NBER WORKING PAPER SERIES

THE SIZE AND 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, Revised February 2021

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 David Atkin, David Baqaee, Irene Brambilla, Ariel Burstein, Steven Davis, Manuel García-Santana, Peter Klenow, Virgiliu Midrigan, Stephen Redding, Diego Restuccia, Robert Shimer, Chad Syverson, Gabriel Ulyssea, Venky Venkateswaran, Daniel Xu and those from participants at the 2017 Society for Economic Dynamics Conference; 2017 meeting of the European chapter of the Econometric Society; 2017 NBER Productivity, Development, and Entrepreneurship workshop; the 2016 Trade and Integration Network of LACEA; the 2020 Ridge Forum in Growth and Development in Macroeconomics; 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 Size and 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, Revised February 2021
JEL No. E24,L24,O47

ABSTRACT

We develop a framework that uses price and quantity information on both establishments' outputs and inputs to assess the roles, on establishment dynamics and welfare, of technical 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 and 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 and sales growth but are important in explaining welfare losses from revenue productivity dispersion.

Marcela Eslava
Department of Economics
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 heterogeneity of firm and establishment size, growth, and a host of attributes correlated with size (e.g., productivity, exports, survival). What are the sources of firm/establishment size and 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 a specific set of attributes of productive units: 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 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]

We bring these approaches together to jointly identify the contribution of different dimensions of productivity and different sources of size wedges to firm size and growth, and consequently to welfare. To do so, we develop a unified conceptual, measurement and estimation structure that integrates these different dimensions of data and establishment attributes. Our framework nests the Hsieh and Klenow (2009) model on the production side and that proposed by Hottman, Redding and Weinstein (2016) on the demand side. Our framework takes advantage of data on output and input prices and quantities to measure establishment-level 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 to establishment size and growth, and ultimately to welfare, while also allowing for wedges between the data and the behavior predicted by the model.

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 establishments with data on quantities and prices, at the detailed product class, for outputs and inputs. We follow individual plants for up to thirty years (1982-2012). The long time coverage allows us to investigate the role of

[1]The misallocation literature is extensive. Prominent examples are Restuccia and Rogerson (2008,2017); 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 using an indirect approach with only revenue and expenditure data.

different attributes over 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 Weinstein (2020), our deflator allows for product turnover and changing appeal across products within the establishment. 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 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, and doing so disaggregating sectors. We introduce an estimation technique that jointly estimates the two functions for each sector, bringing together insights from recent literature on estimating production functions based on output and input use data and proxy methods, 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 technical efficiency and demand to be correlated, even within establishments 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 and life cycle sales growth across plants, and to welfare. We allow for residual wedges, which 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 not only the main driver of heterogeneity in sales and sales growth in our data, but also the single measured dimension with greatest

[2]For production function estimation using proxy methods, 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.

impact on welfare. A dominant role of demand shocks in accounting for sales and sales 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 as stemming from the dynamic appeal in superstar plants, while rapidly contracting technical efficiency is the outstanding characteristic of the worst performers. Love of variety implies welfare gains from heterogeneity in demand shifters and technical efficiency, both of which exhibit considerable dispersion. However, shutting down variability in efficiency has a negligible effect on welfare, while shutting down variability in demand shifters contracts welfare to 1/5 of its actual value.

Heterogeneous input prices and markups play modest roles in explaining cross sectional sales heterogeneity but are important in explaining welfare. Heterogeneous input prices dampen sales growth variability by 5.5% while markups reduce it by 1%. For a subperiod where we can quality-adjust wages, we find that the contribution of the wage component of input prices is further reduced by almost two thirds after this adjustment. Despite their modest contribution to cross sectional variability of sales and sales growth, welfare increases by almost 30% when dispersion in either markups or input prices is shut down. The greater importance of markups in welfare relative to cross sectional variability arises 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 frictions or distortions, it is the departure from frictionless and distortionless labor markets that matters for welfare.

We also find that negatively correlated residual wedges reduce plant revenue variance by 11 percentage points over the first twenty years of life. Residual wedges are positive for plants with lowest productivity and negative for those in the top productivity quartile. Consistent with findings in the misallocation literature, and despite the fact that we separately account for dispersion in markups and input prices, these correlated wedges have important welfare implications. Welfare increases by almost 20% when the dispersion in these residual wedges is eliminated.

The relative importance of different sources of heterogeneity varies considerably over the life cycle. Technical efficiency and residual wedges (negatively correlated with fundamentals) play a relatively more important role for young plants relative to older ones in the decomposition of sales and sales growth variance. In particular, negative wedges that dampen actual sales are particularly marked for young plants in the top quartile of predicted sales (predicted on the basis of productivity, input prices and markups). For entering establishments in this group, residual wedges dampen sales by 0.7 log points, while the corresponding figure is 0.4 at age 10 and 0.3 at age 20.

Our contribution to the literature is multi-fold. First, we bridge the gap between distinct approaches to the study of drivers of establishment 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, but puts them together in a unifying framework and unpacks subcomponents of each. The HK framework focuses on size-to-productivity composite wedges that can be decomposed into separate components of our framework: idiosyncratic markups, quality-adjusted input prices and residual wedges. Their productivity measure is in itself a composite of demand factors and technical efficiency. In turn, the cost component in the HRW framework is a residual bundle of technical efficiency, input costs and residual wedges, which we identify separately.

Unbundling these different attributes sheds light on crucial features of business performance and its contribution to welfare. While composite HK wedges are a significant drag on both sales variability and welfare, a large fraction of these effects is explained by markup and input price dispersion, rather than other sources of wedges. Recent contributions in the misallocation literature have focused on some such other sources of HK wedges, 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). We find that the welfare effect of shutting down dispersion in input prices and markups is twice as large as that of shutting down residual wedges, while input prices plus markups explain -7% of the -17.7% contribution of HK wedges to sales growth. 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.

In addition, the composite HK productivity measure is dominated by demand shifters relative to technical efficiency: demand dispersion contributes over seven times more than technical efficiency to the dispersion of sales growth. Moreover, dispersion in demand is crucial for welfare via love of variety, but technical efficiency dispersion plays a negligible welfare role. In turn, cost factors play a more important role for sales growth variability and welfare than would be attributed by the HRW approach alone, because the negative correlation between technical efficiency and wedges mutes the composite contribution. The -2.7% contribution of the composite HRW "cost" residual to the variance of sales growth in our data reflects a positive contribution of 8.3% of cost factors (13.77% of technical efficiency and -5.44% from input prices), and an additional drag of -11.1% from residual wedges, which are not inherently a cost/supply side factor.

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

[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).

information at the micro level 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 establishment heterogeneity relative to that in De Loecker et al (2020), who recover a measure of market power without imposing structure on the demand side. The need to take a stance on demand in our context also reflects the more general fact that the interpretation of markups depends on the market structure. As we show, a residual approach using cost shares of revenue to identify market power leads to measures of market power that vary with structural markups, structural wedges and specification error, including unmeasured variability in factor elasticities. Compared with our structural markups, such measures of market power are much more weakly correlated with demand shifters and revenue, and more strongly correlated with technical efficiency. Our approach to markups is closer to that in HRW or Edmond, Midrigan and Xu (2019), which tie 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 important and especially prevalent for young businesses, but also that dimensions internal to businesses are even more important than wedges to explain differential firm growth. 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 custumers 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. Moreover, the results suggest that growth based on reducing barriers to quality differentiation is more conducive to welfare gains than that based on reducing dispersion in technical efficiency across businesses.

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 outputs in 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.

Clearly, our application is to an economy where distortions presumably play a much larger role in the US. Indeed, our results on the welfare effects of composite wedges for the Colombian manufacturing sector are in the broad range found by the literature that applies the HK method to developing countries, including those in Latin America. We thus see as likely that the relative role we find for Colombia for the different components of composite wedges (input price variability, markups and residual wedges) applies more widely to similar countries. At the same time, we also find quantitatively similar results for the relative role of cost vs. demand and markup components to those found with data for the US by HRW, which is an indication that our results on the decomposition of the relative role of demand vs. efficiency and cost factors shed light on that role for a variety of environments.

Like much of the literature on productivity dispersion and misallocation, our analysis focuses on establishment-level data and variation. Given that we integrate and explore demand side variation including the quantification of markups, consideration of the relationship between establishments and their parent firms is potentially of interest. Unlike for the U.S., Colombia is dominated by single establishment firms.[6] Only about 7 percent of establishments and 16 percent of employment in the Colombian data are accounted for by multi-plant firms. This implies that the distinction between establishments and firms is much less critical in our context.

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 results on the drivers of size and growth dispersion are presented in section 5. Section 6 examines the value added of our approach relative to prior approaches nested in our framework, in terms of characterizing the sources of size and growth dispersion. The welfare analysis is presented in section 7. Section 8 concludes by providing a more comprehensive view on the implications of our analysis,

[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 104% 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 -2.7% (with a remaining -1.2% and 104% contribution markups and demand shifters).

[6]In the U.S., 80 percent of manufacturing employment is accounted for by establishments owned by multi-unit etablishment firms.

and on open questions for future research.

2 Decomposing establishment growth into fundamentals vs wedges

We start with a simple model of plant optimal behavior given plant attributes or "fundamentals", to derive the relationship that should be observed between size growth and growth in those attributes as a plant ages. We use the words "plant" and "establishment" interchangeably. We also permit establishment size to be impacted by wedges. The attributes we measure are: 1) the efficiency of the establishment's productive process (which we term $TFPQ$ as in Foster, Haltiwanger and Syverson, 2008, though we generalize the concept to producers of heterogeneous goods); 2) a demand shock;[7] 3) unit prices for inputs, in particular material inputs and labor; 4) markups. The conceptual framework below defines each of these components.

In the model, the establishment chooses its size optimally given $TFPQ$, demand shocks, input prices, and wedges. In the spirit of an accounting exercise the framework remains silent about the sources of these attributes, and rather asks how the establishment adjusts its size given those fundamentals at time t, and contingent on survival to that time.[8] 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 size and growth of surviving establishments up to any given age, but include robustness analysis of the determinants of survival in appendix H, which shows that our main results are robust to consideration of selection issues. We 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

[7] 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.

[8] 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. 2018; 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.

distortions arising from the business climate.[9] In the HK framework wedges can also arise from idiosyncratic variability in input prices and markups, but we explicitly account for these sources of heterogeneity. This shortcut enables us to use a simple static model of optimal input determination to frame our analysis of size and growth between birth and any given age. We permit the wedges or distortions to vary by establishment 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 discuss below the potential sources of the variation in input prices even after adjusting for quality.

2.1 Plant Optimization

Consider an establishment 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 establishment's technical efficiency, which we term $TFPQ$ following Foster, Haltiwanger and Syverson (2008). A_{ft} has an aggregate and an idiosyncratic component (A_t and a_{ft}). 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 establishment obtains from a unit of a basket of inputs. Establishment f may be uni- or multi-product. Section 2.2 below discusses the definition of output Q for multi-product establishments.

We use a CES preference structure (specified in more detail below) that yields demand at the establishment 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 establishments . D_{ft} has aggregate and idiosyncratic components $D_t = P_t$ and d_{ft}, respectively.[10]

[9]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. (2020), 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.

[10]$D_t = P_t \left(\frac{E_t}{P_t}\right)$ where E_t is aggregate (sectoral) expenditure, and the aggregate (sectoral) price index is given by $P_t = \left(\sum_{f=1}^{N_F} d_{ft}^{\sigma} P_{ft}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$. N_F is the number of establishments in the sector.

Establishment appeal d_{ft} is measured from equation (2) as the variation in establishment price holding quantities constant, beyond aggregate effects. We refer to d_{ft} generically as the establishment'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 establishment sales are closely linked to consumer welfare. Consequently, the distribution of establishment sales growth is the central focus of our analysis of the establishment growth distribution.

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

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

taking as given A_{ft}, D_{ft}, and unit costs of the composite input, denoted C_{ft}. There may be idiosyncratic wedges τ_{ft}, that lead to a gap between an establishment's actual scale and that which would be implied by the static model given its fundamental attributes.[11] 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".[12] Financing constraints may similarly limit the ability of the establishment 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.[13] The resulting τ_{ft} may be correlated with plant fundamentals themselves. 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.[14]

[11] 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.

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

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

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

We allow establishments to hold market power, so that an establishment's market share may be non-negligible. This also implies that, in choosing its optimal scale, an establishment 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 establishments stems from market power (i.e., establishments 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 establishment'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. In our application, the demand and production parameters are constant across establishments within sectors (at the three digit level of the ISIC revision 3 classification for Colombia, of which there are 22 manufacturing sectors). An establishment's relevant market is defined as the group of producers of the plant's most important CPC 3-digit product, of which there are 112 such groups, so that s_{ft} is f's revenue share in its CPC 3-digit group.

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 sales and life-cycle growth of sales as functions of fundamentals (D_{ft}, A_{ft}, and C_{ft}), wedges τ, and parameters:

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

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 χ_t and $\widehat{\chi}_t$. Crucially, $\chi_{ft} = (1-\tau_{ft})^{\gamma\kappa_1}$ captures idiosyncratic wedges, so we refer to $\chi_{ft}^{1-\frac{1}{\sigma}} = (1-\tau_{ft})^{\gamma\kappa_1\left(1-\frac{1}{\sigma}\right)}$ as a "sales wedge". The second line is obtained by dividing each optimal outcome in period t by its optimal level at birth ($t=0$)(see Appendix B).[15]

[15]There is some slight abuse of notation here as t is used for calendar time and then for every firm

Equation system (5) is the focus of our analysis of the distribution of establishment revenue and establishment revenue growth. We start with the level and growth of (idiosyncratic) attributes that we can measure. Among these,d_{ft}, a_{ft},μ_{ft}, w_{ft}, pm_{ft} are, respectively, idiosyncratic demand shocks, $TFPQ$, markups, and shocks to wages and material input prices. The wedges that an establishment faces may be age-specific.

2.2 CES Demand Structure

In this subsection, we show that the establishment-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 establishment-level output in a manner that allows for within establishment 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 establishment-level price index. As long as different products within an establishment 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 establishments we allow establishment output Q_{ft} to be a CES composite of individual products $Q_{ft} = \left(\sum_{\Omega_t^f} d_{fjt} q_{fjt}^{\frac{\sigma_w - 1}{\sigma_w}} \right)^{\frac{\sigma_w}{\sigma_w - 1}}$, where q_{fjt} is period t sales of good j produced by establishment f, the weights d_{fjt} reflect consumers' relative preference for different goods within the basket offered by establishment f, σ_w is the elasticity of substitution between products within 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 establishments and another for products within establishments. Consumer's utility in this general CES structure in period t is given by:

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* ($t = 0$).

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

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

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

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

where p_{fjt} is the price of q_{fjt}, and I_t is the set of establishments in period t. We refer to d_{fjt} and d_{ft} as, respectively product (within establishment) and establishment appeal or demand shocks, defined as in equations (6) and (7). They correspond to the weight, in consumer preferences, of product fj in establishment f's basket of products, and of establishment f in the set of establishments. Given normalizations in equation (9), product appeal d_{fjt} captures the valuation of attributes specific to good fj relative to other goods produced by establishment f, while establishment appeal d_{ft} captures attributes that are common to all goods provided by establishment f, such as the establishment's customer service and average quality of establishment f's products, in a constant utility framework. Both establishment and product appeal may vary over time besides varying across establishments.

Equation (7) defines real output for an establishment in this multiproduct framework. In a multiproduct-establishment 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 establishment as an aggregate of single-product outputs, in which each product receives a weight equal to its appeal to custumers, relative to that of other products within the establishment. Given (9), this real output measure is normalized by the average appeal of products within the establishment. The crucial relevant assumption here is that products within establishments are not perfect substitutes so that tracking product turnover and changing product appeal within establishments is critical for measuring establishment-level output.

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

$$q_{fjt} = d_{ft}^{\sigma} d_{fjt}^{\sigma_w} \left(\frac{P_{ft}}{P_t}\right)^{-\sigma} \left(\frac{p_{fjt}}{P_{ft}}\right)^{-\sigma_w} \frac{E_t}{P_t} \quad (10)$$

$$R_{ft} = Q_{ft} P_{ft} = d_{ft}^{\sigma} P_{ft}^{1-\sigma} \frac{E_t}{P_t^{1-\sigma}} \quad (11)$$

where

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

Dividing (11) by P_{ft} and solving for P_{ft},[16] we obtain

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

where the establishment-level price index is given by:

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

Given the nested CES demand, the establishment will charge the same markup on all products.[17]

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

[16] We follow Redding and Weinstein (2020) in our treatment of product entry and exit. They don't formally model the decisions to add and subtract 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.

[17] See Appendix S2 of Hottman et. al. (2016). In this nested environment the producer's optimization problem can be decomposed into two steps. The producer first chooses the composite index of products. It then chooses individual products to minimize the composite total cost subject to the optimal level of producer-level output. It is optimal for the producer 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.

(2011), Fieler, Eslava and Xu (2016), and others. Foster et al (2016), in turn, interpret establishment appeal as capturing the strength of the business' client base.

Establishments that produce multiple products matter in our framework for three reasons. First, our cost/production structure is at the establishment-level. That is, we specify the cost/production function as being based on total output of the establishment 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 use data are at the establishment level. Our view is that if one queried most establishments (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 multiple costs are shared across products (i.e., there is joint production). That is, an establishment is not simply a collection of separable lines of production. A second reason that establishments 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, establishments are our object of interest because they are clearly relevant empirical objects. Third, there may be cannibalization between products of the same establishment. For these reasons, we specify an establishment-level profit maximization problem but one that recognizes multi-product producers for purposes of measuring establishment-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.[18] Over 90% of plants in the AMS (i.e. over 90% manufacturing plants in Colombia with size over the inclusion threshold) belong to single-plant firms, so that the distinction between plants and firms is not as crucial in our context as it is in others.

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

[18]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.

[19]Plant IDs in the survey were modified in 1992 and 1993. To follow establishments over that period, we use the official correspondence that maps one into the other.The correspondence seems to

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

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).

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 we do not observe from birth.[21] We restrict all of our analyses to plants

be imperfect (as suggested by apparent high exit in 92 and high entry in 93), but even for actual continuers that are incorrectly classified as entries or exits, our age variable is correct (see further below).

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

[21] See Appendix 1.2 for details.

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. There is also exit in our sample, at a rate of approximately 7% per year. Our analysis includes both stayers and exiters, which we examine separately in Appendix H.

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 establishment by properly deflating revenue with the exact plant level price index, $P_{ft} = \left(\sum_{\Omega_t^f} d_{fjt}^{\sigma_w} p_{fjt}^{1-\sigma_w}\right)^{\frac{1}{(1-\sigma_w)}}$. Since the index depends on unobservable σ_w and $\{d_{fjt}\}$ and thus cannot be constructed readily from observables, we use Redding and Weinstein's (2020) CES Unified Price Index (CUPI) approach. Redding and Weinstein (2020) and Appendix A show that the CUPI is the appropriate empirical analogue of 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:

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

$\Omega_{t,t-1}^f$ 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. That is, we take into account product entry and exit. $\lambda_{ft}^{QRW} = \prod_{\Omega_{t,t-1}^f} \left(\frac{s_{fjt}^*}{s_{fjt-1,\Omega_{t,t-1}^f}^*}\right)^{\frac{1}{\left\|\Omega_{t,t-1}\right\|}}$ 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.[22] The derivation of the CUPI price index from our theoretical price index

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

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

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

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

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. P_{fB}, $\overline{P^*_{ft}}$, $\left(\Lambda^{Q}_{ft} \right)$ can be directly calculated in the data. σ_w is estimated as explained in the following section. Details are provided in Appendix A.[23]

We also 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.

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 unbiased estimates of σ and σ_w. We impose a set of moment conditions that requires

[23] 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.

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). 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_w, we are unable to properly construct P_{ft}, and thus $Q_{ft} = \frac{R_{ft}}{P_{ft}}$ (see section 3.2). We therefore rely on P_{ft}'s two separate components from equation 16: $\overline{P^*_{ft}}$ and Λ^Q_{ft}. We define

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

and proceed in three steps to address this limitation:

1. (This step only sketched here, details provided in the following subsection) Jointly estimate the coefficients of the production function (1), the demand function (13), and σ_w using $Q^*_{ft} = \frac{R_{ft}}{P_{fB}\overline{P^*_{ft}}} = Q_{ft} * \left(\Lambda^Q_{ft}\right)^{\frac{1}{\sigma_w-1}}$ and $\overline{P^*_{ft}} = \frac{P_{ft}\left(\Lambda^Q_{ft}\right)^{\frac{-1}{\sigma_w-1}}}{P_{fB}}$ as the respective dependent variables / regressors of these two functions . We carry Λ^Q_{ft} as a separate regressor in each equation to deal with potential biases induced by the–at this point–still partial estimation of revenue deflators. In particular, not explicitly accounting for changes in product quality and variety within the plant leads to "quality " and "variety" biases in the estimation of production function coefficients, as described by De Roux et al (2020). Λ^Q_{ft} explicitly accounts for those changes, freeing our estimates from those biases. We similarly introduce separately M^*_{ft} and Λ^M_{ft} in the production function (where $M^*_{ft} = \frac{\text{materials expenditure}}{PM_{fB}\overline{PM^*_{ft}}}$, and Λ^M_{ft} is the adjustment factor for the prices of materials analogous to Λ^Q_{ft} see Appendix A). The joint estimation is conducted separately for each three-digit sector. The parameters $\{\alpha,\ \beta,\ \phi,\ \sigma,$ and $\sigma_w\}$ used in the analysis are those estimated in this step.

2. Use the estimated elasticity $\widehat{\sigma_w}$ for the respective three-digit sector to obtain $P_{ft} = P_{fB} * \overline{P^*_{ft}} * \left(\Lambda^Q_{ft}\right)^{\frac{1}{\widehat{\sigma}_w-1}}$ 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 estimat-

ing $TFPQ_{ft}$ 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. More generally, our application only considers idiosyncratic (within sector-year) variation.

We now explain step 1 in detail.

4.1 Joint production-demand function estimation

We want to 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

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

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

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

and

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

In practice, we estimate the parameters of (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 and 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, 2016) and Eslava et al (2004, 2013) 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 elasticities of substitution across and within establishments. We rely on flexible laws of motion for $TFPQ$ and D_{ft}:

$$\begin{aligned} lnA_{ft} &= \pi_0^A + \pi_1^A lnA_{ft-1} + \pi_2^A lnA_{ft-1}^2 + \pi_3^A lnA_{ft-1}^3 + \xi_{ft}^A \\ \ln D_{ft} &= \pi_0^D + \pi_1^D \ln D_{ft-1} + \pi_2^A \ln D_{ft-1}^2 + \pi_3^A \ln D_{ft-1}^3 + \xi_{ft}^D \end{aligned}$$

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

$$E \begin{bmatrix} \ln M_{ft-1}^* \times \xi_{ft}^A \\ \ln L_{ft} \times \xi_{ft}^A \\ \ln K_{ft} \times \xi_{ft}^A \\ \ln D_{ft-1} \times \xi_{ft}^A \\ \ln A_{ft-1} \times \xi_{ft}^D \\ \ln A_{ft} \\ \ln D_{ft} \end{bmatrix} = 0 \tag{22}$$

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. Meanwhile, the conditions that D_{ft-1} must be orthogonal to ξ_{ft}^A while A_{ft-1} must be orthogonal to ξ_{ft}^D identify σ and σ_w, following the logic that the slope of the demand function can be inferred taking advantage of shocks to supply.[24]

More precisely, in the production side we assume that materials are freely adjusted while the demand for capital and labor is assumed quasi-fixed. Thus, in (22) we require 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.[25] 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 De Loecker et. al. (2016) in treating L as quasi-fixed for purposes of estimation.

As for the assumptions that identify demand elasticities, Foster et al (2008, 2016) and Eslava et al (2013) relied on the logic that shocks to production identify the demand curvature, but imposed orthogonality between demand and technology shocks

[24]Production elasticities are initialized at MCO estimates, while σ is initialized at the estimate from an IV regression where $TFPQ_{ft}$ is used as an instrument in the demand equation. Using this initial estimate for σ for each sector, σ_w is initialized at a level such that $\frac{\sigma_w}{\sigma}$ equals the $\frac{\sigma_w}{\sigma}$ ratio for the median sector in Hottman, Redding and Weinstein (2016).

[25]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.

in levels (A_{ft} and D_{ft}). This effectively precludes the possibility that establishments endogenously invest in quality when they perceive better returns (as would be the case with higher $TFPQ$), or that they acquire technologies that increase production costs to produce better quality.[26] Hottman, Redding and Weinstein (2016) and Broda and Weinstein (2006, 2010) address these concerns by imposing orthogonality between double-differenced demand and supply shocks (double differencing over time and varieties). Orthogonality between the double-differenced shocks may still be a strong assumption if, even within product groups, changes in quality require changes in production technologies. Given our ability to specify demand and production separately using the price and quantity data of both output and inputs, we impose $E(\ln D_{ft-1} \times \xi_{ft}^{A})$ and $E(\ln A_{ft-1} \times \xi_{ft}^{D})$ which permit a correlation between $TFPQ$ and demand even over time within the plant. While we are still taking advantage of shocks to the supply curve to identify elasticities on the demand side, we only require that *innovations* in technical efficiency in period t be orthogonal to demand in levels in $t-1$, and that *innovations* in demand in period t be orthogonal to $TFPQ$ in levels in $t-1$, where these innovations come from a very flexible law of motion for TFP and D_t.

Notice also that $TFPQ$ obtained as a residual from quality-adjusted Q is stripped of apparent changes in productivity related to within-establishment appeal changes, eliminating a source of correlation between appeal and technical 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).[27]

We implement this estimation separately for each three digit sector of ISIC revision 3, adapted for Colombia (CIIU-AC by its acronym in Spanish). There are 22 manufacturing sectors at this level. The estimated factor and demand elasticities are summarized in Table 1 and listed in Appendix L. Our results reveal close to constant scale in production at the three-digits sector level for most sectors. The estimated elasticities of substitution across products within the establishment and across establishments stand at averages (over sectors) of 3.5 and 1.92, respectively, with substantial cross-sector variation (see Appendix I). The revenue function curvature parameter stands at an average 0.47 with a maximum of 0.68, by contrast to the 0.67 curvature parameter implied by usual assumptions of CRS in production, CES demand and an elasticity of substitution of 3. The level of σ is crucial in determining the role played by wedges (see Appendix C).

It is encouraging that we obtain plausible factor elasticities for most sectors at the

[26] 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.

[27] 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.

three digits sector level. Proxy methods for the estimation of production functions 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.[28]

Table 1: Factor and demand elasticities

	β	α	ϕ	σ_w	σ	σ_w/σ	γ	$\gamma(1-1/\sigma)$
Average	0.36	0.15	0.51	3.50	1.92	1.82	1.02	0.46
Min	0.16	0.04	0.10	2.15	1.20	1.76	0.92	0.17
Max	0.60	0.40	0.70	4.73	2.66	1.99	1.22	0.68

5 Results

We use the within-plant estimated demand elasticity $\widehat{\sigma_w}$ to construct $\ln P_{ft} = \ln\left(P_{fB}\overline{P^*_{ft}}\right) + \frac{1}{\widehat{\sigma_w}-1}\ln\Lambda^Q_{ft}$ and subsequently recover $Q_{ft} = \frac{R_{ft}}{P_{ft}}$. We proceed in an analogous way to construct pm_{ft} and M_{ft}.[29] We then use Q_{ft}, M_{ft} and P_{ft} to obtain the residuals A_{ft} and D_{ft}. We use the estimated σ (at the three digit level of ISIC revision 3) to obtain the markup $\mu_{ft} = \frac{\sigma}{(\sigma-1)}\frac{1}{\left(1-s_{ft}\right)}$. For markup estimation, we use plant f's market share s_{ft} as its revenue share in its relevant product group, defined at the three digit group of the product classification. Products are classified according to the international CPC classification. There are 112 product groups at the CPC three digit level (while our "sectors" classification, defined using Colombian ISIC three digit level, has 22 classes), with an average number of plants close to 50, and a median of 17.

From this point, we work only with the within-sector variability of all variables of interest. In particular, we deviate all outcome variables (revenue, employment, capital, materials, output prices and input prices) from sector*year effects. 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. It is this variables deviated from sector*year effects that we use when building life cycle growth for any variable ($\frac{Z_{ft}}{Z_{0t}}$ for each variable Z for each variable Z).[30]

[28] If fully unconstrained, our joint estimation does deliver a negative factor labor elasticity for textiles (sector 321). We assign to that sector the average elasticities across 3 digit sectors of 2 digit sector 32, corresponding to textiles, apparel and leather industries.

[29] 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. For each plant, we use for materials the same elasticity of substitution used for outputs.

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

5.1 Plant attributes

Table 2 presents basic summary statistics for (the idiosyncratic component of) sales and our estimates of output, output prices, $\ln A_{ft}$, $\ln D_{ft}$, the sales wedge (in logs), markups and input prices. The sales wedge, $\ln \chi_{ft}^{1-\frac{1}{\sigma}} = \ln (1-\tau_{ft})^{\gamma\kappa_1\left(1-\frac{1}{\sigma}\right)}$, is obtained as a residual from equation 5, since we have measures of all other components.[31]

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 is more relevant in our context). To the extent that quality is more difficult to produce, demand shocks and technical efficiency may be negatively correlated. This is indeed, though weakly, the case in our estimates, also consistent with Forlani et al. (2018). Also especially interesting is the negative and strong correlation of wedges with $TFPQ$ and demand shocks, suggesting that the plants with the best fundamentals face greater barriers, i.e. correlated wedges. (It is worth noting that sales, demand shocks, $TFPQ$ and sales wedges exhibit an important degree of persistence, but that this is much lower for wedges than for the rest: 0.76 for wedges compared to over 0.93 for the rest, see Appendix E.) These basic correlation patterns are echoed in our decompositions below.

Though markups display relatively modest variation across plants, they are positively correlated with $TFPQ$, D and wages, besides plant size. They are also positively correlated with our sales wedge $(1-\tau_{ft})^{\gamma\kappa_1\left(1-\frac{1}{\sigma}\right)}$. It is natural to ask how our approach to measuring the markup compares to the approach to markups in De Loecker and coauthor's work (2012, 2016, 2020). Denoting the markup calculated with such approach as μ^{DL}, Appendix B shows that under the demand and production structure in this paper the following relationship holds:

$$\mu_{ft}^{DL} \equiv \frac{\theta_{ft}^{v}}{\frac{C_{ft}^{v} V_{ft}}{P_{ft} Q_{ft}}} = \frac{\mu_{ft}}{(1-\tau_{ft})}$$

where V_{ft} is some variable input, C_{ft}^{v} its unit cost and θ_{ft}^{v} is the output elasticity for the variable input. We calculate $\widehat{\mu}_{ft}^{DL} = \frac{\phi}{\frac{pm_{ft} M_{ft}}{R_{ft}}}$, referred to as "Markup DL" in Table 2. $\widehat{\mu}_{ft}^{DL}$ displays a very modest (positive) correlation with μ_{ft}, and is strongly negatively correlated with the sales wedge. Its correlation with sales is negative, and

[31] χ_t is no longer relevant once we focus solely on within sector*year variation.

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	Markup DL	Sales Wedge
Sales	1.426	1.000									
Output	1.568	0.902	1.000								
Output prices	0.647	0.011	-0.409	1.000							
TFPQ	0.775	0.217	0.506	-0.720	1.000						
Demand Shock	0.840	0.921	0.675	0.375	-0.064	1.000					
Input prices	0.609	-0.041	-0.103	0.157	0.260	0.022	1.000				
Average wage	0.422	0.609	0.528	0.062	0.144	0.585	-0.001	1.000			
Markup	0.116	0.505	0.455	0.002	0.112	0.465	-0.013	0.343	1.000		
Markup DL	0.633	-0.083	-0.045	0.091	0.235	0.025	-0.086	0.114	0.041	1.000	
Sales Wedge	0.409	-0.147	-0.106	-0.068	-0.420	-0.160	0.002	0.046	0.144	-0.534	1.000
Lagged Demand Shock	0.832	0.883	0.654	0.350	-0.085	0.956	0.017	0.581	0.465	0.012	-0.089

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.

that with demand shocks very weak. Compared with μ_{ft}, $\widehat{\mu}_{ft}^{DL}$ displays much weaker correlations with sales and demand shifters, and a stronger correlation with $TFPQ$. The variance of $\widehat{\mu}_{ft}^{DL}$ (within sector*year) is more than twice the variance of $\frac{\mu_{ft}}{(1-\tau_{ft})}$. This discrepancy can be accounted for by specification or measurement error in θ_{ft}^{v}, or by a factor-specific wedge as discussed in appendix B. Such a factor specific wedge is already implicitly incorporated in the sales wedge which captures all factors that account for any discrepancy between actual and model consistent sales. However, factor-specific wedges enter into factor-specific first order conditions beyond the sales wedge since such wedges impact not only the overall scale of establishment activity but also the factor mix as well. We don't emphasize the role of such factor-specific wedges since our objective is to characterize the determinants of size and growth in terms of overall sales. [32]

[32] Hsieh and Klenow (2009) allow for such factor-specific wedges but like this paper focus much of their attention on the composite wedge impacting the scale of activity of businesses. Blackwood, Haltiwanger and Wolf (2020) explore these issues in more depth. In terms of the current setting, if we make explicit the factor-specific wedge then this yields the expression $\frac{\mu_{ft}(1-\tau_{ft}^{v})}{(1-\tau_{ft})}$ to the right of the equal sign in equation (5.1). Using a structural decomposition such as we use below (and described in detail in appendix G), we find that about 65% of the variance of markup DL is accounted for by this factor-specific distortion. More discussion of the possible sources of this factor-specific wedge is provided in appendix B.

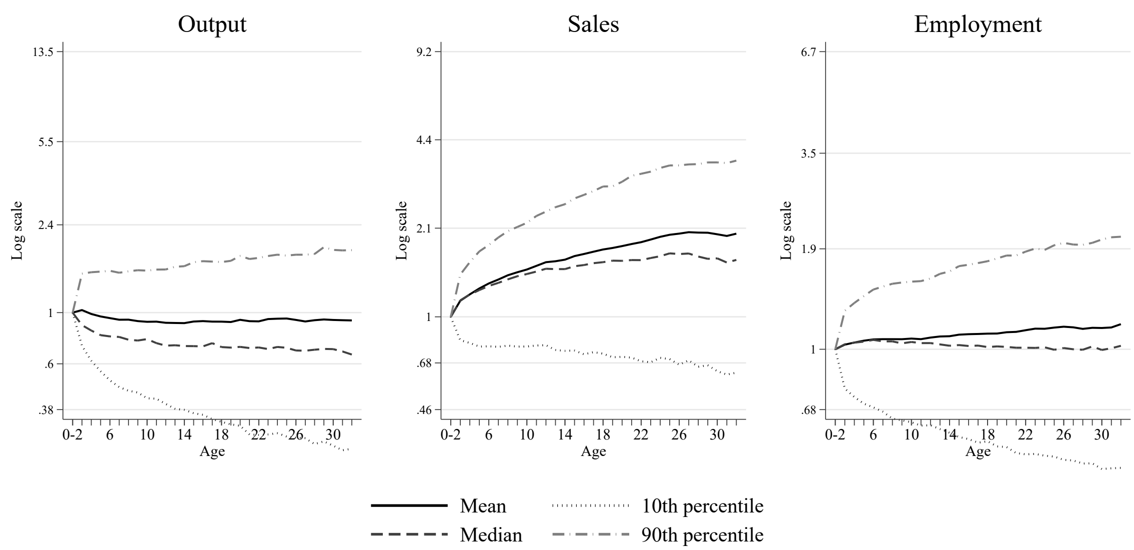
5.2 Growth over the life cycle

We build growth of variable Z over the life cycle of a plant at a given age as $\frac{Z_{f,age}}{Z_{f,0}}$ where Z_{f0} is the level of Z at f's birth, calculated 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). A plant's age in year t is the difference between the current year, t

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 right 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.[33]

Figure 1: Distribution of life cycle growth

Current to initial



Idiosyncratic components

These average growth dynamics, however, hide considerable heterogeneity. Median growth (dashed line) falls under mean growth for all panels, highlighting the fact that it is a minority of fast-growing plants that drive mean growth. Related, the distribution of plant growth is highly skewed. It is this heterogeneity and its welfare implications that we aim to explain in the analysis below.

[33] For revenue and employment, we have $\frac{R_{fa}}{R_{f0}} = 1.59$ 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 $a = 25$.

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 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).

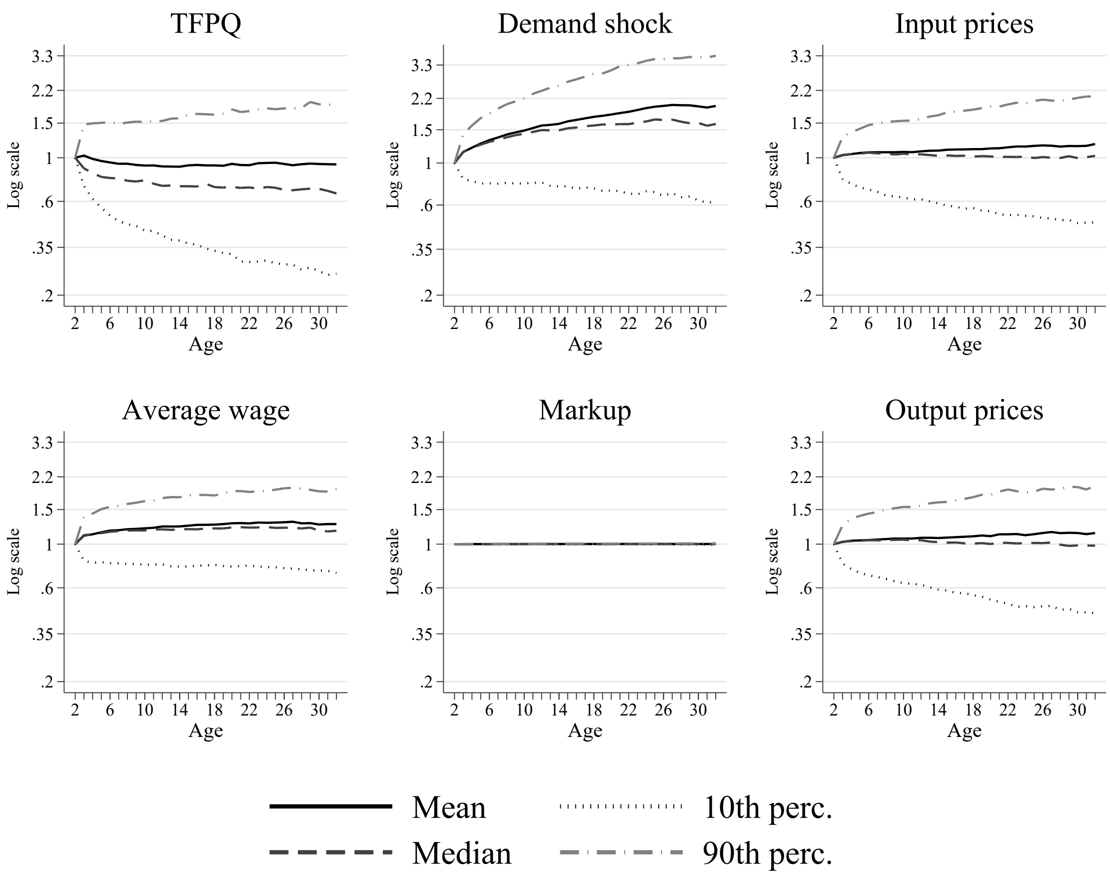
Figure 2 displays the life cycle growth of $TFPQ$ and demand shocks, markups, material input prices and wages. The average growth of demand shocks dominates that of input prices, and both dominate the average growth of $TFPQ$ and markups over the life cycle. By age 25, $TFPQ$ has not grown compared to 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 $TFPQ$ and that of demand shocks. At age 3, the correlation is 0.07, moving to -0.08 at age 10 and -0.12 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 establishment 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 revenue and revenue growth into its sources

We now decompose the variance of R_{ft} and $\frac{R_{ft}}{R_{f0}}$ 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:

Figure 2: Distribution of fundamentals

Current to initial



Idiosyncratic components

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

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 $(\chi_{ft})^{1-\frac{1}{\sigma}}$ in (23) is calculated as a residual, since all of the other components are either measured or estimated. $\widehat{\chi}_t = 1$ since we focus solely on within sector*year variation. $\frac{\chi_{ft}}{\chi_{f0}}$ captures life cycle growth in wedges, including distortions

Table 3: Decomposition of sales, unweighted and revenue weighted by ages.

	Panel A: Unweighted							
	Levels decomposition				Growth decomposition			
	Weighted avg. ages	Age 3	Age 10	Age 20	Weighted avg. ages	Age 3	Age 10	Age 20
TFPQ	0.098	0.117	0.094	0.088	0.138	0.219	0.142	0.109
Demand shock	1.025	1.031	1.030	1.023	1.039	1.025	1.043	1.041
Material prices	0.003	0.005	-0.001	0.006	-0.018	-0.022	-0.017	-0.019
Wages	-0.052	-0.053	-0.053	-0.054	-0.037	-0.046	-0.037	-0.036
Markup	-0.025	-0.024	-0.027	-0.024	-0.012	-0.010	-0.011	-0.013
Sales wedge	-0.049	-0.076	-0.043	-0.040	-0.111	-0.166	-0.120	-0.081
	Panel B: Revenue weighted							
	Levels decomposition				Growth decomposition			
	Weighted avg. ages	Age 3	Age 10	Age 20	Weighted avg. ages	Age 3	Age 10	Age 20
TFPQ	0.121	0.145	0.123	0.136	0.210	0.256	0.207	0.101
Demand shock	1.098	1.075	1.074	1.088	1.060	1.027	1.070	1.013
Material prices	0.003	0.021	-0.009	0.014	-0.009	0.009	-0.029	0.010
Wages	-0.086	-0.079	-0.070	-0.130	-0.041	-0.083	-0.039	-0.011
Markup	-0.178	-0.083	-0.149	-0.200	-0.057	-0.038	-0.042	-0.074
Sales wedge	0.043	-0.078	0.031	0.091	-0.164	-0.172	-0.167	-0.039

from regulations, adjustment costs, and other factors, and measurement error. Because these sales wedge simply reflects the gap between actual growth and that predicted by measured attributes through the lens of our model, it reflects 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.

Results are presented in Table 3.[34] We find that the structural contribution of fundamentals, rather than residual wedges, explains the bulk of cross sectional variation in sales and sales growth over the life cycle. Taken together, fundamentals in fact account for more than 100% of the variance of sales and sales growth across plants within a sector (a fact we turn to further below). Averaging over ages, we find contributions to the variance of revenue level of the demand shock, $TFPQ$, input prices, markups and structural wedges, respectively equal to 1.03, 0.1, -0.05, -0.03, -0.05. In terms of life cycle grow the wedges make a stronger (negative) contribution compensated by a larger contribution of $TFPQ$ and a weaker markup role. In particular, the contributions of demand shock, $TFPQ$, input prices, markups and structural wedges, are respectively

[34]We implement the variance decomposition by ages. See Appendix G for results.

equal to 1.04, 0.14, -0.06, -0.01 and -0.11. In the revenue weighted decomposition, the difference in levels vs growth in the contribution of markups, $TFPQ$ and wedges is exacerbated (Panel B of Table 3).

The most outstanding feature of these results is the overwhelming role of the demand shock: its weight in the variance decomposition is over seven times as important as $TFPQ$ to explain idiosyncratic sales in levels and growth, on both unweighted and a revenue-weighted bases. It is even larger compared to the contributions of input prices and markups. 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 $TFPQ$. The negative correlation between $TFPQ$ and demand shocks also plays a role in these patterns.

Input prices and markups make much smaller, but not negligible, contributions, with markups being particularly important in levels and with revenue weighting.The minor contribution of markups to the variance of sales growth, especially on an unweighted basis reflects 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, however, which explains the more significant role of markups for the variance of revenue levels, especially on a weighted basis. We also show in section 7 that the large markups of large plants have crucial welfare implications.

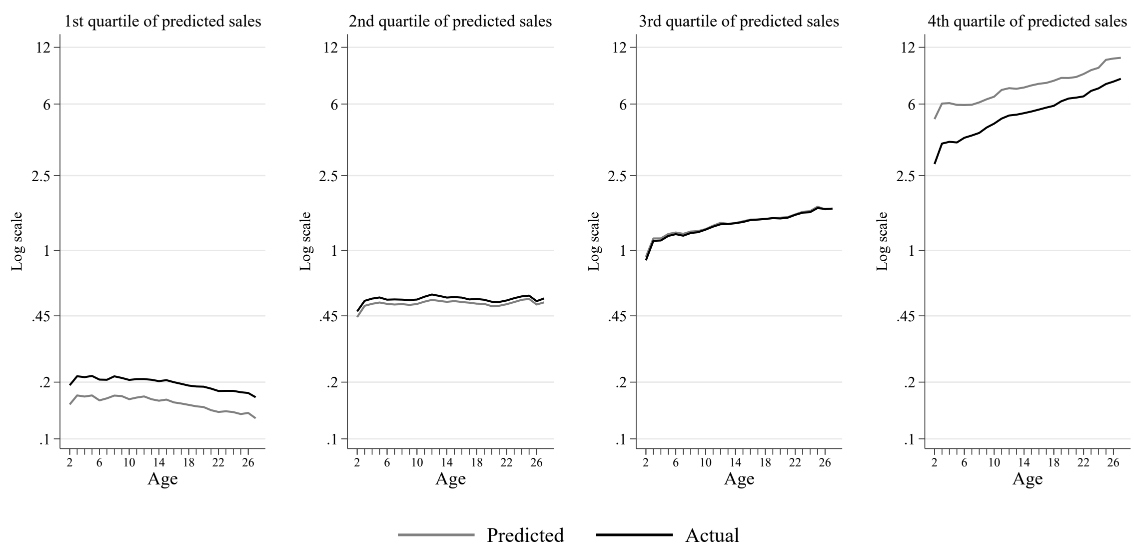
The dominance of demand-side fundamentals over supply side in explaining the variance in sales resonates with recent findings in the literature (Hottman et al. 2016, Foster et al. 2016). It is, however, noticeable that this finding survives the expansion of the measurement framework to explicitly account 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 the host of attributes that we measure (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 adjusting wages, 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.

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 measured attributes on sales, implying that high-productivity high-appeal plants sell and grow less relative to low-productivity and appeal plants than their respective productivity and appeal would imply. The effect is quantitatively large, especially for revenue growth: pooling ages, sales growth dispersion is dampened by about 11% 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.

Figure 3: Actual sales vs. sales predicted based on observed attributes, by age

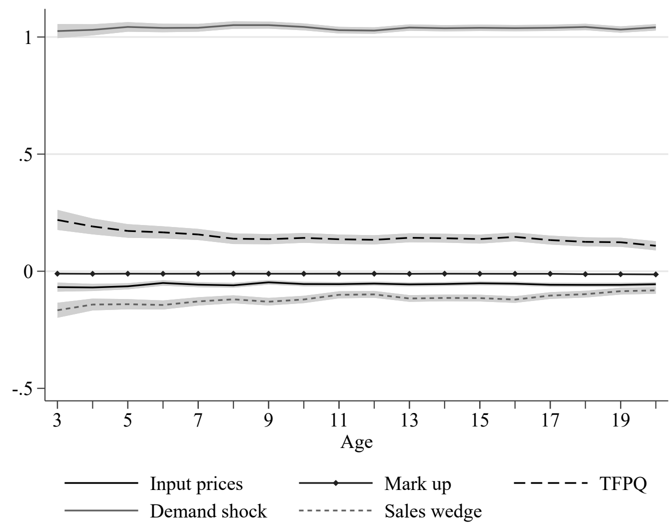
By quartiles of predicted sales



Idiosyncratic components

Figure 3 shows the mechanics behind the negative contribution of structural wedges: the average gap between actual revenue (black solid line) and revenue explained by measured attributes (grey solid line) is positive for plants with low predicted size and negative for those in the highest percentiles of predicted size. Predicted revenue corresponds to revenue in equation (23) setting $\chi_{ft} = 0$. Figure 3 implies that it is plants with weak fundamentals that face positive residual wedges while those with strongest fundamentals face the opposite situation, in particular at young ages. A similar pattern is observed if we predict growth based solely on technical efficiency and demand shifters, though in that case the gap between predicted and actual size is larger in absolute value, given the dampening effect of markup and input price variability jointly.

Figure 4: Variance decomposition of the life cycle growth of sales, by age



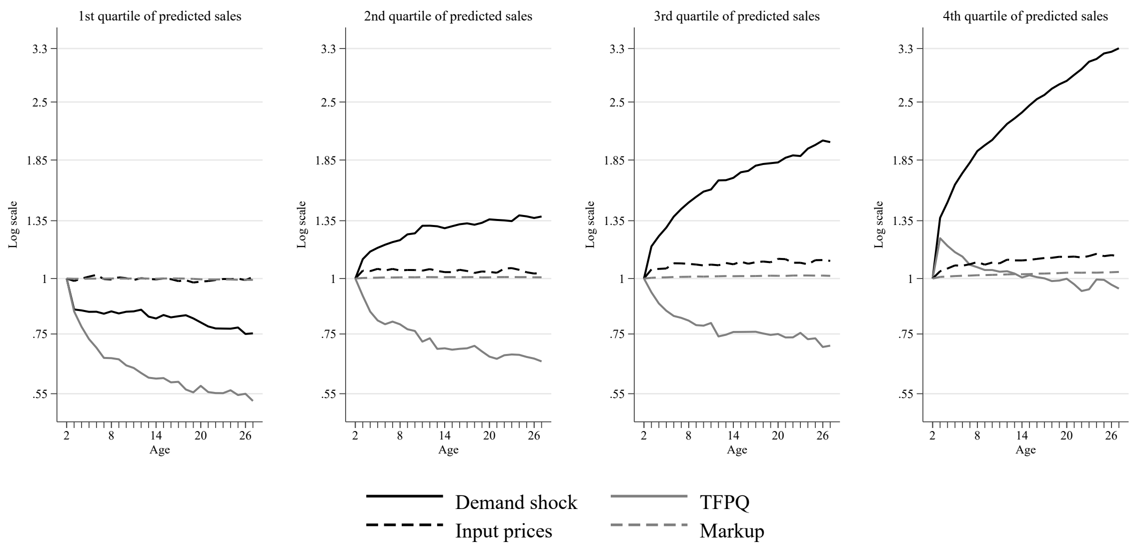
Idiosyncratic components

The contributions of the different attributes to sales and the life cycle growth of sales vary depending on the horizon of growth considered (Table 3 and Figure 4). Demand becomes increasingly important compared to $TFPQ$ over longer horizons. The ratio of contributions to the growth of revenue of demand relative to $TFPQ$ is close to 4.7 at age 3, but by age 20 it has more than doubled, to 9.6. This is because the correlation between sales growth and $TFPQ$ growth decreases for older plants, while that between sales and demand remains fairly stable . 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. Figure 3 shows that the decreasing importance of residual wedges over the life cycle is driven primarily by the top quartile of predicted sales. These top performers face a negative wedge that is much larger at younger ages, decreasing in absolute value from 0.7 log points at birth to 0.4 at age 10 and 0.27 at age 25.

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.

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, the most outstanding feature of bottom quartile plants is weak $TFPQ$ growth.[35]

Figure 5: Life cycle growth of measured attributes
By quartiles of predicted sales growth



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 into a measure of productivity 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.[36] They start from revenue, which in our notation is given by: $R_{ft} = D_{ft}Q_{ft}^{1-\frac{1}{\sigma}} =$

[35]We conduct a similar decomposition for the growth of output, rather that sales, finding similar results. See Appendix G.

[36]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

$D_{ft}\left(A_{ft}X_{ft}^{\gamma}\right)^{1-\frac{1}{\sigma}}$. With estimates of γ and σ one can obtain the composite shock $TFPQ\_HK$ solely from revenue and input data as:

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

Revenue can then be expressed as:

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

That is, the HK residual wedge is a *composite* measure of wedges, the "HK wedge"= $\left((1-\tau_{ft})\frac{(1-\tau_{ft})}{C_{ft}\mu_{ft}}\right)^{\gamma}$, 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 $X_{ft} = \left(\frac{D_{ft}A_{ft}^{1-\frac{1}{\sigma}}\gamma}{C_{ft}\mu_{ft}\left(1-\tau_{ft}\right)^{-1}}\right)^{\frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}}$ we obtain $TFPR\_HK_{ft} = \frac{C_{ft}\mu_{ft}}{\gamma\left(1-\tau_{ft}\right)}$, so $TFPR\_HK$ variability reflects variation not only τ, but also in markups and input prices.[37] 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 6, left panel, contrasts the by-age decomposition of life-cycle growth 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 HK *composite* wedges in the two-way (HK) decomposition is due to the contribution of variable input prices and markups (6.6% out of the 18% assigned to wedges in 6a). 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 6, 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 6 is that using $TFPQ\_HK$

---

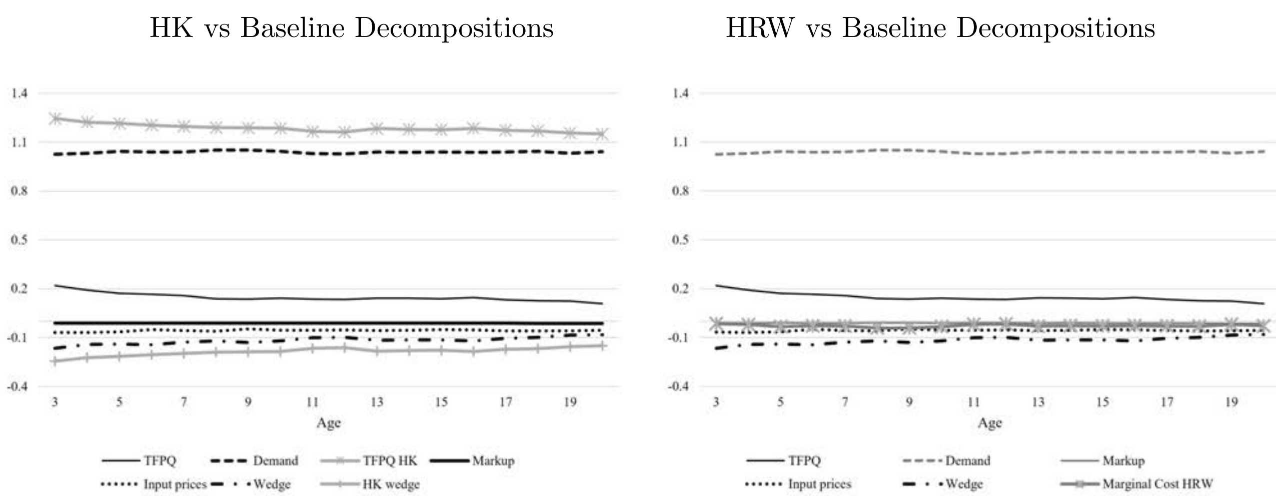
explore properties of $TFPQ\_HK$ constructed from revenue and input data compared to $TFPQ$ and demand shocks constructed from price and quantity data.

[37] 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 $\gamma = 1$.

misses the much larger contribution of demand relative to $TFPQ$, and the changing relative contribution of these two dimensions over the life cycle. As stated the relative contribution of demand to $TFPQ$ doubles from age 3 to age 20, something the HK decomposition is silent about. Figure 5 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 a 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 for both demand and production. 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 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 estimated contribution of wedges is different if σ is allowed to vary compared to imposing a common σ equal to the average of the variable σ.

Figure 6: Hsieh-Klenow and Hottman-Redding-Weinstein decompositions using the same elasticities 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 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.

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 also used in our framework, they decompose sales into the contributions of observed prices and demand shocks obtained 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 and the implied specification of markups, the contribution of the demand shock and markup are, by construction, the contributions one would obtain in their conceptual approach for given substitution elasticity.[38] 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 6 illustrates the by-age decomposition of revenue growth 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 and negligible, 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 input cost, efficiency 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 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.

[38] We refer here to 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: $\left(pm_{ft}^{-\phi\kappa_1} w_{ft}^{-\beta\kappa_1} \mu_{ft}^{-\gamma\kappa_1} \chi_{ft}\right) = 1$. This replicates the HK exercise that accounts for the productivity loss implied by composite wedges; as noted in HK's Appendix II, welfare is synonymous with productivity if demand shocks are explicitly accounted for as we do in our framework. 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.[39]

In different additional 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

[39] Hsieh and Klenow (2009) also use Cobb-Douglas aggregation across sectors. Baqaee and Farhi (2020) show the importance of taking into account the input-output structure of the economy in aggregating across sectors. 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.

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, described two paragraphs above, that replicates HK measurement of welfare costs of wedges). Columns 1 and 2 show a large gap: for the average sector, actual welfare is about 50% of its efficient level, i.e. what it would be in absence of HK wedges. This is in the broad range found by HK for India and China, and by Busso et al (2013) for Latin America, applying the same methodology. The figure is somewhat lower (45%) on a revenue-weighted basis, because it is in the largest sectors where wedges are largest. Large sectors tend to display high elasticities of substitution, which 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.[40]

Table 4: Welfare relative to HK efficient welfare, different scenarios with sector-level parameters

		Average Sector (1)	Average Sector - Revenue Weighted (2)
		Panel A: Actual to HK Efficient Welfare	
		0.496	0.453
		Panel B: Counterfactual to HK Efficient Welfare	
Plant attribute set to counterfactual level (constant mean value =1)	D+TFPQ (TFPQ_HK)	0.122	0.079
	Demand Shock	0.093	0.054
	TFPQ	0.525	0.450
	Input prices + Markup	0.818	0.696
	Input prices	0.635	0.592
	Markup	0.630	0.525
	Sales wedge	0.587	0.559

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

[40]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.

is shut down. On an unweighted basis, this reduces welfare from 50% to 12% of its efficient level. More interestingly, this is completely attributable to the effect of shutting down quality (or D) heterogeneity.[41] The impact on welfare of shutting down $TFPQ$ variability is negligible. The welfare analysis highlights once again the differential role of $TFPQ$ vs. demand, now in terms of welfare.

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 prices and markups explains most 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 82% of its efficient level for the average sector. Shutting down variability in the residual wedge has a more modest impact of moving welfare to 59% of the efficient level.

Interestingly, in contrast to the fact that input prices play a much more important role than markups in explaining cross sectional variability, both factors are similarly relevant to determine welfare. In both cases collapsing variability in the attribute increases welfare from 0.5 to 0.63 of the efficient level. The reason for the contrast with the cross sectional results in Table 3 is the combination of two factors: 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 the top panel of Table 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 "unweighted" cross-plant variation. In fact, the bottom panel of Table 3, displaying the revenue-weighted decomposition, already shows a much larger role for markups than in the top panel. 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.

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

[41] In results reported in Appendix L, we have also explored the effect of shutting down variability over the life cycle.

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

	Largest	Second	Third
Average	1.84	1.11	1.05
10	1.07	1.01	0.98
25	1.13	1.04	1.01
50	1.27	1.09	1.04
75	1.60	1.16	1.07
90	2.33	1.25	1.12
95	3.45	1.32	1.15
99	11.09	1.50	1.22
Max	81.08	1.83	1.39
N	713	713	713

There are 23 sectors in 31 years ($23 \times 31 = 713$) which gives one observation per sector*year.

construct, for that subperiod, quality-adjusted wages using an approach analogous to that of we use to build quality-adjusted materials and output prices.[42]

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 -0.028 to -0.01, compensating it with a reduced positive contribution of $TFPQ$ (Panel A). That is, quality heterogeneity explains almost a third of the dampening role of unadjusted wages over the variance of sales, suggesting that plant growth is accompanied by skill growth, thereby increasing cost and curbing revenue growth. The remaining -0.010 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 example, market segmentation due to search frictions can enhance monopsony power.

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

[42] That is, labeling quality adjusted wages as $\widehat{w}_{ft}$ and denoting the set of the three skill categories in the data as Ω^w, the wage index is given by $\ln \frac{\widehat{w}_{ft}}{\widehat{w}_{ft-1}} = \sum_{j \in \Omega^w} \ln \left(\frac{w_{fjt}}{w_{fjt-1}} \right)^{\frac{1}{3}} + \frac{1}{\sigma_w - 1} \ln \lambda_{ft}^{w,QRW}$

where $\lambda_{ft}^{w,QRW} = \prod_{j \in \Omega^w} \left(\frac{s^w_{fjt}}{s^w_{fjt-1}} \right)^{\frac{1}{3}}$ and s^w_{fjt} is the share of skill class j in f's payroll at time t. We then build 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.

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

		Unadjusted wage		Q-adj. Wage
		1982-2012	2000-2012	2000-2012
Panel A: Decomposition of life cycle growth of sales				
TFPQ		0.138	0.165	0.146
Demand Shock		1.039	1.042	1.042
Pm		-0.018	-0.010	-0.010
Wages		-0.037	-0.028	-0.010
Markup		-0.012	-0.010	-0.010
Sales wedge		-0.111	-0.159	-0.158
Panel B: Actual to HK Efficient Welfare (sector level parameters, average sector)				
		0.496	0.511	0.517
Panel C: Counterfactual to HK Efficient Welfare (sector level parameters, average sector)				
Plant attribute set to counterfact. level	D+TFPQ (TFPQ_HK)	0.122	0.116	0.118
	Demand Shock	0.093	0.090	0.091
	TFPQ	0.525	0.563	0.587
	Input prices + Markup	0.818	0.781	0.782
	Input prices	0.635	0.644	0.639
	Markup	0.630	0.611	0.619
	Sales wedge	0.587	0.607	0.621

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 establishments, rather than because of the extent to which it reflects heterogeneity in the quality of the labor input. We don't further explore such issues in our analysis but the finding that input prices matter even after quality adjustment, and that this is the case especially for welfare, suggests this is an important area for future research.

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 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 our structural wedges, which we find to be highly negatively correlated with productivity, both in terms of efficiency and quality. 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 establishment level prices on outputs and inputs to study the role of variation in technology and markups at the plant-level. Recent analyses by De Loecker, 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, Harun Alp, Nicholas Bloom, and William R. Kerr, 2018. "Innovation, Reallocation, and Growth." American Economic Review, 108 (22): 3450-91.

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 K. Khandelwal, and Adam Osman. 2019. "Measuring Productivity: Lessons from Tailored Surveys and Productivity Benchmarking." AEA Papers and Proceedings, 109: 444-49.

Atkin, David, Amit Khandelwal and Adam Osman, 2017, "Exporting and establishment 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*, 135(1): 105-163

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.

Bloom, Nicholas and John Van Reenen, 2007. "Measuring and Explaining Management Practices Across establishments 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*, 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 establishments export more? Product quality and Colombian plants”, *Journal of Development Economics*, Vol. 80, No. 1, pp. 160-178.

Busso Matias & Madrigal Lucia & Pagés Carmen, 2013. ”Productivity and resource misallocation in Latin America,” The B.E. Journal of Macroeconomics, De Gruyter, vol. 13(1), pages 1-30, June.

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 , 109 (7): 2531-67

De Loecker, J. and F. Warzynski. 2012. “Markups and establishment-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, 135(2):561-644

De Roux, Nicolas, Marcela Eslava, Santiago Franco and Eric Verhoogen, 2020. ”Estimating Production Functions in Differentiated-Product Industries with Quantity Information and External Instruments”. NBER Working Paper No. 28323

Decker, Ryan, John Haltiwanger, Ron Jarmin and Javier Miranda. 2018. ”Changing Business Dynamism and Productivity: Shocks vs. Responsiveness.” *American Economic Review*, December, 110(12): 3952–399.

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, establishment 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 establishment 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 establishment 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) "Establishment 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 establishments 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 Costs of Misallocation." Journal of Economic Perspectives, 31 (3): 151-74.

Web Appendix (Not intended for publication) for: The Life-cycle Growth of Plants: The Role of Productivity, Demand and Wedges.

Marcela Eslava* and John Haltiwanger†

February 8, 2021

1 Appendix A: price indices

1.1 CUPI price index

Our baseline results use Redding and Weinstein's (2018) CUPI price indices at the plant level as deflators. Here, we follow Redding and Weinstein (2018) to derive the CUPI index in the context of our model. The change in prices from one period to the next in our model is:

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

Defining as $\Omega_{t,t-1}^f$ the set of goods that is common to both periods, and multiplying both the numerator and the denominator by $\left(\sum_{\Omega_{t,t-1}^f} d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w} * \sum_{\Omega_{t,t-1}^f} d_{fjt}^{\sigma^w} p_{fjt}^{1-\sigma^w} \right)^{\frac{1}{1-\sigma^w}}$ we obtain:

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

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

$$\frac{P_{ft}}{P_{ft-1}} = \left(\frac{\sum_{\Omega_t^f} d_{fjt}^{\sigma^w} p_{fjt}^{1-\sigma^w}}{\sum_{\Omega_{t,t-1}^f} d_{fjt}^{\sigma^w} p_{fjt}^{1-\sigma^w}} \frac{\sum_{\Omega_{t,t-1}^f} d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w}}{\sum_{\Omega_{t-1}^f} d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w}} \frac{\sum_{\Omega_{t,t-1}^f} d_{fjt}^{\sigma^w} p_{fjt}^{1-\sigma^w}}{\sum_{\Omega_{t,t-1}^f} d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w}} \right)^{\frac{1}{1-\sigma^w}} \quad (2)$$

$$= \frac{\lambda_{ft-1,\Omega_{t,t-1}^f}}{\lambda_{ft,\Omega_{t,t-1}^f}} \left(\frac{\sum_{\Omega_{t,t-1}^f} d_{fjt}^{\sigma^w} p_{fjt}^{1-\sigma^w}}{\sum_{\Omega_{t,t-1}^f} d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w}} \right)^{\frac{1}{1-\sigma^w}} \quad (3)$$

where $\lambda_{ft-1,\Omega_{t,t-1}^f} = \left(\frac{\sum_{\Omega_{t,t-1}^f} d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w}}{\sum_{\Omega_{t-1}^f} d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w}} \right)^{\frac{1}{1-\sigma^w}}$ and $\lambda_{ft,\Omega_{t,t-1}^f} = \left(\frac{\sum_{\Omega_{t,t-1}^f} d_{fjt}^{\sigma^w} p_{fjt}^{1-\sigma^w}}{\sum_{\Omega_t^f} d_{fjt}^{\sigma^w} p_{fjt}^{1-\sigma^w}} \right)^{\frac{1}{1-\sigma^w}}$.

Furthermore, since

$$s_{fjt} = \frac{p_{fjt} q_{fjt}}{R_{ft}} = \frac{p_{fjt}^{1-\sigma^w} \left(d_{fjt}^{\sigma^w} \right)}{P_{fjt}^{1-\sigma^w}} \quad (4)$$

we have that:

$$\lambda_{ft-1,\Omega_{t,t-1}^f} = \left(\sum_{\Omega_{t,t-1}^f} \frac{d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w}}{\sum_{\Omega_{t-1}^f} d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w}} \right)^{\frac{1}{1-\sigma^w}} = \left(\sum_{\Omega_{t,t-1}^f} s_{fjt-1} \right)^{\frac{1}{1-\sigma^w}}$$

That is, $\left(\lambda_{ft-1,\Omega_{t,t-1}^f} \right)^{1-\sigma^w}$ is the share of period $t-1$ expenditures devoted to goods that are common to both periods. Similarly, $\left(\lambda_{ft,\Omega_{t,t-1}^f} \right)^{1-\sigma^w}$ is the share of period t expenditure devoted to goods common to both periods.

With this, the change in prices between the two periods (equation (1)) can be written:

$$\frac{P_{ft}}{P_{ft-1}} = \left(\frac{\sum_{\Omega_{t,t-1}^f} s_{fjt}}{\sum_{\Omega_{t,t-1}^f} s_{fjt-1}} \right)^{\frac{1}{\sigma^w-1}} \frac{P_{ft}^*}{P_{ft-1,\Omega_{t,t-1}^f}^*} = \left(\lambda_{ft}^{Qfee} \right)^{\frac{1}{\sigma^w-1}} \frac{P_{ft}^*}{P_{ft-1,\Omega_{t,t-1}^f}^*} \quad (5)$$

where $P_{ft}^* = \left(\sum_{\Omega_{t,t-1}^f} d_{fjt}^{\sigma^w} p_{fjt}^{1-\sigma^w} \right)^{\frac{1}{1-\sigma^w}}$ is a period t price index for the basket of goods common to t and $t-1$ for firm f, and $P_{ft-1,\Omega_{t,t-1}^f}^* = \left(\sum_{\Omega_{t,t-1}^f} d_{fjt-1}^{\sigma^w} p_{fjt-1}^{1-\sigma^w} \right)^{\frac{1}{1-\sigma^w}}$

is a period $t-1$ price index for that same basket. Term $\frac{\sum_{\Omega^f_{t,t-1}} s_{fjt}}{\sum_{\Omega^f_{t,t-1}} s_{fjt-1}} = \lambda^{Qfee}_{ft}$ is the Feenstra adjustment for changing varieties, expressed in terms of observables.

Moreover, the Marshalian demands, given by $q_{fjt} = d^{\sigma^w}_{ft} d^{\sigma^w}_{fjt} \left(\frac{P_{ft}}{P_t}\right)^{-\sigma^w} \left(\frac{p_{fjt}}{P_{ft}}\right)^{-\sigma^w} \frac{E_t}{P_t}$, imply

$$s^*_{fjt} = \frac{d^{\sigma^w}_{ft} d^{\sigma^w}_{fjt} \left(\frac{P_{ft}}{P_t}\right)^{-\sigma^w} \frac{p^{1-\sigma^w}_{fjt}}{P^{-\sigma^w}_{ft}} \frac{E_t}{P_t}}{\sum_{\Omega^f_{t,t-1}} d^{\sigma^w}_{ft} d^{\sigma^w}_{fjt} \left(\frac{P_{ft}}{P_t}\right)^{-\sigma^w} \frac{p^{1-\sigma^w}_{fjt}}{P^{-\sigma^w}_{ft}} \frac{E_t}{P_t}} = \frac{d^{\sigma^w}_{fjt} p^{1-\sigma^w}_{fjt}}{\left(P^*_{ft}\right)^{1-\sigma^w}}$$

and

$$s^*_{fjt-1,\Omega^f_{t,t-1}} = \frac{d^{\sigma^w}_{ft-1} d^{\sigma^w}_{fjt-1} \left(\frac{P_{ft-1}}{P_{t-1}}\right)^{-\sigma^w} \frac{p^{1-\sigma^w}_{fjt-1}}{P^{-\sigma^w}_{ft-1}} \frac{E_{t-1}}{P_{t-1}}}{\sum_{\Omega^f_{t,t-1}} d^{\sigma^w}_{ft-1} d^{\sigma^w}_{fjt-1} \left(\frac{P_{ft-1}}{P_{t-1}}\right)^{-\sigma^w} \frac{p^{1-\sigma^w}_{fjt-1}}{P^{-\sigma^w}_{ft-1}} \frac{E_{t-1}}{P_{t-1}}} = \frac{d^{\sigma^w}_{fjt-1} p^{1-\sigma^w}_{fjt-1}}{\left(P^*_{ft-1,\Omega^f_{t,t-1}}\right)^{1-\sigma^w}}$$

Dividing s^*_{fjt} by $s^*_{fjt-1,\Omega^f_{t,t-1}}$ and rearranging, we obtain

$$\left(\frac{p_{fjt}}{p_{fjt-1}}\right) = \left(\frac{P^*_{ft}}{P^*_{ft-1,\Omega^f_{t,t-1}}}\right) \left(\frac{s^*_{fjt}}{s^*_{fjt-1,\Omega^f_{t,t-1}}}\right)^{\frac{1}{1-\sigma^w}} \left(\frac{d_{fjt}}{d_{fjt-1}}\right)^{-\frac{\sigma^w}{1-\sigma^w}}$$

Given this, for plant-product weights $\omega_{ft} = \frac{1}{\left\|\Omega^f_{t,t-1}\right\|}$ such that $\sum_{\Omega^f_{t,t-1}} \omega_{ft,t-1} = 1$ we can write,

$$\begin{aligned} &\sum_{\Omega^f_{t,t-1}} \ln \left(\frac{p_{fjt}}{p_{fjt-1}}\right)^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}} \\ = \quad &\ln \left(\frac{P^*_{ft}}{P^*_{ft-1,\Omega^f_{t,t-1}}}\right) + \frac{1}{(1-\sigma^w)} \sum_{\Omega^f_{t,t-1}} \ln \left(\frac{s^*_{fjt}}{s^*_{fjt-1,\Omega^f_{t,t-1}}}\right)^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}} \\ &+\frac{\sigma^w}{\sigma^w-1} \sum_{\Omega^f_{t,t-1}} \ln \left(\frac{d_{fjt}}{d_{fjt-1}}\right)^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}} \end{aligned}$$

where the first right-hand-side term takes into account that $\sum_{\Omega^f_{t,t-1}} s^*_{fjt,\Omega^f_{t,t-1}} = 1$. Shocks d_{fjt} have been defined relative to plant appeal, d_{ft}, such that $\prod_{\Omega^f_{t,t-1}} d_{fjt}^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}} = 1$, with the implication that

$\sum_{\Omega^f_{t,t-1}} \ln\left(\frac{d_{fjt}}{d_{fjt-1}}\right)^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}} = 0$. Notice that this normalization still allows for a distribution of product appeal that varies over time. [1]

The consecutively common good price index growth $\left(\frac{P^*_{ft}}{P^*_{ft-1,\Omega^f_{t,t-1}}}\right)$ therefore corresponds to

$$\ln\left(\frac{P^*_{ft}}{P^*_{ft-1,\Omega^f_{t,t-1}}}\right) = \sum_{\Omega^f_{t,t-1}} \ln\left(\frac{p_{fjt}}{p_{fjt-1}}\right)^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}} - \frac{1}{(1-\sigma^w)} \sum_{\Omega^f_{t,t-1}} \ln\left(\frac{s^*_{fjt}}{s^*_{fjt-1,\Omega^f_{t,t-1}}}\right)^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}}$$

$$= \sum_{\Omega^f_{t,t-1}} \ln\left(\frac{p_{fjt}}{p_{fjt-1}}\right)^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}} - \frac{1}{(1-\sigma^w)} \ln\lambda^{QRW}_{ft}$$

The term $\ln\lambda^{QRW}_{ft} = \sum_{\Omega^f_{t,t-1}} \ln\left(\frac{s^*_{fjt}}{s^*_{fjt-1,\Omega^f_{t,t-1}}}\right)^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}}$ adjusts for changes in appeal for continuing products, addressing the consumer valuation bias. Plugging into equation (5), we obtain

$$\ln\frac{P_{ft}}{P_{ft-1}} = \sum_{\Omega^f_{t,t-1}} \ln\left(\frac{p_{fjt}}{p_{fjt-1}}\right)^{\frac{1}{\left\|\Omega^f_{t,t-1}\right\|}} - \frac{1}{(1-\sigma^w)}\left(\ln\lambda^{QRW}_{ft} + \ln\lambda^{Qfee}_{ft}\right) \quad (6)$$

[1] This is by contrast to empirical price indices that weight across products with variable weights $\omega_{fjt} \neq \omega_{ft}$, such as the commonly used Sato-Vartia approach (Sato, 1974, Vartia, 1974, Feenstra, 2004). Under such variable weights the assumption $\sum_{\Omega^f_{t,t-1}} \ln\left(\frac{d_{fjt}}{d_{fjt-1}}\right)^{\omega_{fjt}} = 0$ does not hold. The fact that traditional approaches using variable weights ignore this term leads to what Redding and Weinstein (2017) have called the "consumer valuation bias" the traditional empirical approaches to economically motivated price indices.

We similarly obtain a measure of materials by deflating material expenditure by plant-level price indices for materials, pm_{ft}, using information on individual prices and quantities of material inputs. We construct pm_{ft} using an analogous approach to that used to construct output prices. The underlying assumption is that M_{ft}, the index of materials quantities used, is a CES aggregate of individual inputs. As is the case with output prices, until we have an estimate of the elasticity of substitution, we can only build a consecutively-common-basket price index $\overline{pm^*_{ft}}$ for plant f, and carry an adjustment factor $\Lambda^M_{ft} = \Lambda^{MRW}_{ft}\Lambda^{Mfee}_{ft}$ for which we later adjust prices. In particular, we deflate materials expenditures to obtain $M^*_{ft} = \frac{\text{materials expenditure}_{ft}}{pm_{fB}\overline{pm^*_{ft}}} = M_{ft} * \left(\Lambda^M_{ft}\right)^{\frac{1}{\sigma^w-1}}$. Once we have obtained an estimate of the elasticity of substitution we calculate $pm_{ft} = pm_{fB} * \overline{pm^*_{ft}} * \left(\Lambda^M_{ft}\right)^{\frac{1}{\sigma^w-1}}$, which is one of the fundamentals on the cost side in our growth decomposition. We use this price index as deflator for materials expenditure to obtain our $TFPQ$ measure. We use for inputs the same elasticity of substitution estimated for outputs. We recognize that using the same elasticity for inputs and outputs is a strong assumption, but find that it does no affect our results in an important way. In particular, we find in Appendix I that using a Divisia price index (with updated input mix each period) generates about the same contribution for materials prices in sales and output volatility as the UPI. The Divisia materials price index does not depend on the elasticity of substitution, suggesting that this assumption is not critical for our results.

1.2 Initializing a plant's CUPI price index

A plant's price index is constructed as

$$P_{ft} = P_{fB} * \overline{P^*_{ft}} * \left(\Lambda^Q_{ft}\right)^{\frac{1}{\sigma_w-1}}$$

The initial level P_{fB}, where B is the base year for plant f, is constructed as: $P_{fB} = P_{base,B} \prod_{\Omega^f_B} \left(\frac{p_{fjB}}{\overline{p_{jB}}}\right)^{s_{fjB}}$, where $\overline{p_{jB}}$ is the geometric average of the price of product j in year B across plants, year B is the first year in which plant f is present in the survey, and $P_{base,B}$ is an overall base. We use 1982 as the base year, so $P_{base,1982} = 1$. For plants with $B \neq 1982$, $P_{base,B}$ is set equal to the geometric mean of the price index across plants that we observe prior to year B. Notice that our approach takes advantage of cross sectional variability across plants for any given product or input j. In the plant's base year B, $\left(\frac{P_{fjB}}{P_{jB}}\right) = 1$ for the average producer of product j. For other plants, it will capture dispersion in price levels around that average.[2]

1.3 Sato-Vartia indices

The Sato-Vartia approach (used in the results in Appendix C) is an alternative way of computing $\ln\left(\frac{P^*_{ft}}{P^*_{ft-1,\Omega^f_{t,t-1}}}\right)$, using weights $\omega^{SV}_{fjt,t-1} = \frac{\frac{(s^*_{fjt}-s^*_{fjt-1,t})}{\ln s^*_{fjt}-\ln s^*_{fjt-1,t}}}{\sum\limits_{\Omega^f_{t,t-1}}\left(\frac{(s^*_{fjt}-s^*_{fjt-1,t})}{\ln s^*_{fjt}-\ln s^*_{fjt-1,t}}\right)}$ and imposing $-\frac{\sigma}{\sigma-1}\sum\limits_{\Omega^f_{t,t-1}}\ln\left(\frac{d_{fjt}}{d_{fjt-1}}\right)^{\omega^{SV}_{fjt}} = 0$. That is, $\ln\left(\frac{P^*_{ft}}{P^*_{ft-1,\Omega^f_{t,t-1}}}\right)^{SV} = \sum\limits_{\Omega^f_{t,t-1}}\ln\left(\frac{p_{fjt}}{p_{fjt-1}}\right)^{\omega^{SV}_{fjt}}$.

Notice, in the derivation above, that when using variable weights $\omega_{fjt} \neq \omega_{ft}$, the assumption $\sum\limits_{\Omega^f_{t,t-1}}\ln\left(\frac{d_{fjt}}{d_{fjt-1}}\right)^{\omega_{fjt}} = 0$ would not hold. In the Sato-Vartia case, since product demand shocks $\frac{d_{fjt}}{d_{fjt-1}}$ are positively correlated with the weights $\omega^{SV}_{fjt,t-1}$ (Redding-Weinstein, 2020), $\sum\limits_{\Omega^f_{t,t-1}}\ln\left(\frac{d_{fjt}}{d_{fjt-1}}\right)^{\omega^{SV}_{fjt}} > 1$ and the consumer valuation bias would be positive. That is, the Sato Vartia approach likely overstates price inflation for the common goods produced by

[2] We deal with excessive noise from partial year reporting and other sources by eliminating outliers. In particular, in any given year we consider only products that represent at least 2% of sales of the respective plant. Shares are re-calculated accordingly for this restricted basket. We also winsorize the 2% tails at each step of the process of building price indices. In particular, we winsorize $\frac{\sum_{\Omega^f_{l,l-1}} s_{fjl}}{\sum_{\Omega^f_{l,l-1}} s_{fjl-1}}$; $\prod\limits_{\Omega^f_{t,t-1}}\left(\frac{s^*_{fjt}}{s^*_{fjt-1,\Omega^f_{t,t-1}}}\right)^{\frac{1}{\|\Omega_{t,t-1}\|}}$; $\frac{p_{fjt}}{p_{fjt-1}}$; $\frac{P^*_{ft}}{P^*_{ft-1,\Omega^f_{t,t-1}}}$; $\frac{P_{ft}}{P_{ft-1}}$.

We also winsorize adjustment factors at the 5% level. Extreme changes in the baskets of goods, where common $(t,t-1)$ products represent a negligible share of revenue in either t or $t-1$ imply extreme values for $\ln\Lambda^Q_{ft}$. These extreme changes may partly reflect measurement error in an enviroment where baskets of goods are auto-reported into relatively wide product components.

plant f in both $t-1$ and t. Such overstatement of price inflation implies understatement of quantity growth and therefore $TFPQ$.

1.4 Törnqvist index

Appendix C also presents results using Törnqvist indices, first imposing a basket of goods that is fixed over the life cycle and constant weights for them, and then imposing constant baskets only over consecutive periods (the "divisia" case). Törnqvist indices for the growth of prices of plant f at time t are constructed as $\frac{P_{ft}}{P_{ft-1}} = \prod_{\Omega^f_{t,t-1}} \left(\frac{p_{fjt}}{p_{fjt-1}}\right)^{\overline{s}_{fjt}}$.

In the constant baskets of goods version of the Törnqvist index, $\Omega^f_{t,t-1} = \Omega^f$ is a basket of all products ever produced (or materials ever used) by plant f, and $\overline{s}_{fj}$ is the average share of j in that basket of products (or materials) plant f produces over the whole period. In this approach, the plant level index is initialized at $\ln P_{fA} = \sum_{\Omega^f} s_{fj} \left(\ln p_{fjA} - \ln \overline{p_{jA}}\right)$. If product j is not produced (or used as input) in years t or $t-1$ (or both), $\Delta \ln(P_{fjt})$ is inputed at the average growth of the price of that product (or input) for other plants within the sector. If no plant in the sector produces that good in t, then the average over all plants is used, independent of sector.

The divisia version of the Törnqvist index is similar, but $\Omega^f_{t,t-1}$ is the basket of goods produced by f in either t or $t-1$ and $\overline{s}_{fjt}$ is the average share of product j in plant $f's$ sales over t and $t-1$.

Törnqvist prices with a constant basket of products do not quality-adjust prices in any way, for either product turnover of changing quality of surviving products. Compared to this version, all versions allowing for evolving baskets of goods (including divisia) have the advantage of capturing evolving expenditure shares over time and therefore quality-adjusting prices, but the disadvantage of being more biased by errors from product coding and coarse aggregation, which are more likely in our context than in that of prices from scan bar codes (Hottman et al 2016). Compared to our baseline estimation with UPI prices, even versions with changing baskets quality-adjust in a less precise (i.e. not exact) manner, by imposing restrictions to the extent to which appeal many vary.

2 Appendix B: firm's problem

Firm chooses X_t to solve:

$$\underset{\{X_{ft}\}}{Max} \quad \pi_{ft} = (1-\tau_{ft}) R_{ft} - C_{ft}X_{ft} = D_{ft}A_{ft}^{1-\frac{1}{\sigma}} X_{ft}^{\gamma\left(1-\frac{1}{\sigma}\right)} - C_{ft}X_{ft}$$

where $R_{ft} = P_{ft}Q_{ft} = D_{ft}Q_{ft}^{1-\frac{1}{\sigma}} = d_{ft}E_t^{\frac{1}{\sigma}} P_t^{1-\frac{1}{\sigma}} Q_{ft}^{1-\frac{1}{\sigma}}$ and $Q_{ft} = A_{ft}X_{ft}^{\gamma}$. Optimal input demand is

$$X_{ft} = \left(\frac{(1-\tau_{ft}) D_{ft}A_{ft}^{1-\frac{1}{\sigma}}\gamma}{\mu_{ft}C_{ft}} \right)^{\frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}} \tag{7}$$

Proof. If the firm has market power, then $\frac{\partial P_t}{\partial X_{ft}} \neq 0$. The first order condition for the firm is then given by

$$\begin{aligned}
(1-\tau_{ft})\left(1-\frac{1}{\sigma}\right)\left(\frac{R_{ft}}{Q_{ft}} + \frac{R_{ft}}{P_{ft}}\frac{\partial P_t}{\partial Q_{ft}}\right)\frac{\partial Q_{ft}}{\partial X_{ft}} &= C_{ft} \\
(1-\tau_{ft})\left(1-\frac{1}{\sigma}\right)\frac{R_{ft}}{Q_{ft}}\left(1 + \frac{Q_{ft}}{P_{ft}}\frac{\partial P_t}{\partial Q_{ft}}\right)\frac{\partial Q_{ft}}{\partial X_{ft}} &= C_{ft} \\
(1-\tau_{ft})\left(\frac{\sigma-1}{\sigma}\right) D_{ft}Q_{ft}^{-\frac{1}{\sigma}} (1-s_{ft}) \frac{\partial Q_{ft}}{\partial X_{ft}} &= C_{ft} \\
\frac{(1-\tau_{ft})}{\mu_{ft}} D_{ft}Q_{ft}^{-\frac{1}{\sigma}} \left(\gamma A_{ft}X_{ft}^{\gamma-1}\right) &= C_{ft} \qquad (8) \\
\frac{(1-\tau_{ft}) D_{ft}A_{ft}^{1-\frac{1}{\sigma}}\gamma}{\mu_{ft}C_{ft}} &= X_{ft}^{1-\gamma\left(1-\frac{1}{\sigma}\right)} \qquad (9)
\end{aligned}$$

Where the third line uses Sheppard's lemma $\left(-\frac{\partial P_t}{\partial Q_{ft}}\frac{Q_{ft}}{P_t} = s_{ft}\right)$, and the fourth line uses $\mu^{-1} = 1 - \left(\frac{1}{\sigma} + \left(\frac{\sigma-1}{\sigma}\right) s_{ft}\right) = \frac{\sigma-1-(\sigma-1)s_{ft}}{\sigma} = \frac{(\sigma-1)\left(1-s_{ft}\right)}{\sigma}$ (see Appendix D). Therefore $X_{ft} = \left(\frac{\left(1-\tau_{ft}\right) D_{ft}A_{ft}^{1-\frac{1}{\sigma}}\gamma}{\mu_{ft}C_{ft}} \right)^{\frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}}$. ■

Suppose $X_{ft} = K_{ft}^{\frac{\alpha}{\gamma}} L_{ft}^{\frac{\beta}{\gamma}} M_{ft}^{\frac{\phi}{\gamma}}$ where K, L and M are, respectively, capital, labor and material inputs, and $\gamma = \alpha + \beta + \phi$. Consequently, C_{ft} is itself a

Cobb Douglas aggregate of factor prices: $C_{ft} = r_t^{\frac{\alpha}{\gamma}} w_{ft}^{\frac{\beta}{\gamma}} pm_{ft}^{\frac{\phi}{\gamma}}$. We note that we do not have information on the rental price of capital, which we assume constant across plants (within a sector). Consequently,

$$\frac{X_{ft}}{X_{f0}} = \left(\frac{d_{ft}}{d_{f0}}\right)^{\kappa_1} \left(\frac{a_{ft}}{a_{f0}}\right)^{\kappa_2} \left(\frac{pm_{ft}}{pm_{f0}}\right)^{-\frac{\phi}{\gamma}\kappa_1} \left(\frac{w_{ft}}{w_{f0}}\right)^{-\frac{\beta}{\gamma}\kappa_1} \kappa_t \widehat{\kappa}_{ft} \quad (10)$$

$$\frac{L_{ft}}{L_{f0}} = \left(\frac{d_{ft}}{d_{f0}}\right)^{\kappa_1} \left(\frac{a_{ft}}{a_{f0}}\right)^{\kappa_2} \left(\frac{pm_{ft}}{pm_{f0}}\right)^{-\phi\kappa_2} \left(\frac{w_{ft}}{w_{f0}}\right)_t^{-\kappa_1+(\alpha+\phi)\kappa_2} \vartheta_t \vartheta_{ft} \quad (11)$$

$$\frac{Q_{ft}}{Q_{f0}} = \left(\frac{d_{ft}}{d_{f0}}\right)^{\gamma\kappa_1} \left(\frac{a_{ft}}{a_{f0}}\right)^{\kappa_1} \left(\frac{pm_{ft}}{pm_{f0}}\right)^{-\phi\kappa_1} \left(\frac{w_{ft}}{w_{f0}}\right)^{-\beta\kappa_1} \chi_t \chi_{ft} \quad (12)$$

where $\kappa_1 = \frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}$; $\kappa_2 = \left(1 - \frac{1}{\sigma}\right)\kappa_1$; $\chi_t = \left(\frac{D_t}{D_0}\right)^{\kappa_1} \left(\frac{A_t}{A_0}\right)^{1+\kappa_2} \left(\frac{C_t}{C_0}\right)^{-\kappa_1} \left(\frac{r_t^{-\alpha\kappa_1}}{r_0^{-\alpha\kappa_1}}\right)$ captures aggregate growth between birth and age t, and $\chi_{ft} = \frac{\left(1-\tau_{ft}\right)^{\gamma\kappa_1}}{\left(1-\tau_{f0}\right)^{\gamma\kappa_1}}$ captures residual variation from wedges, and the unobserved user cost of capital; $\widehat{\kappa}_{ft} = \chi_{ft}^{\frac{1}{\gamma}}$ and $\chi_t = \kappa_t^{\frac{1}{\gamma}} \left(\frac{A_t}{A_0}\right)^{-1}$. We have used the fact that $1+\gamma\kappa_1\left(1 - \frac{1}{\sigma}\right) = \kappa_1$.

Moreover, since $R_{ft} = D_{ft}Q_{ft}^{1-\frac{1}{\sigma}}$ and $1 + \gamma\kappa_1\left(1 - \frac{1}{\sigma}\right) = \kappa_1$ then

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

Notice also that 8 can be re-written as:

$$\begin{aligned} \frac{(1-\tau_{ft})}{\mu_{ft}} D_{ft} Q_{ft}^{-\frac{1}{\sigma}} \theta_{ft}^v &= C_{ft} \\ \frac{(1-\tau_{ft})}{\mu_{ft}} P_{ft} \theta_{ft}^v &= \frac{C_{ft}X_{ft}}{Q_{ft}} \\ \frac{\theta_{ft}^v}{\frac{C_{ft}X_{ft}}{P_{ft}Q_{ft}}} &= \frac{\mu_{ft}}{(1-\tau_{ft})} \end{aligned}$$

where θ_{ft}^v is the output elasticity of a variable factor V and we have used $P_{ft} = D_{ft}Q_{ft}^{-\frac{1}{\sigma}}$. The expression $\frac{\theta_{ft}^v}{\frac{C_{ft}X_{ft}}{P_{ft}Q_{ft}}}$ is markup "a-la-De Loecker", which is here shown to equal a ratio between the model's markup, μ_{ft}, and the revenue wedge $(1 - \tau_{ft})$.

Empirically, we estimate $\widehat{\mu}_{ft}^{DL} = \frac{\phi}{\frac{pm_{ft}M_{ft}}{R_{ft}}}$, because M is the only variable factor in our estimations. As discussed in the main text, using this expression for the markup μ_{ft}^{DL} yields a discrepancy between the ratio of estimated output elasticity to the cost share of revenue and the ratio of the model's markup and the revenue wedge. This discrepancy can be accounted for by a factor-specific wedge yielding:

$$\widehat{\mu}_{ft}^{DL} = \frac{\phi}{\frac{C_{ft}X_{ft}}{P_{ft}Q_{ft}}} = \frac{\mu_{ft}(1 - \tau_{ft}^{v})}{(1 - \tau_{ft})}$$

The factor-specific wedge is already implicitly incorporated in the sales wedge as the latter is a composite wedge measure that captures any source of discrepancy between actual and sales implied by the static model based on model parameters and fundamentals. Factor-specific wedges will have an impact on scale but also will impact factor mix. This implies that there is an additional potential impact of a factor-specific wedge on first-order conditions for individual inputs. This factor-specific wedge may have a variety of sources including factor-specific frictions and wedges, measurement and specification issues. The latter includes, for example, differences in the actual vs. estimated factor elasticity for the variable factor. If we use equation (2) and the type of structural decomposition presented in appendix G, we find that about 65% of measured $\widehat{\mu}_{ft}^{DL}$ is accounted for by the factor-specific wedge.

3 Appendix C: Sensitivity to Revenue Curvature

To assess the contribution of $TFPQ\_HK_{ft}$ and *composite* wedges to sales growth, we first calculate $TFPQ\_HK_{ft} = R_{ft}^{1/(1-\frac{1}{\sigma})}/X_{ft}^{\gamma}$ using our estimates of σ, ϕ, β, α, and the implied $X = M_{ft}^{\frac{\phi}{\gamma}} L_{ft}^{\frac{\beta}{\gamma}} K_{ft}^{\frac{\alpha}{\gamma}}$. We call this calculation $TFPQ\_HK_{ft}$ "unconstrained", since we use detailed parameter estimates that would be hard to obtain if one were constrained by the lack of plant-level data on prices. We also build an estimate of $TFPQ\_HK_{ft}$ "constrained" where, following usual practice, we impose monopolistic competition, $\gamma = 1$, ϕ, β, α equal to the corresponding cost shares, and a constant number for σ (as in the macro misallocation literature).[3] While in the unconstrained case M is the materials quantitities index built deflating with our UPI plant-level deflators for materials, in the constrained one it is materials expenditure deflated with the PPI.

Table C1. Decomposition of sales under baseline and constrained fundamentals

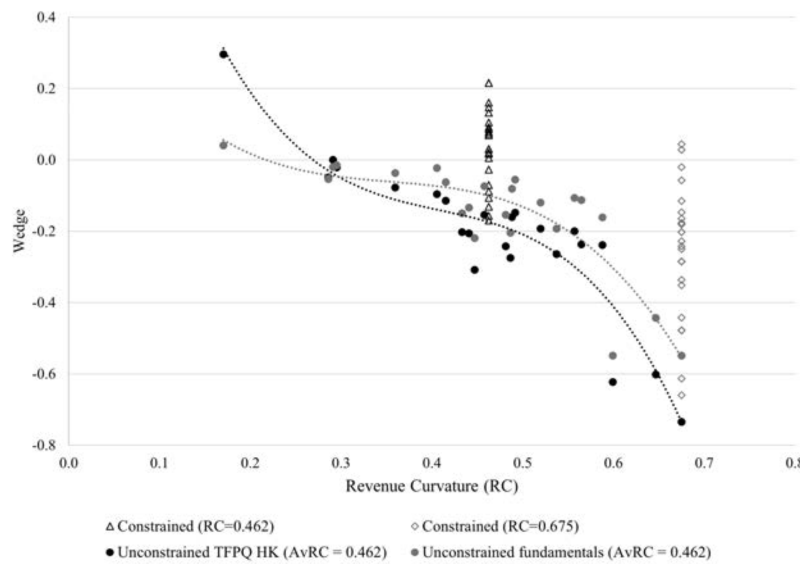
	Structural				
	(1)	(2)	(3)	(4)	(5)
TFPQ HK unconstrained	1.182				
TFPQ HK constrained			0.917	1.161	1.178
TFPQ		0.138			
Demand shock		1.039			
ln Input prices		-0.018			
ln Average wage		-0.037			
ln Markup		-0.012			
Sales Wedge		-0.111			
Composite (HK) Sales Wedge	-0.182		0.083	-0.161	-0.178
Avg Rev Curvature	**0.462**	**0.462**	**0.462**	**0.666**	**0.675**
Max Rev Curvature	**0.675**	**0.675**	**0.462**	**0.666**	**0.675**

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

[3]Under these assumptions, $\mu_{it} = \mu = \frac{\sigma}{\sigma-1}$. Since cost minimization implies $\beta = \frac{w_{ft}L_{ft}}{Cost}\gamma = \frac{w_{ft}L_{ft}}{R_{ft}}\mu$ (Hall, 1994), we impose for each sector $\beta = \frac{\sum_f w_{ft}L_{ft}}{\sum_f R_{ft}}\frac{\sigma}{\sigma-1}$, calculating β first for each year and then averaging over years. We proceed similarly for ϕ, and then obtain $\alpha = 1 - \beta - \phi$.

Figure C1 and Table C1 depict the contribution of wedges for different levels of σ and, therefore, of the curvature of the revenue function. The estimated contribution of wedges to sales is higher when the revenue curvature parameter $\gamma\left(1-\frac{1}{\sigma}\right)$ is high (i.e. curvature is low), and that the increase is nonlinear: in sectors when $\gamma\left(1-\frac{1}{\sigma}\right)$ is close to 1, wedges tend to dominate the contribution of fundamentals.

Figure C1: Wedges vs Revenue Curvature (by 3-digit sector)



4 Appendix D: markups

The firm's (potentially variable) markup after the distortion, $\mu_{ft} = \frac{P_{ft}}{mc_{ft}\left(1-\tau_{ft}\right)^{-1}}$, is given by:

$$\mu_{ft} = \frac{1}{1-\left(\frac{1}{\sigma}-\left(\frac{\sigma-1}{\sigma}\right)s_{ft}\right)} = \frac{\sigma}{\left(\sigma-1\right)\left(1-s_{ft}\right)} \tag{13}$$

Proof. $\underset{Q_{ft}}{Max}\left(1-\tau_{ft}\right)P_{ft}Q_{ft}-CT$ leads to first orden condition $\left(P_{ft}+Q_{ft}\frac{dP_{ft}}{dQ_{ft}}\right) = \frac{mc_{ft}}{\left(1-\tau_{ft}\right)}$. Dividing by P_{ft} we obtain $\frac{1}{\mu_{ft}} = 1+\frac{Q_{ft}}{P_{ft}}\frac{dP_{ft}}{dQ_{ft}} = 1-\ \epsilon^{-1}$ (where we have denoted $\epsilon_{ft} \equiv -\frac{Q_{ft}}{P_{ft}}\frac{dP_{ft}}{dQ_{ft}}$), so that

$$\mu_{ft} = \left(\frac{\epsilon_{ft}}{\epsilon_{ft}-1}\right) \tag{14}$$

In turn, under $Q_{ft} = d_{ft}^{\sigma} P_{ft}^{-\sigma} \frac{E_t}{P_t^{1-\sigma}}$ and its implication that $P_{ft} = d_{ft} Q_{ft}^{-\frac{1}{\sigma}} \left(\frac{E_t}{P_t^{1-\sigma}}\right)^{\frac{1}{\sigma}} = d_{ft} Q_{ft}^{-\frac{1}{\sigma}} \left(\frac{Q_t}{P_t^{-\sigma}}\right)^{\frac{1}{\sigma}}$ and allowing for market power so that $\frac{dP_t}{dQ_{ft}} \neq 0$, the inverse of the demand elasticity as perceived by the firm ($\epsilon_{ft}^{-1} \equiv -\frac{dP_{ft}}{dQ_{ft}}\frac{Q_{ft}}{P_{ft}}$) is:

$$\begin{aligned}
\epsilon_{ft}^{-1} &= -\left(\frac{\partial P_{ft}}{\partial Q_{ft}} + \frac{\partial P_{ft}}{\partial P_t}\frac{\partial P_t}{\partial Q_{ft}}\right)\frac{Q_{ft}}{P_{ft}} && (15)\\
&= -\left(-\frac{1}{\sigma}\frac{P_{ft}}{Q_{ft}} + \left(\frac{\sigma-1}{\sigma}\right)\frac{P_{ft}}{P_t}\frac{\partial P_t}{\partial Q_{ft}}\right)\frac{Q_{ft}}{P_{ft}} \\
&= \left(\frac{1}{\sigma} - \left(\frac{\sigma-1}{\sigma}\right)\frac{\partial P_t}{\partial Q_{ft}}\frac{Q_{ft}}{P_t}\right) \\
&= \left(\frac{1}{\sigma} + \left(\frac{\sigma-1}{\sigma}\right)s_{ft}\right) && (16)
\end{aligned}$$

where the last line uses Sheppard's lemma: $-\frac{\partial P_t}{\partial Q_{ft}}\frac{Q_{ft}}{P_t} = s_{ft}$.

Equations (14) and (16) together imply $\mu_{ft}^{-1} = 1-\epsilon_{ft}^{-1} = 1-\left(\frac{1}{\sigma} + \left(\frac{\sigma-1}{\sigma}\right)s_{ft}\right) = \left(\frac{\sigma-1}{\sigma} - \left(\frac{\sigma-1}{\sigma}\right)s_{ft}\right)$, so that

$$\begin{aligned}
\mu_{ft} &= \frac{1}{1-\left(\frac{1}{\sigma} + \left(\frac{\sigma-1}{\sigma}\right)s_{ft}\right)} \\
\mu &= \frac{\sigma}{\sigma-1} \text{ if } s_{ft} = 0
\end{aligned}$$

■

The markup $\mu_{ft} = \frac{\sigma}{(\sigma-1)\left(1-s_{ft}\right)}$ is increasing in the firm's market share. Thus, the markup is itself a function of fundamentals:

$$\begin{aligned}
s_{ft} &= \frac{P_{ft}Q_{ft}}{E_t} = \frac{D_{ft}Q_{ft}^{1-\frac{1}{\sigma}}}{E_t} = \frac{D_{ft}A_{ft}^{1-\frac{1}{\sigma}}X_{ft}^{\gamma\left(1-\frac{1}{\sigma}\right)}}{E_t} && (17)\\
&= \frac{D_{ft}A_{ft}^{1-\frac{1}{\sigma}}}{E_t}\left(\frac{\gamma\left(1-\tau_{ft}\right)\left(1-\frac{1}{\sigma}\right)D_{ft}A_{ft}^{1-\frac{1}{\sigma}}}{C_{ft}\mu_{ft}\left(\frac{\sigma-1}{\sigma}\right)}\right)^{\frac{\gamma\left(1-\frac{1}{\sigma}\right)}{1-\gamma\left(1-\frac{1}{\sigma}\right)}} && (18)
\end{aligned}$$

so that

$$s_{ft}\left(\frac{\sigma-(\sigma-1)s_{ft}}{\sigma-(\sigma-1)s_{ft}-1}\right)^{\frac{\gamma\left(1-\frac{1}{\sigma}\right)}{1-\gamma\left(1-\frac{1}{\sigma}\right)}}=\frac{D_{ft}^{\frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}}A_{ft}^{\frac{1-\frac{1}{\sigma}}{1-\gamma\left(1-\frac{1}{\sigma}\right)}}}{E_t}\left(\frac{\gamma\left(1-\tau_{ft}\right)\left(1-\frac{1}{\sigma}\right)}{C_{ft}\left(\frac{\sigma-1}{\sigma}\right)}\right)^{\frac{\gamma\left(1-\frac{1}{\sigma}\right)}{1-\gamma\left(1-\frac{1}{\sigma}\right)}}$$

The LHS is increasing in s and the RHS is increasing in D and A, and decreasing in τ and C. Thus, s_{ft} and the markup are increasing in D and A, and decreasing in τ and C.

5 Appendix E: Persistence in Fundamentals and Endogenous Innovation

Firm choices depend on productivity components such as D and A, which is the sense in which these are fundamentals. We take them as given when a firm chooses its size, but note that our results should help guide future work, both theoretical and empirical, about the specific drivers of measured productivity. To further understand the nature of $TFPQ$ vs. demand shock, and potential mechanisms through which businesses accumulate each of them, we have studied the relationship between these fundamentals and reported innovation efforts. The Colombian Manufacturing Survey can be merged with the Innovation Survey at the level of the firm (tax ID). Since 2006 firms report number of innovations by type, defined by categories named "product", "process", and "organizational" innovation. They also report innovation expenditures, unfortunately not broken down in the same categories.

Results from our structural decomposition of growth show that, given fundamentals, high-fundamentals plants are being implicitly taxed while low-fundamentals plants are implicitly subsidized (by the environment, not necessarily by the government). Causality in the opposite direction is also likely: technical efficiency and product-plant appeal, while partly determined by exogenous stochastic dynamics (as in, e.g., Hopenhayn (1992) and Hopenhayn and Rogerson (1993)), partly also result from endogenous investments to improve performance (as in Acemoglu et. al., 2017, or Aw, Roberts and Xu, 2011). In the latter class of models, firms invest in future fundamentals (e.g. via R&D expenditure) to the extent that they expect high returns from such investments. High fundamentals plants should, therefore, invest more in a context with persistence in fundamentals. Since wedges make future profitability less dependent in fundamentals, they should reduce the incentive to invest given by high fundamentals, especially if wedges are negatively correlated with fundamentals (e.g. HK, 2014). Wedges may also have a direct effect on investment if, for instance, the presence of fixed costs of production implies that a subsidy directly increases the chances of surviving to enjoy the returns from R&D.

Table E1 presents an OLS analysis of the persistence in wedges, and the role of lagged wedges for the evolution of sales, output, $TFPQ$ and demand.[4] Wedges are standardized to facilitate interpretation. Both structural wedges (upper panel) and reduced form wedges exhibit considerable positive persistence, though less so in the case of structural wedges. This is consistent with structural wedges in part reflecting non-convex adjustment costs. Such cost generate a wedge that is correlated with fundamentals and that only persists up to the moment in which the benefit of adjusting overcomes its fixed cost.

Wedges are persistent, with an AR1 coefficient of 0.76 (column 1 of Table E1). As in models of endogenous fundamentals, contemporaneous fundamentals and wedges correlate with higher *lagged* wedges (higher implicit lagged subsidies), even after controlling for persistence in fundamentals, but wedges do not account for much variation in outcomes and fundamentals. For example, a one standard deviation increase in lagged structural wedges yields a 0.07 increase in $TFPQ$ and a 0.02 increase in demand. These are small effects relative to the standard deviations of $TFPQ$ and demand reported in Table 2 (0.78 and 0.84, respectively). [5] In turn, as hypothesized, the interaction effect between the lagged dependent variable and lagged structural wedges (negatively correlated with lagged fundamentals, as seen above) is negative. That is, while higher lagged structural wedges boost outcomes and fundamentals, they correlate with reduced persistence in outcomes and fundamentals. But, the interacted effects are also very small.

[4]As background, the standard deviation of reduced-form (uncorrelated) wedges lies in the same ball-park as the standard deviation of $TFPQ$ and demand (Table 2), while that of structural wedges doubles that (all in log points).

[5]Lagged wedges also have modest impact on current output and sales.

Table E1. Wedge and Fundamental persistence

VARIABLES	(1) Sales Wedge (subsidy)	(2) Output	(3) Sales	(4) TFPQ	(5) Demand shock
	Structural sales wedge (both orthogonal and correlated sources)				
Lagged Dependent Variable		0.985***	0.988***	0.936***	0.984***
		(0.001)	(0.001)	(0.001)	(0.001)
Lagged sales wedge (subsidy, standarized)	0.757***	0.038***	0.0434***	0.067***	0.024***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Lagged sales wedge (subsidy, standarized)*Lagged DV		-0.013***	-0.013***	-0.011***	-0.014***
		(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.035***	-0.012***	-0.021***	-0.015***	-0.013***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	114,231	114,231	114,231	114,231	114,231
R-squared	0.561	0.931	0.933	0.803	0.928
Sector*Time FE	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

6 Appendix F: details for the joint estimation of production and demand functions

As in proxy methods for the estimation of the production function, the joint estimation of production and demand is preceded by a first stage that ensures that $TFPQ$ can be proxied by an observable factor, in this case materials, which is conditionally monotonic in $TFPQ$. The free input M_{ft} is a function of $TFPQ_{ft}$, conditional on quasi-fixed inputs. The FOC for materials is

$$
\begin{aligned}
M_{ft} &= \frac{\phi(1-\tau_{ft})R_{ft}}{pm_{ft}}(1-1/\sigma) \\
&= \frac{\phi(1-\tau_{ft})P_{ft}Q_{ft}}{pm_{ft}}(1-1/\sigma) \\
M_{ft}^{1-\phi} &= \frac{P_{ft}A_{ft}K_{ft}^{\alpha}L_{ft}^{\beta}(1-\tau_{ft})(\phi\frac{\sigma-1}{\sigma})}{pm_{ft}}
\end{aligned}
$$

Within a sector, ϕ and σ display no variability. We thus re-write

$$\ln M_{ft} = h\left(\ln A_{ft}, \ln K_{ft}, \ln L_{ft}, \ln \frac{\overline{P^*_{ft}} P_{fB}}{\overline{PM^*_{ft}} PM_{fB}}, \ln \Lambda^Q_{ft}, \ln \Lambda^m_{ft},\ \ln s_{ft}\right)$$

We have included s_{ft} since we do not observe τ but know that all firm choices that ultimately feed into s_{ft} are a function of τ (we have measures for all the other variable terms in the material's FOC). In particular, we condition on a flexible polinomial on s_{ft} rather than τ_{ft}. Furthermore, we have used $P_{ft} = \overline{P^*_{ft}} P_{fB} \left(\Lambda^Q_{ft}\right)^{\frac{1}{\sigma-1}}$ and $pm_{ft} = \overline{PM^*_{ft}} PM_{fB} \left(\Lambda^M_{ft}\right)^{\frac{1}{\sigma-1}}$. Inverting, we obtain

$$\ln A_{ft} = h^{-1}\left(\ln M_{ft}, \ln K_{ft}, \ln L_{ft}, \ln \frac{\overline{P^*_{ft}} P_{fB}}{\overline{PM^*_{ft}} PM_{fB}}, \ln \Lambda^Q_{ft}, \ln \Lambda^m_{ft}, \ln s_{ft}\right) \equiv h^{-1}\left(\overrightarrow{W}\right).$$

Incorporating this expression, recognizing that Q_{ft} is subject to measurement error and other shocks not observed by either the econometrician or the firm at the time of making input choices, and denoting by $\widehat{Q_{ft}} = Q_{ft}\varepsilon_{ft}$ measured Q_{ft}, we write:

$$\begin{aligned}
\widehat{Q_{ft}} &= \alpha \ln K_{ft} + \beta \ln L_{ft} + \phi \ln M_{ft} + h^{-1}\left(\overrightarrow{W}\right) + \varepsilon_{ft} \\
&\quad \text{so that} \\
\widehat{Q^*_{ft}} &= \alpha \ln K_{ft} + \beta \ln L_{ft} + \phi \ln M_{ft} - \frac{1}{\sigma-1}\ln \Lambda^Q_{ft} + \frac{\phi}{\sigma-1}\ln \Lambda^M_{ft} \quad (19) \\
&\quad + h^{-1}\left(\overrightarrow{W}\right) + \varepsilon_{ft}
\end{aligned}$$

where ε_{ft} is measurement error, and the "$*$" refers to the fact that we are estimating the transformed $Q^*_{ft} = \frac{R_{ft}}{P^*_{ft}}$ rather than $Q_{ft} = \frac{R_{ft}}{P_{ft}}$.

In the first stage we proxy productivity and eliminate measurement error by estimating 19 through a flexible third-degree polinomial $\varphi^*\left(\overrightarrow{W}\right)$ estimated via OLS and obtaining the predicted $\widehat{\varphi}^*\left(\overrightarrow{W}\right)$.

We then estimate the system of demand and production functions replacing $\ln Q^*_{ft}$ with $\varphi^*\left(\overrightarrow{W}\right)$ in the production function. We use GMM methods and rely on the moment conditions presented in the main text for identification. Our estimates of production coefficients are initialized at the respective

OLS estimates of the production function augmented with Λ^Q_{ft} and Λ^M_{ft} regressors (coefficients for Λ^Q_{ft} and Λ^M_{ft} also freely estimated by OLS). Our σ estimate is initialized through an IV estimation of demand function, where the instrument for Q is the residual from the OLS production function. The IV procedure follows the spirit of Foster et al (2008), though only for initialization.

We obtain a negative elasticity of production to labor in sector 321. We assign to 321 factor elasticities in production and elasticities of substitution equal to the averages of these parameters for the other sectors in activity 32 (i.e. sectors 322, 323 and 324).

7 Appendix G: Variance decomposition

This appendix explains the structural and reduced form variance decompositions presented in Figures 5 and 6. We follow a two stage procedure, similar to that in Hottman et al. (2016).

7.1 Structural decomposition

The structural decomposition for sales and sales growth is guided by:

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

$$\frac{R_{ft}}{R_{f0}} = \left(\frac{d_{ft}}{d_{f0}}\right)^{\kappa_1} \left(\frac{a_{ft}}{a_{f0}}\right)^{\kappa_2} \left(\frac{pm_{ft}}{pm_{f0}}\right)^{-\phi\kappa_2} \left(\frac{w_{ft}}{w_{f0}}\right)^{-\beta\kappa_2} \left(\frac{\mu_{ft}}{\mu_{f0}}\right)^{-\gamma\kappa_2} \left(\widehat{\chi}_t \chi_{ft}\right)^{1-\frac{1}{\sigma}} \tag{20}$$

1. Guided by the above equation, we obtain $\ln \chi_{ft}$ as a residual from the following equation:

$$\ln \frac{R_{ft}}{R_{f0}} = \beta_D \ln\left(\frac{d_{ft}}{d_{f0}}\right) + \beta_A \ln\left(\frac{a_{ft}}{a_{f0}}\right) + \beta_\mu \ln \frac{\mu_{ft}}{\mu_{f0}} + \beta_M \ln\left(\frac{pm_{ft}}{pm_{f0}}\right) + \beta_w \ln\left(\frac{w_{ft}}{w_{f0}}\right) + \ln\left(\chi_{ft}\right)^{\left(1-\frac{1}{\sigma}\right)} \tag{21}$$

where $\beta_D = \kappa_1$; $\beta_A = \kappa_2$; $\beta_\mu = -\gamma\kappa_2$; $\beta_M = -\phi\kappa_2$; $\beta_w = -\beta\kappa_2$; $\kappa_1 = \frac{1}{1-\gamma\left(1-\frac{1}{\sigma}\right)}$; $\kappa_2 = \left(1-\frac{1}{\sigma}\right)\kappa_1$. We calculate these parameters using our estimates of factor elasticities in technology and the elasticity of substitution.

Because we use these parameters that stem from the structure of the model, we label the residual as a "structural" wedge. The fundamentals d_{ft}, a_{ft}, pm_{ft} and w_{ft} correspond to the idiosyncratic components of demand, technology and input price shocks, estimated as already described ($D_{ft} = D_t d_{ft}$ and so on).

2. We then estimate the following equations:

$$\begin{aligned}
\beta_D \ln\left(\frac{d_{ft}}{d_{f0}}\right) &= \rho_{0,D} + \rho_D \ln\frac{R_{ft}}{R_{f0}} + \nu_{ft,D} \\
\beta_A \ln\left(\frac{a_{ft}}{a_{f0}}\right) &= \rho_{0,A} + \rho_A \ln\frac{R_{ft}}{R_{f0}} + \nu_{ft,A} \\
\beta_\mu \ln\left(\frac{g\left(s_{ft}\right)}{g\left(s_{f0}\right)}\right) &= \rho_{0,\mu} + \rho_\mu \ln\frac{R_{ft}}{R_{f0}} + \nu_{ft,A} \\
\beta_M \ln\left(\frac{pm_{ft}}{pm_{f0}}\right) &= \rho_{0,M} + \rho_M \ln\frac{R_{ft}}{R_{f0}} + \nu_{ft,M} \\
\beta_w \ln\left(\frac{w_{ft}}{w_{f0}}\right) &= \rho_{0,w} + \rho_w \ln\frac{R_{ft}}{R_{f0}} + \nu_{ft,w} \\
\ln\widehat{\chi_{ft}} &= \rho_{0,v} + \rho_v \ln\frac{R_{ft}}{R_{f0}} + \nu_{ft,v}
\end{aligned} \tag{22}$$

We now prove that the contribution of each fundamental to the variance of sales equals the ratio of its covariance with sales to the variance of sales multiplited by its structural parameter in equation 21. Also that, by the properties of OLS, the contribution of the different factors considered add up to 1. We conduct the proof for the two-covariance case for simplicity

For any given log-linear equation (such as 21):

$$Y_f = \beta_1 X_{1f} + \beta_2 X_{2f} + \varepsilon_i \tag{23}$$

If one estimates by OLS The set of equations

$$\beta_1 X_{1f} = \gamma_{1,0} + \gamma_1 Y_f + \nu_{1i} \tag{24}$$

$$\beta_2 X_{2f} = \gamma_{1,0} + \gamma_2 Y_f + \nu_{2i} \tag{25}$$

and

$$\varepsilon_f = \gamma_{\varepsilon,0} + \gamma_\varepsilon Y_f + \nu_{\varepsilon f} \tag{26}$$

The estimated parameters for $j = \{1, 2\}$ are:

$$\begin{aligned} \hat{\gamma}_j &= \frac{Cov(\beta_j X_{jf}, Y_f)}{Var(Y_f)} = \beta_j \frac{Cov(X_{jf}, Y_f)}{Var\,(Y_f)} \\ &= \beta_j Corr(X_{ji}, Y_f) \left(\frac{Var(X_{jf})}{Var\,(Y_f)} \right)^{\frac{1}{2}} \end{aligned}$$

Since $\varepsilon_f = Y_f - (\beta_1 X_{1f} + \beta_2 X_{2f})$, $\hat{\gamma}_\varepsilon$ can be re-written as:

$$\begin{aligned} \hat{\gamma}_\varepsilon &= \frac{Cov(Y_f - (\beta_1 X_{1f} + \beta_2 X_{2f}), Y_f)}{Var\,(Y_f)} \\ &= \frac{Var\,(Y_f) - \beta_1 Cov(X_{1f}, Y_f) - \beta_2 Cov(X_{2f}, Y_f)}{Var\,(Y_f)} = 1 - \hat{\gamma}_1 - \hat{\gamma}_2 \end{aligned}$$

Results for this decomposition are reported in Table 3 and Figure 3 of the main text. Figure 3 is reproduced in the top left panel of Figure G1. We conduct an analogous decomposition for output growth, following the corresponding equation in the main text, and report its results in the bottom left panel of Figure G1.

7.2 Reduced form decomposition

The reduced form decomposition follows a procedure analogous to the one just described, but where the first stage estimates an OLS coefficient for each fundamental rather than imposing the coefficients imposed by our structure. In particular, the first stage estimates by OLS

$$\begin{aligned} \ln \frac{R_{ft}}{R_{f0}} &= \beta^r_D \ln\left(\frac{d_{ft}}{d_{f0}}\right) + \beta^r_A \ln\left(\frac{a_{ft}}{a_{f0}}\right) + \beta^r_\mu \ln \frac{\mu_{ft}}{\mu_{f0}} \\ &\quad + \beta^r_M \ln\left(\frac{pm_{ft}}{pm_{f0}}\right) + \beta^r_w \ln\left(\frac{w_{ft}}{w_{f0}}\right) + \ln\left(\chi_{ft}\right)^{\left(1-\frac{1}{\sigma}\right)} \end{aligned} \tag{27}$$

where the "r" index in each coefficient stands for "reduced form". Once OLS estimates of each of these coefficients are obtained, the second stage is implemented as in the structural decomposition, replacing each β_x with β_x^r, where x stands for any fundamental. Results of this decomposition are reported in Figure G1, right panels.

7.3 Decomposition by ages

To conduct the decomposition by ages, we expand equations 21 and 22 to include interactions with the different age groups. Suppose there are two mutualy exclusive groups: B and C. We redefine the equation 21 as:

$$Y_f = \beta_1 X_{1f} + \beta_2 X_{2f} + \varepsilon_i \tag{28}$$

$$\ln \frac{Q_{ft}}{Q_{f0}} = \beta_{1,C} X_{1f} d_{Cf} + \beta_{1,B} X_{1f} d_{Bf} \tag{29}$$

$$+\beta_{2,C} X_{2f} d_{Cf} + \beta_{2,B} X_{2f} d_{Bf} + \varepsilon_i \tag{30}$$

where $d_{Cf} = 1$ if f belongs to group C (say, an age), and everything else as defined previously.

The new decomposition equation for, say, X_1 will be given by:

$$\beta_{1,C} X_{1f} d_{Cf} + \beta_{1,B} X_{1f} d_{Bf} = \gamma_{C1} Y_f d_{Cf} + \gamma_{B1} Y_f d_{Bf} + \nu_{1f} \tag{31}$$

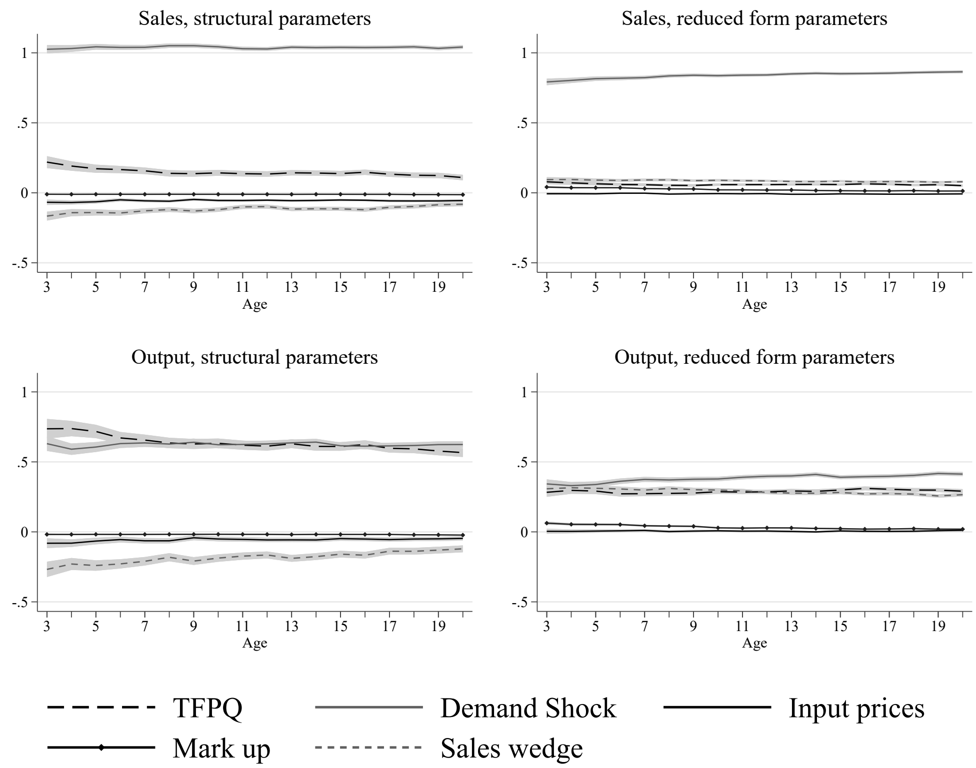
$$\varepsilon_f = \gamma_{C\varepsilon} Y_f d_{Cf} + \gamma_{B\varepsilon} Y_f d_{Bf} + \nu_{\varepsilon f} \tag{32}$$

Just as before $\hat{\gamma_{C1}} + \hat{\gamma_{C\varepsilon}} = \hat{\gamma_{B1}} + \hat{\gamma_{B\varepsilon}} = 1$.

8 Appendix H: Selection

By construction we focus on survivor growth: growth from birth to age a of plants that have survived to age a. However, because we are able to follow life cycle growth directly at the plant level–by contrast to cross sectional comparisons of cohorts–the usual concern that selection drives average growth because size at the initial age is biased downwards by exits-to-be does not apply. We compare plant i's size at age a to i's own birth size. It is the case, still, that plants that eventually exit may grow slower than others before they exit and, in that sense, even true life-cycle average growth is affected by selection: if the exiting plant had instead continued to the following age, average growth would have been lower. Figure H1 illustrates that this is

Figure G1: Life cycle growth variance decomposition by age



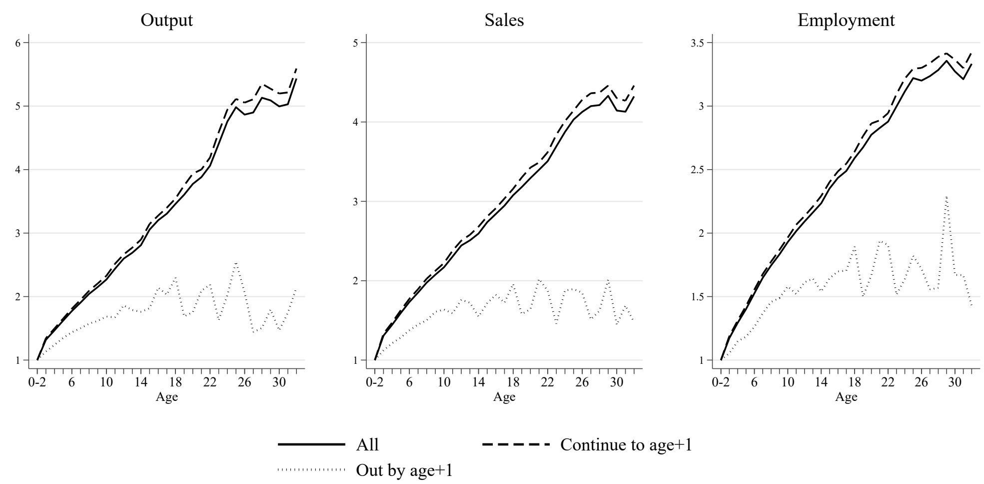
Idiosyncratic components

indeed the case, since the life-cycle growth of plants that exit in the next period does depart significantly, downwards, from that of continuers. But, this growth of plants that exit only affects marginally the overall average. That is, the average patterns described in the previous paragraph are mainly driven by continuous plants (plants of age t that continue on to age $t+1$). Still, in this section, we also explore how fundamentals relate to selection vs. continuer growth.

Figure H2 illustrates average growth of fundamentals separately for plants that continue for at least one additional year and those that exit the following year. The most noteworthy difference is much poorer growth in demand shocks for plants about to exit compared to those that will continue, sugges-

Figure H1: Life cycle growth

Current to initial



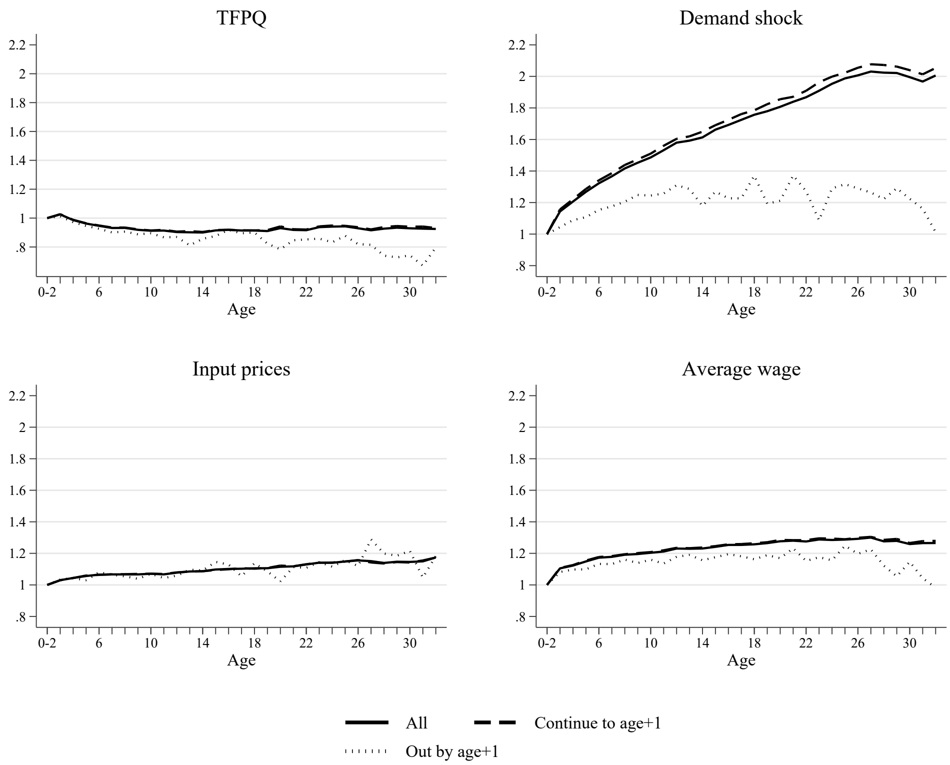
Idiosyncratic components

tive of demand side fundamentals being particularly important determinants of exit.

Figure H3 further carries our decomposition of drivers of output growth for these two groups of plants. We present three-year moving averages because the patterns for plants about to exit are noisy.[6] Fundamentals still play an important role for exiters in explaining their growth from birth to the moment in which they are about to exit. Despite demand shocks being the dimension where most marked differences are observed between exits-to-be and continuers, especially for young ages (Figure H2), $TFPQ$ tends to play a slightly more significant (at least more sustained) role in explaining growth up to age t for plants about to exit compared to continuers, likely capturing the extremely poor $TFPQ$ behavior of exits-to-be.

[6]Since each point (age) in a figure for plants about to exit contains the plants that will exit at age+1, the plants included in a given line are different for each age. This explains the noisy patterns.

Figure H2: Life cycle growth of fundamentals: exiters vs. continuers

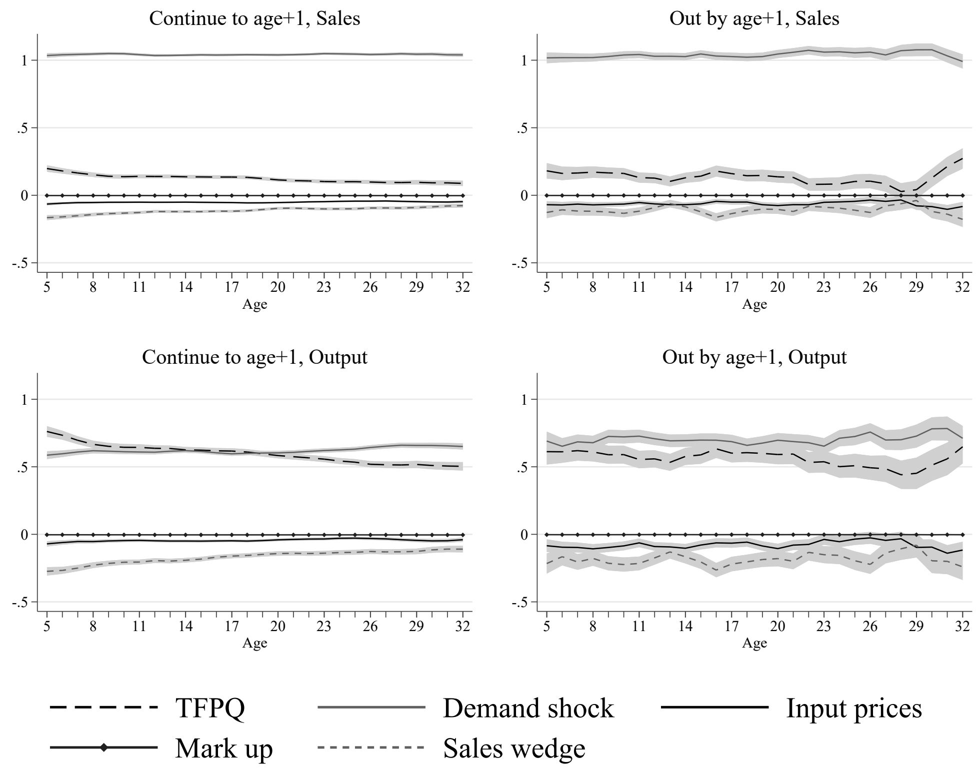


Idiosyncratic components

8.1 Appendix I: The value of Quality Adjustment

UPI plant price indices adjust real output for intra-firm quality/appeal differences. Moreover, in the context of UPI prices, sales measure output that is additionally adjusted for cross-plant quality differences. We now compare results to what would be obtained under two alternatives to price measurement. First, we implement a "statistical" approach based on Törnqvist indices for a constant basket of goods within the plant or, alternatively, on the divisia price index that allows that basket to change and uses average t, $t-1$ expenditure shares. We implement a second alternative approach, using prices based

Figure H3: Life cycle growth decomposition by age



Idiosyncratic components

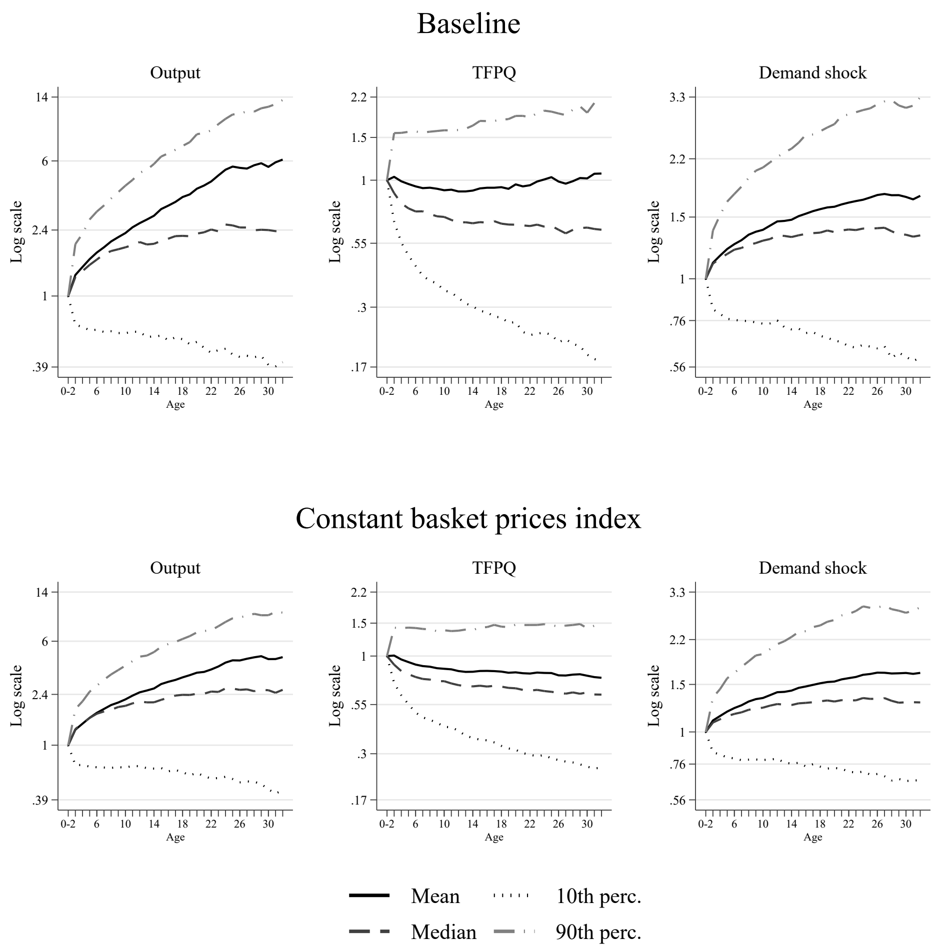
on the insights offered by Sato (1976), Vartia (1976) and Feenstra (1994). The Sato-Vartia approach is economically motivated but keeps appeal shifters and baskets of goods constant over two consecutive periods, implying a much slower quality adjustment for both continuing products and those that enter and exit. The Feenstra adjustment for changing varieties incorporated into our UPI approach can also be added to the Sato-Vartia index to adjust for changing baskets of goods over consecutive periods (it was, in fact, originally implemented by Feenstra, 2004, within the Sato-Vartia approach). The UPI, meanwhile, allows for both changing baskets of goods and varying appeal shifters, dimensions of flexibility which respectively deal with the "consumer valuation bias" and the "variety bias" (Redding and Weinstein, 2020). (For a detailed discussion of each of these alternatives, contrasted with the UPI, see Appendix A, and Redding and Weinstein, 2020).

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

If the quality mix within the plant improves over time, plant-level quality adjusted price indices will grow less than unadjusted ones, as a result yielding less deflated output growth and less $TFPQ$ growth. This composes with overall plant quality growth to imply economically motivated aggregate prices that grow less than unadjusted ones. Not properly adjusting plant-level prices for quality changes biases estimated idiosyncratic output and technical efficiency growth downwards because such estimates will ignore the part of price increases that reflects increasing valuation of goods and the services of plants to their costumers, and thus mistakenly translate those price increases into welfare decreases for given expenditure. Figure I1 shows that output and $TFPQ$ growth are dampened when revenue is deflated with price indices that do not adjust for quality.

Figure I1: Distribution of output and fundamentals life cycle growth

Alternative price indices



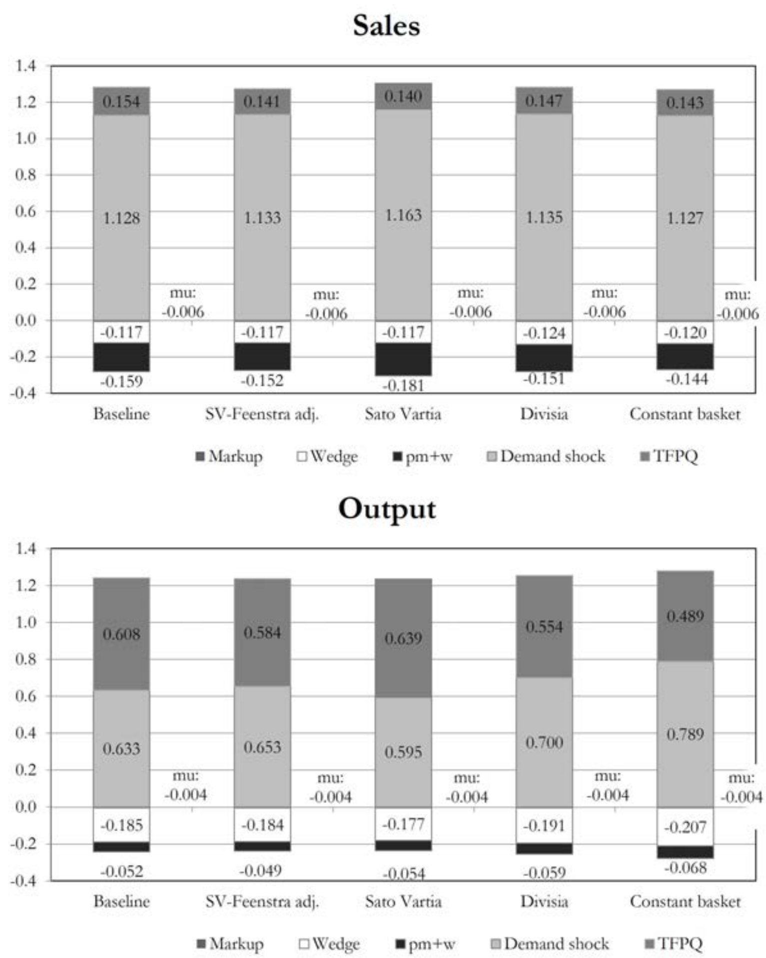
Idiosyncratic components

Figure I2 displays growth decomposition results using alternative price indices. Adjusting output for quality changes assigns a larger weight to technical efficiency, $TFPQ$, and a lesser role to demand or wedges, in explaining output life cycle growth . While with constant-weights-Törnqvist-indices $TFPQ$ and demand are estimated to contribute roughly equally to growth,

$TFPQ$ is assigned progressively more relative importance as one moves to the Sato-Vartia and then to the UPI approaches, in particular in the output decomposition. Quality adjusting prices matters much more in decomposing output than for sales because, beyond the more precise measurement of fundamentals when quality is adjusted for, the measure of output itself is affected by price indices. In addition, quality adjusting materials input prices plays more of a modest role than quality adjusting output prices.

Figure I2: Life-cycle growth variance decomposition

Structural parameters, alternative price indices



9 Appendix K: Hottman, Redding and Weinstein framework accounting explicitly for wedges

Our framework closely follows the modeling of the demand side in Hottman, Redding and Weinstein (2016). On the cost side, however, they model total costs rather than efficiency and input prices individually, and do so at the product level rathe than the firm level. They also abstract from wedges. Expanding HRW's framework to include wedges explicitly, and focusing on the case of uniproduct firms where their approach and ours are equivalent, the firm solves:

$$\underset{Q_{ft}}{Max} \left(1-\tau_{ft}\right) P_{ft} Q_{ft} - CT_{ft}\left(Q_{ft}\right)$$

where $CT_{ft}\left(Q_{ft}\right)$ is total cost as a function of output. Profit maximization leads to first orden condition $\left(P_{ft} + Q_{ft}\frac{dP_{ft}}{dQ_{ft}}\right) = \frac{\frac{\partial CT_{ft}}{\partial Q_{ft}}}{\left(1-\tau_{ft}\right)}$, so that at the optimum

$$\mu_{ft} = \frac{P_{ft}}{\frac{\partial CT_{ft}}{\partial Q_{ft}}\left(1-\tau_{ft}\right)^{-1}} \tag{33}$$

. The associated optimal markup is given by (see appendix D):

$$\mu_{ft} = \frac{1}{1-\left(\frac{1}{\sigma}+\left(\frac{\sigma-1}{\sigma}\right)s_{ft}\right)} \tag{34}$$

Moreover, our demand structure is the same as in HRW. The implied demand function in the case of a uniproduct firm is:

$$Q_{ft} = d_{ft}^{\sigma}\left(\frac{P_{ft}}{P_t}\right)^{-\sigma}\frac{E_t}{P_t} \tag{35}$$

$$or$$

$$R_{ft} = d_{ft}^{\sigma}\left(\frac{P_{ft}}{P_t}\right)^{1-\sigma} E_t \tag{36}$$

$$\frac{P_{ft}}{P_t} = d_{ft}^{\frac{\sigma}{\sigma-1}} s_{ft}^{\frac{1}{1-\sigma}} \tag{37}$$

where $R_{ft} = P_{ft}Q_{ft}$ is firm sales and $s_{ft} = \frac{R_{ft}}{E_{ft}}$ is the firm's share in aggregate (sector) sales. Equation 35 is HRW's equation (5) for the uniproduct case (where $d_{ft} = \varphi_{ft}^{\frac{\sigma-1}{\sigma}}$ and φ_{ft} is the notation used in HRW. Equation 37 is obtained by direct manipulation of 36.

Replacing the optimal markup rule 33 into 36 HRW decompose firm sales into:

$$R_{ft} = d_{ft}^{\sigma} \frac{E_t}{P_t^{1-\sigma}} \left(\mu_{ft} \frac{\frac{\partial CT_{ft}}{\partial Q_{ft}}}{1 - \tau_{ft}} \right)^{1-\sigma} \tag{38}$$

which is equivalent to HRW's equation (16). To see the equivalence, notice that in the uniproduct case $\frac{\partial CT_{fjt}}{\partial Q_{fjt}} = \frac{\partial CT_{ft}}{\partial Q_{ft}}$ (where j is a product and HRW have denoted by $\widetilde{\gamma}_{ft}$ the average marginal cost across products of a firm), and that $d_{ft} = \varphi_{ft}^{\frac{\sigma-1}{\sigma}}$. Firm sales variability can thus be decomposed into variation attributable to : 1) an aggregate component; 2) firm idiosyncratic demand d_{ft}; 3) firm markup; 4) a distortion-adjusted marginal cost $\frac{mc_{ft}}{(1-\tau_{ft})}$.

HRW's empirical procedure is as follows:

1) Estimate the demand function 35, in differences with respect to aggregates and over time, to obtain σ and decompose price (observable) into d_{ft} (not observable) and s_{ft} (observable).

2) Estimate the markup μ_{ft} based on observables, using 34.

3) With these components decompose the idiosyncratic variation of sales from equation 38 into the contributions of d_{ft}, μ_{ft} and the residual component: $\frac{\frac{\partial CT_{ft}}{\partial Q_{ft}}}{(1-\tau_{ft})}$. This is a distortion-adjusted marginal cost component, which HRW do not further decompose into its $\frac{\partial CT_{ft}}{\partial Q_{ft}}$ and $(1 - \tau_{ft})$ components.

10 Appendix L: Supplementary results

Production function coefficients by sector are shown in Table L1.

Counterfactual analysis of the impact of life cycle of wedges and fundamentals on welfare.

Table L1. Factor and demand elasticities by sector

Sector	β	α	ϕ	σ_w	σ	σ_w/σ	γ	$\gamma(1-1/\sigma)$
311	0.27	0.08	0.70	3.38	1.86	1.82	1.06	0.49
313	0.28	0.04	0.67	4.34	2.18	1.99	0.99	0.54
321	0.19	0.11	0.63	4.02	2.24	1.79	0.94	0.51
322	0.16	0.11	0.65	4.63	2.55	1.81	0.92	0.56
323	0.22	0.13	0.58	3.21	1.77	1.81	0.93	0.41
324	0.20	0.10	0.67	4.22	2.40	1.76	0.97	0.56
331	0.25	0.13	0.59	3.19	1.75	1.83	0.97	0.42
332	0.29	0.07	0.62	2.87	1.58	1.82	0.98	0.36
341	0.36	0.11	0.57	2.15	1.20	1.79	1.04	0.17
342	0.53	0.21	0.26	2.52	1.40	1.80	1.01	0.29
351	0.43	0.28	0.37	4.73	2.66	1.78	1.08	0.68
352	0.39	0.19	0.49	3.35	1.83	1.82	1.07	0.49
355	0.59	0.09	0.39	4.17	2.28	1.83	1.07	0.60
356	0.38	0.14	0.54	2.52	1.38	1.82	1.06	0.29
362	0.60	0.40	0.10	3.01	1.69	1.78	1.09	0.45
369	0.50	0.21	0.37	4.51	2.51	1.80	1.07	0.65
371	0.57	0.17	0.48	3.11	1.67	1.86	1.22	0.49
381	0.29	0.13	0.54	2.63	1.45	1.81	0.95	0.30
382	0.40	0.08	0.50	3.33	1.81	1.83	0.98	0.44
383	0.32	0.08	0.60	3.29	1.85	1.78	1.00	0.46
384	0.28	0.10	0.62	4.49	2.45	1.84	0.99	0.59
385	0.42	0.21	0.33	3.65	2.00	1.82	0.96	0.48
390	0.33	0.17	0.49	3.21	1.77	1.81	0.99	0.43
Average	0.36	0.15	0.51	3.50	1.92	1.82	1.02	0.46

Table L2: Counterfactual welfare - relative to HK efficient welfare. Average sector, sector-level parameters

		Specific plant attributes set to constant mean value (1)	Specific attributes of high life-cycle growth plants (>P75) set to average life cycle growth for the rest (2)	Specific attributes of low life cycle growth plants (<P25) set to average life cycle growth for the rest (3)
Benchmark: Actual to HK Efficient Welfare		0.496	0.496	0.496
Plant attribute set to counterfactual level	D+TFPQ (TFPQ_HK)	0.122	0.207	0.416
	Demand Shock	0.093	0.159	0.433
	TFPQ	0.525	0.471	0.462
	Input prices + Markup	0.818	0.587	0.311
	Input prices	0.635	0.553	0.468
	Markup	0.630	0.511	0.302
	Sales wedge	0.587	0.539	0.468

Table L3. Decomposition of DeLoecker Markups

	Structural	Reduced
Markup	0.009	0.001
Sales wedge	0.345	0.280
Residual	0.646	0.719

Table L4: Sector classifications (3 digit ISIC)

Sector	Description	Observations
311	Food manufacturing (311 and 312).	24013
313	Beverage industries (313) and Tobacco indsutries (314).	2024
321	Manufacture of textiles.	6440
322	Manufacture of wearing apparel, except footwear.	16818
323	Manufacture of leather and products of leather, leather substitutes and fur, except footwar and wearing apparel.	2444
324	Manufacture of footwear, except vulcanized or moulded rubber or plastic footwear.	5929
331	Manufacture of wood and wood products, except furniture.	3250
332	Manufacture of furniture and fixtures, except primarily of metal.	7201
341	Manufacture of paper and paper products.	3075
342	Printing, publishing and allied industries.	7205
351	Manufacture of industrial chemicals.	2200
352	Manufacture of other chemical products (352); Petroleum refineries (353); Manufacture of miscellaneous products of petroleum and coal (354).	7157
355	Manufacture of rubber products.	1543
356	Manufacture of plastic products not elsewhere classified.	8694
362	Manufacture of pottery, china and eathenware (361) and Manufacture of glass and glass products (362).	1630
369	Manufacture of structural clay products.	4455
371	Basic metal industries (371 and 372).	1842
381	Manufacture of cutlery, band tools and general hardware.	10287
382	Manufacture of machinery except electrical.	7041
383	Manufacture of electrical machinery, apparatus, appliances and supplies.	3802
384	Manufacture of transport equipment.	3768
385	Manufacutre of professional and scientific, and measuring and controlling equipment not elsewhere classified, and of photographic and optical goods.	957
390	Other manufacturing industries.	3152

Table L5. Sector classifications for first 15 sectors at 3 digit CPC

Sector	Description	Observations
211	Meat and meat products	2685
212	Prepared and preserved fish	223
213	Prepared and preserved vegetables	279
214	Fruit juices and vegetable juices	127
215	Prepared and preserved fruit and nuts	807
216	Animal and vegetable oils and fats (216); Cotton linters (217); Oil-cake and other residues resulting from the extraction of vegetable fats or oils; flours and meals of oil seeds or oleaginous fruits, except those of mustard; vegetable waxes, except triglycerides; degras; residues resulting from the treatment of fatty substances or animal or vegetable waxes (218)	757
221	Processed liquid milk and cream	368
229	Other dairy products	1918
231	Grain mill products	2981
232	Starches and starch products; sugars and sugar syrups n.e.c	179
233	Preparations used in animal feeding	1019
234	Bakery products	8309
235	Sugar	324
236	Cocoa, chocolate and sugar confectionery	835
237	Pasta, macaroni, noodles, couscous and similar farinaceous products	484

Table L6. Distribution of 3 digit CPC sector sizes

	Min	P25	P50	P75	Max	Average
Observations in sector	32	305	484	1191	18322	1183