the different strands of the literature decomposing sources of sales volatility. Using price and quantity data on outputs alone yields an understatement of cost and technology factors in accounting for sales volatility because wedges and cost/technology factors are lumped together, and wedges in our context turn out to be negatively correlated with fundamentals. Alternatively, using revenue and input data alone overstates the contribution of wedges for sales volatility for multiple reasons including mis-estimation of the relevant demand and output elasticities and missing the distinction between input price shocks and wedges.In our results, input price variability explains most of the contribution of wedges that would be estimated without explicit account of input prices. Moreover, the estimated contribution of wedges depends crucially on the curvature of the revenue function. Our ability to estimate of the parameters of the revenue function for the particular application and for each sector turns out to be quantitatively important: when imposing common parameters, wedges are substantially underestimated for sectors where the revenue function has close to constant returns to scale and overestimated for sectors where the revenue function displays more marked curvature. Because the bias is quantitatively larger in the former cases, imposing common parameters leads to an underestimation of the role of wedges.

Post entry growth is highly dispersed and skewed in our data, as it is in other contexts (e.g. Decker et. al. (2014,2016)). Our focus is on decomposing the substantial variance in growth across plants at different stages of the life cycle. By age 25, top quartile plants have multiplied their sales by a factor of 7.6 relative to their birth; for this group, sales would have grown

(2008)

ten-fold in the absence of wedges. Of growth explained by technology and demand-side fundamentals in this group, more than two thirds is attributable to rapidly growing product demand/appeal and the remaining 1/3 to rising technical efficiency. These patterns contrast with the lower quartiles of sales growth. Most noticeably, while superstar plants are held back by wedges especially early in their life cycle, plants in the lowest two quartiles are propped up by wedges. That is, wedges are strongly size-correlated. Moreover, the third quartile exhibits sales growth from product appeal/demand that is less than half of the top quartile and no growth attributable to rising technical efficiency. Plants in the lowest two quartiles barely exhibit appeal growth and display sharp efficiency decreases. Our findings of a dominant role of demand shocks in accounting for life cycle growth is consistent with Foster, Haltiwanger and Syverson (2008, 2016). However, our results imply (at least for Colombia) that this is being driven by the superstar plants (top quartile) where rising demand is especially dominant.

The wide dispersion in sales and output growth is mostly accounted for by dispersion in fundamentals, rather than wedges, with $TFPQ$ and demand shocks both playing a crucial role. Pooling all ages and allowing wedges to be correlated with fundamentals, measured fundamentals account for more than 100% of the variability of revenue and output relative to birth level, reflecting the mentioned fact that wedges dampen growth relative to what is implied by fundamentals. For sales growth, over 90% of the combined contribution of $TFPQ$ and demand is from between plant variation in demand/appeal. The much greater contribution of demand/appeal to sales volatility over the life cycle is consistent with the Hottman et. al. (2016) finding of a dominant role for demand in accounting for the variance of sales in the cross section. We show this finding emerges naturally from our theoretical model where sales (PQ) is directly interpretable as a quality/appeal adjusted measure of output (Q).

Correlated wedges such as the ones we find may have different sources including the adjustment costs, financial frictions and size dependent distortions. If measured wedges completely reflected factors correlated with fundamentals then a reduced form regression of output and sales growth on fundamentals would account for all of the variation. However, when we estimate such reduced form regressions and conduct an accompanying reduced-form decomposition of output and sales growth on fundamentals we find an important residual that accounts for about 50% of output growth volatility and 40% of sales growth volatility over the life cycle. This noise component of wedges is more important for young as opposed to mature plants. Thus, young plants face both greater dampening of growth volatility due to correlated wedges and more random volatility of growth due to uncorrelated wedges.

Our research brings together previous approaches that have estimated the contribution of subsets of the determinants of growth that we consider. The richness of the data allows us to rely on a structure that is less restrictive than that in each of those individual approaches. A large literature starting with the seminal work by Hsieh and Klenow (2009) investigates the role of wedges using data on revenue and input payments, imposing Cobb-Douglas technology with constant returns to scale, homogeneous input prices, and a CES demand structure under monopolistic composition. Under these assumptions, a composite measure of technical efficiency and appeal can be inferred from revenue data, and all dispersion in average revenue products of inputs is attributed to wedges. By incorporating price and quantity data for both inputs and outputs, we can measure efficiency and appeal independently, and relax the assumptions of constant returns to scale in production and monopolistic competition, allowing average revenue product variation from sources other than wedges and incorporating idiosyncratic markups. We are also able to estimate demand elasticities directly and allow for heterogeneity in demand and production parameters across finely defined sectors, an ability that turns out to be crucial for quantitative results on the role of wedges.

Information on P and Q has been previously used by Hottman et al (2016) to assess the role of cost-side vs. demand-side fundamentals in revenue growth, in the context of structure where the demand plus supply factors account for sales volatility completely (i.e. there is no margin for wedges). While we rely heavily on their nested plant-product approach to model demand and markups, the richness of our data allows for a much more detailed analysis of the supply side. In particular, our research adds to theirs by combining price and quantity output data with input use data, which opens the door for wedges between fundamentals and outcomes, and allows decomposing marginal costs into technology and input prices.

Price and quantity information in combination with data on input use has been previously used by De Loecker et al (2016) to decompose prices into marginal costs and markups. Markups, expressed as the ratio between a flexible factor's elasticity in production and its cost share, are obtained taking advantage of the fact that the detailed quantity data allows the authors to properly estimate production (rather than revenue) factor elasticities. To deal with the difficulty of aggregating quantities of different products, the authors estimate production functions relying solely on the information for uniproduct plants. Our research contributes to that literature by providing an approach to aggregate multiple product lines into a measure of output for a multiproduct business. In doing so, our approach highlights the need for relying on explicit assumptions about the structure of demand (preferences) to measure output for a multiproduct firm. Specifying preferences over heterogeneous products allows the researcher not only to define the concept of output for multiproduct firms, but also to establish proper comparisons across businesses producing different goods, even in the case of uniproduct businesses.

Much of the focus in trying to understand the reasons behind slow post-entry productivity growth, and consequenty the focus on designing policy interventions to address such slow growth, has been on dimensions external to the business (our wedges), such as institutions that discourage growth. Our results highlight that, alongside wedges, the dimensions internal to businesses are at least as important. On this internal side, the focus has frequently been on efforts conducive to improvements in technical efficiency. For instance, research on managerial practices that impact productivity has focused on production processes and employee management (e.g. Bloom and Van Reenen, 2007; Bloom et al. 2016). Our approach highlights the multidimensional character of growth drivers that are internal to the business, including the appeal to costumers and input prices potentially affected by its decisions. Our results align with those in Atkin et al (2016) and Atkin et al (2019) in pointing at quality as crucial driver of business growth, and at the fact that quality improvements may impose costs in terms of technical efficiency.

The paper proceeds as follows. Section 2 presents our conceptual framework, defining each of the plant fundamentals that we characterize, and our approach to decompose growth into contributions of those fundamental sources as well as wedges. We then explain the data used in our empirical work, and the approach we use to measure fundamentals, including the joint estimation of the parameters of production and demand, respectively in sections 3 and 4 . Results and comparisons of our results with previous approaches are presented in section 5. Section 6 examines the robustness of our results to using previous approaches and discusses the value added of ours. Section 7 concludes.

2 Decomposing firm growth into fundamentals vs wedges

We start with a simple model of firm optimal behavior given firm fundamentals, to derive the relationship that should be observed between size growth and growth in fundamentals as a firm ages. We also permit firm size to be impacted by wedges. For consistency with the literature on business dynamics, in our theoretical analysis we refer to a business as a "firm", even though the unit of observation for our empirical work is an establishment or plant. The main fundamentals we consider are the efficiency of the firm's productive process (which we term $TFPQ$ as in Foster, Haltiwanger and Syverson, 2008) and a demand shock. The conceptual framework below makes clear what we mean by each of these, and the sense in which they are "fundamentals". Beyond measuring $TFPQ$ and demand shocks, we observe unit prices for inputs, in particular material inputs and labor.

In the model, the firm chooses its size optimally given $TFPQ$, demand shocks, input prices and wedges. As a result, growth over its life cycle is driven by growth in each of them. This is the basis of our analysis. In the spirit of a growth accounting exercise the framework remains silent about the sources of growth of fundamentals, and rather asks how the firm adjusts its size given those fundamentals, and contingent on survival.[6] However, we do explore the relationship between fundamentals and wedges. In the appendix, we also explore the relationship between proxies for investment in innovation and lagged fundamentals in our robustness analysis below. We

[6] For instance, the seminal models of Hopenhayn (1992) and Melitz (2003), and much of the work that has since followed in Macroeconomics and Trade. Endogenous productivity-quality growth has made its way to these models more recently (e.g. Atkenson and Burstein, 2010; Acemoglu et al. 2014; Hsieh and Klenow, 2014; Fieler, Eslava, and Xu, 2016). The firm's efforts to strengthen demand may include investments in building its client base (Foster et al., 2016), and adding new products and/or improving the quality of its pre-existing product lines. Those to strangthen $TFPQ$ may include better management of the production process (e.g. Bloom and Van Reenen, 2007) or acquiring better machines. The results of our decomposition shed light on the relative role and characteristics of each of these accumulation processes, useful for providing guidance about future research that explores the determinants of these fundamentals. We also do not formally model the exit decision in the analysis below. Formally, adding this margin would be straightforward as each period the firm would choose whether or not to continue based on present discounted value considerations net of any fixed cost of operations (which we do not explicitly model). Our analysis, contingent on the stay decision, would still be valid.

focus on decomposing the determinants of surviving firms up to any given age but include robustness analysis of the determinants of survival in appendix H. Appendix H shows that our main results are robust to consideration of selection issues.

We don't explicitly model adjustment frictions but take the shortcut in recent literature on misallocation to permit wedges or distortions between frictionless static first order conditions and actual behavior (e.g. Hsieh and Klenow, 2009). Such distortions and wedges might capture factors such as adjustment frictions, technological frictions, and distortions arising from regulation.[7] This shortcut enables us to use a simple static model of optimal input determination to frame our analysis of growth between birth and any given age. We permit the wedges or distortions to vary by firm age which could be viewed as a proxy for permitting adjustment frictions to vary by firm age.

2.1 Firm Optimization

Consider a firm indexed by f, that produces output Q_{ft} using a composite input X_{ft} to maximize its profits, with technology

$$Q_{ft} = A_{ft} X_{ft}^{\gamma} = a_{ft} A_t X_{ft}^{\gamma} \tag{1}$$

A_{ft} is the firm's technical efficiency, $TFPQ$, which has an aggregate and an idiosyncratic component (A_t and a_{ft}), while γ is the returns to scale (in production) parameter. Equation (1) defines a_{ft} as the (idiosyncratic) efficiency of the productive process: how much output the firm obtains from a unit of a basket of inputs. Firm f may be uni- or multi-product. Section 2.2 below discusses the definition of output Q for multi-product firms.

We use a CES preference structure (specified in more detail below) that yields demand at the firm level to be given by:

$$P_{ft} = D_{ft} Q_{ft}^{-\frac{1}{\sigma}} = D_t d_{ft} Q_{ft}^{-\frac{1}{\sigma}} \tag{2}$$

[7] This shortcut has limitations as the idiosyncratic distortions that we permit don't provide the discipline that formally modeling dynamic frictions imply. See, e.g., Asker, Collard-Wexler and DeLoecker (2014), Decker et. al. (2017) and Haltiwanger, Kulick and Syverson (2018). But it has the advantage in subsuming in a simple measure different types of frictions and distortions.

where D_{ft} is a demand shifter, and σ is the elasticity of substitution between firms . D_{ft} has aggregate and idiosyncratic components $D_t = P_t \left(\frac{E_t}{P_t}\right)^{1/\sigma}$ and d_{ft}, respectively. E_t is aggregate (sectoral) expenditure, and the aggregate (sectoral) price index is given by $P_t = \left(\sum_{f=1}^{N_F} d_{ft}^{\sigma} P_{ft}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ where N_F is the number of firms in the sector.

Firm appeal d_{ft} is measured from equation (2) as the variation in firm price holding quantities constant, beyond aggregate effects. We refer to d_{ft} generically as the firm's (idiosyncratic) demand shock, intuitively capturing quality/appeal. Notice also that, multiplying (2) by Q_{ft} :

$$R_{ft} = D_t d_{ft} Q_{ft}^{1-\frac{1}{\sigma}} = D_t \left(Q_{ft}^{Q}\right)^{\frac{\sigma-1}{\sigma}} \tag{3}$$

where Q_{ft}^{Q} is quality-adjusted output defined as $d_{ft}^{\frac{\sigma}{\sigma-1}} Q_{ft}$. The idiosyncratic component of sales is, thus, driven by quality adjusted output. Using the CES preference structure discussed in more detail below, from which demand equation 2 can be derived, it is apparent that idiosyncratic firm sales are closely linked to consumer welfare. As a result, the distribution of firm sales growth is the central focus of our analysis, although we also apply our analysis to real output.

Putting together technology and demand, the firm chooses its scale X_{ft} to maximize profits[8]

$$\underset{X_{it}}{Max} \ \left(1-\tau_{ft}\right) P_{ft} Q_{ft} - C_{ft} X_{ft} = \left(1-\tau_{ft}\right) D_{ft} A_{ft}^{1-\frac{1}{\sigma}} X_{ft}^{\gamma\left(1-\frac{1}{\sigma}\right)} - C_{ft} X_{ft}$$

taking as given A_{ft}, D_{ft}, and unit costs of the composite input, C_{ft}. There may be idiosyncratic revenue wedges τ_{ft}, that create a gap between a firm's actual scale and that which would be implied by its fundamentals.[9] Such wedges capture, for instance, adjustment costs, product-specific tariffs, financing constraints and size-dependent regulations or taxes. Adjustment

[8]Recall this is the characterization of the optimal size conditional on the firm deciding to operate in the current period.

[9]As in Restuccia and Rogerson, 2009 and Hsieh and Klenow, 2009. Further below, we also consider factor-specific distortions that, for given choice of X_{it}, affect the relative choice of a given input with respect to others.

costs break the link between actual adjustment and the "desired adjustment".[10] Financing constraints may similarly limit the ability of the firm to undertake optimal investments, and force it to remain smaller than optimal and even potentially exit the market during liquidity crunches even if its present discounted value is positive.[11] The resulting τ_{ft} may be randomly distributed across plants or correlated with plant fundamentals themselves. By their very nature, adjustment costs and financing constraints are typically correlated with plant fundamentals. Size-dependent regulations are a prominent example of correlated wedges, though certainly not the only one.[12] In estimating the role of wedges as determinants of life-cycle growth, we distinguish between wedges that are orthogonal to fundamentals and those potentially correlated with them.

We allow firms to hold market power, so that a firm's market share may be non-negligible. This also implies that in choosing its optimal scale, a firm does not take as given the aggregate price index, P_t. Under these conditions and the CES demand structure developed in section 2.2, variability in markups across firms stems from market power (i.e., firms take into account their impact on sectoral prices):

$$\mu_{ft} = \frac{\sigma}{(\sigma - 1)} \frac{1}{(1 - s_{ft})} \tag{4}$$

Where μ_{ft} is the firm's markup and $s_{ft} = \frac{R_{ft}}{E_t}$ (proof: Appendix D). As in Hsieh and Klenow (2009, 2014), marginal cost is defined inclusive of wedges, so that $\mu_{ft} = \frac{P_{ft}}{\frac{\partial CT_{ft}}{\partial Q_{ft}}(1-\tau)^{-1}}$ where CT is total cost.

Profit maximization yields optimal input demand $X_{ft} = \left(\frac{\left(1-\frac{1}{\sigma}\right) D_{ft} A_{ft}^{1-\frac{1}{\sigma}} \gamma}{C_{ft}\left(1-\tau_{ft}\right)^{-1}} \right)^{\frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}}$, which is then used to obtain optimal output and sales as functions of fundamentals (D_{ft}, A_{ft}, and C_{ft}), wedges τ, and parameters. Subsequently dividing each optimal outcome in period t by its optimal level at birth ($t = 0$), we obtain (see Appendix B for a proof):[13]

[10] See, for instance, Caballero, Engel and Haltiwanger (1995, 1997), Eslava, Haltiwanger, Kugler, and Kugler (2010).

[11] Gopinath et al. (2017), Eslava et al. (2018)

[12] E.g. Garcia-Santana and Pijoan-Mas (2014) and Garicano et al. (2016).

[13] There is some slight abuse of notation here as t is used for calendar time and then for every firm we create our life cycle measures by dividing its outcomes and determinants at

$$\frac{Q_{ft}}{Q_{f0}} = \left(\frac{d_{ft}}{d_{f0}}\right)^{\gamma\kappa_1} \left(\frac{a_{ft}}{a_{f0}}\right)^{1+\gamma\kappa_2} \left(\frac{pm_{ft}}{pm_{f0}}\right)^{-\phi\kappa_1} \left(\frac{w_{ft}}{w_{f0}}\right)^{-\beta\kappa_1} \left(\frac{\mu_{ft}}{\mu_{f0}}\right)^{-\gamma\kappa_1} \chi_t \chi_{ft} \quad (5)$$

$$\frac{R_{ft}}{R_{f0}} = \left(\frac{d_{ft}}{d_{f0}}\right)^{\kappa_1} \left(\frac{a_{ft}}{a_{f0}}\right)^{\kappa_2} \left(\frac{pm_{ft}}{pm_{f0}}\right)^{-\phi\kappa_2} \left(\frac{w_{ft}}{w_{f0}}\right)^{-\beta\kappa_2} \left(\frac{\mu_{ft}}{\mu_{f0}}\right)^{-\gamma\kappa_2} \left(\widehat{\chi}_t \chi_{ft}\right)^{1-\frac{1}{\sigma}} \quad (6)$$

where we have further assumed $X_{ft} = K_{ft}^{\frac{\beta}{\gamma}} L_{ft}^{\frac{\alpha}{\gamma}} M_{ft}^{\frac{\phi}{\gamma}}$, so that C_{ft} is the corresponding Cobb-Douglas aggregate of the growth of different input prices, , among which two are observed in the data: the price of material inputs, Pm_{ft}, and average wage per worker, W_{ft}. We allow for potential factor-specific wedges, lumped with revenue wedges and measurement error in χ_{ft}.[14] As noted above, d_{ft} and a_{ft} are the idiosyncratic components of D_{ft} and A_{ft}. Similarly, pm_{ft} and w_{ft} are the idiosyncratic components of Pm_{ft} and W_{ft}. Aggregate components are lumped into $\chi_t = \left(\frac{D_t}{D_0}\right)^{\gamma\kappa_1} \left(\frac{A_t}{A_0}\right)^{1+\gamma\kappa_2} \left(\frac{C_t}{C_0}\right)^{-\gamma\kappa_1}$.

Equations (5) and (6) are the focus of our analysis. We start with the growth of (idiosyncratic) fundamentals that we can measure. Among these,$\frac{d_{ft}}{d_{f0}}$, $\frac{a_{ft}}{a_{f0}}$, $\frac{\mu_{ft}}{\mu_{f0}}$, $\frac{w_{ft}}{w_{f0}}$, $\frac{pm_{ft}}{pm_{f0}}$ are, respectively, life cycle growth in idiosyncratic demand shocks, $TFPQ$, markups, and shocks to wages and material input prices. Crucially, χ_{ft} captures idiosyncratic wedges, including those stemming from τ_{ft}, τ_{ft}^M, and τ_{ft}^L, from the unobservability of the user cost of capital, and from residual variation from noise in fundamentals not observed by the firm at the time of choosing its scale in each period. The wedges that a firm faces may be age-specific, and thus de-couple life-cycle growth in output from the growth of fundamentals.[15] Idiosyncratic wedges to the

some given age by those outcomes and determinants at birth. We use the ratio of these variables at *age* t to *age at birth* ($t = 0$).

[14] $\chi_{ft} = \frac{\delta_{ft}^{\gamma\kappa_1} \alpha_{ft}^{1+\gamma\kappa_2} \zeta_{ft}^{-\gamma\kappa_1} (1-\tau_{ft})^{\gamma\kappa_1} \left(1+\tau_{ft}^M\right)^{-\phi\kappa_1} \left(1+\tau_{ft}^L\right)^{-\beta\kappa_1} r_{ft}^{\frac{-\alpha\kappa_1}{\gamma}}}{\delta_{f0}^{\gamma\kappa_1} \alpha_{f0}^{1+\gamma\kappa_2} \zeta_{0t}^{-\gamma\kappa_1} (1-\tau_{0t})^{\gamma\kappa_1} \left(1+\tau_{f0}^M\right)^{-\phi\kappa_1} \left(1+\tau_{f0}^L\right)^{-\beta\kappa_1} r_{f0}^{\frac{-\alpha\kappa_1}{\gamma}}}$
where δ_{ft}, α_{ft}, and ζ_{ft} capture measurement error in, respectively, demand, technology and input price shocks, and τ^L and τ^M are, respectively, wedges specific to labor and materials with respect to capital.

[15] Some young firms may, for instance, have more dificulty in accessing financing, or face greater adjustment costs than their older counterparts. Also, many startups enjoy benefits that older firms do not face. This is the case, as an example, of small young firms in Colombia who at times have been exempted from specific labor taxes.

use of materials and labor relative to capital, τ_{ft}^{M} and τ_{ft}^{L} , may stem from elements such as factor-specific adjustment costs, and subsidies/taxes to the use of one input.

2.2 CES Demand Structure

In this subsection, we show that the firm-level demand structure used above is consistent with single-product producers as well as multiproduct producers using a CES preference structure. Taking into account multiproduct producers is important in our context to be able to define and measure firm-level output in a manner that captures within firm changes in product mix and product appeal over time. The theoretical structure is such that we can measure output as revenue deflated with an appropriate firm-level price index. As long as different products within a firm are not perfect substitutes, that price index reflects product turnover and changing product appeal across existing products. To accomplish this we use the UPI approach developed by Redding and Weinstein (2017) but also build on insights of Hottman et. al. (2016).

Specifically, in the context of multiproduct firms we allow firm output Q_{ft} to be a CES composite of individual products $Q_{ft} = \left(\sum_{\Omega_t^f} d_{fjt} q_{fjt}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, where q_{fjt} is period t sales of good j produced by firm f, the weights d_{fjt} reflect consumers' relative preference for different goods within the basket offered by firm f, and Ω_t^f is the basket of goods produced by f in year t. In particular, consumers derive utility from a composite CES utility function, with a CES layer for firms and another for products within firms. Consumer's utility in this general CES structure in period t is given by:

$$U\left(Q_{1t},...,Q_{Nt}\right) = \left(\sum_{I_t} d_{ft} Q_{ft}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \tag{7}$$

$$\text{where } Q_{ft} = \left(\sum_{\Omega_t^f} d_{fjt} q_{fjt}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \tag{8}$$

$$s.t. \quad \sum_{f=1}^{N_{Ft}} \sum_{\Omega_t^f} p_{fjt} q_{fjt} = E_t; \tag{9}$$

$$\prod_{\Omega_t^f} d_{fjt}^{\frac{1}{\left\|\Omega_t^f\right\|}} = 1; \prod_{I_t} d_{it}^{\frac{1}{\|I_t\|}} = 1 \tag{10}$$

where p_{fjt} is the price of q_{fjt}, and I_t is the set of firms in period t. We refer to d_{fjt} and d_{ft} as, respectively product (within firm) and firm appeal or demand shocks, defined as in equations 7 and 8: the weight, in consumer preferences, of product fj in firm f's basket of products, and of firm f in the set of firms. Given normalizations in equation (10), product appeal d_{fjt} captures the valuation of attributes specific to good fj relative to other goods produced by the firm, while firm appeal d_{ft} captures attributes that are common to all goods provided by firm f,, such as the firm's customer service and average quality of firm f's products, in a constant utility framework. Both firm and product appeal may vary over time.

Equation (8) defines real output for a firm in this multiproduct framework. As Hottman et al (2016) explain, in a multiproduct-firm context it is not possible to define real output in absence of assumptions about demand. The concept of real output "in theory equals nominal output divided by a price index, but the choice of price index is not arbitrary: it is determined by the utility function" (Hottman et al., 2016, page 1349). We define the real output of a multi-product firm as an aggregate of single-product outputs, in which each product receives a weight equal to its appeal to costumers, relative to that of other products within the firm. Given (10) this real output measure is normalized by the average appeal of products within the firm. The crucial relevant assumption here is that products within firms are not perfect substitutes so that tracking product turnover and changing product appeal within firms is critical for measuring firm-level output.

We assume the elasticity of substitution to be the same between and within firms in a sector. This assumption implies we have a special case of a nested CES with a nest for firms and another for products. Assuming the same elasticity simplifies the analysis substantially by abstracting from within firm cannibalization effects in a multi-product firm setting as explored by Hottman et. al. (2016). As discussed above, our firms still recognize their influence on the aggregate (sectoral) price level as they change their scale yielding the firm-level variation in the markup. This simplifying assumption also implies that in our estimation we can estimate the between firm elasticity of substitution and then apply it for our measurement of firm-level price indices.

Consumer optimization implies that the period t demand for product fj and the firm revenue are, respectively, given by

$$q_{fjt} = d_{ft}^{\sigma} d_{fjt}^{\sigma} \left(\frac{P_{ft}}{P_t} \right)^{-\sigma} \left(\frac{p_{fjt}}{P_{ft}} \right)^{-\sigma} \frac{E_t}{P_t} \tag{11}$$

$$R_{ft} = Q_{ft} P_{ft} = d_{ft}^{\sigma} P_{ft}^{1-\sigma} \frac{E_t}{P_t^{1-\sigma}} \tag{12}$$

where

$$P_{ft} = \left(\sum_{\Omega_t^f} d_{fjt}^{\sigma} p_{fjt}^{1-\sigma} \right)^{\frac{1}{(1-\sigma)}} \tag{13}$$

, and that

$$P_{ft} = D_{ft} Q_{ft}^{-\frac{1}{\sigma}} = D_t d_{ft} Q_{ft}^{-\frac{1}{\sigma}} \tag{14}$$

Equation (14) comes from dividing (12) by P_{ft} and solving for P_{ft}.[16] The implied firm-level price index is given by:

[16] We follow Redding and Weinstein (2016) in our treatment of product entry and exit. They don't formally model the decisions to add and substract products but rationalize the entry and exit of products through assumptions on the patterns of product specific demand shocks. That is, they assume products enter when the product specific demand shock switches from zero to positive and exits when the reverse occurs. We rationalize product entry and exit in the same manner. We consider multi-product plants mostly for the purpose of obtaining a plant-level price deflator that takes into account changing multi-product activity.

$$P_{ft} = \left(\sum_{\Omega_t^f} d_{fjt}^{\sigma} p_{fjt}^{1-\sigma} \right)^{\frac{1}{(1-\sigma)}} \tag{15}$$

Observe that (14) is identical to (2). This consistency is important as we use (15) to construct firm-level prices (using the UPI framework of Redding and Weinstein (2017) to express this price index in terms of observables). It is also useful to note that in using (12) one obtains the analogous interpretation of measured firm appeal (d_{ft}) used by Hottman et al (2016): d_{ft} captures sales holding prices constant. This is akin to quality as defined by Khandelwal (2010), Hallak and Schott (2011), Fieler, Eslava and Xu (2016), and others. Foster et al (2016), in turn, interpret firm appeal as capturing the strength of the business' client base.

Given our assumption of the same elasticity of substitution between and within firms a natural question is whether firms still *matter* in this context. Firms do matter for two reasons. First, our cost/production structure is at the firm-level. That is, we specify the cost/production function as being based on total output of the firm rather than product specific cost/production functions as in Hottman et. al. (2016). We make this assumption for more than the convenience that our input and input price data are at the firm level. Our view is that if one queried most firms (in our case – really plants) to specify input costs (capital, labor, materials and energy) on a product specific basis they would be unable to do so since costs are shared across products (i.e., there is joint production). That is, a firm is not simply a collection of separable lines of production. A second reason that firms matter here is firms may be large enough in the market so that we depart from monopolistic competition as firms don't take the sectoral output price as given. For these reasons, we specify a firm-level profit maximization problem but one that recognizes multi-product producers for purposes of measuring firm-level price deflators and in turn output.[17]

It is easily shown in our setting that we obtain the identical solution for optimal firm-level output as in (5) if we maximize firm-level profits with respect to each product defining profits as revenue (the sum of revenue from each product) minus total costs (which varies with the total output – and

[17] A limitation of our approach is we do not model the endogenous entry and exit of new products but follow Redding and Weinstein (2017) as noted by assuming new products arrive exogenously when d_{fjt} goes from zero to positive and exits when d_{fjt} goes to zero.

in turn inputs) of the firm.[18] This differs from Hottman et. al. (2016) who specify a product specific cost function. In their setting, firms matter only from the demand side – both because of differences in elasticities of substitution within vs. between firms and also because of possible market power effects on sectoral prices.

3 Data

3.1 Annual Manufacturing Survey

We use data from the Colombian Annual Manufacturing Survey (AMS) from 1982 to 2012. The survey, collected by the Colombian official statistical bureau DANE, covers all manufacturing establishments (=plants) belonging to firms that own at least one plant with 10 or more employees, or those with production value exceeding a level close to US$100,000. Our sample contains 23,292 plants over the whole period, with 7,670 plants in the average year.

Each establishment is assigned a unique ID that allows us to follow it over time. Since a plant's ID does not depend on an ID for the firm that owns the plant, it is not modified with changes in ownership, and such changes are not mistakenly identified as plant births and deaths. [19]

Surveyed establishments are asked to report their level of production and sales, as well as their use of employment and other inputs, their purchases of fixed assets, and the value of their payroll. We construct a measure of plant-level wage per worker by dividing payroll into number of employees, and obtain the capital stock using perpetual inventory methods, initializing at book value of the year the plant enters the survey. Sector IDs are also reported, at the 3-digit level of the ISIC revision 2 classification.[20] Since 2004, respondents are also asked about their investments in innovation, with biannual frequency, in a separate "innovation and development" survey.

[18] We also specify that wedges are at the firm-level and scale or factor specific.

[19] Plant IDs in the survey were modified in 1992 and 1993. To follow establishments over that period, we use the official correspondence that maps one into the other.The correspondence seems to be imperfect (as suggested by apparent high exit in 92 and high entry in 93), but even for actual continuers that are incorrectly classified as entries or exits, our age variable is correct (see further below).

[20] The ISIC classification in the survey changed from revision 2 to revision 3 over our period of observation. The three-digit level of disaggregation of revision 2 is the level at which a reliable correspondence between the two classifications exists.

A unique feature of the AMS, crucial for our ability to decompose fundamental sources of growth, is that inputs and products are reported at a detailed level. Plants report separately each material input used and product produced, at a level of disaggregation corresponding to seven digits of the ISIC classification (close to six-digits in the Harmonized System). For each of these detailed inputs and products, plants report separately quantities and values used or produced, so that plant-specific unit prices can be computed for both individual inputs and individual outputs. The average (median) plant produces 3.56 (2) products per year and employs 11.17 (9) inputs per year (Table 2).

Plant-specific unit prices on inputs imply that we directly observe idiosyncratic input costs for individual materials. Furthermore, by taking advantage of product-plant-specific prices, we can produce plant-level price indices for both inputs and outputs, and as a result generate measures of productivity based on output, estimate demand shocks, and consider the role of input prices in plant growth. Details on how we go about these estimations are provided in section 4. Our product level data are not at the detailed UPC code level of Hottman et. al. (2016), but we observe them at the plant-by-product-by-year level, which offers key advantages relative to other data sources. Unlike UPC codes, our product-level information is available by plant (physical location of production) rather than the aggregate firm, and is jointly observed with input use by that plant. And, unlike transactions data for imports (used, for instance by Feenstra, 2004, and Broda and Weinstein, 2006), we observe them not only at the product level (at similar levels of disaggregations with respect to imports transactions data) but by producer at a physical location.

Importantly for this study, the plant's initial year of operation is also recorded–again, unaffected by changes in ownership–. We use that information to calculate an establishment's age in each year of our sample. Though we can only follow establishments from the time of entry into the survey, we can determine their correct age, and follow a subsample from birth. Based on that restricted subsample, we generate measurement adjustment factors that we then use to estimate life-cycle growth even for plants that we do not observe from birth.[21] We restrict all of our analyses to plants born after 1969. Our decomposition results are in general robust to using the subsample observed from birth rather than the full sample, although estimated with less

[21] See Appendix 1.2 for details.

precision and for a shorter life-span. About a third of plants in our sample are observed from birth.

3.2 Plant-level prices built from observables

A crucial feature of our theoretical framework is that it allows the evolution of the plant size distribution to respond to changes in relative product appeal, both within the plant and across plants. Output can be adjusted for appeal (or quality) differences across products within the firm by properly deflating revenue with the exact plant level price index, $P_{ft} = \left(\sum_{\Omega_t^f} d_{fjt}^{\sigma} p_{fjt}^{1-\sigma}\right)^{\frac{1}{(1-\sigma)}}$. Since the index depends on unobservable σ and $\{d_{fjt}\}$ and thus cannot be constructed readily from observables, we use Redding and Weinstein's (2017) Unified Price Index (UPI) approach as the appropriate empirical analogue or our theoretical price index. The UPI adjusts prices to take into account the evolution of the distribution of in-plant product appeal shifters, emanating both from changes in appeal for continuing products and the entry/exit of products.

In particular, the UPI logs change in $f's$ price index is given by:

$$\ln \frac{P_{ft}}{P_{ft-1}} = \sum_{\Omega_{t,t-1}} \ln \left(\frac{p_{fjt}}{p_{fjt-1}}\right)^{\frac{1}{\left\|\Omega_{t,t-1}\right\|}} + \frac{1}{\sigma-1}\left(\ln \lambda_{ft}^{QRW} + \ln \lambda_{ft}^{Qfee}\right) \tag{16}$$

where $\Omega_{t,t-1}^f$ is the set of goods produced by plant f in both period t and $t-1$. $\lambda_{ft}^{Qfee} = \frac{\sum_{\Omega_{t,t-1}^f} s_{fjt}}{\sum_{\Omega_{t,t-1}^f} s_{fjt-1}}$ is Feenstra's (2004) adjustment for within-plant appeal changes from the entry/exit of products. $\lambda_{ft}^{QRW} = \prod_{\Omega_{t,t-1}} \left(\frac{s_{fjt}^*}{s_{fjt-1,\Omega_{t,t-1}^f}^*}\right)^{\frac{1}{\left\|\Omega_{t,t-1}\right\|}}$ is Redding-Weinstein's adjustment for changes in relative appeal for continuing products within the plant, which deals with consumer valuation bias that affects traditional approaches to the empirical implementation of theory motivated price indices.[22] The derivation of the UPI price index is presented

[22] Sato (1976) and Vartia (1976) show how the theoretical price index can be implemented empirically under the assumption of invariant firm appeal shocks and constant baskets of goods. Feenstra (2004) derives an empirical adjustment of the Sato-Vartia approach that takes into account changing baskets of goods, keeping the assumption of a constant firm appeal distribution for continuing products. It is this last assumption that the UPI relaxes.

in Appendix A. The derivation requires imposing the normalization that $\sum_{\Omega_{t,t-1}^{f}} \ln d_{fjt}^{\frac{1}{\left\|\Omega_{t,t-1}\right\|}} = 0$. That is, the UPI adjusts for relative appeal changes within the plant, while average appeal changes for the plant are captured by d_{ft}.

Building recursively from a base year B and denoting $\overline{P_{ft}^{*}} = \prod_{l=B+1}^{t}\left[\prod_{\Omega_{t,t-1}}\left(\frac{p_{fjt}}{p_{fjt-1}}\right)^{\frac{1}{\left\|\Omega_{t,t-1}\right\|}}\right]$, $\Lambda_{ft}^{QRW} = \prod_{l=B+1}^{t}\left[\left(\lambda_{fl}^{QRW}\right)\right]$ and $\Lambda_{ft}^{Qfee} = \prod_{l=B+1}^{t}\left[\left(\lambda_{fl}^{Qfee}\right)\right]$, we obtain:

$$\begin{aligned} P_{ft} &= P_{fB} * \overline{P_{ft}^{*}} * \left(\Lambda_{ft}^{QRW}\Lambda_{ft}^{Qfee}\right)^{\frac{1}{\sigma-1}} \qquad (17) \\ &= P_{fB} * \overline{P_{ft}^{*}} * \left(\Lambda_{ft}^{Q}\right)^{\frac{1}{\sigma-1}} \end{aligned}$$

where P_{fB} is the plant-specific price index at the plant's base year B. We initialize each plant's price index at P_{fB}, which takes into account the average price level in year B and the deviation of plant $f's$ product's prices from the average prices in the respective product category in that year. Details are provided in Appendix A.

From (17), to move from our calculated $\overline{P_{ft}^{*}}$ to the exact price index P_{ft}, we need to adjust for the factor $\left(\Lambda_{ft}^{Q}\right)^{\frac{1}{\sigma-1}}$, which depends on σ. In turn, the estimation of σ requires information on P_{ft} (see section 4). We thus work initially with $\overline{P_{ft}^{*}}$ and carry the adjustment factor $\left(\Lambda_{ft}^{Q}\right)^{\frac{1}{\sigma-1}}$ into the derivations of section 4, where its contribution to price variability is flexibly estimated. In particular

$$Q_{ft}^{*} = \frac{R_{ft}}{P_{fB}\overline{P_{ft}^{*}}} = Q_{ft} * \left(\Lambda_{ft}^{Q}\right)^{\frac{1}{\sigma-1}} \qquad (18)$$

We take advantage of this expression in estimating both the production and demand functions using observables. We similarly obtain a measure of materials by deflating material expenditure by plant-level price indices for materials, pm_{ft}, using information on prices and quantities of material inputs at the detailed product class level. We construct pm_{ft} using an analogous approach to that used to construct output prices. See Appendix A for details.

In an alternative approach against which we compare our baseline quality-adjusted prices (adjusted for quality differences within the firm), we examine the robustness of our results to using "statistical" price indices based on either constant baskets of goods, or on divisia approaches, and to the Sato-Vartia-Feenstra approach. These are discussed in section 6.3.

4 Estimating $TFPQ$ and demand shocks

Calculating $TFPQ$ and demand shocks requires estimating the production and demand functions, 1 and 14. Once the coefficients of these functions have been estimated, $TFPQ$ is the residual from 1 and the demand shock is the residual from 14.

We implement a joint estimation procedure. Jointly estimating the two equations allows us to take full advantage of the information to which we have access to separate supply from demand in the data. As a result, we can estimate production rather than revenue elasticities, even for multiproduct plants, and simultaneously obtain an unbiased estimate of σ. We impose a set of moment conditions that requires less structure overall, and weaker restrictions on the covariance between $TFPQ$ and demand shocks, than other usual estimation methods of the demand-supply system. This is in part possible thanks to the fact that we have access to price and quantity information for both inputs and outputs. Data on inputs informs the estimation directly about the production side, thus allowing us to separate it from demand under weaker restrictions than if we only used information on prices and quantities for outputs (as in, for instance, Broda and Weinstein, 2006, or Hottman, Redding and Weinstein, 2016). On the production side, data on prices allows us to properly both production revenue elasticities.

Beyond the usual simultaneity biases and restrictions on supply vs demand , the estimation of 1 and 14 faces the problem that, until we have an estimate of σ, we are unable to properly construct P_{ft}, and thus $Q_{ft} = \frac{R_{ft}}{P_{ft}}$ (see section 3.2). We therefore need to rely on P_{ft}'s two separate components: $\overline{P^*_{ft}}$ and Λ^Q_{ft}. We proceed in three steps to address this limitation (details provided further below):

1. Jointly estimate the coefficients of the production function 1 and the

demand function 14, using $Q^*_{ft} = \frac{R_{ft}}{P_{fB}\overline{P^*_{ft}}} = Q_{ft} * \left(\Lambda^Q_{ft}\right)^{\frac{1}{\sigma-1}}$ and $\overline{P^*_{ft}} = P_{ft}\Lambda^Q_{ft}$ as the respective dependent variables / regressors of these two functions. We carry Λ^Q_{ft} as a separate regressor in each equation to deal with potential biases from the measurement error induced by the–at this point–still partial estimation of revenue deflators. Similarly introduce separately M^*_{ft} and Λ^M_{ft} in the production function (where $M_{ft} = \frac{\text{materials expenditure}}{PM_{fB}\overline{PM^*_{ft}}}$, and Λ^M_{ft} is the adjustment factor for the prices of materials analogous to Λ^Q_{ft} see footnote 20). The joint estimation is conducted separately for each three-digit sector.

2. Use the estimated demand elasticity $\widehat{\sigma}$ for the respective three-digit sector to obtain $P_{ft} = P_{fB} * \overline{P^*_{ft}} * \left(\Lambda^Q_{ft}\right)^{\frac{1}{\widehat{\sigma}-1}}$ and subsequently $Q_{ft} = \left(\frac{R_{ft}}{P_{ft}}\right)$. Proceed in an analogous way to obtain a quantity index for materials, M_{ft}.

3. Using P_{ft}, Q_{ft}, M_{ft} (now properly estimated) and the estimated coefficients of the production and demand functions, obtain residuals $TFPQ_{ft}$ and D_{ft}. We note that, in estimating $TFPQ_{ft}$ and D_{ft} as residuals at this stage, we first deviate P_{ft}, Q_{ft}, M_{ft}, L_{ft} and K_{ft} from sector*year effects, so that from this stage on, only idiosyncratic variation in $TFPQ_{ft}$ and D_{ft} is considered.

We now explain step 1 in detail.

4.1 Joint production-demand function estimation

We jointly estimate the log production and demand functions:

$$\ln Q_{ft} = \alpha \ln K_{ft} + \beta \ln L_{ft} + \phi \ln M_{ft} + \ln A_{ft} \tag{19}$$

and

$$\ln P_{ft} = \alpha - \frac{1}{\sigma} \ln Q_{ft} + \ln D_{ft} \tag{20}$$

where $Q_{ft} = \left(\frac{R_{ft}}{P_{ft}}\right)$. Using $Q^*_{ft} = \frac{R_{ft}}{P_{fB}\overline{P^*_{ft}}} = Q_{ft} * \left(\Lambda^Q_{ft}\right)^{\frac{1}{\sigma-1}}$ (equation 18), these two equations can be rewritten:

$$\ln Q^*_{ft} = \alpha \ln K_{ft} + \beta \ln L_{ft} + \phi \ln M^*_{ft} + \frac{1}{\sigma-1} \ln \Lambda^Q_{ft} - \frac{\phi}{\sigma-1} \ln \Lambda^M_{ft} + \ln A_{ft} \tag{21}$$

and

$$\ln \overline{P^*_{ft}} = \alpha - \frac{1}{\sigma}\left(\ln Q^*_{ft} + \ln \Lambda^Q_{ft}\right) + \ln D_{ft} \tag{22}$$

We estimate 21 and 22, which are transformations of the original production and demand functions, rather than those original forms.

The usual main concern in estimating these functions is simultaneity bias. In the production function, this is the problem that factor demands are chosen as a function of the residual A_{ft}. A standard approach to deal with this problem is the use of proxy methods, as in Ackerberg, Caves and Frazer (2015, ACF henceforth), De Loecker and Warzinski (2012) and many others. In the demand function, simultaneity arises because both price and quantity respond to demand shocks. Usual estimation approaches rely on assumptions regarding orthogonality between demand and supply shocks at some particular level. Foster et al (2008) use $TFPQ$ estimated at a previous stage as an instrument for Q in the demand function, effectively imposing orthogonality between the levels of $TFPQ$ and demand shocks. Broda and Weinstein (2006) and Hottman, Redding and Weinstein (2016) impose orthogonality between double-differenced demand and marginal cost shocks.

We build on these approaches to estimate 21 and 22, but take advantage of the unique access to prices and quantities on both inputs and outputs, and the consequent possibility of jointly estimating the two equations, to relax the assumptions about covariance between demand and supply shocks that identify the elasticity of substitution. We rely on flexible laws of motion for both $TFPQ$ and demand shocks:

$$\begin{aligned} lnA_{ft} &= \pi^A_0 + \pi^A_1 lnA_{ft-1} + \pi^A_2 lnA^2_{ft-1} + \pi^A_3 lnA^3_{ft-1} + \xi^A_{ft} \\ lnD_{ft} &= \pi^D_0 + \pi^D_1 lnD_{ft-1} + \pi^D_2 lnD^2_{ft-1} + \pi^D_3 lnD^3_{ft-1} + \xi^D_{ft} \end{aligned}$$

That is, ξ^A_{ft} is the stochastic component of the innovation to $TFPQ$. Given this structure, our identification of production and demand elasticities (α, β, ϕ, σ) uses standard GMM procedures, imposing the following set of moment conditions (further details provided in Appendix F):

$$E\begin{bmatrix} lnM^*_{ft-1} \times \xi^A_{ft} \\ lnL_{ft} \times \xi^A_{ft} \\ lnK_{ft} \times \xi^A_{ft} \\ lnD_{ft-1} \times \xi^A_{ft} \\ \ln A_{ft} \\ \ln D_{ft} \end{bmatrix} = 0 \tag{23}$$

As in ACF-based methods, we assume that, depending on whether inputs are freely adjusted or quasi-fixed, they respond to stochastic innovations to $TFPQ$ contemporaneously or with a lag, respectively. We assume that materials are freely adjusted while the demand for capital and labor is assumed quasi-fixed. Thus, in 23 we impose lagged materials demand to be orthogonal to current $TFPQ$ innovations, while L and K are required to be contemporaneously orthogonal to ξ^A_{ft}. The assumption that K is quasi-fixed is standard, as is that indicating that M adjusts freely.[23] L is also assumed quasi-fixed in our context because important adjustment costs have been estimated for the Colombian labor market (e.g. Eslava et al. 2013). In fact, when we estimate factor elasticities allowing L to adjust freely results are frequently implausible (e.g. negative estimated elasticities of production to labor), yielding further support to our assumption.

The condition that D_{ft-1} must be orthogonal to ξ^A_{ft} identifies σ. Orthogonality between demand and technology shocks in levels has been used to identify demand elasticities by Foster et al (2008, 2016) and Eslava et al (2013), following the logic that the slope of the demand function can be inferred taking advantage of shocks to supply. However, assuming orthogonality in levels (that is, between A_{ft} and D_{ft}) has been criticized on the basis that firms may endogenously invest in quality when they perceive better returns (potentially because they have higher $TFPQ$) and that demand shifters may be correlated with $TFPQ$ shocks if greater quality is more difficult to produce.[24] Hottman, Redding and Weinstein (2016) and Broda and

[23] For $\ln M_{ft-1}$ to be useful in the identification of ϕ, it must be the case that input prices are highy persistent. The AR1 coefficient for log materials prices is 0.95 in our sample.

[24] R&D decisions that are endogenous to current profitability and affect future profitability, for instance, are present in Aw, Roberts and Xu, 2011. Their framework does not separately identify the demand and technology components of profitability, but both

Weinstein (2006, 2010) partly address these criticisms by imposing orthogonality only between double-differenced demand and supply shocks (double differencing over time and varieties). Imposing the orthogonality of the double-differenced shocks is still a strong assumption. Given our ability to specify demand and production separately given the price and quantity data of both output and inputs, we impose $E(\ln D_{ft-1} \times \xi_{ft}^{A})$ which permits a correlation between changes in $TFPQ$ and demand even within the plant. While we are still taking advantage of shocks to the supply curve to identify the elasticity of demand, we only require that innovations in technical efficiency in period t be orthogonal to demand shocks in $t-1$.

Notice also that $TFPQ$ obtained as a residual from quality-adjusted Q is stripped of apparent changes in productivity related to within-firm appeal changes, eliminating a source of correlation between appeal and efficiency stemming from measurement error. Moreover, since we use plant-specific deflators for both output and inputs, our estimation is not subject to the usual bias stemming from unobserved input prices (De Loecker et al. 2016).[25]

We implement this estimation separately for each three digits sector of ISIC revision 3.[26] We obtain plausible factor elasticities for almost all sectors at the three digits sector, which is an encouraging sign of the suitability of our method and data since proxy methods are usually implemented in estimations at the two-digit level, and frequently yield implausible results–in particular negative estimated factor coefficients for several sectors–at finer

could plausibly respond dynamically. In turn, the idea that quality is more costly to produce appears in Fieler, Eslava, and Xu (2018), to characterize cross sectional correlations between quality and size.

[25]De Loecker et al (2016), use plant-level deflators for output but not for inputs. This induces a bias stemming from unobserved input price heterogeneity, that they address by including plant level output prices as controls in their estimation of the production function, under the assumption that output prices enter the determination of input prices. Furthermore, they address the within-plant aggregation issue by constraining their estimation of the production function to uniproduct plants, where output quantity is observed and well defined. The issue of how to properly compare units of output of different products across plants, however, remains unresolved. Our approach points that appeal shifters D_{fj} (and thus quality adjustment of output *across plants)* addresses this issue.

[26]More precisely, we use the official Colombian-adapted ISIC (CIIU for its Spanish acronym), revision 3. The data are originally codified using ISIC revision 2 until 1997 and revision 3 from 1998 onwards. We use the official correspondence tables to obtain a consistent codification over time. At the three digit level the correspondence is almost one-to-one. To solve the few cases in which it is not, we lump together a few sectors We end up with 23 three-digits sectors.

Table 1. Factor and demand elasticities						
Sector	**lnL**	**lnK**	**lnM**	**sigma**	**Returns to scale in production**	**Returns to scale in revenue**
Average	0,45	0,20	0,44	3,10	1,09	0,63
Min	0,12	0,05	0,01	1,23	0,95	0,23
Max	0,91	0,57	0,75	7,59	1,29	0,90

levels of disaggregation. Still, if fully unconstrained, our joint estimation does deliver implausible results for a few sectors. In particular, the unconstrained estimation yields increasing revenue returns to scale for four (out of 23) three-digits sectors, and negative factor coefficients in production for two sectors. We thus further constrain returns to scale in revenue to be 0.9 or less.[27] We test and discuss the robustness of our results to changing this constraint in sensitivity analysis below. Revenue returns to scale estimated or imposed in the literature usually range between 0.67 and 0.85. In HK, the combination of CRS in production, CES demand and an elasticity of substitution of 3 implies a returns to scale parameter of 0.67 in the revenue function.

The estimated factor and demand elasticities are summarized in table 1 and listed in Appendix I. Our results reveal slightly increasing returns to scale in production at the three-digits sector level for most sectors. The estimated elasticity of substitution stands at an average of 3.15, and varies substantially across sectors, from 1.23 for plastics to 7.59 in processed food. Returns to scale in revenue stand at an average 0.63 (0.7 ignoring sectors that hit the 0.9 bound). While our average estimated curvature is not far from that imposed by HK, there is substantial dispersion across three-digits sectors. We show below how ignoring this heterogeneity surprisingly dampens the estimated contribution of wedges to sales variability.

[27] Only sectors for which this is violated in the uncostrained estimation are re-estimated imposing the constraint. We still obtain a negative coefficient for labor in production for one sector and an elasticity of substitution below one for another sector. For these two sectors, we impose the full set of factor and substitution elasticites estimated for the closest sectors. We also conduct robustness analysis in appendix C.

5 Results

5.1 Outcome growth over the life cycle

We use the estimated demand elasticity $\widehat{\sigma}$ to construct $\ln P_{ft} = \ln\left(P_{fB}\overline{P^*_{ft}}\right) + \frac{1}{\widehat{\sigma}-1}\ln\Lambda^Q_{ft}$ and subsequently recover $Q_{ft} = \frac{R_{ft}}{P_{ft}}$. We proceed in an analogous way to construct pm_{ft} and M_{ft}.[28] To build idiosyncratic life cycle growth in revenue, $\frac{R_{ft}}{R_{0t}}$, we first deviate revenue from sector*year effects and then obtain the ratio of current to initial (idiosyncratic) revenue. All other utcome variables, in particular employment, capital, materials, output prices and input prices are also stripped from sector*year effects before building life cycle growth ($\frac{Z_{ft}}{Z_{0t}}$ for each variable Z). Also, when building $TFPQ$, D, and μ we only exploit idiosyncratic (i.e. within sector*year) variation in the levels of outcomes. That is, from this point, we will be dealing exclusively with the idiosyncratic component of life cycle growth, for both outcome and fundamental variables.[29]

We define *age* as the difference between the current year, t , and the year when the plant began its operations, and define the plant's revenue (or other outcome) level at birth R_{f0} as the average for ages 0 to 2. By averaging over the plant's first few years in operation we deal with measurement error coming, for instance, from partial-year reporting (e.g. if the plant was in operation for only part of its initial year).

The solid black lines in Figure 1 present mean growth from birth for output, sales and employment. As in the rest of figures throughout the paper, we use a logarithmic scale. The average establishment in our sample grows by a factor of 2.3 in terms of output by age 10, and almost 6 times by age 25.[30] Average life-cycle revenue growth is more modest, growing four-fold rather than six-fold by age 25. For comparison with existing literature on life-cycle growth, the lower panel presents analogous results for employment: $\frac{L_{ft}}{L_{0t}}$. By age 10 the average establishment has almost doubled it employment,

[28] I.e. we use the same measurement approach incorporating multi-materials inputs to construct the plant-level deflator for materials, and use it to deflate expenditures in materials to arrive at materials inputs. We use the same elasticity of substitution at the sectoral level for this purpose.

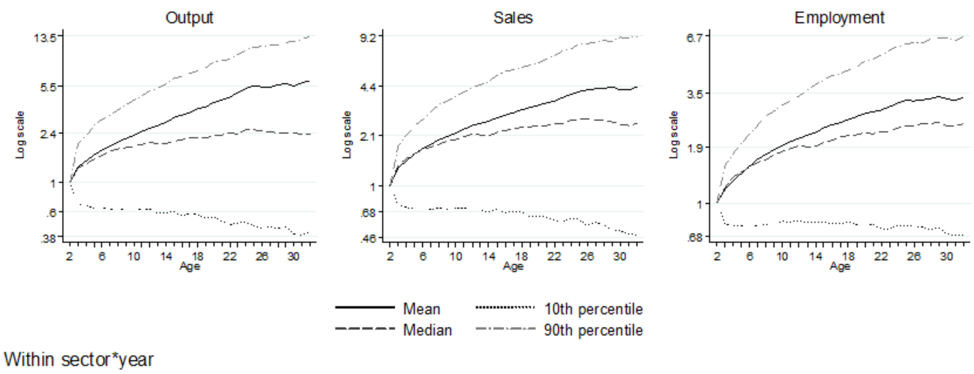
[29] We also winsorize life cycle growth for each variable at 1% and 99% to eliminate outliers that may drive the results of our decompositions.

[30] More precisely, $\frac{Q_{fa}}{Q_{f0}} = 1.63$ when $a = 5$, $\frac{Q_{fa}}{Q_{f0}} = 2.35$ when $a = 10$, and $\frac{Q_{fa}}{Q_{f0}} = 5.57$ when $a = 25$.

and 25 years after birth employment it has grown more than three-fold.[31]

Figure 1: Distribution of life cycle growth

Current to initial variables



These average growth dynamics, however, hide considerable heterogeneity. Median growth (dashed line) falls under mean growth for all panels, highlighting the fact that it is a minority of fast-growing plants that drive mean growth. Related, the distribution of plant growth is highly skewed, displaying a much more marked gap for the 90th-50th percentiles than for the 50th-10th. By age five, for instance, while the average plant has multiplied its output at birth by a 1.63 factor, the plant in the 90th percentile has multiplied it by 2.76, the median plant by 1.51, and the plant in the 10th percentile has shrank to 63% of its original size. At age ten the 90th percentile of life cycle similarly more than doubles the median (4.32 rather than 1.91). Employment and sales growth are characterized by similarly wide dispersion and marked skewness.

Eslava et al. (2018) show that, though dispersion in life-cycle growth across Colombian manufacturers is large and highly skewed towards a dynamic top decile, both dispersion and skewness fall short of that observed in the U.S. This is consistent with the view that less developed economies are characterized by less dynamic post-entry growth. Hsieh and Klenow (2009) and Buera and Fattal (2014) attribute such cross-country differences to institutions that fail to encourage investments in productivity and healthy market selection in developing economies. Identifying the role that specific

[31] For revenue and employment, we have $\frac{R_{fa}}{R_{f0}} = 1.6$ and $\frac{L_{fa}}{L_{f0}} = 1.4$ when $a = 5$, $\frac{R_{fa}}{R_{f0}} = 2.17$ and $\frac{L_{fa}}{L_{f0}} = 1.93$ when $a = 10$, and $\frac{R_{fa}}{R_{f0}} = 4.03$ and $\frac{L_{fa}}{L_{f0}} = 3.22$ when $a = 25$.

institutions play is an interesting area of future research.[32]

We emphasize that we can measure life cycle growth directly using longitudinal data for each plant, rather than relying on cross-cohort comparisons. This approach addresses some of the usual selection concern in the literature of business' life cycle growth. Still, we can only characterize and decompose growth for survivors. Appendix H describes life-cycle growth for exits-to-be, showing that the patterns in Figure 1 are mainly driven by plants that will survive (so the exit bias is small).

5.2 TFPQ and demand shocks

As indicated, $TFPQ_{ft}$ and D_{ft} are recovered as residuals from, respectively, the production function (1) and the demand function (14), using the estimated factor and demand elasticities reported in Table 1, and deviating $Q_{ft,}$ L_{ft}, M_{ft}, and K_{ft} from sector*year effects previously, so that $TFPQ_{ft}$ and D_{ft} contain only idiosyncratic variation. Table 2 presents basic summary statistics for (the idiosyncratic component of) sales and our estimates of output, output prices, $\ln A_{ft}$, $\ln D_{ft}$ and input prices.[33] Idiosyncratic dispersion in sales, output, output prices, $TFPQ$, demand and input prices are all large.

$TFPQ$ is strongly negatively correlated with output prices, which is intuitive to the extent that more efficient production allows charging lower prices. This was also found by Eslava et al. (2013) for an earlier period and using a different approach to measure $TFPQ$ and D. Interestingly, Foster et. al. (2008,2016) find similar correlations between prices and fundamentals using US data for a selected number of commodity-like products. By contrast with those products, endogenous quality may be more relevant in our context.

To the extent that quality is more difficult to produce, demand shocks and technical efficiency may be negatively correlated. This is indeed the case in our estimates. Output exhibits strong positive correlations with $TFPQ$ and demand while sales is especially positively correlated with demand. These

[32] Within-country changes in institutions, either across businesses or over time (or both) offer a fruitful ground for such exploration, to the extent that they keep constant other factors potentially influencing business dynamics, from the macroeconomic environment to business culture. We undertake that exploration for Colombia, taking advantage of changes in import tariffs, in a separate paper.

[33] As explained above, $TFPQ$ and demand shocks are obtained using only the idiosyncratic components of Q, prices and inputs.

Table 2. Descriptive statistics

Panel A. Number of plants, number of products and materials per plant-year									
Number of plants		Number of products per plant				Number of materials per plant			
Total	Avg. year	Avg.	P25	P50	P75	Avg.	P25	P50	P75
23,292	7,670	3.56	1	2	5	11.17	5	9	14

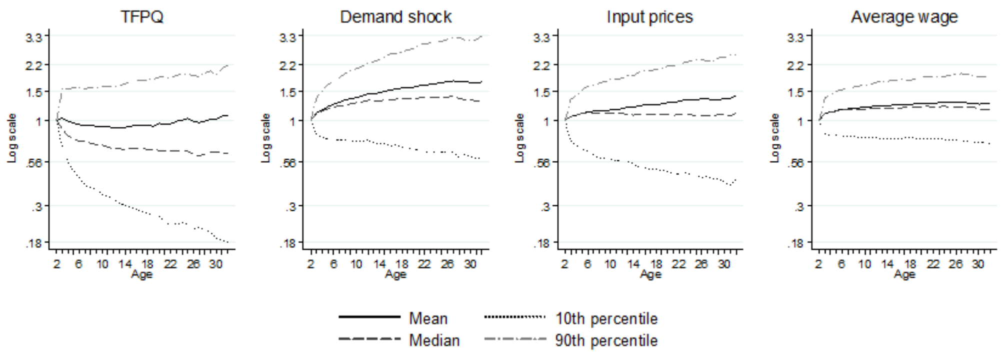
Panel B.Standard deviations and correlation coefficient for outcomes and fundamentals (within sector*year, all variables in logs)								
	Standard Deviation	Sales	Output	Output prices	TFPQ	Demand Shock	Input prices	Average wage
Sales	1.438	1						
Output	1.611	0.89	1					
Output prices	0.736	0.007	-0.451	1				
TFPQ	0.874	0.135	0.464	-0.752	1			
Demand Shock	0.758	0.722	0.42	0.493	-0.243	1		
Input prices	0.693	-0.036	-0.095	0.136	0.155	0.045	1	
Average wage	0.414	0.603	0.517	0.045	0.099	0.477	0.003	1

basic correlation patterns remain true for within-plant correlations, and are echoed in our growth decompositions below. Forlani et al. (2018) also find $TFPQ$ and demand to be negatively correlated.

The within sector*year distributions of the evolution over the life cycle of fundamentals are displayed in Figure 2, including the life cycle growth of $TFPQ$ and demand shocks, $\widehat{A_{ft}}$ and $\widehat{D_{ft}}$, as well as that of material input prices and wages. The average growth of demand shocks dominates that of input prices, and both dominate the average growth of $TFPQ$ over the life cycle. By age 25, $TFPQ$ has barely grown on compared to birth on average, while the demand shifter has grown on average close to two-fold. Part of what is driving the contradicting patterns in Figure 2 is the evolution of the negative correlation between the life cycle growth of $TFPQ$ and that of demand shocks. At age 3, the correlation is -0.152, at age 10, -0.264 and by age 20, -0.324. The rapid rise of product appeal/quality over the life cycle comes at the cost of dampening the growth of $TFPQ$. The interplay between output prices and demand shocks is also interesting: with growing output over the life cycle, downward sloping demand would imply that the plant would have to charge ever shrinking prices over its life cycle, unless the appeal of f to costumers changed over time. We do not observe such fall in output prices, signaling increasing ability of the firm to sell more at given prices. By construction, this is what the life cycle growth of the demand shock, $\widehat{D_{ft}}$, captures.

Figure 2: Distribution of fundamentals

Current to initial



Within sector*year

5.3 Decomposing growth into fundamental sources

We now decompose the variance of $\frac{R_{ft}}{R_{f0}}$ and $\frac{Q_{ft}}{Q_{f0}}$ into contributions associated with different fundamental sources, most notably $TFPQ$ and demand shocks (equations (5) and (6)). We follow a two stage procedure, similar to that in Hottman et al. (2016), but implement two variants of it: a structural decomposition and a reduced form decomposition. We summarize each in this section. Details are provided in Appendix G.

Structural decomposition: As shown in Appendix G, the contribution of each (log) fundamental to the variance of (log) sales equals the ratio of its covariance with sales to the variance of sales, multiplied by its structural parameter in equation 6, reproduced below:.

$$\frac{R_{ft}}{R_{f0}} = \left(\frac{d_{ft}}{d_{f0}}\right)^{\kappa_1} \left(\frac{a_{ft}}{a_{f0}}\right)^{\kappa_2} \left(\frac{pm_{ft}}{pm_{f0}}\right)^{-\phi\kappa_2} \left(\frac{w_{ft}}{w_{f0}}\right)^{-\beta\kappa_2} \left(\frac{\mu_{ft}}{\mu_{f0}}\right)^{-\gamma\kappa_2} \left(\widehat{\chi}_t \chi_{ft}\right)^{1-\frac{1}{\sigma}} \tag{24}$$

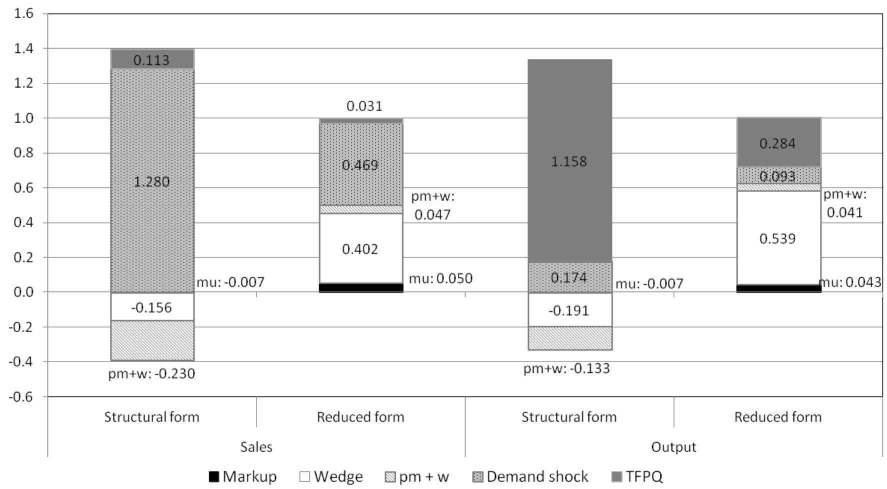
where $\kappa_1 = \frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}$, $\kappa_2 = \left(1-\frac{1}{\sigma}\right)\kappa_1$, and γ and σ have been estimated as explained above. The term $\left(\widehat{\chi}_t \chi_{ft}\right)^{1-\frac{1}{\sigma}}$ is calculated as a residual, since all of the other components are either measured or estimated. From equation 6 , error term $\ln \chi_{ft}$ captures life cycle growth in wedges, including distortions from regulations, adjustment costs, and other factors, and measurement error. Because these wedges simply reflect the gap between actual growth and that predicted by fundamentals through the lens of our model, they reflect all sources for such gaps, including some that may be correlated with fundamentals themselves. Thus, these wedges may imply exacerbated growth if plants with better fundamentals also exhibit higher wedges than plants with worse fundamentals, or dampened growth in the opposite case. We conduct an analogous decomposition for output, following equation (5).

The first bar of Figure 3 depicts the result of this decomposition, pooling across ages, and reporting the contributions of material prices and wages together to simplify the figure. We find that the structural contribution of fundamentals explains the bulk of sales growth over the life cycle. Taken together, fundamentals in fact account for more than 100% of the variance of growth across plants within a sector (a fact we turn to further below). The demand shock is ten times as important as $TFPQ$ to explain idiosyncratic sales growth (or quality adjusted output growth). Input prices make smaller, but far from negligible, contributions. This reflects the fact that, pooling across ages, the covariance of demand shocks growth with sales growth is almost five-fold that between $TFPQ$ growth and sales growth (Table 3). The significant negative correlation between $TFPQ$ and demand shocks undoubtedly plays a role in this fact. In the case of markups growth, its contribution to the variance of sales growth is minimal, not even visible in the graph, reflecting market shares concentrated around zero in most sectors.mantribution of $TFPQ$ for output growth volatility as compared to sales is not surprising, the fact that demand shocks still account for almost 20% of real output growth volatility is interesting, especially in a context where real output growth has been adjusted for within plant changes in product mix and quality.

The dominance of demand-side fundamentals over supply side in explaining the variance in sales resonates with recent findings in the literature (Hottman et al. 2016, Foster et al. 2016). It is, however, noticeable that this finding survives the expansion of the measurement framework to explicitly account for wedges. The availability of price and quantity data together with data on input use, rare in the literature and enabled by the richness of the Colombian data, is crucial to identify wedges from the gap between actual growth and that predicted by fundamentals (see detailed discussion in section 6).

Input prices, especially that of labor, also play a dampening role for the variability of sales. This is consistent with Table 2 that shows a positive correlation between input prices and wages in particular with $TFPQ$ and demand. The variation in wages across plants might reflect many factors. For example, it may reflect the geographic segmentation of labor markets as

Figure 3: Life-cycle growth variance decomposition



well as institutional barriers in the labor market. However, the correlations in Table 2 with the accompanying dampening implications suggest that some of this might reflect rent sharing but it also might reflect unmeasured quality differences. We deal with quality differences for materials inputs by building a quality-adjusted deflator, but not for labor, which is not broken down by skill categories in the Manufacturing Survey for the long period covered by our estimations. To address the relative importance of these two possible sources of sales variance arising from wages, we take advantage of data on broad skill categories available for 2000-2012 and construct quality-adjusted wages and a quality-adjusted labor input given by the payroll deflated with our adjusted wages. Skill categories are production workers without tertiary education, production workers with tertiary education and administrative workers. Implementing our decomposition with this alternative measure of wages rather than the average wage per worker (Table J1, appendix J) reduces the negative contribution of wages for 2000-2012 from -0.128 to -0.058, suggesting that increasing labor quality explains about half the dampening role of wages over the variance of sales. Moreover, consistent with this interpretation, we find that accounting for labor quality reduces the positive contribution of $TFPQ$ by about the same amount as the decrease in the negative contribution of wages. In turn, there is virtually no impact on the contribution of wedges, demand or other factors.

A striking feature of these results is that the wedge contributes negatively to the variance of life cycle growth of both output and sales (or quality adjusted output). That is, the different sources of wedges captured in this term dampen the effect of fundamentals growth on outcome growth, implying that high-productivity high-appeal plants grow less relative to low-productivity and appeal plants than their respective fundamentals would imply. The effect is quantitatively large: sales dispersion is dampened by about 15% with respect to that implied by fundamentals. The corresponding figure for output growth is about 20%. That is, Colombian manufacturing plants face significant size-correlated wedges that de-link actual growth from the fundamental attributes of plants.

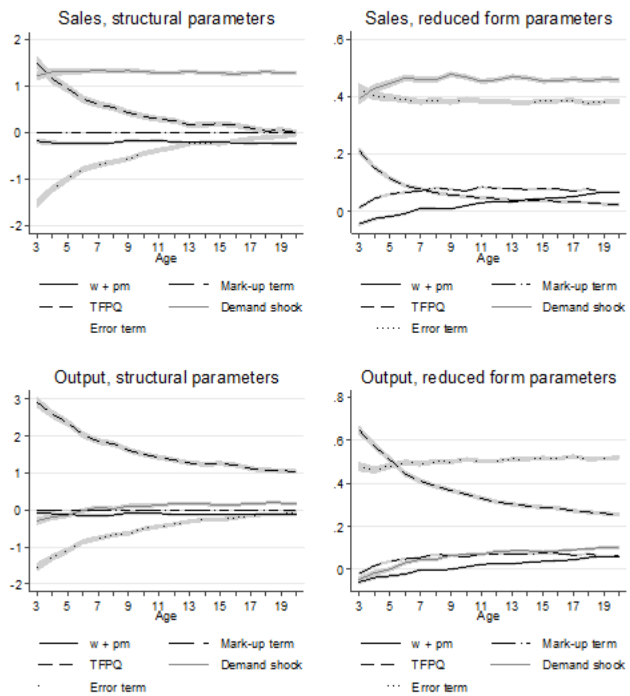
The contributions of these different factors to sales and output life cycle growth vary significantly depending on the horizon of growth considered. The left panels of Figure 4 display results of the structural decomposition separately for different ages.[34] For both sales and output, demand becomes increasingly important compared to $TFPQ$ over longer horizons. This is because, although the covariance between sales growth and both $TFPQ$ growth and the growth of demand shocks increases as plants age, the latter does so at a much faster speed (Table 3). These patterns echo the increasing negative correlation between $TFPQ$ and demand shocks over the life cycle. Wedges, interestingly, play a more important dampening role at the youngest ages. That is, wedges dampen output and sales variability compared to that implied by fundamentals more among young plants than among older ones (left panels of Figure 4). Appendix H shows that these general patterns are robust to selection, in the sense of being similar for survivors-to-be and exits-to-be. $TFPQ$ plays a relatively more important role vis-a-vis demand for the latter than the former.

Figure 5 shows the mechanics behind the negative contribution of structural wedges: the gap between actual growth (black solid line) and that explained by fundamentals (grey solid line) is positive for plants with low predicted growth and negative for those in the highest percentiles of predicted growth.[35] Predicted growth corresponds to growth in equation (24)

[34] To conduct the decomposition by ages, we expand equations the decomposition equations to include interactions with the different age groups. See Appendix G for details.

[35] The 1% tails on both sides are excluded from the figure because they tend to dominate the scale of the figure, rendering it useless to illustrate the point. Figure 6 shows that the outliers in the distribution of predicted growth are not generated by extreme estimates of fundamentals, but by the fact that, with high returns to scale in revenue for some sectors, the model would predict extreme sizes for the best performing plants in those sectors.

Figure 4: Life-cycle growth variance decomposition by age



Within sector*year.

setting $\chi_{ft} = 0$. Figure 5 implies that plants with weak growth in fundamentals are implicitly subsidized while those with strongest fundamentals are implicitly taxed, especially at young ages.

Figure 6 indicates that plants in the highest percentiles of predicted growth have both higher demand and higher $TFPQ$ than those with low predicted growth. Interestingly, the superstar plants (those in the upper quartile of growth in fundamentals) differ from the rest most clearly in terms of the growth of demand. For the rest of the distribution, $TFPQ$ growth is at least as important as demand growth to explain the difference between the worst and the not-so-bad plants.

Since the error term in equation (24) reflects both wedges to profitability that may be correlated to fundamentals and others that are not, it is interesting to uncover the full contribution of fundamentals, bringing together that

Appendix I shows the equivalent of Figure 5 without eliminating 1% tails.

Table 3: moments of the distribution of life cycle growth for sales, demand shocks and TFPQ and structural coefficients of the decomposition of growth (pooling across ages and sectors)

	Age=all	Age=5	Age=10	Age=20
Cov(TFPQ,Revenue)	0.062	0.040	0.071	0.087
Cov(Demand, Revenue)	0.274	0.062	0.190	0.351
Var(Revenue)	0.688	0.156	0.468	0.859
Var(TFPQ)	0.527	0.145	0.403	0.676
Var(D)	0.241	0.054	0.171	0.333
Structural coefficients (average sector)	Kappa1	Kappa2	Gamma	Sigma
	3.806	2.598	1.080	3.151

implied by our model and that stemming from the impact of fundamentals on our structural wedges. Wedge sources potentially correlated with fundamentals may arise from size-dependent policies, adjustment costs and endogenous financial constraints. Wedges that are orthogonal to fundamentals may come from horizontal regulations and measurement error. To decompose the role of orthogonal vs. correlated wedges, we estimate the full contribution of fundamentals by implementing the following reduced form decomposition:

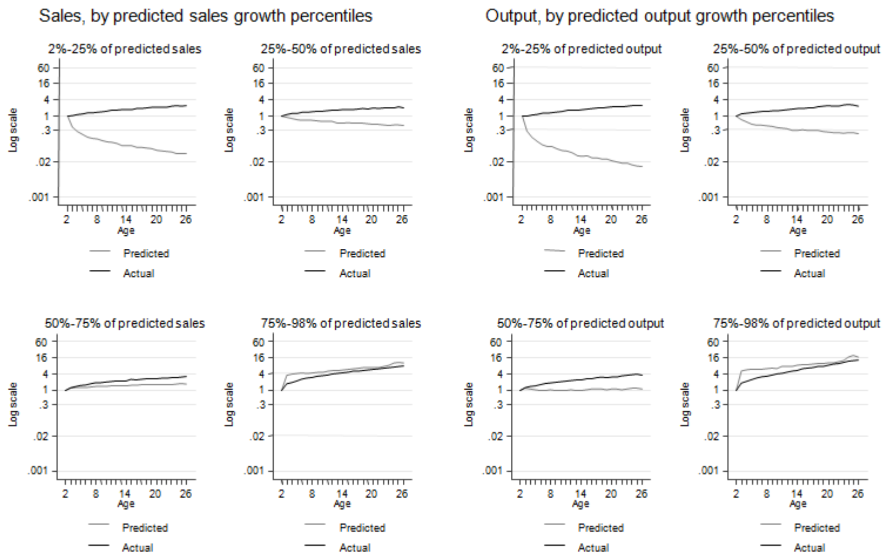
Reduced form decomposition: The contribution of each (log) fundamental to the variance of (log) sales equals the ratio of its covariance with sales to the variance of sales, multiplied by its reduced form parameter in the following equation, estimated by OLS:

$$\begin{aligned}\ln \frac{R_{ft}}{R_{f0}} &= \beta_d^r \ln\left(\frac{d_{ft}}{d_{f0}}\right) + \beta_a^r \ln\left(\frac{a_{ft}}{a_{f0}}\right) + \beta_m^r \ln\left(\frac{pm_{ft}}{pm_{f0}}\right) \\ &\quad + \beta_w^r \ln\left(\frac{w_{ft}}{w_{f0}}\right) + \beta_\mu^r \ln\left(\frac{\mu_{ft}}{\mu_{f0}}\right) + \varepsilon_{ft}\end{aligned}$$

The residual term of this OLS estimation is orthogonal to the fundamentals by construction, and thus captures only uncorrelated wedges. As a result, the reduced form decomposition assigns to each fundamental the role it plays directly (i.e. its "structural" role) and also that it plays indirectly through its effect on wedges and its correlation with other fundamentals. Covariances between fundamentals are assigned equally to the contribution of the different fundamentals. [36]

[36] We find that the structural wedge has a correlation of -0.30 with $TFPQ$ and -0.13 with,

Figure 5: Contribution of fundamentals to life-cycle growth



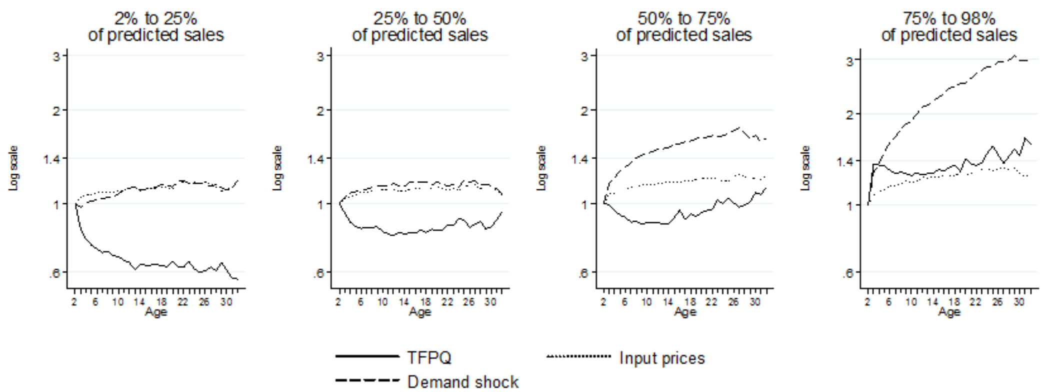
Idiosyncratic components

Results of this alternative exercise are shown in columns 2 and 4 of Figure 3, and in the right panels of Figure 4. The uncorrelated wedge term contributes positively to the variance of outcome growth. In particular, it explains 40% of sales growth dispersion and 53% of output growth dispersion. It is also interesting that, in transiting from the reduced form to the structural decomposition, the contribution of $TFPQ$ grows by (proportionally) more than that of demand shocks. To the extent that (negatively) correlated distortions are reflected in our structural wedges but not in the reduced form ones, this suggests that such distortions are most strongly correlated with $TFPQ$, distorting the return to technical efficiency more than that to quality/appeal.

demand shocks consistent with our interpretation of the structural wedges being negatively correlated with fundamentals. In contrast, the reduced form wedge has essentially zero correlation with the fundamentals.

Figure 6: Life-cycle fundamentals growth

By predicted sales growth percentiles



6 Robustness and the Value Added from Building Up Jointly from P, Q and inputs data

6.1 Value added of bringing P and Q data to the Hsieh-Klenow framework

Relative to the literature on wedges vs. fundamentals as determinants of size and growth that build on Hsieh and Klenow (HK, 2009), our approach takes advantage of rich data on prices and quantities at the micro level. HK have shown that, in absence of P and Q data, one can estimate the contribution of wedges relative to fundamentals imposing a set of usual assumptions. Our approach directly builds on HK's, but even within that frameworks there is multi-fold value added of the micro price and quantity data on both outputs and inputs. First, the micro price and quantity data permit measurement of $Q_{ft} = \frac{R_{ft}}{P_{ft}}$ directly, so that a production function (as opposed to a revenue function) and a demand structure can both be estimated to obtain production and demand elasticities. These elasticities are themselves key ingredients to determine the role of fundamentals vs. structural wedges, and are therefore widely used when making inferences about the drivers of business performance. In absense of the ability to estimate them, inferences are frequently based on external estimates that correspond to a context not necessarily relevant to the particular application, are broadly aggregated (e.g. the same elasticity of substitution is used for all sectors) and may not be appropriately specified (e.g. revenue function elasticites or cost shares used in place for production function elasticities). Second, estimation of the production and demand structure naturally yield estimates of $TFPQ_{ft}$ and D_{ft}, so that their individual role can be assessed. Third, the price and quantity data for inputs permits identifying the contribution of idiosyncratic input prices to size and growth. Clearly, then, these detailed P and Q data are necessary if one is interested in learning about the separate roles of A_{ft}, D_{ft} and input prices. But, how important is it to have access to such detailed data to answer questions not related to unpackaging these fundamentals? For instance, does having access to P and Q data lead to a different answer to the question of the contribution of wedges vs. "composite" fundamentals?

The latter question has been object of a long-standing literature, much of which builds on insights from Restuccia and Rogerson (2008) and Hsieh and Klenow (2009, 2014). HK, in particular, have shown that, in absence of P and Q data, one can estimate the contribution of wedges relative to fundamentals imposing a set of usual assumptions. Since our structure closely follows that proposed by HK, we now impose HK's assumptions to estimate the role of a composite fundamentals shock without using P and Q data. We then compare such results to those obtained for the same composite fundamental shock but in a scenario where we relax assumptions on parameters, and use the P and Q data to estimate those parameters. We denote the composite measure of fundamentals, which bundles up our $TFPQ$ and D shocks, as $TFPQ\_HK$.[37]

The starting point of this approach is revenue which in our notation is given by: $R_{ft} = D_{ft}Q_{ft}^{1-\frac{1}{\sigma}} = D_{ft}\left(A_{ft}X_{ft}^{\gamma}\right)^{1-\frac{1}{\sigma}}$. Thus, one can obtain the composite shock $TFPQ\_HK$ solely from revenue and input data as:

$$TFPQ\_HK_{ft} = R_{ft}^{1/(1-\frac{1}{\sigma})}/X_{ft}^{\gamma} = A_{ft}D_{ft}^{\frac{1}{1-\frac{1}{\sigma}}} \tag{25}$$

[37] In the appendix to their paper, HK (2009) show how, in the presence of demand shocks, the measure they call $TFPQ$ is actually a composite of the technology and the demand shock. Our expression for the $TFPQ\_HK$ composite shock is exactly the same as their expression (i.e. $TFPQ\_HK$ in this paper is what is called $TFPQ$ by HK). Haltiwanger, Kulick and Syverson (2018) also explore properties of $TFPQ\_HK$ constructed from revenue and input data compared to TFPQ and demand shocks constructed from price and quantity data.

Optimal input demand can be expressed as a function of this composite shock (see Appendix B), so that revenue $R_{ft} = D_{ft}\left(A_{ft}X_{ft}^{\gamma}\right)^{1-\frac{1}{\sigma}}$ can also be expressed solely in terms of this composite shock and other primitives of the model. Life cycle growth in revenue can then be expressed as:

$$\frac{R_{ft}}{R_{f0}} = \left[\left(\frac{TFPQ\_HK_{ft}}{TFPQ\_HK_{f0}}\right)\left(\frac{(1-\tau_{ft})}{(1-\tau_{f0})}\frac{C_{f0}\mu_{f0}}{C_{ft}\mu_{ft}}\right)^{\gamma}\right]^{\frac{1-\frac{1}{\sigma}}{1-\gamma\left(1-\frac{1}{\sigma}\right)}} \tag{26}$$

where C_{ft} was defined above as a Cobb-Douglas composite of the prices of these inputs: $C_{ft} = pm_{ft}^{\frac{\phi}{\gamma}} w_{ft}^{\frac{\beta}{\gamma}} r_{ft}^{\frac{\alpha}{\gamma}}$.[38] This expression implies that we can decompose life cycle sales into its $TFPQ\_HK$ component and a residual component that will reflect wedges, input cost variation and idiosyncratic markup variation. Such a two-way decomposition is feasible with revenue and input data, so far as estimates of demand and factor elasticities are available.

We now assess the contribution of $TFPQ\_HK_{ft}$ growth to sales growth following the expression in (26), under two alternative approaches:

1. Use our estimates of the elasticities of output with respect to production factors, and the implied returns to scale coefficient γ to obtain $X = M_{ft}^{\frac{\phi}{\gamma}} L_{ft}^{\frac{\beta}{\gamma}} K_{ft}^{\frac{\alpha}{\gamma}}$. Subsequently use our estimated $\frac{1}{\sigma}$ and γ to obtain $TFPQ\_HK_{ft} = R_{ft}^{1/(1-\frac{1}{\sigma})}/X_{ft}^{\gamma}$ and obtain the contribution of this composite shock in (26). We call our estimate of $TFPQ\_HK_{ft}$ under this approach "$TFPQ\_HK_{ft}$ unconstrained".

2. Impose the usual assumptions that $\gamma = 1$ and factor elasticities are cost shares to obtain $X = M_{ft}^{\phi_c} L_{ft}^{\beta_c} K_{ft}^{\alpha_c}$, where the subindex c denotes cost share. Subsequently impose a common demand elasticity to obtain $TFPQ\_HK_{ft} = R_{ft}^{1/(1-\frac{1}{\sigma})}/X_{ft}$ and obtain the contribution of this composite shock in (26). We call our estimate of $TFPQ\_HK_{ft}$ under this approach "$TFPQ\_HK_{ft}$ constrained". We use different values of σ for $TFPQ\_HK$ constrained: 1) $\sigma = 3$ used in Hsieh and Klenow (2009); 2) the σ necessary to replicate returns to scale in revenue equal to the average in Table 1, $\left(1-\frac{1}{\sigma}\right) = 0.638$; 3) the σ necessary to replicate returns to scale as in the maximum permitted in Table 1, $\left(1-\frac{1}{\sigma}\right) = 0.9$.

We implement our two-stage decomposition under these different approaches. Table 4 presents both the structural (upper panel) and reduced

[38] This might also include factor specific wedges.

form (lower panel) versions of the decompositions. [39] Starting with the comparison of columns 1 and 2 for the structural decomposition in the upper panel, it is easily observed that the combined contribution of $TFPQ$ and demand in column 2 (which corresponds to our baseline decomposition of Figure 3) is equivalent to the contribution of unconstrained $TFPQ\_HK$. While this result is by construction, comparing columns 1 and 2 highlights the fact that some of what is attributed to wedges in a two-way decomposition in column 1 is due to the contribution of variable input prices and markups in column 2 (25% out of the 40% assigned to wedges in column 1). Thus, the first important inference is that even with the correctly estimated demand and output elasticities, the composite $TFPQ\_HK$ overstates the contribution of wedges.

The left-panel of Figure 7, which reproduces column 1 of Table 4 by age and is to be compared with the upper left panel of Figure 4, shows that the message that correlated wedges affect young plants the most is still present using the HK approach, since the contribution of input prices and markups does not vary significantly over the life cycle. The underlying reasons for input price and markup variability may well be related to factors that a benevolent central planner could help address, such as idiosyncratic benefits from policy, but they may also reflect deeper features of input and output markets, and as such it is unclear that they arise from "distortions" that policy could address.

Turning to the reduced form decomposition (lower panel) Column 2 shows that the composite shock $TFPQ\_HK$ overstates the contribution of orthogonal wedges as well, not only because it attributes to wedges the contribution of input prices and markups, but also because it lumps together $TFPQ$ and demand, and their joint contribution is dampened by their negative correlation.

Compared to our baseline estimates, the constrained $TFPQ\_HK$ also overstates the contribution of wedges in the structural decomposition, with the size of the bias depending on the magnitude of the returns to scale in revenue (which depends on γ and σ). Comparing across columns 3, 4 and

[39] As before, the structural version in the upper panel imposes the respective parameters σ and γ in the first stage (equation 24) while the reduced form estimates those first-stage parameters via OLS. The parameters imposed in the first stage of the structural decomposition are the estimated ones in the unconstrained version and $\gamma = 1$, $\sigma = 3$ (or other imposed value) in the constrained version.

Table 4. Decomposition of sales under baseline and constrained fundamentals

	Structural				
	(1)	(2)	(3)	(4)	(5)
TFPQ_HK	1.40				
TFPQ_HK constrained			1.243	1.276	2.267
TFPQ		0.11			
Demand shock		1.28			
ln pm		-0.078			
ln wage		-0.152			
ln markup		-0.007			
Wedge	-0.396	-0.155	-0.243	-0.276	-1.267
Average RS in	**0.638**	**0.638**	**0.638**	**0.666**	**0.9**
Max RS in revenue	**0.9**	**0.9**	**0.638**	**0.666**	**0.9**

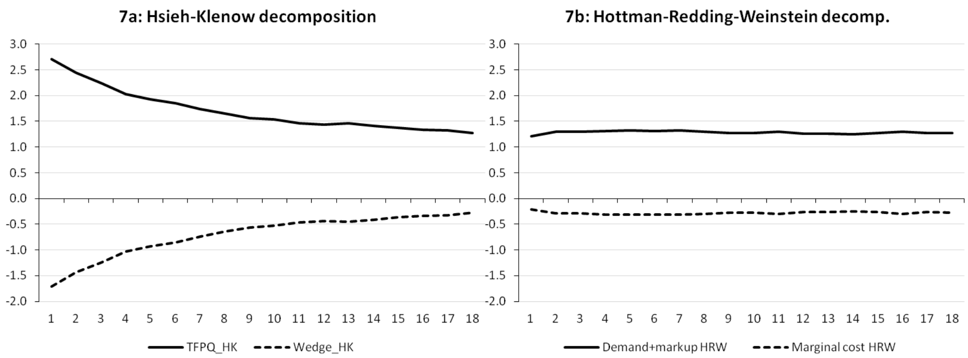
	Reduced				
	(1)	(2)	(3)	(4)	(5)
TFPQ_HK	0.321				
TFPQ_HK constrained			0.642	0.596	0.187
TFPQ		0.031			
Demand shock		0.469			
ln pm		-0.002			
ln wage		0.049			
ln markup		0.05			
Wedge	0.679	0.402	0.358	0.404	0.813
Average RS in	**0.638**	**0.638**	**0.638**	**0.666**	**0.9**
Max RS in revenue	**0.9**	**0.9**	**0.638**	**0.666**	**0.9**

TFPQ_HK is a function of TFPQ, demand shocks, and the elasticity of substitution. The unconstrained version uses the factor and substitution elasticities estimated using P and Q data, reported in Table 1. The constrained version uses cost shares as factor elasticities and imposes a common elasticity consistent with CRS in production and the revenue RS reported in the corresponding column.

5 to column 1, it is clear that under less curvature in the revenue function (higher RS in revenue) the decomposition assigns a more predominant role to wedges.

While the fact that the estimated wedge increases in importance as returns to scale grow is well known, Table 4 highlights the striking magnitude of differences and their non-linearity with respect to changes in sigma: when sigma is such that the curvature of the revenue function approaches constant returns to scale, wedges gain much more weight than with increases in sigma far from this region. This nonlinearity is the reason why the wedge is larger in column 1 compared to column 3 of Table 4, despite the fact that the average curvature is the same in both columns: some sectors exhibit sufficiently high curvature that the role assigned to wedges outweights that of fundamentals (detailed by sector results for different curvatures are shown in Appendix C). Since results for the decomposition depend so closely on the elasticities used to estimate fundamentals, the possibility of estimating elasticities relevant to

Figure 7: Hsieh-Klenow and Hottman-Redding-Weinstein decompositions using the same elasicities used in the baseline decomposition



These figures reproduce the structural decomposition considering, alternatively, the components considered by Hiseh and Klenow (2009, 2014) and Hottman, Redding and Weinstein (2016). The HK decomposition of the left panel is the by-age version of the decomposition in column 1 of table 4, the «unconstrained HK» case, since the factor elasticities in production and elasticity of substitution of the baseline case are used.

the particular context–enabled by the availability of P and Q data–is highly valuable. Appendix C shows that, beyond this nonlinear increase in the role of wedges as returns to scale in revenue increase, the relative importance of different fundamentals is robust to changes in the revenue curvature.

Table 4 thus sends three main messages about the value of P and Q output and input data in our estimation. Using only revenue and input data yields: 1) an overstatement of the contribution of wedges in the structural and reduced form estimation when using the correctly estimated output and demand elasticities; 2) an overstatement of the contribution of wedges when using cost shares for output elasticities and an assumed unique value of the demand elasticity – the magnitude of the latter varies substantially with the assumed demand elasticity; and 3) an inability to identify the distinct contributions of demand, $TFPQ$ and idiosyncratic input prices.

6.2 Value added of bringing input data to the Hottman-Redding-Weinstein framework

The differential contribution of demand vs. cost-side socks to plant sales is explored by Hottman, Redding and Weinstein (HRW, 2016). Using the demand structure that we also impose in our baseline estimation, they decompose sales into the contributions of price (observed) and demand shocks (residual) using the estimated elasticity of substitution, and subsequently decompose price into the contributions of markups–computed as in equation 4–and residual marginal costs. These residual marginal costs thus capture input price variability, technical efficiency, and any other factor not directly modelled in their framework, including wedges.

Since we fully rely on HRW's demand structure, the contribution of the demand shock and markup are, by construction, the contributions one would obtain in their approach. The availability of data on input use and input prices, beyond P and Q data on the output side which their approach already employs, allows us to further decompose their marginal cost component into input prices, $TFPQ$ and wedges. Figure 7b illustrates the by-age decomposition obtained in our data with the HRW approach (to be compared with the upper left panel of Figure 4). As in their results for consumer goods in the US, demand shocks explain the bulk of sales growth variation, and markups play a modest role. But the negative, flat over ages, pattern estimated for the contribution of marginal costs is a combination of the positive contribution of $TFPQ$ and the dampening role of wedges and input prices in the context of our application, each of them negatively correlated with sales. HRW found a relatively minor but positive contribution of cost shocks to the variance of consumer goods sales in the U.S. Differences from our results stem from at least two sources. First, the literature on misallocation has pointed that size-correlated distortions are generally stronger in less developed economies, so that wedges are more likely to represent a drag to the cost component a-la HRW in Colombia than the US. Second, since our strategy to identify the elasticity of substitution and demand shocks imposes weaker restrictions on the correlation between demand shocks and cost shocks than that in HRW, if delivering products with greater quality and appeal to demand requires a greater effort in production and more costly inputs, our approach will take this negative correlation between demand shocks and cost shocks into greater account and thus assign a more dampening role to cost factors over the variance of sales.[40]

[40] HRW impose orthogonality between double-differenced demand and cost shocks at the product level, where differencing takes place over time and with respect to one of the business' products. This identification assumes away the possibility that the business shifts towards higher quality products that imply greater cost. Our baseline identification at the plant level allows the average quality of plant's products to correlate with cost shocks, with findings that imply that greater appeal comes at a higher cost (Table 2)

The lumping together of cost, productivity and wedges also misses the rich life cycle dynamics of each of these factors. Productivity becomes less important as do wedges for older businesses in our baseline framework but this pattern is missed completely in the HRW approach. Relatedly, the increasing magnitude of the inverse correlation between demand and $TFPQ$ over the life cycle is missed in the HRW approach.

6.3 The value of Quality Adjustment

Results discussed so far use UPI prices to deflate output . UPI plant price indices adjust real output for intra-firm quality/appeal differences (see section 3.2). Moreover, in the context of UPI prices, sales measure output that is additionally adjusted for cross-plant quality differences.

We now discuss the empirical role of quality adjustment in more detail. We do so by comparing results to what would be obtained under two alternatives to price measurement. First, we implement a "statistical" approach based on Törnqvist indices for a constant basket of goods within the plant or, alternatively, on the divisia price index that allows that basket to change and uses average t, $t-1$ expenditure shares. We implement a second alternative approach, using prices based on the insights offered by Sato (1976), Vartia (1976) and Feenstra (1994). The Sato-Vartia approach is economically motivated but keeps appeal shifters and baskets of goods constant over two consecutive periods, implying a much slower quality adjustment for both continuing products and those that enter and exit. The Feenstra adjustment for changing varieties incorporated into our UPI approach can also be added to the Sato-Vartia index to adjust for changing baskets of goods over consecutive periods (it was, in fact, originally implemented by Feenstra, 2004, within the Sato-Vartia approach). The UPI, meanwhile, allows for both changing baskets of goods and varying appeal shifters, dimensions of flexibility which respectively deal with the "consumer valuation bias" and the "variety bias" (Redding and Weinstein, 2017). (For a detailed discussion of each of these alternatives, contrasted with the UPI, see 3.2, Appendix A, and Redding and Weinstein, 2017).

In each approach, the aggregation from the plant to the sector level is analogous to the aggregation from the product to the plant level, using weights and shares that correspond to the basket of plants in aggregate expenditure by contrast to the basket of products in plants' sales. For theory-based indices this is directly implied by theory. For statistical indices we impose it for consistency.

If the quality mix within the plant improves over time, plant-level quality adjusted price indices will grow less than unadjusted ones, as a result yielding less deflated output growth and less $TFPQ$ growth. This composes with overall plant quality growth to imply economically motivated aggregate prices that grow less than unadjusted ones. Not properly adjusting plant-level prices for quality changes biases estimated idiosyncratic output and technical efficiency growth downwards because such estimates will ignore the part of price increases that reflects increasing valuation of goods and the services of plants to their costumers, and thus mistakenly translate those price increases into welfare decreases for given expenditure.

Figure 8 depicts aggregate price changes under these four different approaches, (where aggregation is at the 3-digit sector level, reported for the average sector.[41] UPI growth is very similar to price growth using constant baskets in all periods, but the difference is much more marked starting in 1991. On average over 1991-2012, baseline (UPI) price growth is 3.2. percentage points below that of the statistical index with a fixed basket of goods, while for the pre-1991 period the two indices display virtually identical variations.[42] Interestingly, this is precisely the time when market-oriented reforms were implemented. As many other countries in Latin America and around the globe, Colombia undertook wide market-oriented reforms during the 1990s, including unilateral trade opening, financial liberalization, and flexibilization of labor regulations. Figure 8 suggests more quality adjustment starting at that time, broadly consistent, for instance, with findings in Fieler et al. (2018) about the effect of the 1990s trade liberalization on quality in Colombian manufacturing.

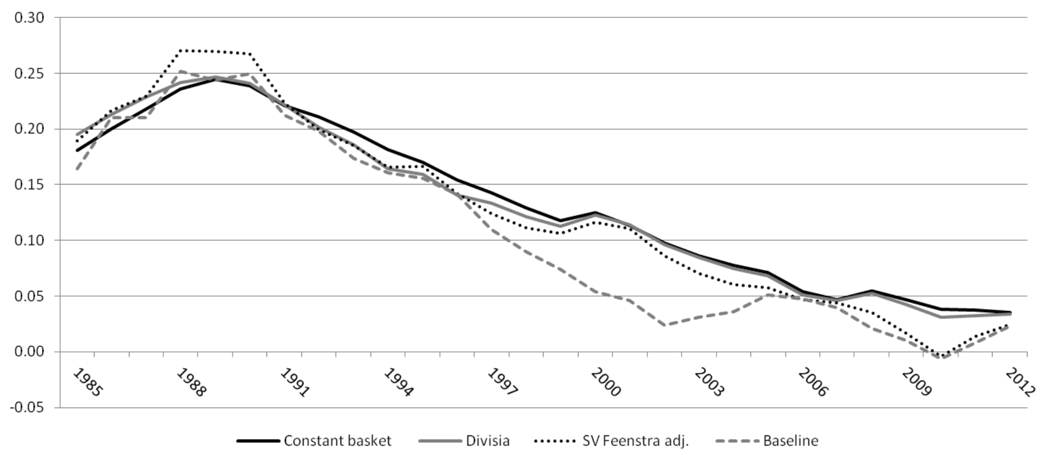
As a result, adjusting output for quality changes assigns a much larger weight to technical efficiency, $TFPQ$, and a lesser role to demand, in explaining output life cycle growth (see Appendix I for detailed results). While with constant-weights-Törnqvist-indices $TFPQ$ and demand are estimated to contribute roughly equally to output growth, $TFPQ$ is assigned progressively more relative importance as one moves to the Sato-Vartia and then to the UPI approaches. But quality adjusting prices matters much more in decom-

[41] Three-year moving averages are shown to smooth out jumps in the series.

[42] The gap between the UPI and the statistical index with a fixed basket is slightly smaller in magnitude compared to that reported by Redding and Weinstein (2017) for the U.S. using data on final consumption goods. They find a gap of close to 5% in aggregate price growth.

Figure 8: Annual changes in the aggregate price index

Average three-digit sector, tree-year moving average



posing output than for sales because, beyond the more precise measurement of fundamentals when quality is adjusted for, the measure of output itself is affected by price indices.

7 Conclusion

Our use of product-level price and quantity data on outputs and inputs for plants enables us to overcome a host of conceptual, measurement and estimation challenges in the literature. However, our findings raise a number of questions and point to important areas for future research. First, our approach has the advantage that wedges are measured as the components of sales and output volatility that cannot be accounted for by fundamentals with the latter estimated independently of measuring wedges. While this is an advantage, wedges are still a residual and therefore a black box. Identifying the specific sources of wedges that dampen output and sales growth especially for young plants is one potential area of research. Since there is an important role for correlated wedges, one natural candidate is adjustment costs that especially impact young businesses. From this perspective, this may include the costs of developing and accumulating organizational capital (such as customer base). Our finding that between-plant differences in demand become more important in accounting for output growth volatility for more mature plants is consistent with this hypothesis.

Size-dependent policies and other characteristics of the regulatory environment are othe set of candidate explanations behind wedges. Colombia is a country that underwent dramatic reforms over our sample period, some of them displaying cross-sectional variability (such as product-specific reductions to import tariffs in the early 1990s), and thus offers fruitful ground for investigating the impact of the regulatory environment on life-cycle dynamics. In prior work, we have explored the effect of these reforms in cross-sectional productivity and factor adjustments, finding that the they have changed adjustment dynamics of factors (see, e.g., Eslava et. al. (2010)), the responsiveness of selection to fundamentals, and within-plant productivity growth (see, e.g., Eslava et. al. (2013)). Moreover, Eslava, Haltiwanger and Pinzón (2018) show that high growth plants have become more prevalent in Colombia from the 1980s to 2000s.

Our findings provide insights into the relative importance of the variance in fundamentals in explaining plant growth, inviting further research into the ultimate sources of the variance in these fundamentals. While our current framework allows for wedges that are correlated with current fundamentals, and in fact we find that they are indeed (inversely) correlated, we do not take explicit account of the likely endogenous response of the variance of fundamentals over the life cycle to past performance and past wedges. Research that sheds light on the endogenous determinants of the variance in the supply side ($TFPQ$) and demand side fundamentals should have a high priority in future research. In exploratory analysis shown in Appendix E we find evidence that $TFPQ$ and demand shocks are highly persistent and part of this persistence reflects that observable indicators of endogenous innovation such as R&D expenditures are increasing in lagged fundamentals. We also find suggestive evidence that wedges influence the evolution of fundamentals but the quantitative impact of lagged wedges on current period fundamentals or current period R&D expenditures is relatively small.

Our research also finds support for the agenda that highlights the importance of quality-adjusting measures of price indices. Our findings in this paper are that, in Colombia, quality-adjusted inflation (of manufacturing products) is about three percentage points lower than the unadjusted indicator. And, interestingly, that this gap grows substantially at the beginning of the nineties, coinciding with wide-spread market reforms, including trade liberalization. Those findings suggest that quality adjustments have become an increasingly important source of welfare gains (partly from trade, as demonstrated in Fieler et al. 2018). Estimating the changing relative importance of the components of fundamentals during these market reforms is explored in Eslava and Haltiwanger (2018).

8 References

Acemoglu, Daron, Ufuk Akcigit, Nicholas Bloom, and William R. Kerr, 2017. "Innovation, Reallocation, and Growth.' NBER Working Paper No. 18993 (revised).

Ackerberg, D. A., Caves, K. and Frazer, G., 2015, "Identification Properties of Recent Production Function Estimators." Econometrica, 83: 2411-2451.

Atkin, David, Amit Khandelwal and Adam Osman, 2019, "Measuring Productivity: Lessons from Tailored Surveys and Productivity Benchmarking" (No. w25471). National Bureau of Economic Research.

Atkin, David, Amit Khandelwal and Adam Osman, 2017, "Exporting and Firm Performance: Evidence from a Randomized Experiment" Quarterly Journal of Economics: Vol. 132 No. 2, Editor's Choice.

Aw, Bee Jan, Mark Roberts and Daniel Xu, 2011. "R&D investment, exporting and productivity dynamics," American Economic Review, 101: 1312-1344.

Bartelsman, Eric, John Haltiwanger and Stefano Scarpetta, 2013. "Cross-Country Differences in Productivity: The Role of Allocation and Selection," American Economic Review, vol. 103(1), pages 305–334.

Bento, Pedro and Diego Restuccia, 2017. "Misallocation, Establishment Size, and Productivity," American Economic Journal: Macroeconomics, Volume 9 (3), July 2017, pp. 267-303.

Bento, Pedro and Diego Restuccia, 2018. "On Average Establishment Size across Sectors and Countries," NBER Working Papers 24968.

Bloom, Nicholas and John Van Reenen, 2007. "Measuring and Explaining Management Practices Across Firms and Countries," The Quarterly Journal of Economics, Oxford University Press, vol. 122(4), pages 1351-1408.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, John Van Reenen, 2016. "International Data on Measuring Management Practices," American Economic Review, American Economic Association, vol. 106(5), pages 152-56, May.

Buera, F.J, and R. Fattal. (2014) "The Dynamics of Development: Entrepreneurship, Innovation, and Reallocation". Mimeo.

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121 (2006), 541-86.

Caballero, Ricardo, Eduardo Engel and John Haltiwanger. 1995. "Plant-level Adjustment and Aggregate Investment Dynamics," *Brookings Papers on Economic Activity* 2: 1-54.

Caballero, Ricardo, Eduardo Engel and John Haltiwanger. 1997. "Aggregate Employment Dynamics: Building from Microeconomic Evidence," *American Economic Review,* 87: 115-137.

De Loecker, J. and F. Warzynski. 2012. "Markups and Firm-Level Export Status," *American Economic Review* 106(2): 2437-2471.

De Loecker, J., P. Goldberg, A. Khandelwal and N. Pavcnik (2016) "Prices, Markups and Trade Reform," Econometrica, Vol. 84, No. 2 (March), 445-510.

Decker, Ryan, John Haltiwanger, Ron Jarmin and Javier Miranda. 2018. "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness." NBER Working Paper No. 24236.

Eslava, M., J. Haltiwanger, A. Kugler and M. Kugler (2013) "Trade Reforms and Market Selection: Evidence from Manufacturing Plants in Colombia," Review of Economic Dynamics, 16, 135-158.

Eslava, M., J. Haltiwanger, A. Kugler and M. Kugler (2010) "Factor Adjustments After Deregulation: Panel Evidence from Colombian Plants," Review of Economics and Statistics, 92, 378-391.

Eslava, M., J. Haltiwanger, A. Kugler and M. Kugler (2004) "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia," Journal of Development Economics, 75 (2), 333-372.

Eslava, M., J. Haltiwanger (2018) "Market Reforms and Life Cycle Growth of Plants," (in process).

Eslava, M. John Haltiwanger and Alvaro-Jose Pinzon (2018) Job Creation in Colombia vs the U.S.: 'Up or out Dynamics' Meets 'The Life Cycle of Plants'. SSRN Working Paper.

Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 84 (1994), 157-177

Fieler, C., M. Eslava and D. Xu (2018) "Trade, Quality Upgrading, and Input Linkages: Theory and Evidence from Colombia," American Economic Review, 108(1), 109-146

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.

Garcia-Santana, M., & Pijoan-Mas, J. (2014). The reservation laws in India and the misallocation of production factors. Journal of Monetary Economics, 66, 193-209.

Garicano, L., Lelarge, C., & Van Reenen, J. (2016). Firm size distortions and the productivity distribution: Evidence from France. American Economic Review, 106(11), 3439-79.

Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. 2017. "Capital Allocation and Productivity in South Europe." Quarterly Journal of Economics 132 (4): 1915-1967.

Hottman, Collin, Stephen J. Redding and David E. Weinstein, 2016. "Quantifying the Sources of Firm Heterogeneity," The Quarterly Journal of Economics, Oxford University Press, vol. 131(3), pages 1291-1364.

Haltiwanger, John, Ron Jarmin, and Javier Miranda (2013) "Who Creates Jobs? Small vs. Large vs. Young." *Review of Economics and Statistics*, 95(2), 347-361.

Haltiwanger, John, Robert Kulick, and Chad Syverson (2018), "Misallocation Measures: The Distortion that Ate the Residual," NBER Working Paper No. 24199.

Hopenhayn, H (2016) "Firm Size and Development." Economía, vol. 17 no. 1, 2016, pp. 27-49.

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

Hsieh, Chang-tai and Peter Klenow (2014) "The life cycle of plants in India and Mexico," *Quarterly Journal of Economics,* 129(3): 1035-84.

Khandelwal, Amit (2010) "The Long and Short (of) Quality Ladders," Review of Economic Studies, Oxford University Press, vol. 77(4), pages 1450-1476.

Redding, Stephen and David E. Weinstein (2016) "A Unified Approach to Estimating Demand and Welfare," NBER Working Papers 22479, National Bureau of Economic Research, Inc.

Restuccia, Diego and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Review of Economic*

Dynamics, 11(October): 707-720.

Restuccia, Diego and Richard Rogerson. 2017. "The Causes and Consequences of Misallocation,"' NBER Working Paper No. 23422, National Bureau of Econonomic Reserach, Inc.