NBER WORKING PAPER SERIES

THE LIFE-CYCLE GROWTH OF PLANTS:
THE ROLE OF PRODUCTIVITY, DEMAND AND WEDGES

Marcela Eslava
John C. Haltiwanger

Working Paper 27184
http://www.nber.org/papers/w27184

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2020

We thank Alvaro Pinzón and Nicolás Urdaneta 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 Virgiliu Midrigan, Venky Venkateswaran, Ariel Burstein, David Atkin, David Baqaee, 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 2019 NBER Summer Institute; the 2016 Trade and Integration Network of LACEA; and seminars at Universitat Pompeu Fabra, Universitat Autónoma de Barcelona, the University of Chicago, UCLA, LSE, Oxford and Universidad de Los Andes. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Life-cycle Growth of Plants: The Role of Productivity, Demand and Wedges
Marcela Eslava and John C. Haltiwanger
NBER Working Paper No. 27184
May 2020
JEL No. E24,L24,O47

ABSTRACT

We develop a framework that uses price and quantity information on both firms' outputs and inputs to assess the roles, on firm dynamics and welfare, of efficiency, input prices, demand/ quality, idiosyncratic markups, and residual wedges. Our strategy nests previous approaches limited by data availability. In our application, demand/quality is found to dominate the cross sectional variability of sales growth, while quality-adjusted input prices and residual wedges play dampening roles, especially at birth. Markups play only a modest role for cross-sectional variability of sales growth but are important in explaining welfare losses from revenue productivity dispersion.

Marcela Eslava
Universidad de los Andes
Calle 19A # 1-37 Este. Bloque W of. 810
Bogotá, Colombia
meslava@uniandes.edu.co

John C. Haltiwanger
Department of Economics
University of Maryland
College Park, MD 20742
and NBER
haltiwan@econ.umd.edu

A data appendix is available at http://www.nber.org/data-appendix/w27184

1 Introduction

A prevalent feature of market economies is wide heterogeneity of firm size, firm growth, and a host of firm attributes correlated with size (e.g., productivity, exports, survival). What are the sources of firm size and firm growth heterogeneity? How does the answer matter for welfare? The macro literature on misallocation studies the role of productivity vs. remaining sources of dispersion, with special focus on wedges that distort the size distribution of activity. Other literatures in macro, trade, and IO have focused on the role of specific attributes of firms: demand (quality), markups, or costs. Hottman, Redding and Weinstein (2016) recently integrated demand, markups and residual costs into an estimation framework, but not wedges (i.e. departures from the model), finding a dominant role for demand attributes. In the face of data constraints, assessing the roles of all of these different margins of firms' primitives simultaneously has not been possible. Productivity and wedges are typically identified from structures that exploit micro data on revenue and input expenditures, while structures that use product-level data on output prices and quantities have been used to identify quality, costs and markups.[1]

Building on these distinct approaches, we develop a unified conceptual, measurement and estimation structure that integrates these different dimensions of data and firm attributes. Our framework takes advantage of data on output and input prices and quantities to measure idiosyncratic demand shifters, markups, and two distinct dimensions of idiosyncratic marginal costs: technical efficiency and quality-adjusted input prices. It accounts for the contribution of each of these attributes of firms to firm size and growth, while also allowing for wedges between the data and the behavior predicted by the model. We apply this framework to the analysis of both growth over the life cycle of manufacturing plants and welfare. Life cycle business growth is crucially related to aggregate productivity growth (Hsieh and Klenow, 2014) and displays wide heterogeneity across businesses (Haltiwanger, Jarmin and Miranda, 2013; Eslava, Pinzón and Haltiwanger, 2019). Our framework also lends itself to analyzing the impact of different sources of heterogeneity across firms on welfare, and we use it to that end.

We use detailed product-level data on quantities and prices for outputs and inputs from the Colombian Annual Manufacturing Survey. This is a uniquely rich census of non-micro manufacturing plants with data on quantities and prices, at the detailed product class, for outputs and inputs within plants. We follow individual plants for up to thirty years (1982-2012). The long time coverage allows us to investigate the

[1] The misallocation literature is extensive. Prominent examples are Restuccia and Rogerson (2008); Hsieh and Klenow (2009, 2014)) ; Guner, Ventura and Xu (2008); Midrigan and Xu (2013); Bartelsman et. al. (2013); Bento and Restuccia (2017); Adamopoulos and Restuccia (2014). Quality is the focus in Brooks (2006); Fieler, Eslava and Xu (2018); Hallak and Schott (2011) Khandelwal (2011); Kugler and Verhoogen (2012); Manova and Zhang (2012). Production efficiency vs. demand is emphasized in Foster, Haltiwanger and Syverson (2008 and 2016), and Eslava et al. (2013). De Loecker and Warzynski (2012) and De Loecker et al (2020) have focused on markups.

determinants of medium- and long-term life cycle growth.

By technology or technical efficiency we refer to a production function residual (Foster, Haltiwanger and Syverson, 2008), where production in multiproduct plants is plant-level revenue deflated with a quality adjusted plant-level deflator. Following Redding and Weinsten (2020), our deflator allows for product turnover and changing appeal across products within the firm. On the demand side, we estimate plant-specific demand function residuals, that identify greater appeal/quality as the ability to charge higher prices per unit of a product (Hottman, Redding and Weinstein, 2016; Khandelwal,2011; Fieler, Eslava and Xu, 2018). Our specification of demand and competition allows for idiosyncratic markups, which can also be calculated using our data. Input costs are directly measured from input price data, separately for materials and labor, also permitting the construction of quality-adjusted input prices.

Our approach requires, and the richness of the data permits, estimating the parameters of the production and demand functions. We introduce an estimation technique that jointly estimates the two functions, bringing together insights from recent literature on estimating production functions based on output and input use data, and literature on estimating demand functions using P and Q data for outputs.[2] The joint estimation ensures consistency and separate identification of demand vs. production parameters. Moreover, the granularity of our data allows estimating different production and demand elasticities for different sectors, and doing so without imposing constant returns to scale. In contrast to much of the literature estimating demand functions in contexts of multiple products, we also allow efficiency and demand to be correlated, even within firms over time.

After estimating plant-specific technical efficiency, demand shifters, markups and quality-adjusted input prices, we measure the contribution of each to the variability of sales growth across plants over the life-cycle. Residual wedges in our framework correspond to the gap between actual size at any point of the life cycle and size implied by the model given measured attributes. Since we explicitly account for idiosyncratic (quality-adjusted) input price and markup variability, the distribution of these wedges is not adequately captured by revenue productivity dispersion (in contrast to the framework proposed by Hsieh and Klenow's 2009, 2014, which is nested in our model).[3]

Dispersion in demand shifters is the main driver of sales growth heterogeneity in our data. Though a dominant role of demand shocks in accounting for life cycle growth has been previously found (Hottman, Redding and Weinstein, 2016; and Foster, Haltiwanger and Syverson, 2016) full-distribution accounting allows us to identify this role

[2] For production function estimation, see, e.g. Ackerberg, Caves and Frazer (2015); De Loecker et al. (2016). For demand function estimation see, e..g. Hottman, Redding and Weinstein (2016); Foster, Haltiwanger and Syverson (2008).

[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 face, as they may stem from constraints more technological than institutional in nature, such as adjustment costs.

as stemming from extremely dynamic appeal in superstar plants. Moreover, our results also point to technical efficiency efforts as a necessary condition for success: rapidly contracting technical efficiency is the outstanding characteristic of the worst performers.

We also find that negatively correlated wedges reduce plant revenue variance by 12 percentage points over the first twenty years of life, while heterogeneous input prices reduce it by an additional 16%. For a subperiod where we can quality-adjust wages, we find that the contribution of the latter is reduced by half after this adjustment.

Correlated wedges are particularly large for plants with lowest productivity growth. By age 20, in the absence of wedges, the top quartile of sales' predicted growth would have grown close to seven-fold, rather than the actual five-fold, while the bottom quartile would have contracted markedly, rather than growing close to 40% as it did. Markup dispersion, in turn, plays a negligible role for sales growth heterogeneity. The relative importance of different sources of heterogeneity varies considerably over the life cycle. For mature plants, most of the variation in life-cycle growth is explained by demand and efficiency, while for younger plants efficiency and wedges (negatively correlated with fundamentals) play a more important role in the decomposition of variance.

A counterfactual analysis of the individual impact of wedges and fundamentals on welfare highlights the value of disentangling the role of technical efficiency vs demand, and that of input prices and markups from residual wedges. Love of variety implies welfare gains from heterogeneity in product demand (which reflects heterogeneity in both product quality and appeal). This welfare effect turns out to be large. While markups are unimportant for explaining cross-plant heterogeneity in growth, they turn out to be crucial for welfare. Most of the welfare losses from revenue productivity dispersion are explained by quality-adjusted input price and markup variability, each being as important as the other. This is because, though markups exhibit little dispersion in the overall cross-sectional distribution, a few plants with very large market shares weigh heavily in aggregate welfare. Quality adjusting wages matters significantly for size dispersion as noted above, but not so much for welfare, which is inherently size-weighted. This suggests that, though an important fraction of wage dispersion reflects quality heterogeneity rather than some type of friction or distortion, it is the departure from frictionless and distortionless labor markets that matters for welfare.

Our contribution to the literature is multi-fold. First, we bridge the gap between distinct approaches to the study of drivers of firm size and growth, which alternatively focus on either productivity vs. wedges, or on the roles of demand, cost and markups. Our framework builds on Hsieh and Klenow (2009, 2014)–henceforth HK–on the supply side, and on Hottman, Redding and Weinstein (2016)–henceforth HRW–on the demand side. Cost factors are found to play a more important role for sales growth variability than would be identified by the HRW approach alone, because the cost component in that framework is a residual that lumps together efficiency, input costs, and residual wedges. The latter are negatively correlated with efficiency so the composite contribution is muted. Pooling ages, the contribution of the composite HRW "cost" residual to the variance of sales growth is -12.2% in our data. Unpacking this, we find it reflects a positive contribution of 15.4% of technical efficiency, a negative contribution of -15.9% from input prices and an additional drag of -11.7% from residual wedges. Importantly, residual wedges are not inherently a cost/supply side factor so the composite HRW "cost" residual may reflect a host of factors as we discuss below. Relative to the implications of the HK decomposition, in turn, our approach yields insights masked by using a composite productivity measure as well as a composite wedge measure. On the one hand, the composite productivity measure is dominated by demand shifters relative to technical efficiency, where the former contributes over ten times more than the latter to the dispersion of sales growth. On the other hand, composite HK wedges are a drag of 40% on sales growth variability pooling ages, but unpacking we find that 25% of this 40% is due to input prices (and markups, but the latter play a negligible role in this decomposition). The remainder is due to residual wedges. Related, idiosyncratic input prices and markups explain about half of the welfare losses from composite HK wedges (with residual wedges explaining the rest).

Within the misallocation literature, recent contributions have increasingly focused on decomposing size-to-productivity wedges into components such as adjustment costs, information frictions, financial frictions, labor market frictions (see, e.g., Asker et al., 2014; David and Venkateswaran, 2018; Midrigan and Xu, 2014; Guner, Ventura and Xu, 2008). In a distinct but related vein, our results highlight that *composite* wedges can be decomposed into idiosyncratic markups, quality-adjusted input prices and residual wedges. Our findings on the importance of input price heterogeneity (even adjusting for quality) point to important sources of such heterogeneity, including frictions in the markets for inputs as well as potentially monopsony power. Sorting this out should be an important topic for future research.

Second, we contribute to the literature on estimating production functions and to that on estimating demand functions.[4] Our joint estimation of the two functions is an important novelty. It highlights the importance of relying on output price and quantity information to distinguish revenue from production parameters, and the usefulness of including information on the production process (inputs, in particular) to distinguish demand from supply elasticities. Moreover, our approach to measuring plant-level production for multiproduct plants underscores the need to take a stance on the structure of demand, not only to measure plant output in the presence of multiple products, but to even define it.

Third, we provide an alternative take on the role of markups in firm heterogeneity relative to that in De Loecker et al (2018), who recover the markup without imposing structure on the demand side. The need to take a stance on demand in our context

[4] For production function estimation, we follow Olley and Pakes (1996); Levinsohn and Petrin (2003), Ackerberg, Caves and Frazer (2015) , and De Loecker et al (2016). For demand estimation, HRW and Foster et al (2008).

also reflects the more general fact that the interpretation of markups depends on the market structure. A residual approach using cost shares of revenue to identify markups lead to markup measures that capture all factors driving cost share heterogeneity across producers. Our approach to markups is closer to that in HRW or Edmond, Midrigan and Xu (2019), which ties the markup to a specific demand system, and it is complemented with our ability to separately measure demand elasticities, input price heterogeneity and other sources of market power.

Finally, our findings contribute to the policy discussion regarding interventions to address the limitations to business growth. Our results highlight that size-to-productivity wedges are especially prevalent for young businesses, but also that dimensions internal to businesses are even more important than wedges. 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.

While the data infrastructure we use is very rich, it faces limitations particularly with respect to the increasingly prevalent use of item-level price and quantity data as in HRW. Our data are at detailed product class level for each establishment but not at the item-level. While this prevents us from drawing the rich insights that emerge from item-level data, the combination of price and quantity data for both outputs and inputs at the product class level within establishments yields crucial new insights that help to bridge the findings of HRW with those from the large literature using revenue and input expenditure data at the establishment level. We also find it reassuring that both qualitatively and quantitatively we generate results on the overall contribution of demand and cost factors to the growth distribution across businesses that are consistent with alternatively implementing the HRW approach.[5] As we note above, our data infrastructure and approach allow us to unpack their composite cost residual into distinct efficiency, input price and residual wedge components.

The paper proceeds as follows. Section 2 presents our framework. 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 . Our main results are presented in section 5. Section 6 examines

[5]Panel A of Table X of HRW shows that demand (combining appeal/scope) accounts for 107% of firm sales growth in their data, compared to our finding of 113% in the Colombian data (averaging across the life cycle). Combined cost factors are a drag of -7% in HRW's application, while if we combine the contributions of efficiency, input prices and residual wedges that we find for Colombia, we account for about -13%.

the the value added of our approach relative to prior approaches. Welfare analysis is presented in section 7. Section 8 concludes by providing a more comprehensive view on the implications of our analysis, and on open questions for future research.

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, though we generalize the concept to producers of heterogeneous goods) and a demand shock.[6] 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 to each given age.[7] However, we do explore the empirical cross-sectional relationship between fundamentals and wedges. In the appendix, we also explore the relationship between proxies for investment in innovation and lagged fundamentals. We focus on decomposing the determinants of growth 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

[6] Hsieh and Klenow (2009, 2014) use the term $TFPQ$ 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.

[7] 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 strengthen $TFPQ$ may include better management of the production process (e.g. Bloom and Van Reenen, 2007) or acquiring better machines.

conclude that our findings for plants that survive up to age t are largely driven by the establishments that survive at least one more year.

We don't explicitly model dynamic 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 costs, information frictions and distortions arising from the business climate.[8] 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.

For developing the theoretical predictions, we treat input prices as exogenous and potentially idiosyncratic for the common composite input. Empirically we consider multiple inputs and make efforts to take into account input heterogeneity through quality adjusting prices. Given that idiosyncratic input prices turn out to play a non-trivial role empirically, we consider below the potential sources of the variation in input prices even after adjusting for quality.

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:

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

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.

[8]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 De-Loecker (2014), Decker et. al. (2017), and David and Venkateswaran (2018). But it has the advantage in subsuming in a simple measure different types of frictions and distortions, including those that capture dynamic considerations.

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 as will become clear in our discussion of demand primitives further below. 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. Consequently, the distribution of firm sales growth is the central focus of our analysis of the firm growth distribution.

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

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

taking as given A_{ft}, D_{ft}, and unit costs of the composite input, denoted 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 the static model given its fundamental attributes.[9] Such wedges capture, for instance, adjustment costs that may be present in terms of changing the scale or mix of inputs or building up a customer base, product-specific tariffs, financing constraints, information frictions, and size-dependent regulations or taxes. Adjustment 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 correlated with plant fundamentals themselves.

[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.

[10] See, for instance, Caballero, Engel and Haltiwanger (1995, 1997), Eslava, Haltiwanger, Kugler, and Kugler (2010) and Asker et. al. (2014).

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

By their very nature, adjustment costs and financing constraints are typically correlated with plant fundamentals. Size-dependent regulations are another prominent example of correlated wedges.[12]

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 $X_{ft} = \left(\frac{D_{ft}A_{ft}^{1-\frac{1}{\sigma}}\gamma}{C_{ft}\mu_{ft}\left(1-\tau_{ft}\right)^{-1}}\right)^{\frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}}$, 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):[13]

$$\frac{R_{ft}}{R_{f0}} = \left(\frac{d_{ft}}{d_{f0}}\right)^{\kappa_1} \left(\frac{a_{ft}}{a_{f0}}\right)^{\kappa_2} \left(\frac{pm_{ft}}{pm_{f0}}\right)^{-\phi\kappa_2} \left(\frac{w_{ft}}{w_{f0}}\right)^{-\beta\kappa_2} \left(\frac{\mu_{ft}}{\mu_{f0}}\right)^{-\gamma\kappa_2} \left(\frac{\widehat{\chi}_t}{\chi_0}\frac{\chi_{ft}}{\chi_{f0}}\right)^{1-\frac{1}{\sigma}} \tag{5}$$

where $\kappa_1 = \frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}$, $\kappa_2 = \left(1-\frac{1}{\sigma}\right)\kappa_1$, and 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 input prices, two are observed in the data: the price of material inputs, Pm_{ft}, and average wage per worker, W_{ft}. 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, from D_t, A_t and C_t are lumped into $\frac{\chi_t}{\chi_0}$ and $\frac{\widehat{\chi}_t}{\chi_0}$. Crucially, $\frac{\chi_{ft}}{\chi_{f0}}$ captures life cycle growth in idiosyncratic wedges, including those stemming from τ_{ft}, from potential factor-specific wedges, and

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

from measurement error and noise in fundamentals not observed by the firm at the time of choosing its scale in each period.[14]

Equation (5) is the focus of our analysis of the distribution of firm growth. We start with the growth of (idiosyncratic) attributes 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. The wedges that a firm faces may be age-specific, and thus de-couple life-cycle growth in output from the growth of those attributes.

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,where two thirds of observations correspond to multiproduct producers. We define and measure firm-level output in a manner that allows for 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 CUPI approach developed by Redding and Weinstein (2020) 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. That is, 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:

---

14 $\frac{\chi_{ft}}{\chi_{f0}} = \frac{\delta_{ft}^{\gamma\kappa_1} \alpha_{ft}^{1+\gamma\kappa_2} \zeta_{ft}^{-\gamma\kappa_1} (1-\tau_{ft})^{\gamma\kappa_1} \left(1+\tau_{ft}^M\right)^{-\phi\kappa_1} \left(1+\tau_{ft}^L\right)^{-\beta\kappa_1} r_{ft}^{\frac{-\alpha\kappa_1}{\gamma}}}{\delta_{f0}^{\gamma\kappa_1} \alpha_{f0}^{1+\gamma\kappa_2} \zeta_{0t}^{-\gamma\kappa_1} (1-\tau_{0t})^{\gamma\kappa_1} \left(1+\tau_{f0}^M\right)^{-\phi\kappa_1} \left(1+\tau_{f0}^L\right)^{-\beta\kappa_1} r_{f0}^{\frac{-\alpha\kappa_1}{\gamma}}}$

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.

$$U\left(Q_{1t},...,Q_{Nt}\right) = \left(\sum_{I_t} d_{ft} Q_{ft}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \quad (6)$$

$$\text{where } Q_{ft} = \left(\sum_{\Omega_t^f} d_{fjt} q_{fjt}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \quad (7)$$

$$s.t. \quad \sum_{f=1}^{N_{Ft}} \sum_{\Omega_t^f} p_{fjt} q_{fjt} = E_t; \quad (8)$$

$$\prod_{\Omega_t^f} d_{fjt}^{\frac{1}{\left\|\Omega_t^f\right\|}} = 1; \prod_{I_t} d_{it}^{\frac{1}{\|I_t\|}} = 1 \quad (9)$$

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 (6) and (7): 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 (9), product appeal d_{fjt} captures the valuation of attributes specific to good fj relative to other goods produced by firm f, 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 besides varying across firms.

Equation (7) defines real output for a firm in this multiproduct framework. 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 (9), 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

$$
\begin{aligned}
q_{fjt} &= d_{ft}^{\sigma} d_{fjt}^{\sigma} \left(\frac{P_{ft}}{P_t}\right)^{-\sigma} \left(\frac{p_{fjt}}{P_{ft}}\right)^{-\sigma} \frac{E_t}{P_t} && (10)\\
R_{ft} &= Q_{ft} P_{ft} = d_{ft}^{\sigma} P_{ft}^{1-\sigma} \frac{E_t}{P_t^{1-\sigma}} && (11)
\end{aligned}
$$

where

$$P_t = \left(\sum_{I_t} d_{ft}^{\sigma} P_{ft}^{1-\sigma}\right)^{\frac{1}{(1-\sigma)}} \tag{12}$$

Dividing (11) by P_{ft} and solving for P_{ft},[15] we obtain

$$P_{ft} = D_{ft} Q_{ft}^{-\frac{1}{\sigma}} = D_t d_{ft} Q_{ft}^{-\frac{1}{\sigma}} \tag{13}$$

where the firm-level price index is given by:

$$P_{ft} = \left(\sum_{\Omega_t^f} d_{fjt}^{\sigma} p_{fjt}^{1-\sigma}\right)^{\frac{1}{(1-\sigma)}} \tag{14}$$

Given the nested CES demand, the firm will charge the same markup on all products.[16]

Observe that (13) is identical to (2). This consistency is important as we use (14) to construct firm-level prices, using the CUPI framework of Redding and Weinstein (2020)

[15] 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.

[16] See Appendix S2 of Hottman et. al. (2016). In this nested environment the firm's optimization problem can be decomposed into two steps. The firm first chooses the composite index of products. It then chooses individual products to minimize the composite total cost subject to the optimal level of firm-level output. It is optimal for the firm to equate the ratio of marginal costs across products to the ratio of marginal utilities. Since consumers maximization yields that the ratio of marginal utilities across products is equal to the ratio of prices this implies the markups must be the same across products. One important difference with Hottman et. al. (2016) is that we don't permit product-specific random cost shocks.

to express this price index in terms of observables. It is also useful to note that in using (11) 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 in our framework 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 some may be large enough in the market that they don't take the sectoral output price as given. That is, we depart from monopolistic competition. At a deeper level, firms are our object of interest because they are clearly relevant empirical objects. 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.

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 17,351 plants over the whole period, with 4,352 plants in the average year.[17]

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. [18]

[17] We have constrained the sample to plants born after 1969, for greater comparability across plants of the section of the life cycle that we characterize.

[18] 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

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.[19]

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 11.15 (9) inputs per year (Table 2).

Plant-specific unit prices on inputs at the product level 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 used by Hottman et. al. (2016), which implies the limitations discussed in the introduction, 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

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).

[19]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.

we do not observe from birth.[20] 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 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 (2020) CES Unified Price Index (CUPI) approach as the appropriate empirical analogue or our theoretical price index. The CUPI 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 CUPI log change in $f's$ price index is given by:

$$\ln \frac{P_{ft}}{P_{ft-1}} = \sum_{\Omega_{t,t-1}^f} \ln \left(\frac{p_{fjt}}{p_{fjt-1}}\right)^{\frac{1}{\left\|\Omega_{t,t-1}^f\right\|}} + \frac{1}{\sigma-1}\left(\ln \lambda_{ft}^{QRW} + \ln \lambda_{ft}^{Qfee}\right) \tag{15}$$

$\Omega_{t,t-1}^f$ 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. $\lambda_{ft}^{QRW} = \prod_{\Omega_{t,t-1}^f} \left(\frac{s_{fjt}^*}{s_{fjt-1,\Omega_{t,t-1}^f}^*}\right)^{\frac{1}{\left\|\Omega_{t,t-1}\right\|}}$ 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.[21] The derivation of the CUPI price index is presented 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 CUPI adjusts for relative appeal

[20] See Appendix 1.2 for details.

[21] 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.

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 $\overline{P^*_{ft}} = \prod_{l=B+1}^{t} \left[\prod_{\Omega_{t,t-1}} \left(\frac{p_{fjt}}{p_{fjt-1}} \right)^{\frac{1}{\|\Omega_{t,t-1}\|}} \right]$, $\Lambda^{QRW}_{ft} = \prod_{l=B+1}^{t} \left[\left(\lambda^{QRW}_{fl} \right) \right]$ and $\Lambda^{Qfee}_{ft} = \prod_{l=B+1}^{t} \left[\left(\lambda^{Qfee}_{fl} \right) \right]$, we obtain:

$$\begin{aligned} P_{ft} &= P_{fB} * \overline{P^*_{ft}} * \left(\Lambda^{QRW}_{ft} \Lambda^{Qfee}_{ft} \right)^{\frac{1}{\sigma-1}} \\ &= P_{fB} * \overline{P^*_{ft}} * \left(\Lambda^{Q}_{ft} \right)^{\frac{1}{\sigma-1}} \end{aligned} \tag{16}$$

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 (16), to move from our calculated $\overline{P^*_{ft}}$ to the exact price index P_{ft}, we need an adjustment for the factor $\left(\Lambda^{Q}_{ft} \right)^{\frac{1}{\sigma-1}}$, 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 $\left(\Lambda^{Q}_{ft} \right)^{\frac{1}{\sigma-1}}$ into the derivations of section 4, where its contribution to price variability is flexibly estimated. In particular

$$Q^*_{ft} = \frac{R_{ft}}{P_{fB}\overline{P^*_{ft}}} = Q_{ft} * \left(\Lambda^{Q}_{ft} \right)^{\frac{1}{\sigma-1}} \tag{17}$$

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.[22]

[22] 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 appendix I. We find that the impact of deflating with quality-adjusted plant-level price indices is more important on the output relative to the input side.

4 Estimating $TFPQ$ and demand shocks

Calculating $TFPQ$ and demand shocks requires estimating the production and demand functions, (1) and (13). Once the coefficients of these functions have been estimated, $TFPQ$ is the residual from (1) and the demand shock is the residual from (13).

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 in multiproduct settings. 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 estimate both production and revenue elasticities.

Beyond the usual simultaneity biases and restrictions on supply vs demand , the estimation of (1) and (13) 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_{ft}^{Q}. 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 (13), using $Q_{ft}^{*} = \frac{R_{ft}}{P_{fB}\overline{P_{ft}^{*}}} = Q_{ft} * \left(\Lambda_{ft}^{Q}\right)^{\frac{1}{\sigma-1}}$ and $\overline{P_{ft}^{*}} = \frac{P_{ft}\left(\Lambda_{ft}^{Q}\right)^{\frac{-1}{\sigma-1}}}{P_{fB}}$ as the respective dependent variables / regressors of these two functions. We carry Λ_{ft}^{Q} 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_{ft}^{M} in the production function (where $M_{ft}^{*} = \frac{\text{materials expenditure}}{PM_{fB}\overline{PM_{ft}^{*}}}$, and Λ_{ft}^{M} is the adjustment factor for the prices of materials analogous to Λ_{ft}^{Q} see Appendix A). 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 $P_{ft} = P_{fB} * \overline{P_{ft}^{*}} * \left(\Lambda_{ft}^{Q}\right)^{\frac{1}{\widehat{\sigma}-1}}$ 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}. 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:

$$\ln Q_{ft} = \alpha \ln K_{ft} + \beta \ln L_{ft} + \phi \ln M_{ft} + \ln A_{ft} \tag{18}$$

and

$$\ln P_{ft} = \alpha - \frac{1}{\sigma} \ln Q_{ft} + \ln D_{ft} \tag{19}$$

where $Q_{ft} = \left(\frac{R_{ft}}{P_{ft}}\right)$. Using (16) and (17), the system can be rewritten:

$$\ln Q^*_{ft} = \alpha \ln K_{ft} + \beta \ln L_{ft} + \phi \ln M^*_{ft} + \frac{1}{\sigma - 1} \ln \Lambda^Q_{ft} - \frac{\phi}{\sigma - 1} \ln \Lambda^M_{ft} + \ln A_{ft} \tag{20}$$

and

$$\ln \left(\overline{P^*_{ft}} P_{fB}\right) = \alpha - \frac{1}{\sigma} \left(\ln Q^*_{ft} + \ln \Lambda^Q_{ft}\right) + \ln D_{ft} \tag{21}$$

We estimate (20) and (21), 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 Olley ad Pakes (1996); Levinsohn and Petrin (2003); 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 demand estimation approaches rely on assumptions regarding orthogonality between demand and supply shocks at some particular level. Foster et al (2008) impose orthogonality between the levels of $TFPQ$ and demand shocks, while in Broda and Weinstein (2006) and Hottman, Redding and Weinstein (2020) double-differenced demand and marginal cost shocks are assumed orthogonal.

We build on these approaches, but take advantage of prices and quantities for both inputs and outputs, and the consequent possibility of jointly estimating (20) and (21), 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 $TFPQ$:

$$lnA_{ft} = \pi_0^A + \pi_1^A lnA_{ft-1} + \pi_2^A lnA_{ft-1}^2 + \pi_3^A lnA_{ft-1}^3 + \xi_{ft}^A$$

That is, ξ_{ft}^A 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):

$$E\begin{bmatrix} lnM_{ft-1}^* \times \xi_{ft}^A \\ lnL_{ft} \times \xi_{ft}^A \\ lnK_{ft} \times \xi_{ft}^A \\ lnD_{ft-1} \times \xi_{ft}^A \\ \ln A_{ft} \\ \ln D_{ft} \end{bmatrix} = 0 \tag{22}$$

As in ACF-based methods, we purge measurement error in a first stage of the estimation (Appendix F) and 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 (22) we impose lagged materials demand to be orthogonal to current $TFPQ$ innovations, while L and K are required to be contemporaneously orthogonal to ξ_{ft}^A. 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). We thus follow DeLoecker et. al. (2016) in treating L as quasi-fixed for purposes of estimation.

The condition that D_{ft-1} must be orthogonal to ξ_{ft}^A identifies σ, following the logic that the slope of the demand function can be inferred taking advantage of shocks to supply. Foster et al (2008, 2016) and Eslava et al (2013) relied on the same logic but imposed orthogonality between demand and technology shocks in levels. This effectively precludes the possibility that firms endogenously invest in quality when they perceive better returns (as would be the case with higher $TFPQ$) and correlations between demand shifters and $TFPQ$ shocks if greater quality is more difficult to produce.[24] Hottman, Redding and Weinstein (2016) and Broda and Weinstein (2006, 2010) address these concerns by imposing orthogonality between double-differenced demand

[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 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.

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 $E(\ln D_{ft-1} \times \xi^A_{ft})$ which permits a correlation between changes in $TFPQ$ and demand even over time 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 digit sector of ISIC revision 3, adapted for Colombia (CIIU by its acronym in Spanish). 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. The revenue function curvature parameter stands at an average 0.63. This parameter in the literature usually ranges 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 revenue curvature parameter of 0.67. While our average estimated curvature of the revenue function is not far from that imposed by HK, there is substantial dispersion across three-digits sectors. We show below how ignoring this heterogeneity dampens the estimated contribution of wedges to sales variability. It is encouraging that we obtain plausible factor elasticities for most sectors at the three digits sector 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 levels of disaggregation.[26]

[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.

[26] Still, if fully unconstrained, our joint estimation does deliver implausible results for a few sectors: negative factor elasticities for some, and implausible curvature parameters of the revenue function for others ($\gamma(1-\frac{1}{\sigma}) > 0.9$ or < 0.1). For those sectors, we assign the average production and demand elasticites of the corresponding two digit sector.

Table 1. Factor and demand elasticities						
Sector	**β**	**α**	**ϕ**	**σ**	**γ**	**γ(1-1/σ)**
Average	0.28	0.12	0.61	3.47	1.01	0.63
Min	0.05	0.06	0.36	1.53	0.91	0.31
Max	0.50	0.29	0.85	7.61	1.08	0.88

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}.[27] 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 outcome 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, as previously stated, 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.[28]

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 attribute) 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. Revenue grows four-fold on average 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 its number of workers, and 25 years after birth employment it has grown more than three-fold.[29]

These average growth dynamics, however, hide considerable heterogeneity. Median

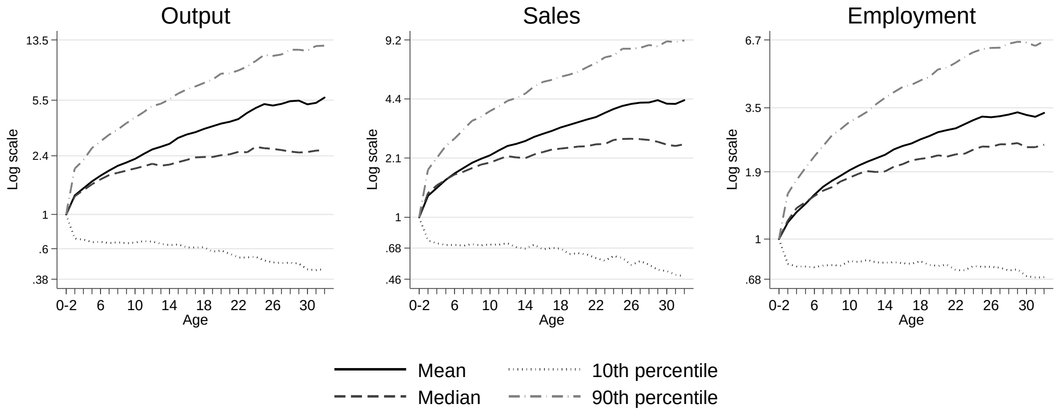
[27] I.e. we use the same measurement approach incorporating multi-materials inputs to construct the plant-level deflator for materials, and use it to deflate expenditures in materials to arrive at materials inputs. We use the same elasticity of substitution at the sectoral level for this purpose.

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

[29] 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 $a = 25$.

Figure 1: Distribution of life cycle growth

Current to initial



Within sector*year.

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.It is this heterogeneity and its welfare implications that we aim to explain in the analysis below.

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 Plant attributes

Table 2 presents basic summary statistics for (the idiosyncratic component of) sales and our estimates of output, output prices, $\ln A_{ft}$, $\ln D_{ft}$, wedges, markups and input prices. We note that we have adjusted materials prices for quality, but have not done the same for wages as yet due to data constraints. In section 7.1 we do quality-adjust wages for a subperiod for which this is possible.

Idiosyncratic dispersion in sales, output, output prices, $TFPQ$, demand, wedges and input prices is large. $TFPQ$ is strongly negatively correlated with output prices, which is intuitive to the extent that more efficient production allows charging lower prices (consistent with findings for Colombia in Eslava et al., 2013, and for commodity like products in the US in Foster et al. 2008, 2016, though 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

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
17,351	4,352	3.56	1	2	5	11.15	5	9	14

Panel B.Standard deviations and correlation coefficient for outcomes and fundamentals (within sector*year, all variables in logs, average sector)

	Standard Deviation	Sales	Output	Output prices	TFPQ	Demand Shock	Input prices	Average wage	Markup	Sales Wedge
Sales	1.426	1.000								
Output	1.588	0.897	1.000							
Output prices	0.672	0.000	-0.427	1.000						
TFPQ	0.841	0.160	0.440	-0.681	1.000					
Demand Shock	0.673	0.743	0.403	0.621	-0.287	1.000				
Input prices	0.640	-0.046	-0.106	0.152	0.327	0.063	1.000			
Average wage	0.422	0.606	0.523	0.054	0.120	0.483	0.001	1.000		
Markup	0.029	0.625	0.566	-0.012	0.104	0.459	-0.034	0.400	1.000	
Sales Wedge	1.076	-0.176	-0.127	-0.070	-0.471	-0.194	0.008	-0.038	-0.056	1.000

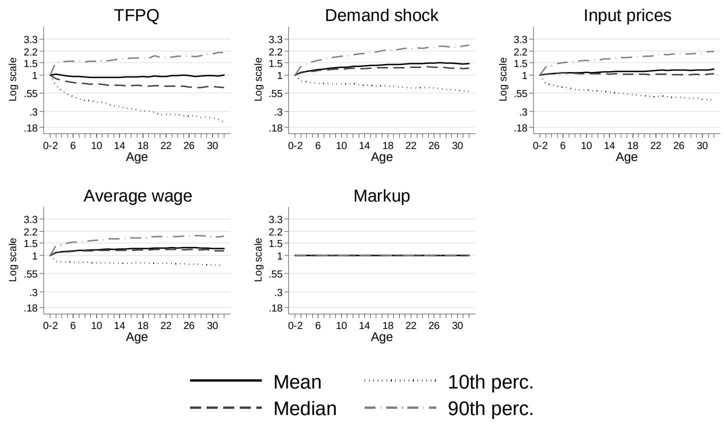
Note: The sample includes fewer plants than the original Manufacturing Survey, especially in the early years of the sample, due to the restriction of plants born after 1969.

negatively correlated. This is indeed the case in our estimates, also consistent with Forlani et al. (2018). Though markups display only modest variation across plants, they are positively correlated with $TFPQ$, D and wages. Especially interesting is the negative and strong correlation of wedges with $TFPQ$ and demand shocks, suggesting that the plants with the best fundamentals are implicitly taxed the most.[30] These basic correlation patterns remain true for within-plant correlations, and are echoed in our growth decompositions below.

The within sector*year distributions of the evolution over the life cycle of fundamentals and wedges are displayed in Figure 2, including the life cycle growth of $TFPQ$ and demand shocks, markups, material input prices and wages. The average growth of demand shocks dominates that of input prices, and both dominate the average growth of $TFPQ$ and markups 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 $TFPQ$-demand patterns in Figure 2 is the evolution of the negative correlation between the life cycle growth of $TFPQ$

[30]Log wedges are residuals: $\ln \chi_{ft}^{level} = \ln \left(\frac{R_{ft}}{d_{jt}^{\kappa_1} a_{jt}^{\kappa_2} pm_{jt}^{-\phi\kappa_2} w_{jt}^{-\beta\kappa_2} \mu_{jt}^{-\gamma\kappa_2}} \right)^{\frac{1}{1-\frac{1}{\sigma}}}$ (see equation 5, where we ignore χ_t since table 2 is based only idiosyncratic variation)

Figure 2: Distribution of Fundamentals
Current to initial



Within sector*year

and that of demand shocks. At age 3, the correlation is -0.191, moving to -0.289 at age 10 and -0.349 at age 20. 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. Markups barely vary over the life cycle and across deciles of the distribution, to the point that the variation is not observable to the naked eye compared to the scale of variation of other fundamentals. As we will see below, when we consider activity-weighted distributions and related measures (e.g., welfare), markups play an important role as a relatively small share of very large plants have very high markups.

5.3 Decomposing growth into fundamental sources

We now decompose the variance of $\frac{R_{ft}}{R_{f0}}$ into contributions associated with different fundamental sources (equation (5)). We follow a two stage procedure, similar to that in Hottman et al. (2016), whose details are provided in Appendix G. As we prove in that Appendix, the contribution of growth in each (log) fundamental to the variance of growth of (log) sales depends on the covariance and relative variances between the two. In particular, the contribution of the life cycle growth of $TFPQ$ to the life cycle growth of sales is given by the product: $\kappa_2 * corr\left(\frac{a_{it}}{a_{i0}}, \frac{R_{it}}{R_{i0}}\right) * \frac{std\left(\frac{a_{it}}{a_{i0}}\right)}{std\left(\frac{R_{it}}{R_{i0}}\right)}$ where κ_2 is the structural parameter associated with $TFPQ$ in the decomposition equation 5, reproduced below:

$$\frac{R_{ft}}{R_{f0}} = \left(\frac{d_{ft}}{d_{f0}}\right)^{\kappa_1} \left(\frac{a_{ft}}{a_{f0}}\right)^{\kappa_2} \left(\frac{pm_{ft}}{pm_{f0}}\right)^{-\phi\kappa_2} \left(\frac{w_{ft}}{w_{f0}}\right)^{-\beta\kappa_2} \left(\frac{\mu_{ft}}{\mu_{f0}}\right)^{-\gamma\kappa_2} \left(\frac{\widehat{\chi}_t}{\chi_0}\frac{\chi_{ft}}{\chi_{f0}}\right)^{1-\frac{1}{\sigma}} \tag{23}$$

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 contribution of other sources of growth is calculated in an analogous manner.

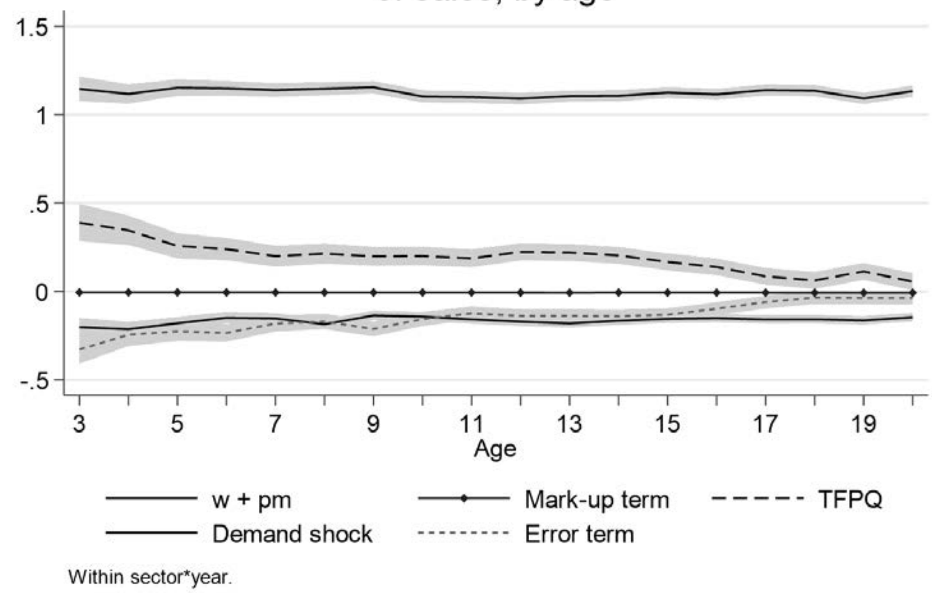
The term $\left(\frac{\widehat{\chi}_t}{\chi_0}\frac{\chi_{ft}}{\chi_{f0}}\right)^{1-\frac{1}{\sigma}}$ in (23) is calculated as a residual, since all of the other components are either measured or estimated. From equation (5) , error term $\ln\frac{\chi_{ft}}{\chi_{f0}}$ captures life cycle growth in 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 implement the variance decomposition by ages.[31] Results are presented in Figure 3. We find that the structural contribution of fundamentals, rather than residual wedges, 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). Averaging over ages, on a weighted basis, we find contributions of the demand shock, $TFPQ$, input prices, and structural wedges, respectively equal to 1.13, 0.15, -0.16 and -0.12. That is, the demand shock is over seven times as important as $TFPQ$ to explain idiosyncratic sales growth . Input prices make smaller, but far from negligible, contributions. Mechanically, this reflects the fact that, for the average sector and pooling across ages, the covariance of demand shocks growth with sales growth is more than six-fold that between $TFPQ$ growth and sales growth, and the coefficient associated with demand growth in equation (23) is also much larger than that for $TFPQ$ (Table 4). The significant negative correlation between $TFPQ$ and demand shocks undoubtedly plays a role in these patterns. The contribution of markups to the variance of sales growth is minimal, not even visible in the graph, reflecting market shares concentrated around zero and barely changing over the life cycle in most sectors. A few plants in some sectors hold large market shares. Though their low numbers imply that these plants do not play an important role in explaining the cross sectional distribution, we show in section 7 that their large markups have crucial welfare implications.

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,

[31] See Appendix G for details.

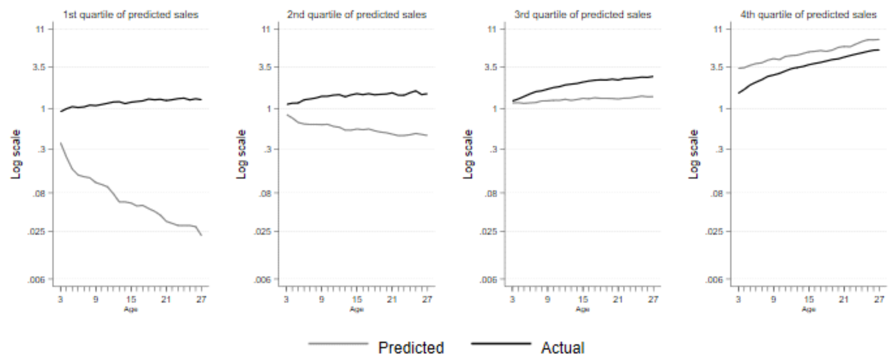
Figure 3. Variance decomposition of the life cycle growth of sales, by age



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, 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, including the geographic segmentation of labor markets as well as institutional barriers or other frictions in the labor market. However, the correlations in Table 2 suggest that wages variability might also rather reflect unmeasured quality differences since, by contrast to material inputs prices we are unable to quality adjust wages for our entire sample period. Section 7.1 explores the role of these different mechanisms for a subperiod in which quality adjustment is possible. Previewing those results, adjusting wages for labor quality reduces the contribution of wage dispersion in accounting for sales growth heterogeneity and increases the contribution of TFPQ. This is not surprising as adjusting for labor quality impacts the measurement of technical efficiency. The effect of quality adjustment, however, is not large even for $TFPQ$ and wages, and does not affect other components, so we proceed with our main full sample results as a baseline that provides robust inferences.

Figure 4: Contribution of fundamentals to life-cycle growth
Sales, by predicted growth percentiles



Idiosyncratic components

Another important feature of these results is that the remaining wedge also contributes negatively to the variance of life cycle growth of sales (or, equivalently, 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: pooling ages, sales dispersion is dampened by about 12% with respect to that implied by fundamentals. 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 the life cycle growth of sales vary significantly depending on the horizon of growth considered. Demand becomes increasingly important compared to $TFPQ$ over longer horizons. This is because the correlation between sales growth and $TFPQ$ growth decreases for older plants, while that between sales and demand remains fairly stable (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 sales variability compared to that implied by fundamentals more among young plants than among older ones (left panels of Figure 4), and this is because their (negative) correlation with sales becomes increasingly loose as plants age.

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. However, $TFPQ$ plays a relatively more important role vis-a-vis demand for the latter than the former.

Table 3. Moments of the distribution of life cycle growth for sales and fundamentals (Average sector, age<=20)

Life Cycle Growth of:	Age = 3 years		Age = 10 years		Age = 20 years	
	St. Dev.	Correlation with sales growth	St. Dev.	Correlation with sales growth	St. Dev.	Correlation with sales growth
Sales	0.393	-	0.676	-	0.910	-
TFPQ	0.369	0.242	0.581	0.152	0.764	0.126
Demand shock	0.210	0.683	0.378	0.696	0.505	0.678
Material prices	0.248	0.035	0.405	0.042	0.538	0.034
Wages	0.224	0.240	0.314	0.285	0.349	0.340
Markup	0.002	0.549	0.005	0.481	0.008	0.448
Wedge	0.645	-0.232	0.955	-0.197	1.164	-0.140
Coefficients	κ1	κ2	γ	σ	ϕ	β
(median sector)	2.546	1.526	1.013	2.496	0.618	0.255

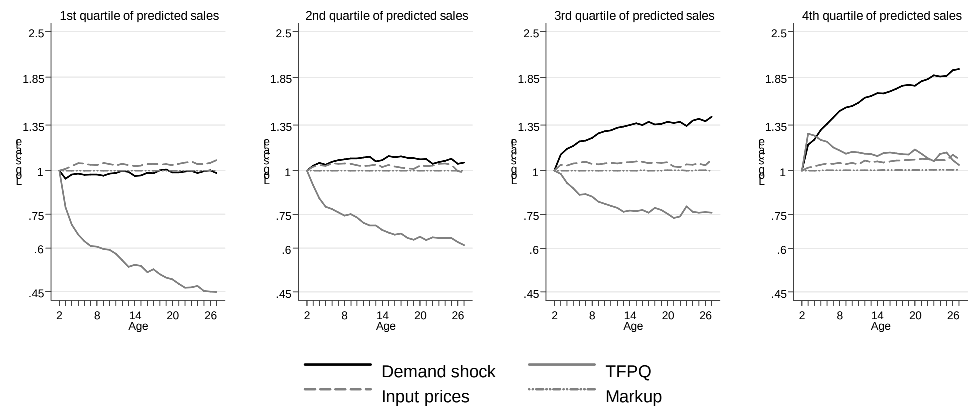
Notes: the top panel of this table presents, standard deviations for the life cycle growth of different measured fundamentals, and coefficients of correlation between them and the life cycle growth of sales, calculated across plants of the average sector. The bottom panel presents coefficients used to calculate loading factors for the contribution of each fundamental in the life-cycle revenue decomposition, for the median sector. The contribution of a given fundamental to life cycle growth is given by the product between the corresponding loading factor, correlation coefficient, and standard deviation, divided by the standard deviation of sales. This calculation of the contribution holds exactly within sectors, so appropriate caution is necessary in comparing Table 3 to Figures 3 and 4, where sectors are pooled together.

Figure 4 shows the mechanics behind the negative contribution of structural wedges: the average 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. Predicted growth corresponds to growth in equation (23) setting $\frac{\chi_{ft}}{\chi_{f0}} = 0$. Figure 4 implies that it is plants with weak growth in fundamentals that are implicitly subsidized while those with strongest fundamentals are implicitly taxed, especially at young ages.

Figure 5 indicates that plants in the highest percentiles of predicted growth have both higher average demand growth and higher average $TFPQ$ growth 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. In the opposite end of the distribution, it is weak $TFPQ$ growth that explains why the bottom quartiles plants are classified as such.

We conduct a similar decomposition for the growth of output, rather that sales, finding very similar results. An exception is the fact that $TFPQ$ plays, by far, the predominant role in explaining the variance of output growth, with a contribution four times as large as that of the demand shock. As in the case of sales, the structural wedge dampens the variability of output growth, with an average contribution of -0.15. Results for the output decomposition are presented in Appendix G along with a related reduced form decomposition.

Figure 5: Life-cycle fundamentals growth
By predicted sales growth percentiles



Idiosyncratic components

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

In absence of data on input and output prices HK decompose revenue and revenue growth into a measure of fundamentals that combines our $TFPQ$ and D shocks, which we label as $TFPQ\_HK$, and a residual wedge that captures all determinants of size other than efficiency and demand.[32] They start from revenue, which in our notation is given by: $R_{ft} = D_{ft} Q_{ft}^{1-\frac{1}{\sigma}} = D_{ft}\left(A_{ft} X_{ft}^{\gamma}\right)^{1-\frac{1}{\sigma}}$. With estimates of γ and σ one can obtain the composite shock $TFPQ\_HK$ solely from revenue and input data as:

$$TFPQ\_HK_{ft} = R_{ft}^{1/(1-\frac{1}{\sigma})} / X_{ft}^{\gamma} = A_{ft} D_{ft}^{\frac{1}{1-\frac{1}{\sigma}}} \tag{24}$$

Life cycle growth in revenue can then be expressed as:

$$\frac{R_{ft}}{R_{f0}} = \left[\left(\frac{TFPQ\_HK_{ft}}{TFPQ\_HK_{f0}}\right)\left(\frac{(1-\tau_{ft})}{(1-\tau_{f0})}\frac{C_{f0}\mu_{f0}}{C_{ft}\mu_{ft}}\right)^{\gamma}\right]^{\frac{1-\frac{1}{\sigma}}{1-\gamma\left(1-\frac{1}{\sigma}\right)}} \tag{25}$$

[32] See the appendix to HK (2009) where they extend their model to account for D shocks. What we label $TFPQ\_HK$ 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.

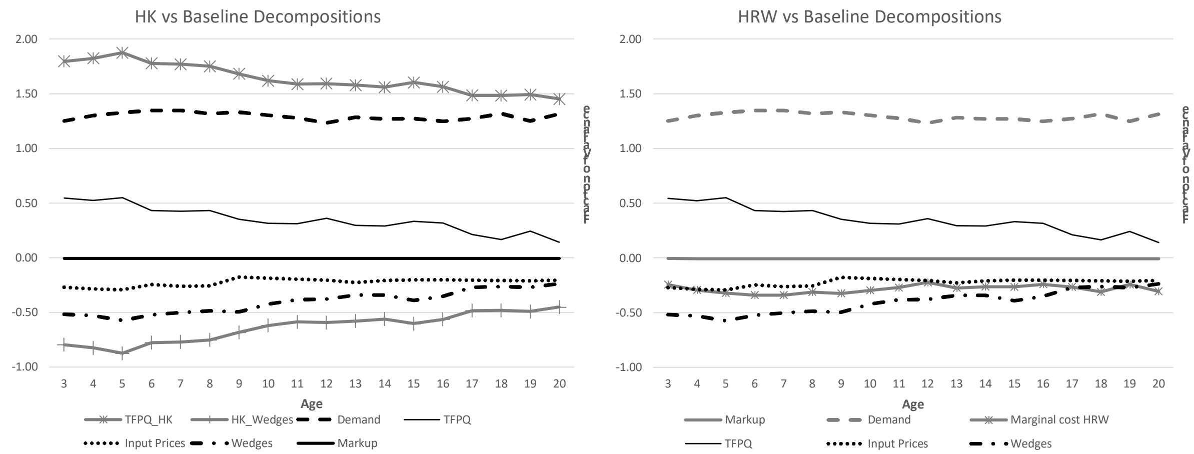
That is, the HK residual wedge is a *composite* measure of wedges, the "HK wedge"= $\left(\frac{(1-\tau_{ft})}{(1-\tau_{f0})}\frac{C_{f0}\mu_{f0}}{C_{ft}\mu_{ft}}\right)^{\gamma}$, just as TFPQ_HK is a *composite* measure of efficiency and demand. A widely used implication of HK's framework is that wedges can be estimated from the idiosyncratic component of $TFPR\_HK = \frac{R_{ft}}{X_{ft}}$. Replacing optimal input demand $X_{ft} = \left(\frac{D_{ft}A_{ft}^{1-\frac{1}{\sigma}}\gamma}{C_{ft}\mu_{ft}(1-\tau_{ft})^{-1}}\right)^{\frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}}$ we obtain $TFPR\_HK_{ft} = \frac{C_{ft}\mu_{ft}}{\gamma(1-\tau_{ft})}$, so $TFPR\_HK$ variability reflects variation not only τ, but also in markups and input prices.[33] We thus observe that the *composite* wedges we obtain from (25) are analogous to those that can be obtained from $TFPR\_HK$ but also that, given the importance of input variability in our data to explain the growth distribution, $TFPR\_HK$ dispersion cannot be used to infer the dispersion of τ.

Figure 7, left panel, contrasts the by-age decomposition using the $TFPQ\_HK$ approach (grey lines) with that of our approach (black lines). To calculate $TFPQ\_HK_{ft} = R_{ft}^{1/(1-\frac{1}{\sigma})}/X_{ft}^{\gamma}$ we use our estimates of σ, ϕ, β, α, and the implied $X = M_{ft}^{\frac{\phi}{\gamma}}L_{ft}^{\frac{\beta}{\gamma}}K_{ft}^{\frac{\alpha}{\gamma}}$. The figure shows that a non-negligible fraction of the variation explained by the *composite* wedges in a two-way decomposition is due to the contribution of variable input prices and markups (25% out of the 40% assigned to wedges in 7a). It is clearly instructive to isolate the contribution of input prices and markups from residual wedges; input price and markup variability may well be related to market distortions but may also reflect structural features (e.g., market segmentation) of input and output markets. Figure 7, however, also 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. Another important insight from Figure 7 is that using $TFPQ\_HK$ misses the changing relative contribution of demand vs. $TFPQ$ over the life cycle; Figure 6 shows that the increasingly dominant role of demand is driven by the upper quartile "superstar" plants while weak $TFPQ$ growth dominates the poorly performing lower quartiles. These insights about the relative role of $TFPQ$ vs. demand, and the relevance of input prices and markups in the HK composite wedge, are not possible in the two-way decomposition based on revenue data.

Another important gain of using detailed P and Q data stems from the ability to estimate sector-specific parameters. Appendix C reports results with the composite $TFPQ\_HK$ approach following the usual practice in the misallocation literature of imposing monopolistic competition; $\gamma = 1$; ϕ, β, α equal to the corresponding cost shares in each sector; and a common σ for all sectors. Results show that the estimated contribution of wedges closely depends on the level and dispersion of σ and of the

[33] If, as originally defined in Foster et al (2008), we rather defined $TFPR$ as $\frac{R_{ft}}{X_{ft}^{\gamma}}$, $TFPR$ dispersion would also reflect A_{ft} and D_{ft} dispersion. Their definition of $TFPR_{ft} = P_{ft}A_{ft}$. $TFPR\_HK_{ft}$ corresponds to HK's definition if $\gamma = 1$.

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 Hsieh and Klenow (2009, 2014) and Hottman, Redding and Weinstein (2016). Components of our baseline decomposition (from Figure 3) are depicted in black if they are not also a component of the HK or HRW decomposition in the respective panel, while components of the HK and HRW are depicted in grey in the respective panel.

implied curvature of the revenue function. The estimated contribution of wedges grows with σ in a manner that is not linear (see Figure C1 in Appendix C). For this reason, the contribution of wedges is much lower when imposing a common σ equal to the average of the σ that we estimate by sectors; because the contribution of wedges is larger in sectors with low curvature, taking into account the dispersion of σ recognizes a more dominant role for wedges. Appendix C provides further support for these claims

Summarizing, there are three main messages about the value of P and Q output and input data in our estimation. Using only revenue and input data (but the internally consistent estimated demand and production elasticities) yields: 1) an inability to identify the distinct contributions of demand and $TFPQ$ which have distinct contributions over the life cycle and over different segments of the distribution of life cycle growth rates; and 2) an inability to isolate the contribution of idiosyncratic input prices and markups from residual wedges. Moreover, most of the literature using the HK methodology assumes the same elasticity of substitution for all sectors, and in general relies on strong assumptions to estimate the coefficients of the production function. Such constraints on measurement and estimation, most importantly ignoring that some sectors exhibit large elasticities of substitution, leads to underestimations of the role of wedges.

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

The differential contribution of demand vs. cost-side shocks 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 observed prices and demand shocks estimated using the estimated elasticity of substitution, and subsequently decompose price into the contributions of markups–computed as in equation (4)–and residual marginal costs:

$$\mu_{ft} = \frac{P_{ft}}{\frac{\partial CT_{ft}}{\partial Q_{ft}}(1-\tau)^{-1}}$$

where CT is total cost. These residual marginal costs, given by $\frac{\partial CT_{ft}}{\partial Q_{ft}}(1-\tau)^{-1}$, thus capture idiosyncratic variation in costs (from input price variability and technical efficiency), as well as wedges. Importantly, wedges are not inherently driven by cost/supply side factors. For example, they could reflect the adjustment costs associated with building up a customer base. See Appendix K for greater details.

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 conceptual approach.[34] 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. The right panel of Figure 7 illustrates the by-age decomposition obtained in our data with the HRW approach (components in grey) vs. our baseline decomposition (components in black, plus demand and markup, which are separately identified in both HRW and our approach). 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. The lumping together of cost, productivity and wedges also misses the rich life cycle dynamics of each of these factors. Technical efficiency becomes less important as do wedges for older businesses in our baseline framework but this pattern is missed completely in the HRW approach. Related, the increasing magnitude of the inverse correlation between demand and $TFPQ$ over the life cycle is missed in the HRW approach.

[34] By this we mean their conceptual approach to the decomposition of sales volatility. Given the differences in their data infrastructure relative to ours, their identification of the demand and supply/cost components is related but distinct from our approach.

7 Welfare implications of heterogeneity in wedges and plant fundamentals

We use $U = \left(\sum_{I_t} d_{ft} Q_{ft}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ in equation 6 to analyze welfare implications of heterogeneous wedges and fundamentals. Replacing equation 1 into this expression after having inserted the optimal expression for X_{ft}, we obtain an expression for welfare that, up to aggregate shocks, can be calculated from plant attributes and wedges that we have estimated:

$$U = \left(\sum_{I_t} d_{ft} Q_{ft}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} = \left(\sum_{I_t} d_{ft} (d_{ft}^{\gamma\kappa_1} a_{ft}^{1+\gamma\kappa_2} pm_{ft}^{-\phi\kappa_1} w_{ft}^{-\beta\kappa_1} \mu_{ft}^{-\gamma\kappa_1} \chi_t \chi_{ft})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \quad (26)$$

where χ_{ft} is the residual $\chi_{ft} = \frac{Q_{ft}}{d_{ft}^{\gamma\kappa_1} a_{ft}^{1+\gamma\kappa_2} pm_{ft}^{-\phi\kappa_1} w_{ft}^{-\beta\kappa_1} \mu_{ft}^{-\gamma\kappa_1}}$. We build a series of counterfactual welfare ratios, where welfare is compared to what its level would be in the hypothetical efficient situation where the composite (HK) wedge is set to one:

$$\frac{U^{count}}{U^{HKeff}} = \frac{\left(\sum_{I_t} d_{ft}^{count} Q_{ft}^{count\,\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}}{\left(\sum_{I_t} d_{ft} Q_{ft}^{HKeff\,\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}} \quad (27)$$

Q_{ft}^{HKeff} is the value obtained by setting the composite HK wedge equal to one: $\left(pm_{ft}^{-\phi\kappa_1} w_{ft}^{-\beta\kappa_1} \mu_{ft}^{-\gamma\kappa_1} \chi_{ft}\right) = 1$. Aggregate shocks χ_t cancel out in expression (27). This measure of welfare is for a single sector. We compute this ratio for the average sector including on a revenue-weighted basis. The latter is equivalent to Cobb-Douglas aggregation across sectors neglecting any between sector aggregation effects that arise from goods in one sector being used in the production of other sectors. The approach of using a nested CES within sectors with multi-product producers and Cobb-Douglas aggregation between sectors is used in HRW.

In different counterfactual cases we set d_{ft}, a_{ft},pm_{ft},w_{ft}, and/or μ_{ft} to 1, keeping χ_{ft} at its actual level. In another case, we set $\chi_{ft} = 1$, keeping the other components at their actual levels. We compare our results for (27) to a benchmark case where the numerator corresponds to actual welfare, that is, all plant attributes are at their actual levels. Table 4 presents the results of this analysis.

Panel A quantifies the welfare gap attributed to the presence of HK wedges (the benchmark case just described). Columns 1 and 2 show a large gap: for the average sector, actual welfare is 28% of its efficient level (what it would be in absence of HK wedges). The figure is much lower (13%) on a revenue-weighted basis, because it is in the largest sectors where wedges are largest. This is the case because large sectors tend to display high elasticities of substitution. The implication of moderate curvature of the revenue function also implies that optimality would shift more resources to the plants with highest composite productivity ($TFPQ\_HK$), while in fact these sectors tend to be large precisely because they are fractioning revenue in a large number of plants. As a result, estimated wedges are large.

Welfare losses associated to HK wedges in columns 1 and 2 of panel A are larger to those usually obtained in the HK-based literature. Columns 3 and 4, show that numbers closer to that tradition are obtained by using constant (average) production and demand parameters in equations (26) and (27), as is frequently done in the literature, especially in the case of elasticities of substitution.[35] Moreover, revenue weighting reduces the size of the estimated welfare gap when using average parameters, while the opposite is true when using sector-level parameters. These findings point again at the fact, discussed above, that wedges (and the associated losses) tend to be underestimated when imposing the same parameters to all sectors, specifically for sectors that have larger-than-average elasticities of substitution, which also tend to be the largest sectors (see also Appendix C). This highlights the quantitative importance of properly estimating the parameters of the revenue function.[36]

Panel B of Table 4 further analyzes the impact of shutting down variability in each of the sources of plant heterogeneity considered in the analysis, individually and in combinations. (Because the different components are correlated with each other, the impact of shutting down two components simultaneously may be larger or smaller than the sum of the impact of shutting them individually). The fact that consumers in our model display love for variety tends to reduce welfare when $TFPQ\_HK$ variability is shut down. On an unweighted basis, this reduces welfare from 28% to 12% of its efficient level. More interestingly, shutting down quality (or D) heterogeneity alone has an impact almost as large as that of shutting down $TFPQ\_HK$, highlighting again the differential role of $TFPQ$ vs. demand, now in terms of welfare.[37]

On a similar note, unpacking HK wedges into their components sheds light on the sources of welfare losses from these wedges. In particular, heterogeneity in input

[35] Plant attributes, however are kept at their measured levels, for which the sector level parameters were used.

[36] While it is not their primary focus, Baqaee and Farhi (2020) find that taking into account heterogeneity in elasticities of substitution is important for their generalized measure of allocative efficiency. A core focus of Baqaee and Farhi (2020) is taking into account the input-output structure of the economy in aggregating the impact of distortions at the economy-wide level. We don't explore such implications but doing so would be of interest in future research with our data infrastructure with price and quantity data for both outputs and inputs at the plant-level.

[37] In results reported in Appendix L, we have also explored the effect of shutting down variability over the life cycle, i.e. maintaining dispersion at birth as the only source of plant heterogeneity. Qualitative results are similar to those of shutting down both sources of heterogeneity in each plant attribute. Interestingly, there are asymmetric effects of collapsing the distribution of life cycle growth at the upper or the lower end of the distribution: forcing fundamentals of high growth plants (75th percentile) to grow at the mean level produces welfare losses that are proportionally larger than the welfare gain from conducting the analogous exercise with low growth plants.

Table 4: Counterfactual welfare relative to HK efficient welfare, setting specific plant attributes to constant mean value (=1).

		Sector-level Parameters		Average Parameters	
		Average Sector (1)	Average Sector - Revenue Weighted (2)	Average Sector (3)	Average Sector - Revenue Weighted (4)
Panel A: Actual to HK Efficient Welfare					
		0.278	0.127	0.363	0.686
Panel B: Counterfactual to HK Efficient Welfare					
Plant attribute set to counterfact. level	D+TFPQ (TFPQ_HK)	0.123	0.075	0.204	0.386
	Demand Shock	0.126	0.108	0.260	0.480
	Input prices + Markup	0.551	0.283	0.883	1.187
	Input prices	0.416	0.240	0.621	1.080
	Markup	0.338	0.144	0.479	0.735
	Wedge	0.490	0.365	0.545	0.537

prices and markups explains an important part of the welfare loss from HK wedges. Collapsing them both to their mean value (of 1), while keeping $TFPQ\_HK$ at its actual value, brings welfare to 55% of its efficient level. Shutting down variability in the residual wedge has a similar impact of moving welfare to 49% of the efficient level. And, this residual wedge has an individual impact that surpasses that of either input prices or markups alone.

Interestingly, though shutting down input prices alone has a larger impact than shutting down markups alone, the latter also has important impact, increasing welfare from 0.28 to 0.34 of the efficient level. This stands in contrast to the results of the decomposition of cross-plant heterogeneity of Figure 3, where markups play a very modest role. The reason is the combination of two facts: 1) while there is little variability in markups because most market shares are close to zero, a few large plants exhibit large markups; 2) the decomposition of Figure 3 explains cross plant dispersion on an unweighted basis, while aggregate welfare in (26), by its very nature, "weights" plants according to their appeal to consumers. Large markups thus play a much more important role in explaining aggregates than in explaining cross-plant variation. Table 5 illustrates that a few plants exhibit market shares well above their sectors' mean shares, despite the very low variability in markups shown in Table 2.

Table 5. Distribution of Largest Plants' Markup Relative to Sector*Year Average

	Pooling		
Percentile	Largest	Second	Third
Average	1.26	1.07	1.05
10	1.04	1.02	1.02
25	1.07	1.04	1.03
50	1.13	1.06	1.04
75	1.24	1.10	1.06
90	1.56	1.14	1.08
95	1.97	1.16	1.10
99	3.21	1.20	1.12

The table presents percentiles across sector-year combinations of the markups, relative to average, of the largest, second largest and third largest plant in the sector.

7.1 Robustness to quality-adjusting wages

Our counterfactual welfare analysis shows that heterogeneity in input prices implies non-negligible welfare loses. Input price heterogeneity, as previously discussed, may reflect input market frictions or accompanying distortions, as well as input heterogeneity. Although we have adjusted materials prices for quality (within the plant) the same is not true of wages, since the data does not break labor into skill categories for the full extent of our estimation period. To address the relative importance of quality heterogeneity for labor, we now take advantage of data on broad skill categories available for 2000-2012. The available skill categories are production workers without tertiary education, production workers with tertiary education and administrative workers. We construct, for that subperiod, quality-adjusted wages using an approach analogous to that of we use to build quality-adjusted materials and output prices, and a quality-adjusted labor input given by the payroll deflated with our adjusted wages. $TFPQ$ is also re-calculated using this quality adjusted input. We conduct our welfare analysis using these adjusted data. For completeness, we also conduct the sales growth decomposition with this adjustment. Results are presented in Table 6.

Implementing our decomposition with this alternative measure of wages rather than the average wage per worker reduces the negative contribution of wages for 2000-2012 from -0.065 to -0.026, compensating it with a reduced positive contribution of $TFPQ$ (Panel A). That is, quality heterogeneity explains about half the dampening role of unadjusted wages over the variance of sales. The remaining -0.026 is our estimate of the dampening effect of dispersion in quality-adjusted wages. The latter may stem from frictions or from distortions in the labor market that accompany such frictions. For

Table 6: Decomposition of life cycle growth and counterfactual welfare analysis to quality-adjustment of wages

		Unadjusted wage		Q-adj. wage
		1982-2012	2002-2012	2002-2012
Panel A: Decomposition of life cycle growth of sales				
TFPQ		0.154	0.287	0.244
Demand shock		1.128	1.141	1.141
ln pm		-0.078	-0.051	-0.051
ln wage (unadjusted)		-0.081	-0.065	-0.026
ln markup		-0.006	-0.005	-0.005
Wedge		-0.117	-0.307	-0.303
Panel B: Actual to HK Efficient Welfare (sector level parameters, average sector)				
		0.278	0.296	0.299
Panel C: Counterfactual to HK Efficient Welfare (sector level parameters, average sector)				
Plant attribute set to counterfact. level	D+TFPQ (TFPQ_HK)	0.123	0.278	0.295
	Demand Shock	0.126	0.262	0.288
	Input prices + Markup	0.551	0.571	0.549
	Input prices	0.416	0.472	0.469
	Markup	0.338	0.332	0.336
	Wedge	0.490	0.491	0.506

example, market segmentation due to search frictions can enhance monopsony power. We don't further explore such issues in our analysis but the finding that input prices matter even after quality adjustment suggests this is an important area for future research.

Interestingly, the welfare effect of input prices and markups is not much changed if wages are quality adjusted (Panels B and C). That is, wage quality adjustment matters significantly for size dispersion, but not so much for welfare, which is size-weighted. This suggests that wage dispersion affects welfare mainly because it captures monopsony power, or other frictions, associated to the largest firms, rather than because of the extent to which it reflects heterogeneity in the quality of the labor input.

8 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, while we are able to attribute a large part of the role of HK wedges to input price and markups dispersion, our remaining wedges are still a black box. Identifying the specific sources of wedges that dampen output and sales growth especially for young plants, beyond input prices and markups that we analyze, is one potential area of research. One natural candidate is adjustment costs that especially impact young businesses. These may include the costs of developing and accumulating organizational capital (such as the 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. Also, the fact that we decompose *composite productivity* into its technical efficiency and demand components yields guidance as to the potential source of wedges dampening growth.

Size-dependent policies and other characteristics of the regulatory environment are another 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. Future work that explored the relationship between regulatory and tariff reform and the evolution of the fundamentals and wedges we identify would be of interest.

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 (inversely) correlated, we do not take explicit account of the 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.

Another interesting area for future research is to explore approaches that take advantage of firm level prices on outputs and inputs to study the role of variation in technology and markups at the plant-level. Recent analyses by DeLoecker, Eeckhout and Unger (2020) highlight the potentially important role of markup dispersion across producers. They present evidence of substantial dispersion in markups across producers using an approach that is flexible on the structure of demand but that has the potential limitation of attributing to markups variation that may come from the structure of technology across producers. Our analysis using plant-level quality adjusted prices, while more restrictive in the sense of imposing a given demand structure, highlights challenges for pursuing this agenda. As we emphasize, even measuring plant-level output and inputs for multi-product plants who use a variety of inputs requires taking a stand on the demand structure. Tackling technology and markup heterogeneity in this multi-product, multi-input environment with ongoing quality change will be a challenge.

9 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.

Adamopoulos , T., and D. Restuccia (2014) "The Size Distribution of Farms and International Productivity Differences," American Economic Review, Volume 104 (6) , pp. 1667-97.

Asker, John, Allan Collard-Wexler and Jan De Loecker. 2014. “Dynamic Inputs and Resource (Mis)Allocation,” Journal of Political Economy, 122(5), 1013-1065.

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.

Baqaee, David and Emmanuel Farhi. 2020,"Productivity and Misallocation in General Equilibrium" Quarterly Journal of Economics:. (forthcoming).

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.

Brooks, Eileen L. (2006). “Why don’t firms export more? Product quality and Colombian plants”, Journal of Development Economics, Vol. 80, No. 1, pp. 160-178.

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.

David, Joel and Venky Venkateswaran. 2018. "The Sources of Capital Misallocation," *American Economic Review*, (forthcoming).

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.

De Loecker, J., J. Eeckhout, and G. Unger (2020). "The Rise of Market Power and the Macroeconomic Implications", *Quarterly Journal of Economics*, (forthcoming).

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

Edmond, C. V. Midrigan and D. Xu (2018) "How costly are markups," NBER Working Paper No. 24800

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. 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.

Gervais, A. (2015) Product quality and firm heterogeneity in international trade. *Canadian Journal of Economics*, 48(3), 1152-1174.

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.

Guner, N., G. Ventura and D. Xu (2008) "Macroeconomic Implications of Size-Dependent Policies," Review of Economic Dynamics, 11 (4), 721-744.

Hallak JC, Schott PK. 2011. Estimating Cross-Country Di erences in Product Quality. Quarterly Journal of Economics, 126(1): 417-74.

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 (2011) "The Long and Short (of) Quality Ladders," Review of Economic Studies, Oxford University Press, vol. 77(4), pages 1450-1476.

Kugler, M. and E. Verhoogen (2012) "Prices, Plant Size, and Product Quality." *Review of Economic Studies*, vol. 79 no. 1, pp. 307-339

Levinsohn, J. and Petrin, A. (2003) "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies* 317-342

Manova K, Zhang Z. 2012. Export Prices across Firms and Destinations. *Quarterly Journal of Economics* 127.

Melitz MJ. 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*. 71: 1695-725.

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *American Economic Review*, 104 (2): 422-58.

Olley, S. and Pakes, A.(1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica* 64:1263-1295

Redding, Stephen and David E. Weinstein (2020) "Measuring Aggregate Price Indexes with Taste Shocks: Theory and Evidence for CES Preferences", *Quarterly Journal of Economics*, 135(1), 503-560, 2020.

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.