Decomposing growth over an establishment's. life cycle: the role of TFPQ, demand and input costs.*

Marcela Eslava[†] and John Haltiwanger[‡]

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE.

May 11, 2016

Abstract

This paper takes advantage of rich microdata on Colombian manufacturing establishments to decompose growth over an establishment's life cycle into that attributable to three fundamental sources of growth: physical productivity, price effects and input costs. Price effects capture both demand shocks (e.g. quality build-up) and idiosyncratic distortions, which we plan to further decompose by focusing on distortions that are conditional on either size or sector. Our approach essentially boils down to a micro-level growth accounting exercise. We find that growth over the life cycle of a Colombian manufacturing establishment is mainly driven by price effects, while physical productivity growth and input costs reductions contribute only marginally. This is true even within narrow sectors and size classes, which leads us to conclude that quality build-up is crucial for dynamic post-entry growth.

Keywords: Life cycle of plants; post-entry growth; TFPQ; demand

JEL codes: O47; O14; O39

*We thank Alvaro Pinzón for superb research assistance.

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

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

1 Introduction

This paper takes advantage of rich microdata on Colombian manufacturing establishments to decompose growth over an establishment's life cycle into that attributable to three fundamental sources of growth: physical productivity, price effects and input costs. Price effects capture both demand shocks (e.g. quality build-up) and idiosyncratic distortions, which we plan to further decompose by focusing on distortions that are conditional on either size or sector.

The growing availability of detailed firm and establishment level data has allowed researchers to dig into the empirical micro foundations behind sluggish aggregate growth in many low- and middle-income economies. A recent strand of the literature has focused on how businesses grow over their life cycle, uncovering patterns that suggest that less developed economies are characterized by business post-entry growth slower than that of startups in developed economies (Hsieh and Klenow, 2014; Buera and Fattal, 2014). These studies argue that slow post-entry growth can be attributed to slow growth of physical productivity—related to poor market selection or poor innovation–, in turn caused by poor market institutions that discourage, or fail to encourage, investments in productivity enhancement and healthy market selection.

Recent work by Foster, Haltiwanger and Syverson (2016), however, argues that a startup's ability to increase its demand may be even more important to ensure rapid growth than its ability to increase the physical efficiency of its production process. They observe that, for US manufacturers in a restricted set of sectors, employment growth from birth to maturity closely tracks demand growth (defined as growth in prices not explained by changes in physical productivity) but not growth in physical productivity (TFPQ). However, for the US the rich data on individual prices necessary to decompose profitability into its TFPQ and price components are only available for a restricted set of industries characterized by product homogeneity.

Establishment level data for the manufacturing sector in Colombia is uniquely rich. Since at least 1982, the Annual Manufacturing Survey has been recording information on all individual products produced by an establishment and all individual material inputs used by the establishment. Establishment level prices can be constructed using this information. Moreover, establishments in the survey (all manufacturing establishments, except for the micro ones) can be followed longitudinally, and the age of the establishment from the time of the start of its operations is reported.

We take advantage of these unique data to assess the relative importance of different plant-level "fundamentals" as determinants of growth over the life cycle of startups in Colombia's manufacturing industry. On the side of fundamentals we can measure TFPQ, output prices, and input prices at the plant level, an unusually rich set of measured fundamentals.

Our analysis contributes to the growing literature on post-entry growth by expanding the set of plant-level determinants of growth to which attention is paid, from the traditional focus on physical productivity to a more comprehensive view that recognizes the importance of demand, including building one's client base, introducing new products and emphasizing product quality and differentiation in existing product lines. In this aspect, we build on the ideas recently proposed by Foster et al. (2016), but expand the empirical reach by enriching the characterization of fundamentals, and widening the sectoral scope to all manufacturing sectors. This new focus on demand, as opposed to productivity, has potential crucial implications for policies aimed at fostering high growth entrepreneurship, where much of the focus has been on boosting efficiency rather than helping startups build their client base.

Following a life-cycle approach, we show that an average establishment in our sample doubles its birth level of employment by around age 12, and multiplies it by four by the time it is 25. Post-entry growth is faster in the most recent decade than in the 1980s. The two decades are separated by a wave of reforms that arguably transformed institutions in the country. But, there is wide dispersion in the patterns of growth across firms, with average growth driven by a small fraction of rapidly growing businesses. Post-entry growth in other outcomes–output, revenues, capital stock, scope, skill composition–follows overall patterns consistent with those just described, though with important quantitative differences. Output, revenues, and in particular the capital stock grow much more rapidly over the life cycle than employment does, while the magnitudes of growth in skill composition and number of products are similar to those observed in employment.

The paper proceeds as follows. We first explain the data used, in section 3. We then characterize growth over establishments' life cycle in terms of employment and other outcomes (Section 4). Section 2 then presents our conceptual approach to decompose outcome growth into fundamental sources. The approach we use to measure fundamentals is described in section 5. Results for our growth decomposition are presented in section 6. Section concludes 7.

2 Decomposing growth into fundamental sources

We start with a very simple static model of firm optimal behavior given firm fundamentals, to derive the relationship that should be observed between size growth and growth in fundamentals as a firm ages. Those fundamentals are the productivity of the firm's productive process (often termed TFPQ in the literature), the unit costs of its inputs, and a residual "price effect" that captures how much the firm is able to charge for a unit of product. Idiosyncratic distortions to the decisions of the firm, stemming for instance from regulations, will be for the time being subsumed in the generic price effects and the input costs, depending on their origin. The firm chooses its size optimally given these fundamentals. As a result, growth over its life cycle will be driven by growth in these different fundamentals. For consistency with the literature on business dynamics, we refer to a business as a "firm", even though the unit of observation for our empirical work is an establishment.

Though for simplicity we take growth of fundamentals as exogenous in describing this conceptual framework, it is clear that the firm may make investments to modify both the physical productivity of its production process and the prices the firm is able to charge at given productivity levels. The firm's efforts to increase prices may include investments in building its client base (Foster et al., 2014), or in adding new products or improving the quality of its pre-existing product lines (Atkeson and Burstein, 2010; Acemoglu et al., 2014). At a later stage in this project, we will also investigate such firm investments–some of which are observed in the data–and to what extent the actual evolution of fundamentals and, ultimately, of firm size can be attributed to them.

Consider a monopolistic competitor that produces output Q_{it} using a composite input X_{it} to maximize its profits, with technology

$$Q_{it} = A_{it} X^{\gamma} \tag{1}$$

A_{it} is the firm's physical total factor productivity, to which we will refer as TFPQ, and γ the returns to scale parameter. The firm is assumed to face a downward sloping (inverse) demand function given by

$$P_{it} = D_{it} Q_{it}^{-\varepsilon} \tag{2}$$

where D_{it} is the residual price effect. This residual price effect reflects both idiosyncratic demand for the plant's products, d_{it}, and distortions to profitability: $D_{it} = d_{it}\left(1 - \tau_{Ri}\right)$, where τ_{Ri} is a firm-specific revenue distortion.

The firm maximizes its profits, taking as given A_{it}, D_{it}, and potentially idiosyncratic unit costs of the composite input, C_{it}. Unit input costs faced by the firm may also be affected by firm-specific input price distortions: $C_{it} = c_{it}\left(1 + \tau_{ci}\right)$. The setup just described closely follows assumptions in Hsieh and Klenow's analyses of manufacturing plants (2009, 2014), except that we relax the assumption of constant returns to scale, and allow unit input costs to vary across plants even in the absence of distortions.

Profit maximization yields optimal input demand of $X_{it} = \left(\frac{\gamma(1-\varepsilon) D_{it} A_{it}^{1-\varepsilon}}{C_{it}}\right)^{\frac{1}{1-\gamma(1-\varepsilon)}}$. Denoting by Z_{i0} the value at birth of variable Z for firm i, growth over the life cycle of the firm can be attributed to growth in the different fundamentals:

$$\frac{X_{it}}{X_{i0}} = \left(\frac{D_{it}}{D_{i0}}\right)^{\delta_1} \left(\frac{A_{it}}{A_{i0}}\right)^{\delta_2} \left(\frac{C_{it}}{C_{i0}}\right)^{\delta_3} \tag{3}$$

where $\delta_1 = \frac{1}{1-\gamma(1-\varepsilon)}$, $\delta_2 = (1-\varepsilon)\,\delta_2$ and $\delta_3 = -\delta_1$. These parameters are constant across firms that produce with the same technoloy and face the same demand elasticity. Equation (3) decomposes growth in (log) firm size into the contribution of (log) fundamentals, and is one of the key focuses of this paper.

One can easily derive analogous decompositions in terms of employment or output, as opposed to the composite input X, and explore the role of the specific dimension of input prices that is observed in the data: the price of material inputs. In particular:

$$\frac{L_{it}}{L_{i0}} = \lambda_{it} \left(\frac{D_{it}}{D_{i0}}\right)^{\lambda_1} \left(\frac{A_{it}}{A_{i0}}\right)^{\lambda_2} \left(\frac{pm_{it}}{pm_{i0}}\right)^{\lambda_2} \tag{4}$$

where pm_{it} is price of material inputs (measurable at the plant level using information from the survey); $\lambda_1 = \frac{1}{1-\gamma(1-\varepsilon)}$; $\lambda_2 = (1-\varepsilon)\lambda_1$; $\lambda_3 = -\phi(1-\varepsilon)\lambda_1$; and λ_{it} is a residual capturing cross-plant dispersion in the evolution of wages and capital unit costs.[1]. Similarly:

$$\frac{Q_{it}}{Q_{i0}} = \chi_{it}\left(\frac{D_{it}}{D_{i0}}\right)^{\gamma\lambda_1}\left(\frac{A_{it}}{A_{i0}}\right)^{1+\gamma\lambda_2}\left(\frac{pm_{it}}{pm_{i0}}\right)^{\gamma\lambda_2} \tag{5}$$

Hsieh and Klenow (2014) have pioneered the empirical exploration of growth over a business' life cycle, showing that average size growth over the life cycle of a manufacturing establishment varies considerably between the US, India and Mexico, with US plants growing much more dynamically than those in Mexico or India, and have attributed those differences to variability in idiosycratic distortions across those economies. We aim to establish a series of facts related to those uncovered by Hsieh and Klenow: 1) The degree of dispersion in growth across plants. 2) The degree to which size increases over a firm's life cycle are attributable to increases in physical productivity, as compared to input costs and price factors. 3) The degree to which dispersion in size growth across plants can be attributed to dispersion in growth in each of these fundamentals. In a companion paper (Eslava and Haltiwanger, 2016) we attempt to uncover the degree to which dispersion in input costs and demand shocks is attributable to distortions from a specific regulation, taking advantage of cross-sectional variability in the dispersion of import tariffs that affect a plant's inputs and those that affect its products.

We also investigate how such dispersion in the growth of fundamentals relates to dispersion in the average product of inputs, as well as in TFPR, two concepts highlighted in Hsieh and Klenow's work. $TFPR$ has been defined by Foster et al (2010) as $TFPR_{it} = P_{it}A_{it}$. Though in Hsieh and Klenow's original work $TFPR$ and the average product of inputs are equivalent, under the generalized production function we are assuming they are not, as shown by Haltiwanger (2016). In particular, and denoting the firm's revenue by R_{it}, $\frac{R_{it}}{X_{it}} = \frac{P_{it}A_{it}X_{it}^{\gamma}}{X_{it}}$. It is clear that $\frac{R_{it}}{X_{it}} = TFPR_{it}$ only if $\gamma = 1$. Under our current assumptions, further, one can show that, in the optimum $\frac{R_{it}}{X_{it}} = \frac{C_{it}}{\gamma(1-\varepsilon)}$ while $TFPR_{it} = \left[\frac{C_{it}^{\varepsilon}D_{it}^{1+\varepsilon\gamma}}{\left(\gamma(1-\varepsilon)A_{it}^{1+\varepsilon\gamma}\right)^{\varepsilon}}\right]^{\frac{1}{1-\gamma(1-\varepsilon)}}$. Therefore, while dispersion in revenue per input is driven solely by dispersion in input costs, and muted when such costs are homogeneous across plants, dispersion in $TFPR$ responds to dispersion not only in unit costs but also in TFPQ and the residual price effect.

3 Data

We use data from the Colombian Annual Manufacturing Survey (AMS) from 1982 to 2012. The survey, collected by the Colombian official statistical bureau

[1] $\lambda_{it} = \left(\frac{\omega_{it}}{\omega_{i0}}\right)^{\frac{1}{1-\varepsilon}-(\alpha+\phi)}\left(\frac{r_{it}}{r_{i0}}\right)^{-\alpha}$

DANE, covers all manufacturing establishments 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. The unit of observation in the survey is the establishment. An establishment is a specific physical location where production occurs.

Each establishment is assigned a unique ID that allows us to follow it over time. Since a plant's ID is not modified with changes in ownership, such changes are not mistakenly identified as births and deaths. Plant IDs in the survey were modified in 1992 and 1993. In theory, there is a one-to-one mapping between the old and the new codification. We use the official correspondence that maps one into the other to follow establishments over that period.[2]

Surveyed establishments are asked to report their level of production and sales, as well as their use of employment and other inputs, and their purchases of fixed assets. Sector ID's are also reported, at the 3-digit level of the ISIC revision 2 classification.[3] Since 2004, respondents are also asked about their investments in innovation, with bi-annual frequency.

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 great level of detail. Plants report separately each 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 individual inputs and products, plants report separately quantities and values used or produced, so that plant-specific prices can be computed for both inputs and outputs. We thus directly observe idiosyncratic input costs. Furthermore, by taking advantage of plant-specific prices, we can use produce measures of productivity based on physical output, and also estimate demand shocks. Details on how we go about these estimations are provided below.

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 actual age, and follow a subsample from birth. We denominate tat subsample, composed of the establishments we observe from birth, as the *life cycle sample*.

With respect to studies that rely on data from economic censuses, one clear limitation of our approach is that we only observe a fraction of establishments from birth, and that fraction is selected: corresponds to establishments born at or beyond a given size. The upside is that we can follow each establishment

[2] It is the case, however, that we see higher "exit rates" in 1991 compared to other years, and also higher entry in 1993. DANE does report having undertaken efforts to improve actual coverage in 1992, which may explain higher entry in 1993. But, it is also likely that the IDs used to map one codification into the others were not fully recorded, leading to failure to follow some continuing plants over that period. Even in those cases, however, the "initial year of operation" variable correctly indicates the actual birth year for each observation.

[3] 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 we are able to consistently assign sector codes over the period.

longitudinally, and do it at higher frequencies (annual, rather that inter-census). We attempt to deal with selection biases using a variety of approaches, from controlling for initial sizes to contrasting our findings for plants observed from birth to analogous figures for all of the other plants in the manufacturing survey. It is important to highlight, also, that the Colombian manufacturing survey is a census of all manufacturing non-micro establishments.

4 Growth over the life cycle

We start by characterizing "outcome" growth over the life cycle of a manufacturing establishment (the left hand side of our growth decomposition). We first follow the recent literature, characterizing employment growth over the life cycle of the establishments that we observe from birth. At the end of this section we include other plant outcomes, such as production, skill composition, and scope. Because of our interest on characterizing longitudinal patterns within plants, our focus is on the sample of plants that we follow from birth, which we call the "life cycle" sample.

4.1 Average life cycle growth

To characterize employment growth for the average establishment in our sample, we estimate a full set of ϕ_{age} coefficients in equation:

$$\frac{L_{it}}{L_{i0}} = \alpha_t + \alpha_s + \sum_{age=3}^{age=20+} \phi_{age} d_{age,i,t} + \varepsilon_{it} \tag{6}$$

where $\frac{L_{it}}{L_{i0}}$ is the ratio between plant $i's$ employment level in year t and the level at plant's birth; $d_{age,i,t}$ is a dummy variable that takes the value of 1 if plant i has age age in year t; and ε_{it} is an estimation error. We control for (three-digit) sector effects and aggregate time effects. We define age as the difference between the current year, t , and the year when the plant began its operations, and define the plant's employment level at birth L_{i0} as the average employment it reported in ages 0 to 2. By averaging over the first few years in operation we deal with measurement error coming from partial-year reporting (if the plant, for instance, was in operation for only part of its initial year) or with fuzzy definitions of a plant's start date.

Figure 1 presents the coefficients associated with different ages in the estimated equation (6). The average establishment in our life cycle sample doubles its size by around age 12, and multiplies it by four by the time it reaches around 25 years. Estimation results for equation (6) are also shown in Table 1, though only up to age 12 to keep the table manageable.

Average growth over the life cycle is driven by both within-plant growth for surviving establishments and by exit. With healthy market selection, it should be the less productive and also smaller plants that are exiting, driving average size up for older ages. Figure 2 indicates that, indeed, the exiting businesses tend to have grown less than their surviving counterparts by the time they exit. For any age, the dotted line in this figure represents $\frac{L_{it}}{L_{i0}}$ for establishments that will exit in the following period, while the dashed line shows the analogous figure for establishments that continue to operate in the following age.

Figure 1 also shows the corresponding cross-sectional patterns. These are calculated by dividing the average employment level of plants of a given age in the overall sample (i.e. not only plants in our life cycle sample) by the average size of plants at birth. The estimated growth dynamics are considerably dampened in the cross sectional approach compared to the longitudinal one, which hints at the importance of being able to follow individual units longitudinally. Cross-sectional comparisons of cohorts, by contrast to our life cycle approach, implicitly give more weight to plants born larger, which results in the dampened cross-sectional dynamics observed in Figure 1.[4]

To give an idea of where these patterns fit in the international spectrum, Figure 3 compares the cross sectional patterns with US cross sectional patterns, calculating the two in an analogous manner. The US data comes from the publicly available information in the Bureau of the Census' Business Dynamics Database, which shows average size for given age categories. We focus solely on statistics for the manufacturing sector and cut the Colombian data in analogous categories for comparability. The period is limited to 2002-2012, which is the time span for which we can assign age tags to all categories in the US data.

The growth speed of the average US establishment basically doubles that in Colombia. For instance, employment in the 16-20 category more than doubles that of the initial category in the US, while it is at about 1.5 times the initial size for Colombia.[5] This is consistent with results in Hsieh and Klenow (2014) indicating that less developed economies are characterized by less dynamic post-entry growth.[6]

Hsieh and Klenow (2009) and Buera and Fattal (2014) attribute such cross-country differences to poor institutions in developing economies, that fail to encourage investments in productivity and healthy market selection. Identifying the actual role that specific institutions play is an interesting area of future

[4] Notice that

$$\frac{\overline{L_{age}}}{\overline{L_0}} = \frac{\sum_{i=1}^{N} L_{i,age}}{\sum_{i=1}^{N} L_{i,0}} = \frac{\sum_{i=1}^{N} \frac{L_{i,age}}{L_{i,0}} * L_{i,0}}{\sum_{i=1}^{N} L_{i,0}} = \sum_{i=1}^{N} \frac{L_{i,age}}{L_{i,0}} \frac{L_{i,0}}{\sum_{i=1}^{N} L_{i,0}}$$

[5] Results for the cross sectional approach with the Colombian data in Figures 1 and 2 are consistent. The corresponding lines look a little different between the two figures because of the use of wider age bins in Figure 2.

[6] Though similar to Hsieh and Klenow's, our numbers for the US are not identical to theirs, even if we focused on the same year, because of several differences in the calculation. We use data from the Business Dynamics Statistics, which directly records the age of an establishment. It also records employment for establishments of all sizes. Meanwhile, Hsieh and Klenow impute age based on previous appearance in Census, and use imputed rather than observed employment for small businesses.

research. Within-country changes in institutions, either across businesses or over time (or both) offer a fruitful ground for such exploration, to the extent that they keep constant other factors potentially influencing business dynamics, from the macroeconomic environment to business culture.

Colombia, as many other countries in Latin America and around the globe, undertook wide market-oriented reforms at the beginning of the 1990s. These included unilateral trade openning, financial liberalization, and flexibilization of labor regulations. Eslava et al (2004, 2013, 2010) present evidence that these reforms did generate changes in business dynamics consistent with a reduction in the distortions to business incentives : allocative efficiency improved, the market selection mechanism was enhanced, capital and labor adjustment became more flexible (though in an apparently capital-biased way). We now ask whether there are noticeable differences pre- vs. post-refom in the life cycle dynamics of Colombian manufacturing plant described above.

Figure 4 depicts the results of allowing the coefficients of equation (6) to vary between the pre- vs. post-reform periods. We define 1982-1992 as pre-reform and 2002-2012 as post-reform. We leave out the years between 1993 and 2001 for two reasons: 1) It is not clear whether firms born just as the reforms are being adopted behave in their first few years as pre-reform or post-reform firms, which in tun also makes it hard to judge whether faster growth after a few years should be attributed to a comparison to poor birth conditions. 2) Between 1997 and 2001, the country went through its deepest recession in 70 years. The 2002-2012 period could be described as one where the new regulatory approach is consolidated and macroeconomic conditions are not that different from those in the eighties.

For each of these two subperiods, we construct a life cycle sample of plants observed from birth within the subperiod, so the maximum age in each case is 10 years. We pool the two samples, and run the regression

$$\frac{L_{it}}{L_{i0}} = \alpha_t + \alpha_s + \sum_{age=1}^{age=20+} \phi_{age} d_{age,i,t} + \sum_{age=1}^{age=20+} \phi_{age} d_{age,i,t} * dpost_t + \varepsilon_{it} \quad (7)$$

where $dpost_t = 1$ if $t \in [2002, 2012]$ and $dpost_t = 0$ otherwise. Results are summarized in Figure 4.

Plants in the post-reform period multiplied their birth size by 1.5 by age 6, reaching a maximum size of around 1.7 times their birth size in their first ten years. By contrast, plants in the pre-reform don't cross the 1.5 mark during their first decade of operation. That is, consistent with theroetical arguments implying slower growth of businesses in environments subject to more stringent distortionary regulations, the average manufacturing establishment grows faster in the post-reform period.

Figure 5 provides suggestive evidence that both improved market selection and starker growth for survivors contributed to greater post-entry dynamism after the reforms. Continuers' growth in the post reform is more marked than pre-reform, slightly but consistently. The gap between continuers and exiters is also more marked in the more recent period.

4.2 Skewness

The analysis above describes the dynamics of growth for an average manufacturing establishment in our sample. But, how typical is the average business? It has been well established that the answer is "not at all" in terms of business size and also inter-anual growth (e.g. Haltiwanger, Jarmin and Miranda, 2013). How about life-cycle growth? How skewed is the distribution of life cycle growth?

Figures 6 and 7 show different moments of the distribution of life-cycle growth.[7] The stark difference between median and mean patterns in Figure 6 (in log scale) highlights the fact that it is a minority of fast-growing plants that drive mean growth. A plant in the 90th percentile grows five times as fast as the median plant. Plants in the 10th percentile, meanwhile, shrink substantially over their life cycle. Figure 7, meanwhile, points that the post-reform is characterized by both faster growth in the 90th percentile, and less contraction in the 10th percentile.

In summary, there is very wide dispersion around the average patterns of growth described above. And, though the arguably more healthy market institutions that prevail in the decade of 2000 compared to twenty years early are associated with more rapid business growth, there seems to be as much dispersion in growth post-reform as there was in the 1980s.

4.3 Other plant outcomes over the life cycle

Figures 8 and 9 shows different measures of "outcome" growth over the life cycle. Plants consolidate over their life cycle in all of the dimensions explored. In terms of value of outputs and inputs, interestingly, employment is the slowest dimension of outcome growth post-entry. Output and revenue both grow about twice as fast as employment for the average plant. This is apparently driven by much faster growth in the use of capital and materials than employment.

At the same time, the average plant becomes more sophisticated as it ages. Over its first decade, it increases the number of 8-digit product categories in which it produces by about 50%. Skill composition also improves in a similar magnitude.

5 Measuring fundamentals

This section explains our approach to measuring the three fundamental dimensions into which we then decompose output and input growth. A key feature of our analysis is the availability of plant level prices, which we use to obtain indicators of quantities produced, and in turn to estimate the physical efficiency

[7] For each age the figures depict the respective moment of the distribution of $\frac{L_{it}}{L_{i0}}$.

of production. We first explain our approach to constructing plant-level output and input prices. Then, we describe our estimation of TFPQ and price effects, using those plant level prices.

5.1 Plant-level prices

Tornqvist indices for the growth of prices of plant j at time t are constructed, from its composite of products or materials h, as $\Delta P_{jt} = \sum_{h=1}^{H_j} \overline{s}_{hj} \Delta \ln(P_{hjt})$,where P_{hjt} is the price charged for product h, or paid for material h, by plant j in year t. $\Delta \ln(P_{hjt}) = \ln P_{hjt} - \ln P_{hjt-1}$ and $\overline{s}_{hj}$ is the average share of h in the basket of products (or materials) of plant j. After obtaining plant-level price changes, the indices for the levels of output (or material) prices for each plant j are constructed recursively as $\ln P_{jt} = \ln P_{ijt-1} + \Delta P_{jt}$.

The price series for each plant is initialized at a given level P_{j0}, where 0 is the base year for plant j. We construct the base price for a plant as: $\ln P_{j0} = \sum_{h=1}^{H_j} \overline{s}_{hj} \left(\ln P_{hj0} - \ln \overline{P_{h0}} \right)$, where year 0 is the first year in which the plant is present in the survey. Notice that this approach takes advantage of cross sectional variability across plants for any given product or input h. In the base year, the price index will be normalized one for the average producer (user) of product h (input h), but will capture the dispersion of other plants for the same product (input) around that average.

Our plant-level input price indices measure unit costs for materials, one of the inputs in the production function. They are, thus, one of the components in growth decomposition (3). $TFPQ$ and demand shocks are other key components. We now explain our approach to measuring them.

5.2 Physical productivity

We use our plant-level output prices to construct physical quantities of output, by deflating the nominal output by the plant-level price index. Physical quantities of materials are similarly constructed using the plant-level materials price index as deflator. With these elements in hand, and direct reports of the number of employees and the stock of physical capital, we estimate TFPQ from the (log) residual of production function (1) (repeated here)

$$Q_{it} = A_{it} X_{it}^{\gamma}$$

where $X = K_{it}^{\frac{\alpha}{\gamma}} L_{it}^{\frac{\beta}{\gamma}} M_{it}^{\frac{\phi}{\gamma}}$. A usual concern in the literature is that we observe values, rather than quantities, of output and inputs, so that the residual from estimating the (revenue) production function cannot be interpreted as A_{it} (or TFPQ). The fact that we use plant-level prices to deflate output and materials deals with this concern.

We estimate the production function for each two-digit sector, using proxy methods. In particular, we follow the approach proposed by Ackerberg, Caves and Frazer (2015, ACF henceforth), combining proxy-methods (Levinsohn-Petrin) with a control function based on GMM insights. Our key identifying assumption is that innovations to productivity are orthogonal to the current capital stock and labor, and to materials lagged one period. Labor market regulations in the country create barriers to immediate employment adjustment, which is why we assume labor to be (at least semi-) fixed.[8]

We also try variants of this approach proposed by De Loecker et al (2015), augmenting the control function to include plant-level prices. Both approaches yield very similar results in our case, so we present results using the ACF approach.[9] We have also examined the robustness of our results to other methods for determining factor elasticities, including using a cost share approach and including in the control functions other variables related to firm behavior (see De Loecker and Warzynski, 2012). We plan to further examine robustness to using Ghandi, Navarro and Rivers (2013) approach to the estimation of production functions. So far, results are broadly consistent as long as estimated returns to scale are not markedly decreasing.

Table 3 presents the results of our estimation of the production function, carried at the two-digit level of ISIC revision 2 (with descriptive statistics presented in Table 2). We estimate returns to scale approximately constant, and a coefficient for labor that tends to double that of capital. In most sectors, the coefficient for materials is estimated to be above 0.5.

5.3 Price effects

Our (log) residual price effect, $\ln D_{it}$, is the residual from estimating the demand function (2)

$$\ln P_{it} = \alpha - \varepsilon \ln Q_{it} + \varpi_{it} \tag{8}$$

That is, $\varpi_{it} = \ln D_{it}$. Estimating (8) by OLS would yield biased results, to the extent that Q is correlated with the residual price for reasons beyond demand shocks. We thus estimate this demand function using IV methods. In particular, we use the physical productivity shock A_{it}as an instrument for the plant's output, as in Eslava et al (2013) and Foster et al (2016). By using A_{it} as an instrument we focus on pure demand: the variability that is orthogonal to supply side shocks. The estimate that we recover for ε_{it} is an unbiased estimate

[8] In other words, materials are declared as a free input while labor is considered a semi-fixed one. Declaring labor, besides materials, as a free input, yields somewhat unplausible results for some sectors. Under that assumption, returns to scale are frequently (i.e. for several sectors and some periods of estimation) estimated to be increasing, and the coefficient for labor shoots up. Such implausible results support our prior that treating labor as a free input is not appropriate in the context in which we carry our estimation.

[9] Unlike De Loecker et al (2015), we deflate not only output but also materials using plant-level deflators. This helps us deal with the biases that they address by including output prices in their control function.

of the ability of the firm to charge a different price when observing a shock to its sold quantity that is unrelated to the efficiency of its production process.

The lower panel of Table 2 presents basic descriptive statistics for TFPQ and D. The degree of dispersion is similar (on a log basis) for both fundamentals. Table 4 characterizes our estimates of TFPQ and D_{it} in terms of their correlations with other variables. As found by Eslava et al. (2013) for an earlier period, TFPQ is negatively correlated with output prices, which is intuitive to the extent that more efficient production allows charging lower prices. Both TFPQ and D are positively correlated with plant size (captured in the table by output). Note also that TFPQ is highly correlated with TFPR . By construction, our estimate for $\log D_{it}$ captures only the part of the price effect that is uncorrelated with TFPQ, so the correlation between TFPQ and D is zero.

Price effects capture a variety of relevant determinants of firm growth. Product quality is a particularly interesting dimension, but it is mingled with distortions from either regulations or other fronts in our D measure. Moreover, it is difficult to think of quality improvements as orthogonal to TFPQ. Improving quality likely requires efforts that are costly in terms of efficiency. We thus recover an estimate of D that does not eliminate the part of price effects correlated with TFPQ. To this end, we begin by noticing that one can obtain a residual Z_{it} from the revenue function:

$$Z_{it} = \frac{R_{it}}{X_{it}^{\gamma(1-\varepsilon)}} = \frac{P_{it}Q_{it}}{\left(K_{it}^{\frac{\alpha}{\gamma}} L_{it}^{\frac{\beta}{\gamma}} M_{it}^{\frac{\phi}{\gamma}}\right)_{it}^{\gamma(1-\varepsilon)}}$$

and that $Z_{it} = A_{it}^{1-\varepsilon} D_{it}$. We thus obtain Z_{it} using observed R_{it} and X_{it} and our estimates of α, β, ϕ, and ε, and then recover $D_{it}^{corr} = \frac{Z_{it}}{A_{it}^{1-\varepsilon}}$, where we use the notation D^{corr} to highlight the fact that this measure of the price effect does not eliminate the correlation with TFPQ. Crucially, we have previously obtained unbiased estimates of factor and demand elasticities, and can now use them to derive an estimate of the demand shock that is not restricted to the part of demand uncorrelated with productive efforts.

Table 4 also displays the correlations between D_{it}^{corr} and our other estimates of fundamentals. D_{it}^{corr} is highly correlated with D_{it}. Its correlation with TFPQ, as expected, is negative.

(TENTATIVE) Khandelwal (2010) defines quality as the price variation not explained by quantities within narrowly defined product categories and years. We plan to follow this approach by calculating quality as the variation of D_{it}^{corr} that is left after considering firm-product-year effects. To pursue this strategy we will re-estimate our production and demand functions at the plant-product level and obtain:

$$D_{ijt}^{q} = D_{ijt}^{corr} - \sigma_{ijt}$$

where σ_{ijt} is a plant-product-time fixed effect.

6 Results: Decomposing growth into fundamental sources

We now decompose output and employment growth as characterized in section 4, into that associated with different fundamental sources: TFPQ, price effects, and input prices (equation 3). The distributions of the evolution of plant fundamentals over the life cycle are displayed in Figure 9, which also shows output prices. As we did before for employment, for a given variable Z each figure depicts statistics for $\frac{Z_{it}}{Z_{i0}}$, on a log scale.

Average TFPQ growth is quite anemic compared to that of price effects. And, both the mean and the median of input prices aer basically static over a plant's life. Improvements in the plant's ability to negotiate input prices as it ages are either null, or eroded by possible increases in the quality (and therefore cost) of inputs used.

There is also considerable dispersion in the growth of both TFPQ and price effects. The 90th-10th gap, however, is much wider (on a log basis) for the price effect. Moreover, even the 10th percentile grows modestly in the case of the price effect, while TFPQ falls markedly for the 10th percentile. Survival is clearly made possible, for this worst performing plants, by growth in prices.

Figure 10 depicts, directly, our output growth decomposition (on a log scale):

$$\frac{Q_{it}}{Q_{i0}} = \chi_{it} \left(\frac{D_{it}}{D_{i0}} \right)^{\gamma\lambda_1} \left(\frac{A_{it}}{A_{i0}} \right)^{1+\gamma\lambda_2} \left(\frac{pm_{it}}{pm_{i0}} \right)^{\gamma\lambda_2} \tag{9}$$

where, as explained in section 2 $\lambda_1 = \frac{1}{1-\gamma(1-\varepsilon)}$; $\lambda_2 = (1-\varepsilon)\lambda_1$; $\lambda_3 = -\phi(1-\varepsilon)\lambda_1$; and where we are using our estimates of γ, ε and ϕ to construct each of the terms of the decomposition. χ_{it} is a residual capturing cross-plant dispersion in the evolution of wages and capital unit costs.

The grey line in Figuer 10 corresponds to $\left(\frac{A_{it}}{A_{i0}}\right)^{1+\gamma\lambda_2}$, while the solid dashed line adds the contribution of materials prices by depicting $\left(\frac{A_{it}}{A_{i0}}\right)^{1+\gamma\lambda_2} \left(\frac{pm_{it}}{pm_{i0}}\right)^{\gamma\lambda_2}$, and the solid black line further adds the contribution of price effects. The dotted line represents outcome growth, $\frac{L_{it}}{L_{i0}}$. The difference between this dotted line and the solid black line is the contribution of unmeasured factors, χ_{it}. Each panel of the figure constructs the decomposition for a given section of the distribution of plant outcome growth. The upper left panel depicts, for each age, the decomposition for the average $\frac{L_{it}}{L_{i0}}$ plant. The upper right panel represents plants in the lowest decile of $\frac{L_{it}}{L_{i0}}$, and the lower left and right panels represent, respectively, the median and the highest decile.

All three fundamental dimensions explored in these figures play a role in explaining output growth, as seen, for instance, in the fact that slow (rapid) growth plants also show particularly slow (rapid) growth in fundamentals, in particular TFPQ and price effects. Plants with growth in the lowest decile of life cycle output growth, which in fact perceive a consistent decline in output , exhibit a negative contribution of TFPQ to growth further deepened by a contraction of the price effect. In turn, rapid growers in the 90th percentile of growth exhibit positive and dynamic contributions of all three fundamentals, and their distance with respect to slow and average performers is particularly marked in the price effect dimension.

These results point at a crucial role of price effects to explain extraordinary growth, with contributions of TFPQ and input prices less important in magnitude but with the expected sign. Table 5 further supports this interpretation. Even after controlling for (three-digit) sector and year effects, which take care of distortions at these levels, price effects contribute the most to the variability of output growth. However, all three fundamentals, especially TFPQ and price effects explain a sizable fraction of the variability in output growth over the life cycle. All of them together explain over 70% of the variability in output growth. This finding is in line with Foster et al's (2016) argument that consolidating a solid client basis is more central to post-entry business growth than physical efficiency gains, and their consistent results for selected US

Figure 11 and the lower panel of Table 5 replicate the decomposition using employment growth, $\frac{L_{it}}{L_{i0}}$, rather than output growth as a dependent variable. The three fundamentals measured here explain much less of the overall variability in emploment growth compared to their contribution towards explaining output growth: the overall R2 falls to 0.425 compared to 0.737 for employment. Adjustment costs likely play an important role in the relatively highly regulated Colombian labor market. Moreover, employment growth tracks even more closely price effects relative to TFPQ. In fact, the contribution of TFPQ to employment growth is negligible

Figure 12 explores pre- vs. post-reform differences in our results for the decomposition. The importance of price effects is particularly stark in the post-reform period.

7 Conclusion

References

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

De Loecker, J., P. Goldberg, A. Khandelwal and N. Pavcnik (2015) "Prices, Markups and Trade Reform,"

Forthcoming Econometrica.

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.

Foster, Lucia, John Haltiwanger, and Chad Syverson (2016) “The slow growth of new plants: learning about demand?” *Economica* 83(329) 91-129.

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.

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

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

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

Restuccia, Diego and Richard Rogerson. 2008. “Policy Distortions and Aggregate Productivity with Heterogeneous Plants,” *Review of Economic Dynamics*, 11(October): 707-720.

Table 1. Employment growth over the life cycle (showing coefficients only up to age 12)

	L_t/L_0. Life cycle sample	L_t. Full sample	L_t/L_0 . Life cycle sample. Pre- vs. post-reform
d_age=0-2		36.33***	
		(2.667)	
d_age=3	1.086***	38.81***	1.010***
	(0.260)	(2.991)	(0.0649)
d_age=4	1.267***	40.00***	1.121***
	(0.278)	(2.882)	(0.0737)
d_age=5	1.362***	41.39***	1.169***
	(0.280)	(2.839)	(0.0750)
d_age=6	1.466***	42.53***	1.232***
	(0.281)	(2.812)	(0.0763)
d_age=7	1.606***	43.94***	1.341***
	(0.282)	(2.791)	(0.0787)
d_age=8	1.718***	45.04***	1.407***
	(0.284)	(2.768)	(0.0820)
d_age=9	1.806***	47.29***	1.344***
	(0.285)	(2.768)	(0.0888)
d_age=10	1.896***	49.00***	1.195***
	(0.287)	(2.764)	(0.105)
d_age=11	1.957***	51.19***	
	(0.289)	(2.762)	
d_age=12	2.043***	53.30***	
	(0.291)	(2.765)	
pos*d_age=3			0.0538
			(0.0796)
pos*d_age=4			0.00937
			(0.0853)
pos*d_age=5			0.124
			(0.0929)
pos*d_age=6			0.189*
			(0.109)
pos*d_age=7			0.175
			(0.116)
pos*d_age=8			0.209*
			(0.124)
pos*d_age=9			0.301**
			(0.138)
pos*d_age=10			0.247
			(0.213)
Observations	38,345		9,727
R-squared	0.237		0.649
Effects	Year and sector	Year and sector	Year and sector

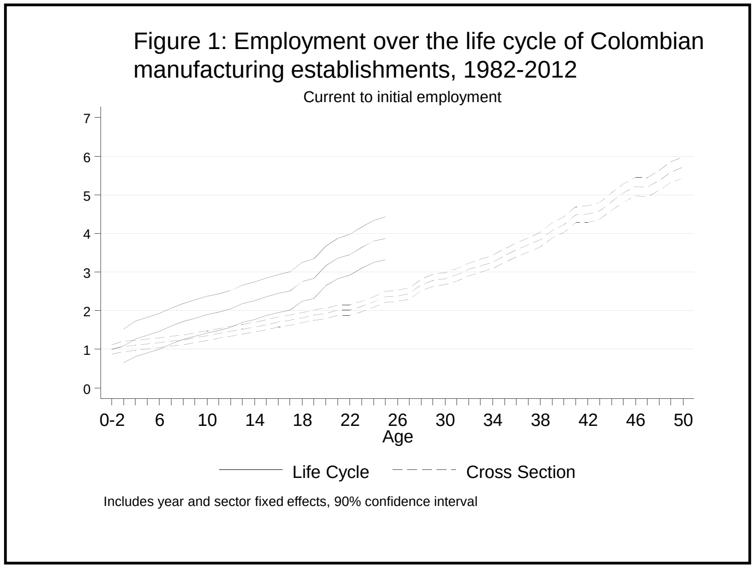


Figure 1: Employment over the life cycle of Colombian manufacturing establishments, 1982-2012
Current to initial employment
0 1 2 3 4 5 6 7
0-2 6 10 14 18 22 26 30 34 38 42 46 50
Age
Life Cycle
Cross Section
Includes year and sector fixed effects, 90% confidence interval

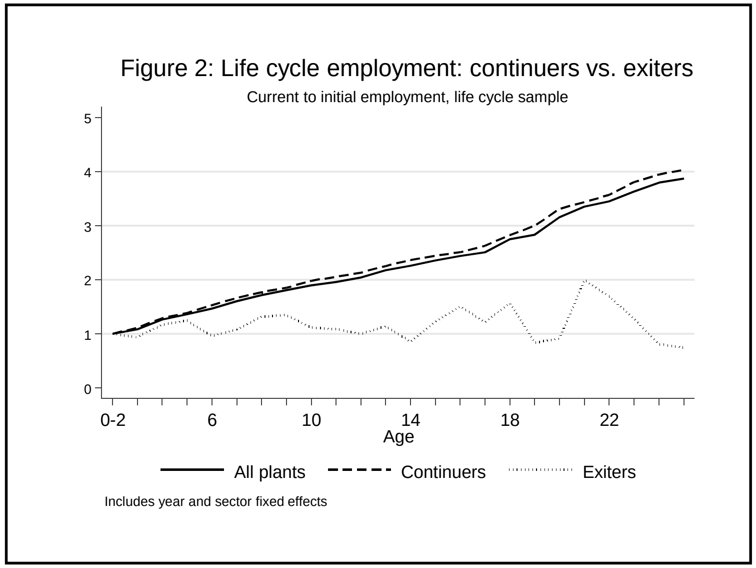


Figure 2: Life cycle employment: continuers vs. exiters
Current to initial employment, life cycle sample
0 1 2 3 4 5
0-2 6 10 14 18 22
Age
All plants
Continuers
Exiters
Includes year and sector fixed effects

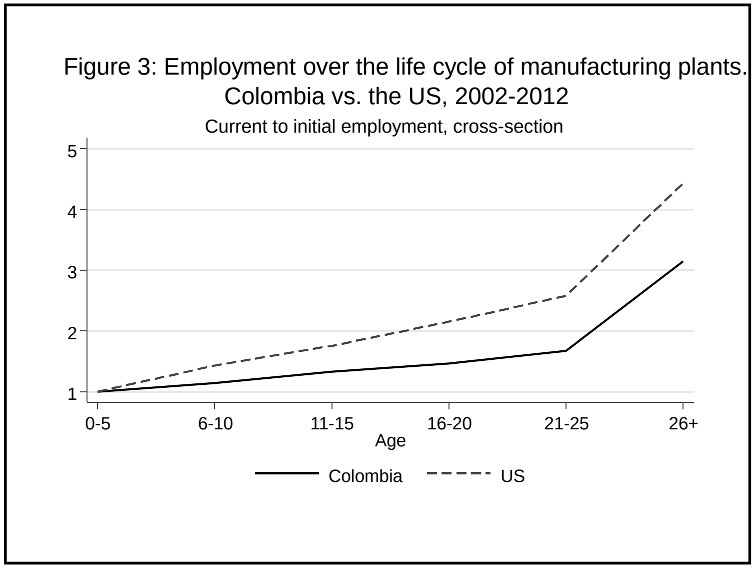


Figure 3: Employment over the life cycle of manufacturing plants.
Colombia vs. the US, 2002-2012
Current to initial employment, cross-section
1 2 3 4 5
0-5 6-10 11-15 16-20 21-25 26+
Age
Colombia
US

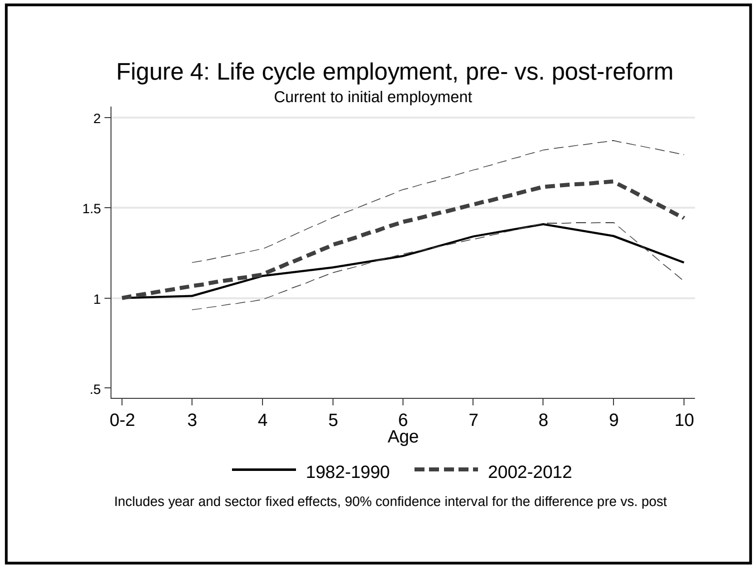


Figure 4: Life cycle employment, pre- vs. post-reform
Current to initial employment
.5 1 1.5 2
0-2 3 4 5 6 7 8 9 10
Age
1982-1990
2002-2012
Includes year and sector fixed effects, 90% confidence interval for the difference pre vs. post

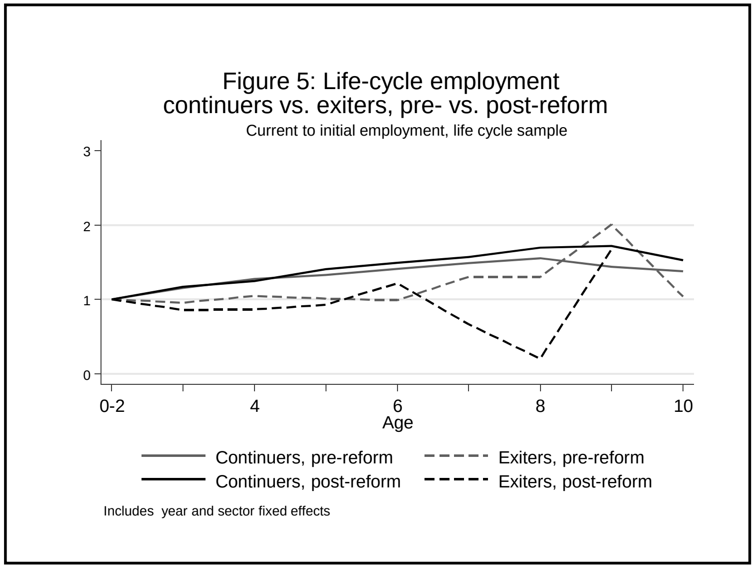


Figure 5: Life-cycle employment
continuers vs. exiters, pre- vs. post-reform
Current to initial employment, life cycle sample
3
2
1
0
0-2
4
6
8
10
Age
Continuers, pre-reform
Exiters, pre-reform
Continuers, post-reform
Exiters, post-reform
Includes year and sector fixed effects

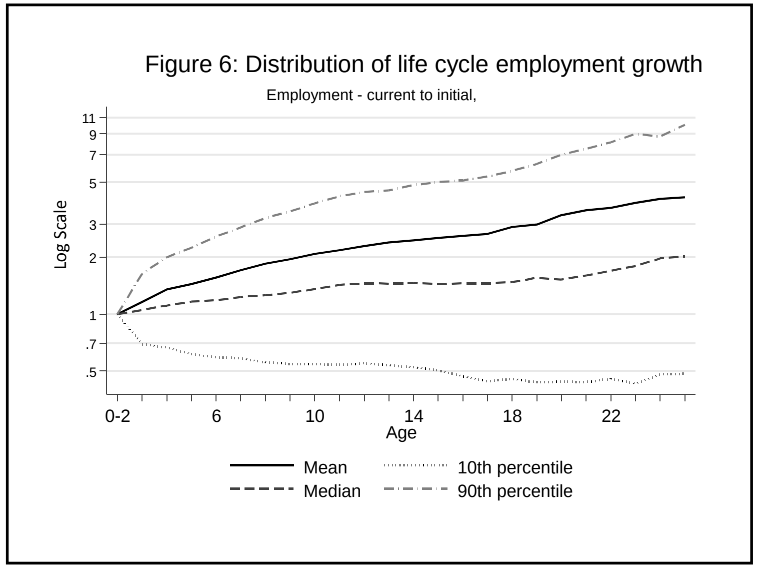


Figure 6: Distribution of life cycle employment growth
Employment - current to initial,
Log Scale
11
9
7
5
3
2
1
.7
.5
0-2
6
10
14
18
22
Age
Mean
10th percentile
Median
90th percentile

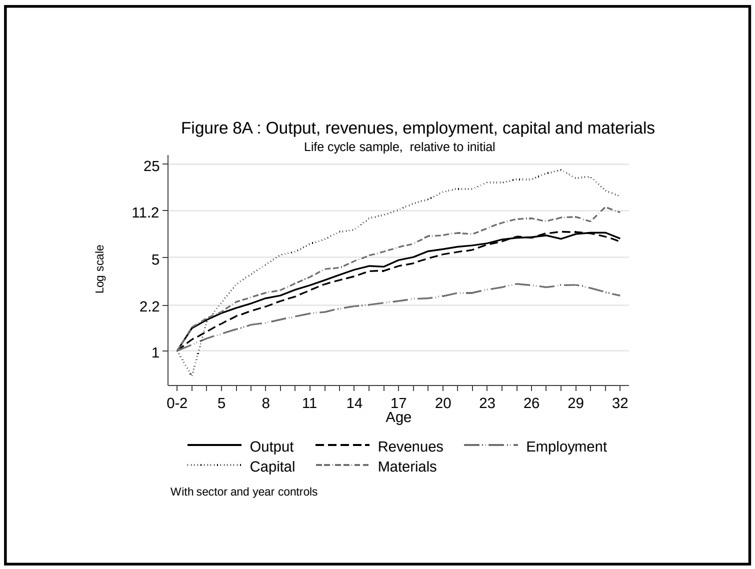


Figure 8A : Output, revenues, employment, capital and materials
Life cycle sample, relative to initial
Log scale
25
11.2
5
2.2
1
0-2
5
8
11
14
17
20
23
26
29
32
Age
Output
Revenues
Employment
Capital
Materials
With sector and year controls

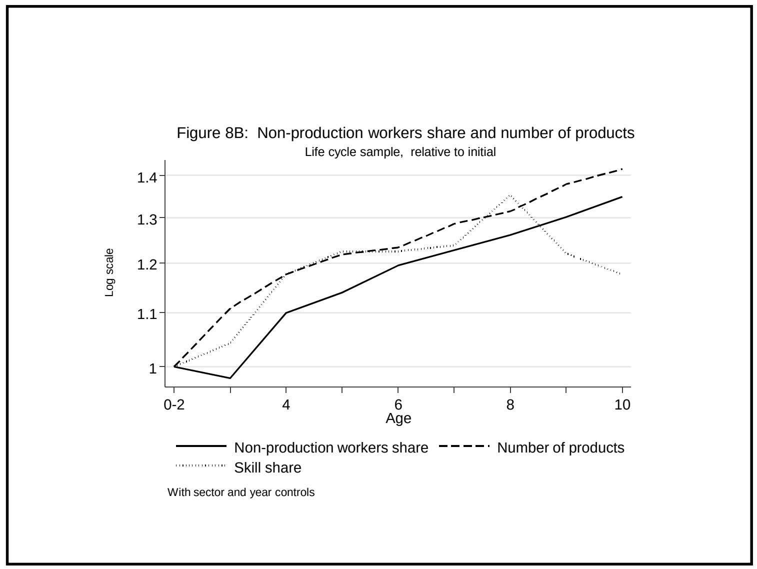


Figure 8B: Non-production workers share and number of products
Life cycle sample, relative to initial
Log scale
1.4
1.3
1.2
1.1
1
0-2
4
6
8
10
Age
Non-production workers share
Number of products
Skill share
With sector and year controls

Figure 7 :Distribution of life cycle employment, pre- vs. post-reform

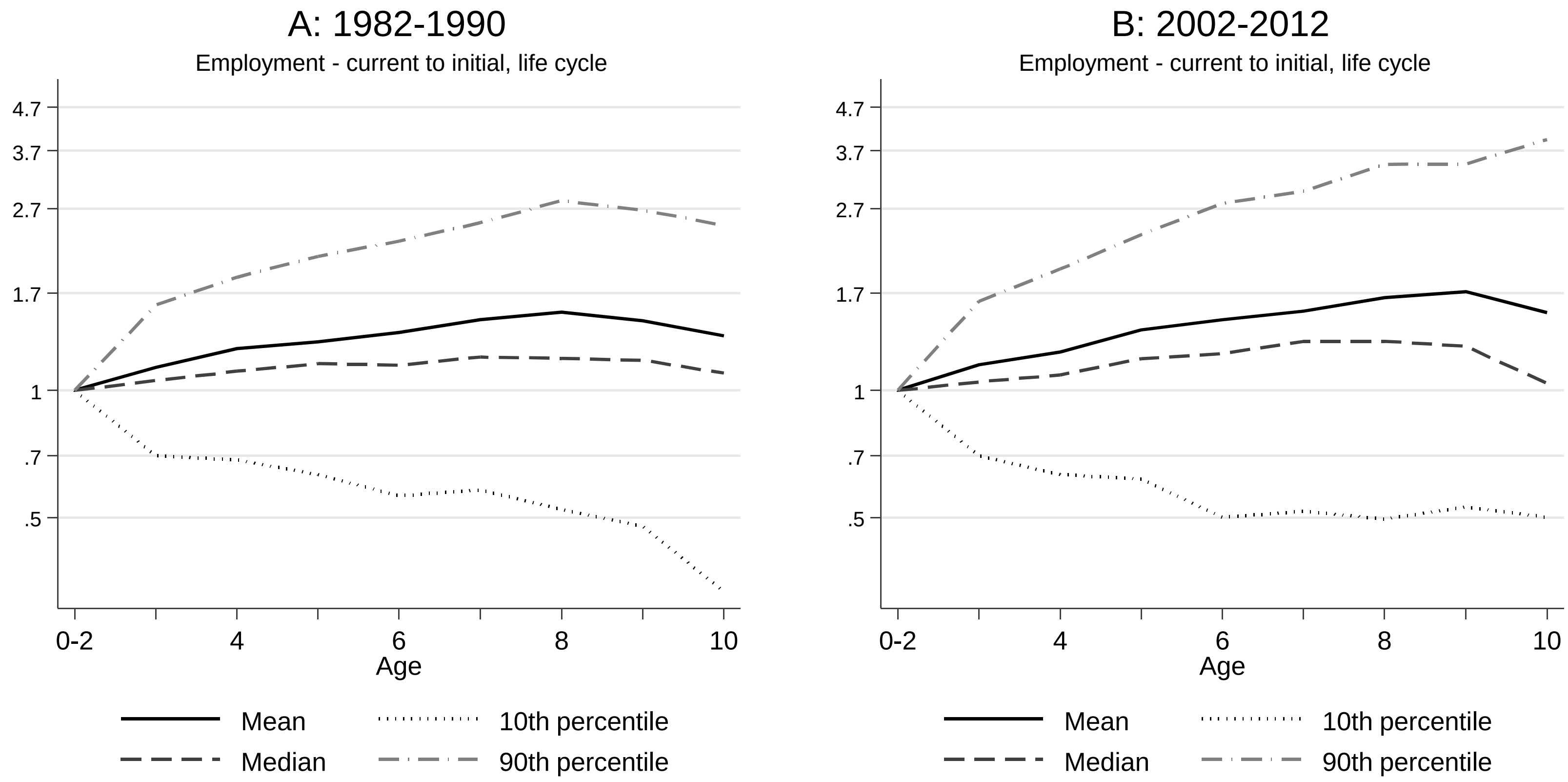


Table 2: Descriptive statistics

	Mean	SD
Ln Output	10.696	1.744
Ln Revenue	11.875	1.642
Employment	59.201	118.642
Ln Employment	3.353	1.112
Ln Capital	10.223	1.971
Ln Input prices	-0.291	0.676
Ln Output prices	-0.045	0.686
Ln TFPQ	2.421	0.986
Ln D	-0.037	1.055
N	40,087	

Table 3: Estimated gross output production function

Sector	Ln(L)	Ln(K)	Ln(MQ)	Returns to scale	αL/αK
Overall	0.327	0.166	0.549	1.043	1.970
31	0.319	0.146	0.639	1.104	2.185
32	0.269	0.134	0.575	0.978	2.008
33	0.250	0.080	0.668	0.998	3.125
34	0.613	0.252	0.214	1.080	2.433
35	0.411	0.214	0.471	1.097	1.921
36	0.481	0.242	0.346	1.070	1.988
37	0.525	0.182	0.463	1.170	2.885
38	0.289	0.099	0.647	1.035	2.919
39	0.434	0.176	0.408	1.017	2.466

*Materials as free input

Table 4: Correlation between plant fundamentals, output and revenue.

	Ln TFPQ	Ln D	Ln D^{corr}	Ln Input Price	Ln Output price	Ln Output
Ln TFPQ	1.000	0.000	-0.117	0.12	-0.611	0.287
Ln D	0.000	1.000	0.591	0.06	0.180	0.887
Ln D^{corr}	-0.117	0.591	1.000	-0.094	0.022	0.581
Ln Input Price	0.12	0.06	-0.094	1.00	0.15	-0.05
Ln Output Price	-0.611	0.180	0.022	0.15	1.000	-0.295
Ln Output	0.287	0.887	0.581	-0.05	-0.295	1.000

Figure 9: Mean, median, 10th and 90th percentiles

Life cycle sample, relative to initial

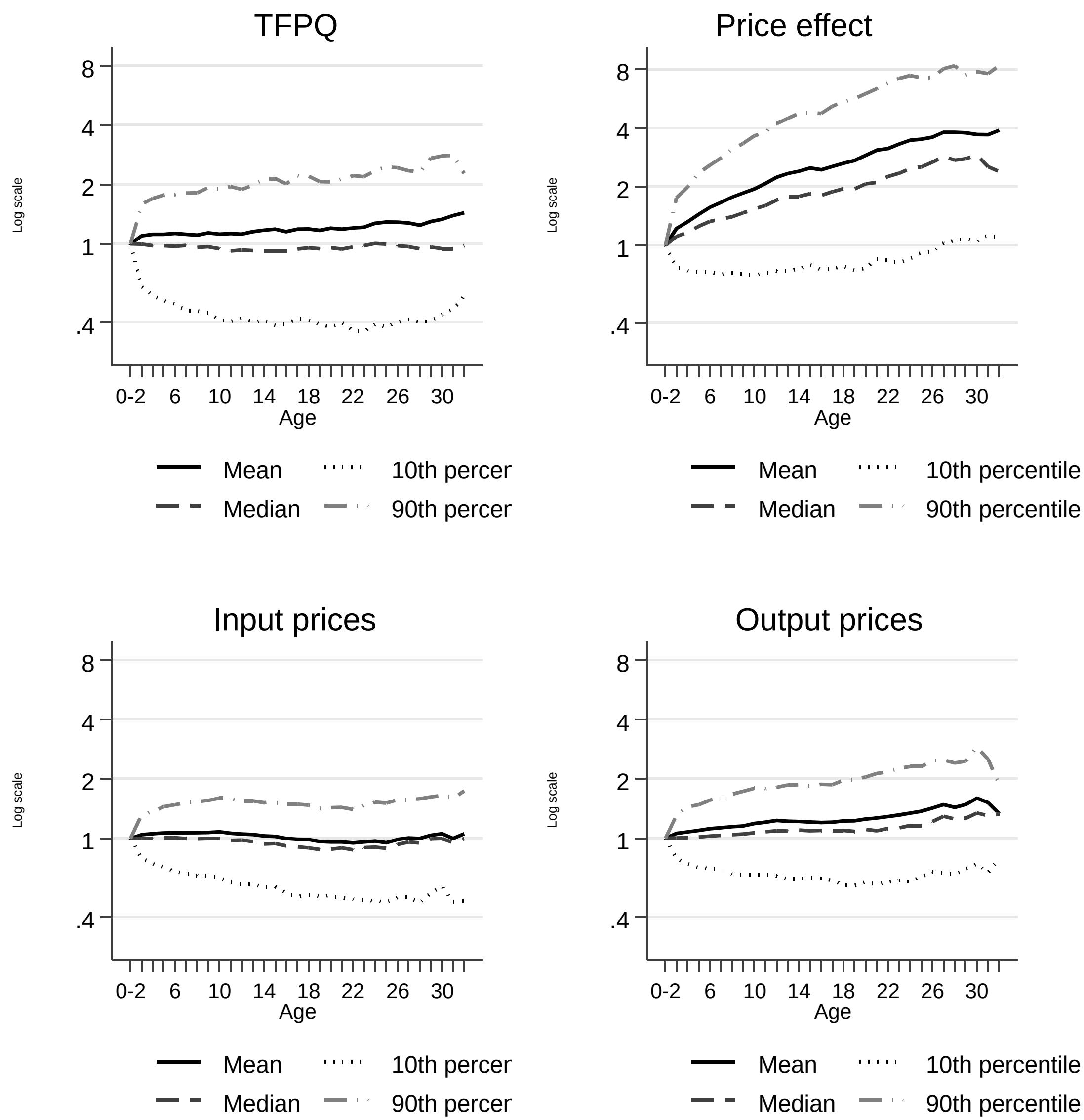
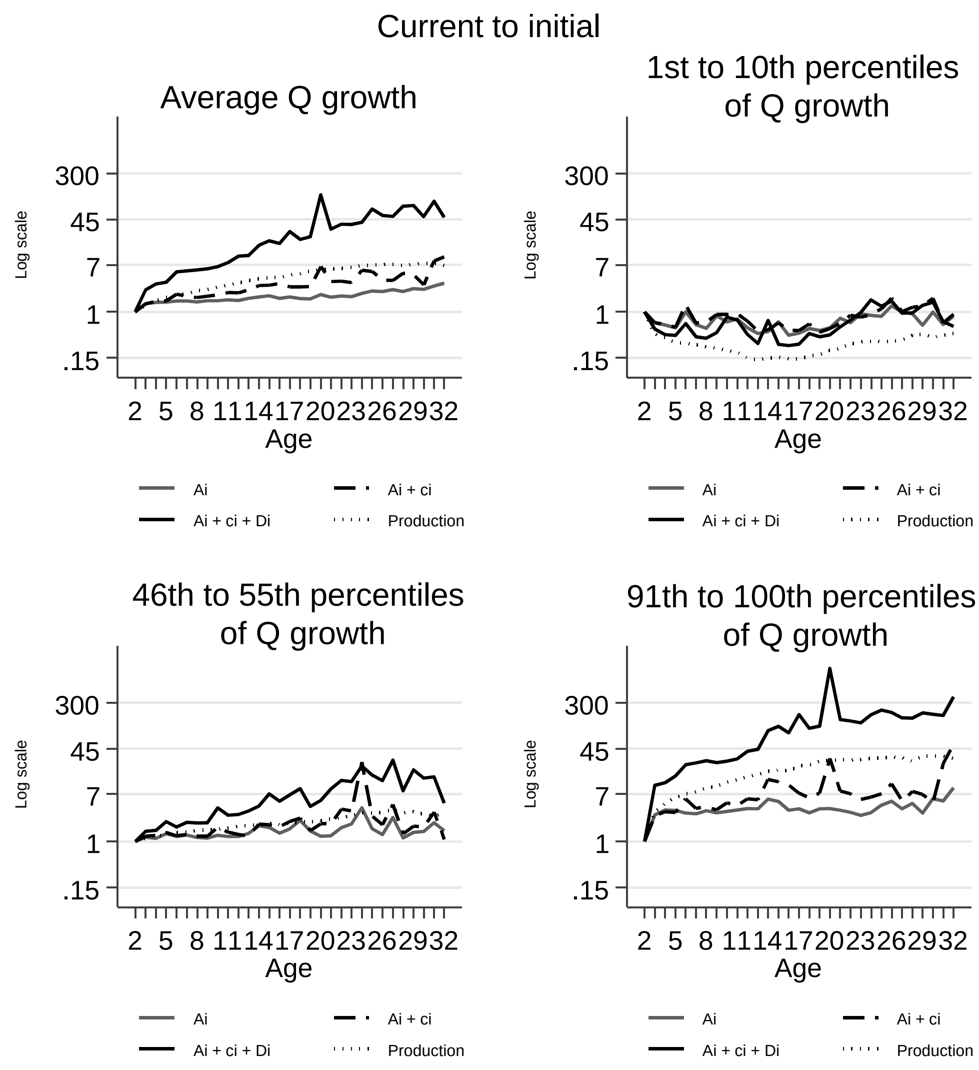


Figure 10: Contribution of TFPQ, input prices and demand shocks to production growth by age

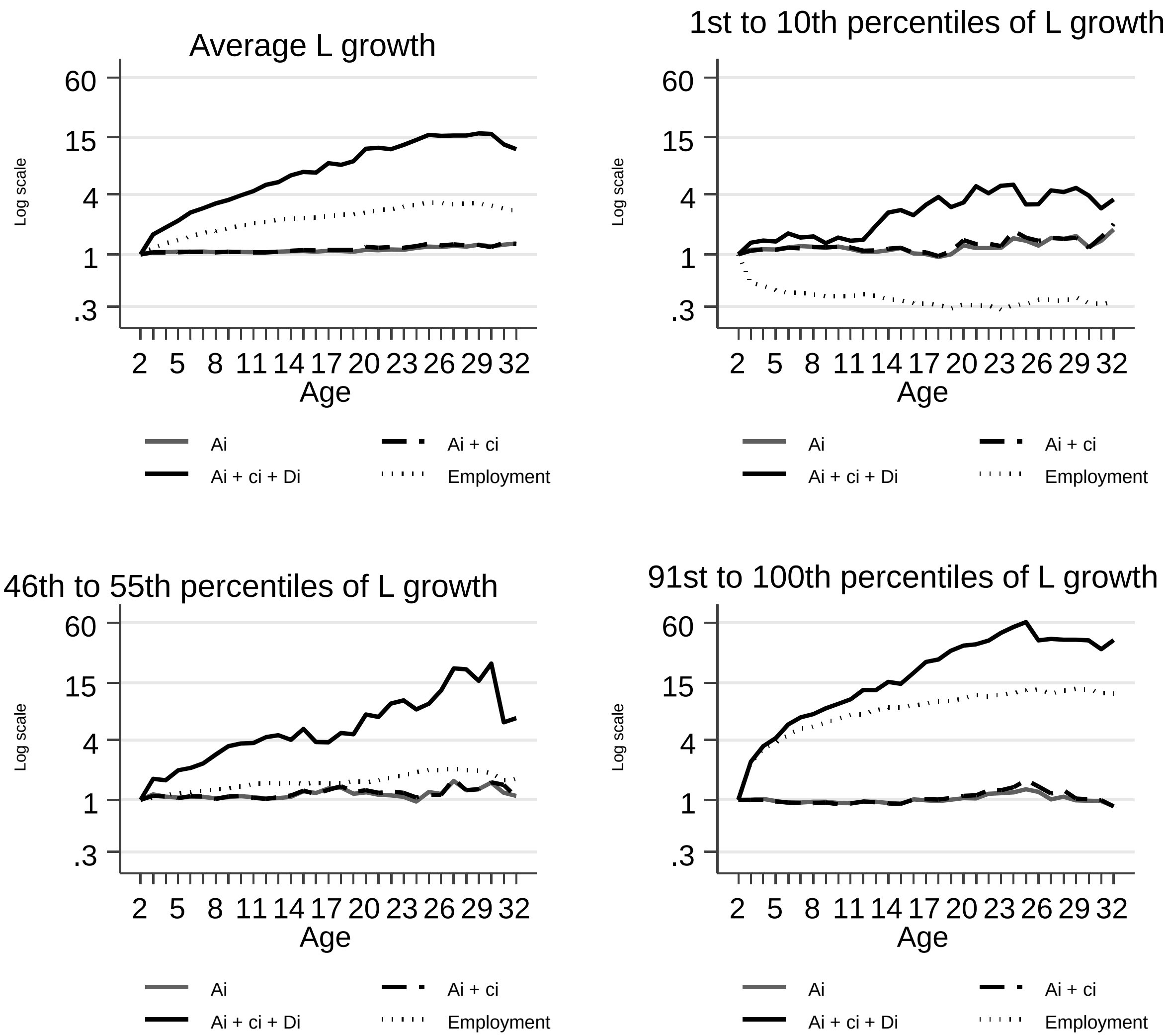


By current to initial output percentiles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Dependent variable: Ln (Output_it/Output_i0)										
TFPQi_t/TFPQ_i0			0.647***	0.659***					0.684***	0.695***
			(0.00791)	(0.00790)					(0.00467)	(0.00462)
D_it/D_i0					1.115***	1.119***			1.136***	1.139***
					(0.00538)	(0.00538)			(0.00427)	(0.00422)
pm_it/pm_i0							-0.139***	-0.122***	-0.485***	-0.477***
							(0.0135)	(0.0135)	(0.00743)	(0.00737)
N	40,718	40,718	38,065	38,065	40,718	40,718	40,272	40,272	37,503	37,503
R-squared	0.038	0.051	0.186	0.202	0.532	0.540	0.043	0.055	0.726	0.737
Sector FE	No	Yes	No	Yes	No	Yes	No	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Dependent variable: Ln (L_it/L_i0)										
TFPQi_t/TFPQ_i0			-0.131***	-0.129***					-0.157***	-0.155***
			(0.00651)	(0.00652)					(0.00521)	(0.00519)
D_it/D_i0					0.698***	0.704***			0.725***	0.730***
					(0.00464)	(0.00462)			(0.00475)	(0.00473)
pm_it/pm_i0							-0.00450	0.0110	-0.0600***	-0.0478***
							(0.00979)	(0.00984)	(0.00828)	(0.00827)
Observations	42,234	42,234	37,938	37,938	40,604	40,604	42,234	42,234	37,352	37,352
R-squared	0.030	0.043	0.042	0.055	0.377	0.391	0.032	0.044	0.410	0.425
Sector FE	No	Yes	No	Yes	No	Yes	No	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

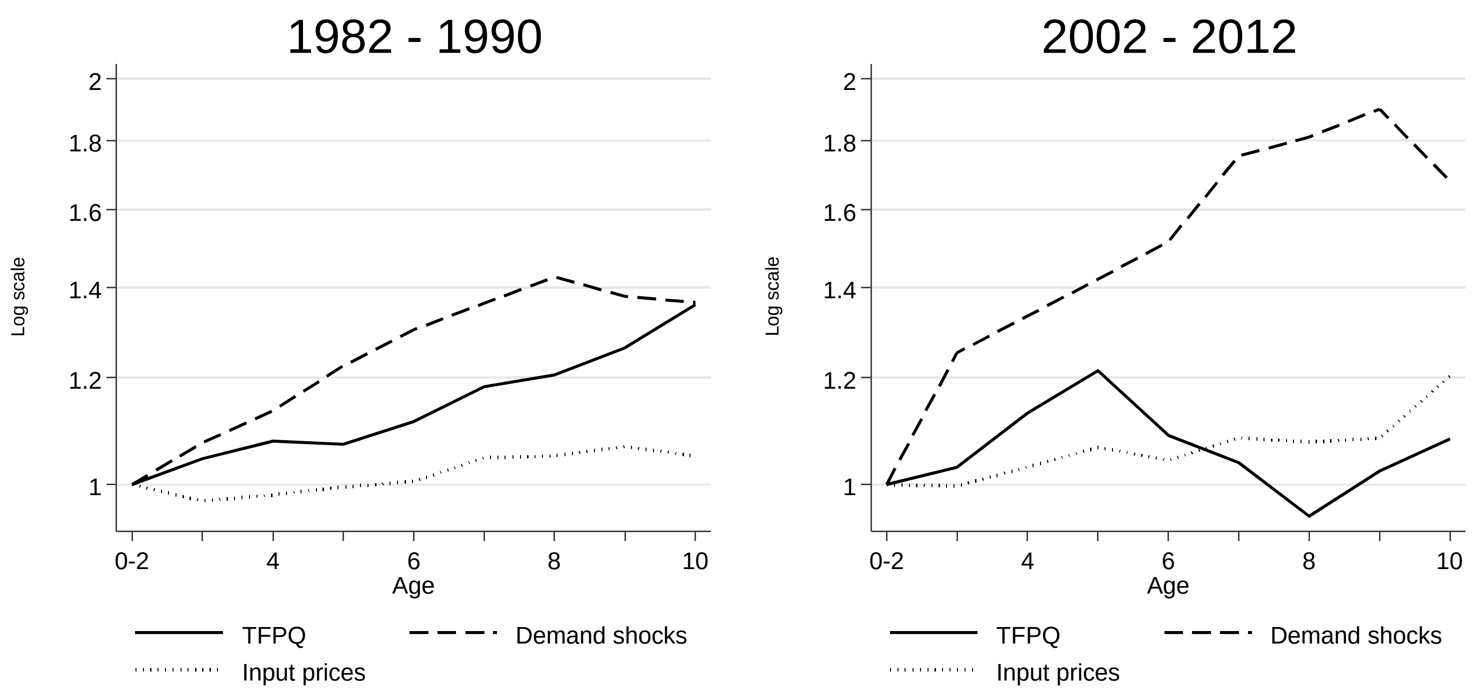
Figure 11: Contribution of TFPQ, input prices and Price effects
to employment growth by age

Current to initial



Current to initial employment percentiles

Figure 12 : TFPQ, demand shocks and input price index by age
Life cycle sample, relative to initial



With sector and year controls