Productivity and Organization in Portuguese Firms*

Lorenzo Caliendo
Yale University

Giordano Mion
University of Sussex

Luca David Opromolla
Banco de Portugal

Esteban Rossi-Hansberg
Princeton University

December 7, 2015

Abstract

The productivity of firms is, at least partly, determined by a firm's actions and decisions. One of these decisions involves the organization of production in terms of the number of layers of management the firm decides to employ. Using detailed employer-employee matched data and firm production quantity and input data for Portuguese firms, we study the endogenous response of revenue-based and quantity-based productivity to a change in layers: a firm reorganization. We show that as a result of an exogenous demand or productivity shock that makes the firm reorganize and add a management layer, quantity based productivity increases by about 4%, while revenue-based productivity drops by more than 4%. Such a reorganization makes the firm more productive, but also increases the quantity produced to an extent that lowers the price charged by the firm and, as a result, its revenue-based productivity.

1 Introduction

A firm's productivity depends on the way it organizes production. The decisions of its owners and managers on how to combine inputs and factors of different types with particular technologies, as well as size, marketing and pricing strategies all determine the production efficiency of a firm. Clearly, decision makers in a firm face many constraints and random shocks. Random innovations or disruptions, regulatory uncertainties, changes in tastes and fads, among many other idiosyncratic shocks, are undoubtedly partly responsible for

*We thank Jan De Loecker, Jakub Kastl, Steve Redding, and seminar participants at various institutions and conferences for useful comments and discussions. Caliendo: lorenzo.caliendo@yale.edu, Mion: g.mion@sussex.ac.uk, Opromolla: luca.opromolla@nyu.edu, Rossi-Hansberg: erossi@princeton.edu. The analysis, opinions, and findings represent the views of the authors and they are not necessarily those of Banco de Portugal.

fluctuations in firm productivity. However, these random –and perhaps exogenous– productivity or demand fluctuations, result in firm reactions that change the way production is organized, thereby affecting its measured productivity. For example, a sudden increase in demand due to a product becoming fashionable can lead a firm to expand and add either a plant, a more complex management structure, a new division, or a new building or structure to its production facilities. Many of these investments are lumpy and, as such, will change the production efficiency of the firm discontinuously.

In this paper we study the changes in productivity observed in Portuguese firms when they reorganize their management structure by adding or dropping layers of management. Consider a firm that wants to expand as a result of a positive demand shock and decides to add a layer of management (say add another division with a CEO that manages the whole firm). The new organization is suitable for a larger firm and lowers the average cost of the firm thereby increasing its quantity-based productivity. Moreover, the switch to an organizational structure fitted for a larger firm also reduces the marginal cost of the firm leading to higher quantities and lower prices. That is, at the moment of the switch, the firm is using a technology that is still a bit big for the size of its market, which reduces revenue-based productivity. The reason for this is that the organizational decision is lumpy. So a change in organization that adds organizational capacity in the form of a new management layer, leads to increases in quantity-based productivity, but reductions in revenue-based productivity through reductions in prices (due to reduction in the marginal cost, and, perhaps also due to reductions in mark-ups). In that sense, the endogenous response of firm productivity to exogenous shocks can be complex and differ depending on the measure of productivity used. Using a recently developed measure of changes in organization we show that these patterns are very much present in the Portuguese data.

Although the logic above applies to many types of organizational changes and other lumpy investments, we explain it in more detail using a knowledge-based hierarchy model that can guide us in our empirical implementation. Furthermore, this model provides an easy way to identify the changes in organization as we explain below. The theory of knowledge-based hierarchies was developed in Rosen (1982), Garicano (2000) and in an equilibrium context with heterogeneous firms in Garicano and Rossi-Hansberg (2006) and Caliendo and Rossi-Hansberg (2012, from now on CRH). In particular, we use the model in CRH since it provides an application of this theory to an economy with firms that face heterogeneous demands for their products. In that paper the authors characterize the pattern of quantity-based and revenue-based productivity as firms reorganize as a result of an exogenous demand or productivity shock.

The basic technology is one that requires time and knowledge. Workers use their time to produce and generate 'problems' or production possibilities. Output requires solving this problems. Workers have knowledge that they use to try to solve these problems. If they know how to solve them, they do, and output is realized. Otherwise they can redirect the problem to a manager one layer above. Such a manager tries to solve the problem and, if it cannot, can redirect the problem to an even higher-level manager. The organizational problem of the firm is to determine, how much does each employee know, how many of them to employ, and how many layers of management to use in production.

Using matched employer-employee data for the French manufacturing sector, Caliendo, Monte and Rossi-Hansberg (2015, from now on CMRH) show how to use occupation data to identify the layers of management in a firm.[1] They show that the theory of knowledge-based hierarchies can rationalize the layer-level changes in the number of employees and wages as firms grow either with or without changing layers. For example, as implied by the theory, a reorganization that adds a layer of management leads to increases in the number of hours employed in each layer but to a *reduction* in the average wage in each preexisting layer. In contrast, when firms grow without reorganizing they add hours of work to each layer and they *increase* the wages of each worker. This evidence shows that when firms expand and contract they actively manage their organization by hiring workers with different characteristics. The Portuguese data exhibits the same patterns that CMRH found for France. Importantly, the detailed input, price and quantity data for Portugal allows us to go a step further and measure the productivity implications of changes in organization.

Measuring productivity well is notoriously complicated and the industrial organization literature has proposed a variety of techniques to do so (see Berry, Levinsohn and Pakes (1995), Olley and Pakes (1996), Levinsohn and Petrin (2003), Wooldridge (2009), and De Loecker and Warzinsky (2012), among others). The first issue is whether we want to measure quantity or revenue-based productivity. The distinction is crucial since the first measures how effective is a firm in transforming inputs and factors into output, while the other also measure any price variation, perhaps related to markups, that result from market power. The ability of firms to determine prices due to some level of market power is a reality that is hard to abstract from. Particularly when considering changes in scale that make firms move along their demand curve and change their desired prices. We find that using a host of different measures of revenue productivity (from value-added per worker to Olley and Pakes, 1996, and Wooldridge, 2009), adding layers is related to *decreases* in revenue-based productivity. As explained above, our theory suggests that this should be the case since firms reduce prices when they expand. However, since firms also received a variety of idiosyncratic demand, markup, and productivity shocks every period it is hard to just directly look at prices to measure this effect.

To measure the effect of organizational change on quantity-based productivity we need a methodology that can account for demand, markup, and productivity shocks over time and across firms.[2] We use the methodology proposed by Forlani et al., (2014), which from now on we refer to as MULAMA. This method makes some relatively strong assumptions on the structure of the production function, as well as firm maximization and competition (Cobb-Douglas and monopolistic competition with some generalizations), but it allows for correlated demand and productivity disturbances. Furthermore, it is amenable to introducing the organizational structure we described above. Note also that since we focus on changes in quantity-based productivity as a result of a firm reorganization we can sidestep the difficulties in comparing quantity-based productivity across horizontally differentiated products. Using this methodology, and quantity data available in the Portuguese data, we find that quantity-based productivity is an increasing function of the

[1] Following CMRH several studies have shown that occupational categories identify layers of management in other datasets. For example, Tåg (2013) for the Swedish data and Friedrich (2015) for the Danish data.

[2] See Marschak and Andrews (1944) and Klette and Griliches (1996) for a discussion of the output price bias when calculating productivity.

quantity produced. However, quantity-based productivity increases significantly only when firms grow by adding layers. Furthermore, when we introduce layers directly in the measurement of firm productivity, adding layers is associated with increases in quantity-based productivity. The finding survives a variety of robustness checks and alternative formulations of the productivity process. For example, we can allow for changes in organization to have a permanent or only a contemporaneous impact on quantity-based productivity. This is our main finding: we link a careful measure of productive efficiency (quantity-based productivity as measured by MULAMA) with an increase in the management capacity of a firm (as measured by the number of layers)[3].

Up to this point we have not addressed the issue of causality. The results above only state that adding layers coincides with declines in revenue-based productivity and increases in quantity-based productivity. Our theory suggests that the relationship is causal, and the fact that it explains the pattern of both revenue-based and quantity-based productivity seems to support this interpretation. So we set out to use instrumental variables to verify this implication empirically. Our methodology suggests a variety of past firm decisions (like capital, past employment, etc.) as valid instrumental variables when organizational change affects productivity only contemporaneously. We show that our results on quantity-based productivity in fact seem to be causal. Our findings with instrumental variables are in general significant, although the estimation is somewhat more noisy, which prevents us from using as rich a set of fixed effects as the one we used in all the other regression results.

In sum, in this paper we show that the organizational structure of firms, as measured by their hierarchical occupational composition, has direct implications on the productivity of firms. As they add organizational layers, their quantity-based productivity increases, although the corresponding expansion decreases their revenue productivity as they reduce prices. This endogenous component of productivity determines, in part, the observed heterogeneity in both revenue and quantity-based productivity across firms. Failure to reorganize in order to grow can, therefore, result in an inability to exploit available productivity improvements. This would imply that firms remain inefficiently small, as has been documented in some developing countries (Hsieh and Klenow, 2014).

The rest of this paper is organized as follows. In Section 2 we provide a short recap of the knowledge-hierarchy theory that we use to guide our empirical exploration and describe its implications for productivity. Section 3 discusses the Portuguese manufacturing data set we use in the paper. Section 4 presents the basic characteristics of Portuguese production hierarchies. In particular, we show that firms with different numbers of layers are in fact different and that changes in the number of layers are associated with the expected changes in the number of workers and wages at each layer. Section 5 presents our main results on revenue-based productivity, as well as the methodology we use to measure quantity-based productivity and our main empirical results on this measure. It also presents a variety of robustness results as well as our results for quantity-based productivity using instrumental variables. Section 6 concludes. The appendix

[3]In a related result, Garcia and Voigtländer (2014) find among new Chilean exporters a reduction in revenue-based productivity and an increase in quantity-based productivity. The mechanism and findings in our paper can be used directly to rationalize their findings since exporting amounts to a firm revenue shock.

includes more details on our data set, a description of all tables and figures, as well as additional derivations and robustness tests of the results in the main text.

2 A Sketch of a Theory of Organization and its Empirical Implications

The theory of knowledge-based hierarchies, initially proposed by Garicano (2000), has been developed using a variety of alternative assumptions (see Garicano and Rossi-Hansberg, 2015, for a review). Here we discuss the version of the technology with homogenous agents and heterogeneous demand developed in CRH.

So consider firm i in period t that faces a Cobb-Douglas technology

$$Q_{it}\left(O_{it}, M_{it}, K_{it}\right) = a_{it} O_{it}^{\alpha_O} M_{it}^{\alpha_M} K_{it}^{\gamma-\alpha_M-\alpha_H} \tag{1}$$

with quantity-based productivity a_{it}, returns to scale given by γ and where O_{it} denotes the labor input, M_{it} material inputs and K_{it} capital. The parameter $\alpha_O \geq 0$ represents the expenditure share on the labor input, $\alpha_M \geq 0$ on materials and $\gamma - \alpha_M - \alpha_O$ on physical capital. The labor input is produced using the output of a variety of different workers with particular levels of knowledge. The organizational problem is embedded in this input. That is, we interpret the output of the knowledge hierarchy as generating the labor input of the firm. Hence, in the rest of this section we focus on the organizational problem of labor and abstract from capital and materials. We return to the other factors in our estimation of productivity below.

Production of the labor input requires time and knowledge. Agents employed as workers specialize in production, use their unit of time working in the production floor and use their knowledge to deal with any problems they face in production. Each unit of time generates a problem, that, if solved yields one unit of output. Agents employed as managers specialize in problem solving, use h units of time to familiarize themselves with each problem brought by a subordinate, and solve the problems using their available knowledge. Problems are drawn from a distribution $F(z)$ with $F''(z) < 0$. Workers in a firm know how to solve problems in an interval of knowledge $[0, z_L^0]$, where the superindex 0 denotes the layer (0 for workers) and the subindex the total number of management layers in the firm, L. Problems outside this interval, are passed on to managers of layer 1. Hence, if there are n_L^0 workers in the firm, $n_L^1 = hn_L^0\left(1 - F\left(z_L^0\right)\right)$, managers of layer one are needed. In general, managers in layer ℓ learn $\left[Z_L^{\ell-1}, Z_L^{\ell}\right]$ and there are $n_L^{\ell} = hn_L^0(1 - F(Z_L^{\ell-1}))$ of them, where $Z_L^{\ell} = \sum_{l=0}^{\ell} z_L^l$. Problems that are not solved by anyone in the firm are discarded. Agents with knowledge z_L^{ℓ} obtain a wage $w\left(z_L^{\ell}\right)$ where $w'\left(z_L^{\ell}\right) > 0$ and $w''\left(z_L^{\ell}\right) \geq 0$. Market wages simply compensate agents for their cost of acquiring knowledge.

The organizational problem of the firm is to choose the number of workers in each layer, their knowledge and therefore their wages, and the number of layers. Hence, consider a firm that produces a quantity O of the labor input. $C_L(O; w)$ is the minimum cost of producing a labor input O with an organization with L

layers[4] at a prevailing wage schedule $w(\cdot)$, namely,

$$C_L(O;w) = \min_{\{n_L^\ell, z_L^\ell\}_{l=0}^L \geq 0} \sum_{\ell=0}^L n_L^\ell w\left(z_L^\ell\right) \tag{2}$$

subject to

$$\begin{aligned} O &\leq F(Z_L^L) n_L^0, && (3)\\ n_L^\ell &= h n_L^0 [1 - F(Z_L^{\ell-1})] \text{ for } L \geq \ell > 0, && (4)\\ n_L^L &= 1. && (5) \end{aligned}$$

The first constraint just states that total production of the labor input should be larger or equal than O, the second is the time constraint explained above, and the third states that all firms need to be headed by one CEO. The last constraint is important since it implies that small firms cannot have a small fraction of the complex organization of a large firm. We discuss bellow the implications of partially relaxing this constraint. The *variable* cost function is given by

$$C(O;w) = \min_{L \geq 0} \{C_L(O;w)\}.$$

CRH show that the average cost function ($AC(O;w) = C(O;w)/O$) that results from this problem exhibits the properties depicted in Figure 1 (which we reproduce from CMRH). Namely, it is U-shaped given the number of layers, with the average cost associated to the minimum efficient scale that declines as the firm adds layers. Each point in the average cost curve in the figure correspond to a particular organizational design. Note that the average cost curve faced by the firm is the lower-envelope of the average cost curves for a given number of layers. The crossings of these curves determine a set of output thresholds (or correspondingly demand thresholds[5]) at which the firms decides to reorganize by changing the number of layers. The overall average cost, including materials and capital, of a firm that is an input price taker will have exactly the same shape (given our specification of the production function in equation (1) under $\gamma = 1$).

Consider the three dots in the figure, which correspond to firms that face different levels of demand as parametrized by λ.[6] Suppose that after solving the corresponding profit maximization using the cost function above, a firm that faces a demand level of λ decides to produce $Q(\lambda)$ (or $q(\lambda)$ in logs). The top panel on the right-hand-side of Figure 1 tells us that it will have one layer with 5 workers and one layer with one manager above them. The figure also indicates the wages of each of them (the height of each bar),

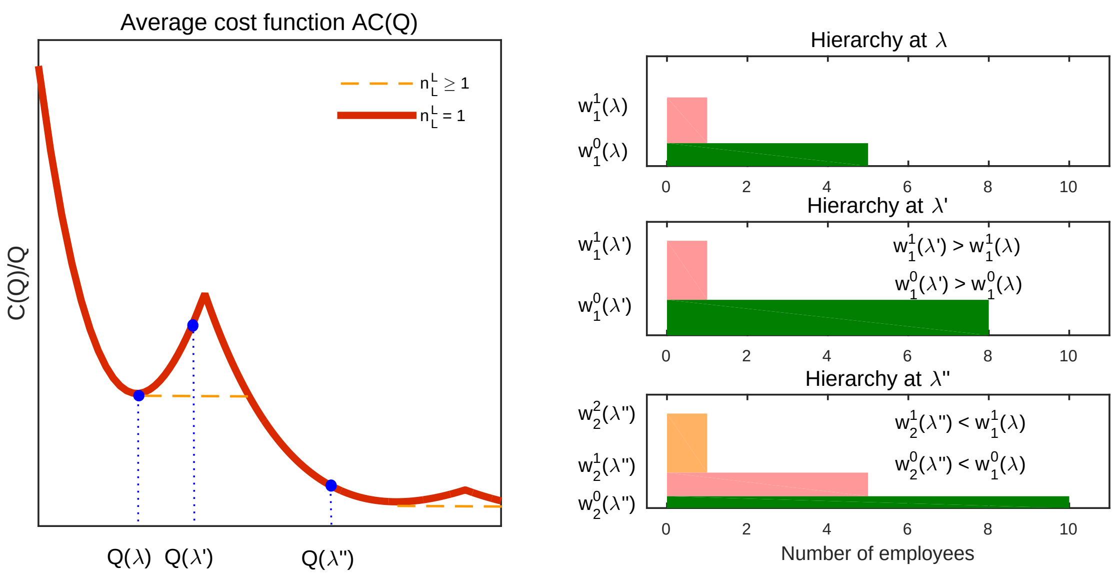
[4] Throughout we refer to the number of layers of the firm by the number of management layers. So firms with only workers have zero layers, firms with workers and managers have 1 layer, etc.

[5] Note that since output increases (decreases) discontinuously when the firm adds (drops) layers, the average cost curve is discontinuous as a function of the level of demand λ.

[6] In our examples here we focus on changes in the level of demand. Later on we will further consider changes in the exogenous component of productivity and changes in markups. Indeed, whatever pushes the firm to change its desired output can affect a firm's organizational structure.

which is increasing in their knowledge. Now consider a firm that as a result of a demand shock expands to $Q(\lambda')$ without reorganizing, that is, keeping the same number of layers. The firm expands the number of workers and it increases their knowledge and wages. The reason is that the one manager needs to hire more knowledgeable workers, who ask less often, in order to increase her span of control. In contrast, consider a firm that expands to $Q(\lambda'')$. This firm reorganizes by adding a layer. It also hires more workers at all preexisting layers. However, it hires less knowledgeable workers, at lower wages, in all preexisting layers. The reason is that by adding a new layer the firm can avoid paying multiple times for knowledge that is rarely used by the bottom ranks in the hierarchy. In the next section we show that all these predictions are confirmed by the data.

Figure 1: Average Cost and Organization



We can also use Figure 1 to show how the organizational structure changes as we relax the integer constraint of the top manager, in (5). First, note that at the minimum efficient scale (MES), which is given by the minimum of the average cost, having one manager at the top is optimal for the firm. So the constraint in (5) is not binding. Hence, relaxing the constraint can affect the shape of the average cost function on segments to the right and to the left of the MES. The reason why average costs rise for quantities other than MES is that firms are restricted to have one manager at the top. Otherwise, the firm could expand the optimal organizational structure at the MES by just replicating the hierarchy proportionally as it adds or reduces managers at the top.

For instance, suppose we allow organizations to have more than one manager at the top, namely $n_L^L \geq 1$. Figure 1 presents dashed lines that depict the shape of the average cost for this case. As we can see, the average cost is flat for segments to the right of the MES up to the point in which the firm decides to add a new layer. At the moment of the switch, the average cost starts falling until it reaches the MES and then it becomes flat again. All the predictions that we discussed before still hold for this case. The only difference is the way in which firms expand after they reach their MES up to the point in which they reorganize. We allow for this extra degree of flexibility when we use the structure of the model and take it to the data.[7]

2.1 Productivity Implications

In the following section we show that firms that grow or shrink substantially do so by adding or dropping management layers. These reorganizations also have consequences on the measured productivity of firms. In the model above quantity-based productivity of a firm in producing the labor input can be measured as the inverse of the average cost at constant factor prices; namely, $Q(\lambda)/\bar{C}(Q(\lambda);C(\cdot;1),1,1)$ where $\bar{C}(Q(\lambda);C(\cdot;w),P_m,r)$ denotes the overall cost function of the firm and P_m and r the price of materials and capital. Note that $Q(\lambda)$ denotes quantity produced and not revenue. Revenue-based productivity is instead given by $P(\lambda)Q(\lambda)/\bar{C}(Q(\lambda);C(\cdot;1),1,1)$ where $P(\lambda)$ denotes the firm's output price.[8]

For most demand systems, under monopolistic competition, the price $P(\lambda)$ will respond to changes in the marginal cost. If the demand system is CES, the change is proportional. CRH show that the marginal cost falls discontinuously when firms adds layers and increases discontinuously when firms drop them. The reason is that by adding layers the firm switches to a technology that is suited for a larger scale of production. That is, a firm can add layers even if demand does not increase enough to make it produce at the minimum efficient scale of the new technology. Hence, the marginal cost of producing a unit of output is lower than at the minimum efficient scale, and we know that the minimum efficient scale of the firm increases with the number of layers (and, at that point, the average and marginal costs are identical).

Quantity-based productivity increases with an increase in λ when the firm adds layers. The reason is that any voluntary increase in layers results in a level of produced quantity that lowers the average costs of the firm. Still, under CES preferences, CRH show that the resulting effect on prices dominates the positive quantity-based productivity effect and results in a discontinuous decrease in revenue-based productivity.[9] This effect in both types of productivity is illustrated in Figure 2 where we consider the effect of a shock in λ that leads to a reorganization that adds one layer of management.

In sum, firms that add layers as a result of a marginal revenue shock increase their quantity discontinuously. The new organization is more productive at the new scale, resulting in an increase in quantity-based

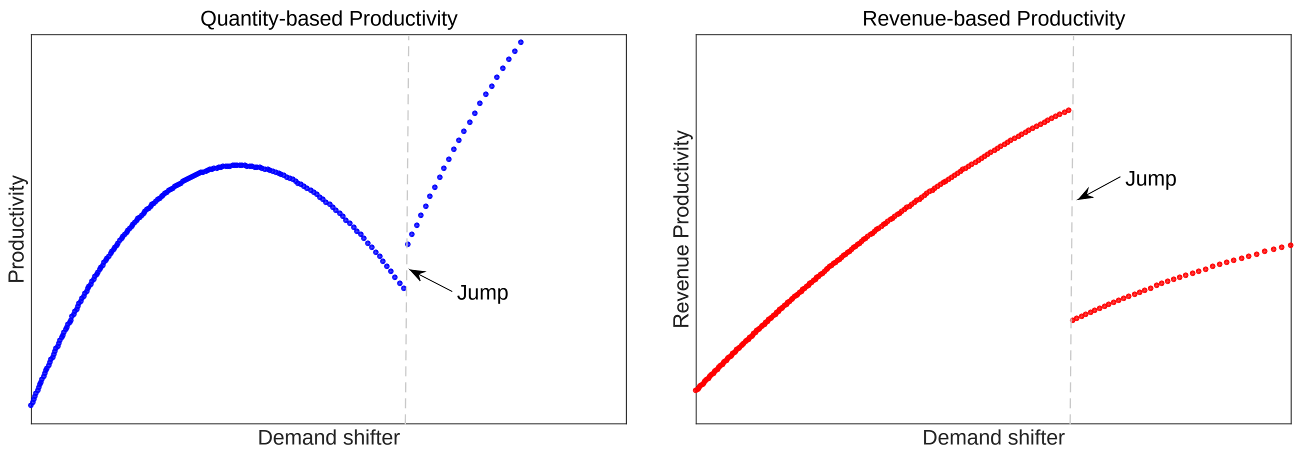
[7] Alternatively, one could also relax the integer constraint by letting $n_L^L \geq \epsilon$, where $1 > \epsilon > 0$. Following the discussion in the main body, in this case, the average cost also has flat segments to the left of the MES up to the point in which it reaches $n_L^L = \epsilon$. At this point the average cost jumps to the level of the MES of the new optimal (and lower) number of layers. Depending on the value of ϵ this will imply that the firm might decide to drop more than one layer. If ϵ is low enough, the average cost curve will be a step function with no smoothing declining segments. The lower is ϵ, the easier it is for the firm to produce less quantity with more layers, and in the limit, as $\epsilon \to 0$, firms converge to $L = \infty$. This case is counterfactual since we observe that in most cases firms expand by adding one layer at the time (see Section 4).

[8] Of course, once we reintroduce materials and capital, changes in O are changes in total production Q instead, and the cost is the total average cost of the firm. However, the implications for revenue-based and quantity-based productivity are the same under our Cobb-Douglas production function specification.

[9] CRH show this result where markups are constant. More generally, the result holds for preferences such that the effect on prices is dominated by changes in marginal costs rather than by changes in markups.

productivity, but the quantity expansion decreases price and revenue-based productivity. When firms face negative shocks that make them drop layers we expect the opposite effects.

Figure 2: Quantity and Revenue Productivity Changes as a Firm Adds Layers



2.1.1 An example

To illustrate the mechanism described above we can use the example of a single-product firm producing aluminium cookware (anonymous given confidentiality requirements). It increased its workforce over time and, in particular, by 27 percent between 1996 and 1998. In the same period exports increased by 170%, and went from representing 10% of the firms sales in 1996 to 16% in 1998. Between 1997 and 1998 the firm reorganized and added a layer of management.

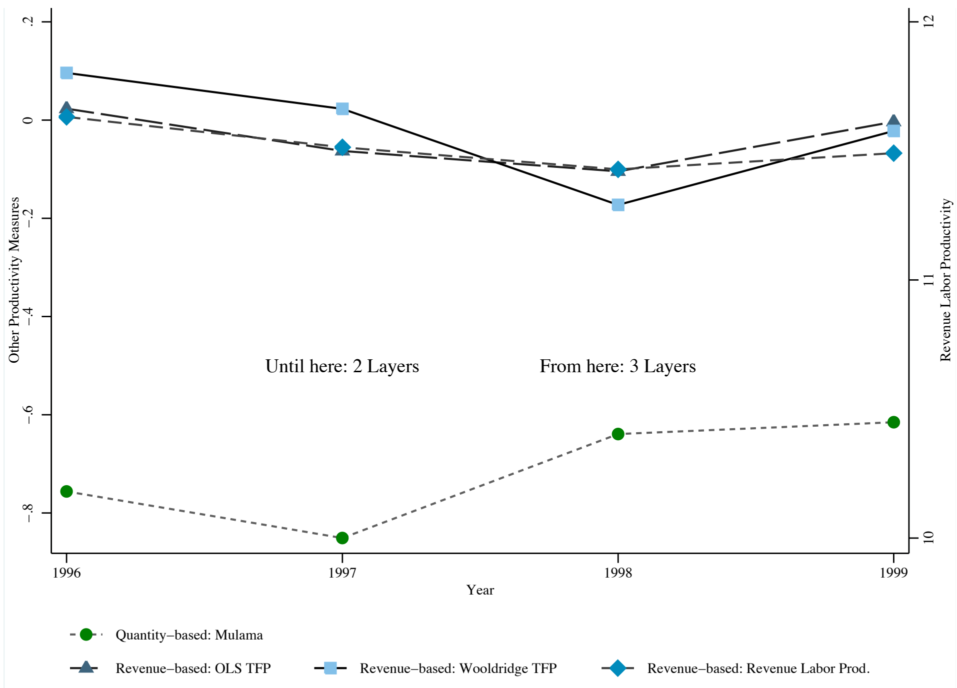
Our firm had a layer of workers and a layer of managers until 1997 and it added a new layer of management in 1998 (so it went from 1 layer to 2 layers of management). Figure 3 plots its quantity-based and revenue-based productivity around the reorganization (we plot 3 alternative measures of revenue-based productivity).[10] The pattern in the figure is typical in our data. The year in which the firm reorganizes its quantity-based productivity clearly jumps up and its revenue-based productivity declines. In contrast, it is hard to see any significant pattern in the changes in these measures of productivity for the year before or the year after adding the extra layer.

Figure 4 shows the corresponding levels of output, prices and revenue for the same firm and time period. The graph shows how, in fact, the increase in quantity-based productivity is accompanied by an increase in quantity, a fairly large decrease in price, and a small increase in revenue. These changes align exactly with our story in which the increase in quantity-based productivity generated by the reorganization (that adds a layer of management) leads to an increase in quantity, a lower marginal cost that leads to a decline in price, and a correspondingly muted increase in revenue and decline in revenue-based productivity. Note

[10] We describe the precise methodology and data used to measure both types of productivity in detail in Section 4.

that quantity in this firms grows not only at the time of the reorganization but before and after it as well. This is consistent with a firm that is progressively moving toward the quantity threshold in which it decides to reorganize. In these other years, demand and productivity shocks do not trigger a reorganization and so we do not see the corresponding decline in price.

Figure 3: An Example of a Firm that Adds Layers: Productivity Measures (logs)



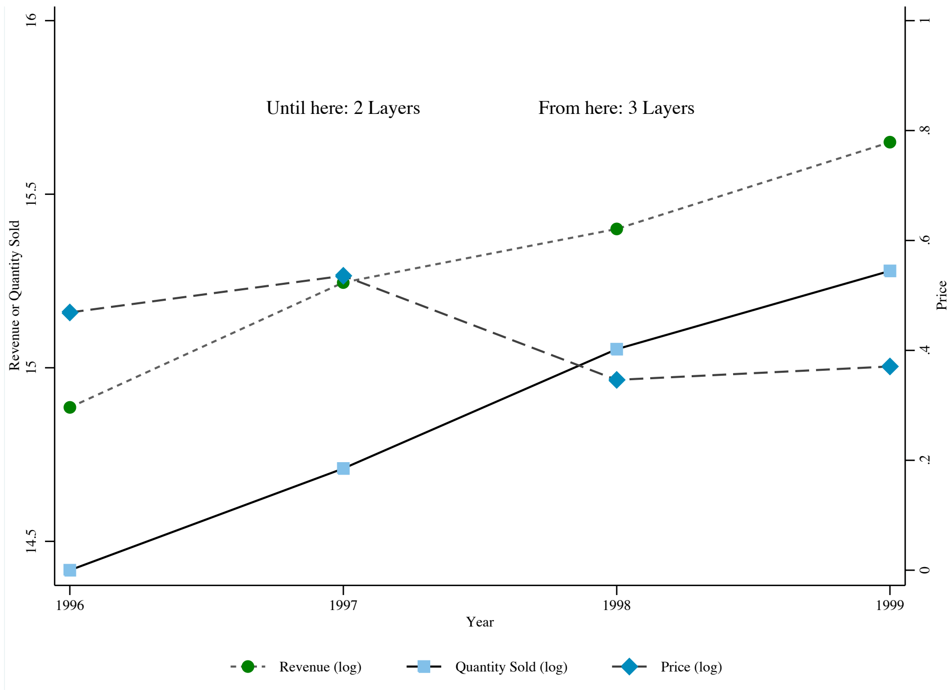
Of course, the case of this firm could be an isolated event in which all these variables happen to align in a way consistent with our interpretation. The rest of the paper is dedicated to present systematic evidence of the ubiquitousness of these exact patterns for quantity and revenue-based productivity as firms reorganize.

3 Data Description and Processing

Our data set is built from three data sources: a matched employer-employee panel data set, a firm-level balance sheet data set, and a firm-product-level data set containing information on the production of manufactured goods. Our data covers the manufacturing sector of continental Portugal for the years 1995-2005.[11] As explained below in detail, the matched employer-employee data virtually covers the universe of firms, while both the balance sheet data set and the production data set only cover a sample of firms. We build

[11]Information for the year 2001 for the matched employer-employee dataset was not collected. Hence, our sample excludes the year 2001 (see Appendix A).

Figure 4: An Example of a Firm that Adds Layers: Output, Price, and Revenue



two nested samples. The largest of them sources information from the matched employer-employee data set for the subset of firms for which we also have balance sheet data. We refer to this sample as Sample 1. It contains enough information to calculate measures of revenue-based productivity at the firm-year-level. The second sample covers a further subset of firms for which we also have production data. This data is necessary to calculate quantity-based productivity at the firm-product-year-level. We refer to this sample as Sample 2. All our revenue-based productivity results below hold similarly well for both samples, although we present mostly results using Sample 2 in order to make results more easily comparable.

Employer-employee data come from *Quadros de Pessoal* (henceforth, QP), a data set made available by the Ministry of Employment of Portugal, drawing on a compulsory annual census of all firms in Portugal that employ at least one worker.[12] Currently, the data set collects data on about 350,000 firms and 3 million employees. Reported data cover the firm itself, each of its plants, and each of its workers. Each firm and each worker entering the database are assigned a unique, time-invariant identifying number which we use to follow firms and workers over time. Variables available in the data set include the firm's location, industry, total

[12] Public administration and non-market services are excluded. *Quadros de Pessoal* has been used by, amongst others, Blanchard and Portugal (2001) to compare the U.S. and Portuguese labor markets in terms of unemployment duration and worker flows; by Cabral and Mata (2003) to study the evolution of the firm size distribution; by Mion and Opromolla (2014) to show that the export experience acquired by managers in previous firms leads their current firm towards higher export performance, and commands a sizeable wage premium for the manager.

employment, and sales. The worker-level data cover information on all personnel working for the reporting firms in a reference week in October of each year. They include information on occupation, earnings, and hours worked (normal and overtime). The information on earnings includes the base wage (gross pay for normal hours of work), seniority-indexed components of pay, other regularly paid components, overtime work, and irregularly paid components. It does not include employers' contributions to social security.[13]

The second data set is *Central de Balanços* (henceforth, CB), a repository of yearly balance sheet data for non financial firms in Portugal. Prior to 2005 the sample was biased towards large firms. However, the value added and sales coverage rate was high. For instance, in 2003 firms in the CB data set accounted for 88.8 percent of the national accounts total of non-financial firms' sales. Information available in the data set includes a firm sales, material assets, costs of materials, and third-party supplies and services.

The third data set is the *Inquérito Anual à Produção Industrial* (henceforth, PC), a data set made available by Statistics Portugal (INE), containing information on sales and volume sold for each firm-product pair for a sample of firms with at least 20 employees covering at least 90 percent of the value of aggregate production. From PC we use information on the volume and value of a firm's production. The volume is recorded in units of measurement (number of items, kilograms, liters) that are product-specific while the value is recorded in current euros. From the raw data it is possible to construct different measures of the volume and value of a firm's production. For the sake of this project we use the volume and value corresponding to a firm's sales of its products. This means that we exclude products produced internally and to be used in other production processes within the firm as well as products produced for other firms, using inputs provided by these other firms. The advantage of using this definition is that it nicely corresponds to the cost of materials coming from the balance sheet data. For example, the value of products produced internally and to be used in other production processes within the firm is part of the cost of materials while products produced for other firms, using inputs provided by these other firms, is neither part of the cost of materials nor part of a firm's sales from the PC data. We aggregate products at the 2-digits-unit of measurement pairs and split multi-products firms into several single product firms using products revenue shares as weights (see Appendix A).

3.1 Occupational structure

To recover the occupational structure at the firm level we exploit information from the matched employer-employee data set. Each worker, in each year, has to be assigned to a category following a (compulsory) classification of workers defined by the Portuguese law.[14] Classification is based on the tasks performed

[13]The Ministry of Employment implements several checks to ensure that a firm that has already reported to the database is not assigned a different identification number. Similarly, each worker also has a unique identifier, based on a worker's social security number. The administrative nature of the data and their public availability at the workplace—as required by the law—imply a high degree of coverage and reliability. It is well known that employer-reported wage information is subject to less measurement error than worker-reported data. The public availability requirement facilitates the work of the services of the Ministry of Employment that monitor the compliance of firms with the law.

[14]Following CMRH we use occupational categories to identify layers of management. In the case of French firms, CMRH use the PCS classification. In this study we use the Portuguese classification (Decreto Lei 121/78 of July 2^{nd} 1978) which is not

and skill requirements, and each category can be considered as a level in a hierarchy defined in terms of increasing responsibility and task complexity. Table A.1 in Appendix A contains more detail about the exact construction of these categories.

On the basis of the hierarchical classification and taking into consideration the actual wage distribution, we partition the available categories into management layers. We assign “Top executives (top management)” to occupation 3; “Intermediary executives (middle management)” and “Supervisors, team leaders” to occupation 2; “Higher-skilled professionals” and some “Skilled professionals” to occupation 1; and the remaining employees, including “Skilled professionals”, “Semi-skilled professionals”, “Non-skilled professionals”, and “Apprenticeship” to occupation 0.

We then translate the number of different occupations present in a firm into layers of management. A firm reporting c occupational categories will be said to have $L = c - 1$ layers of management: hence, in our data we will have firms spanning from 0 to 3 layers of management (as in CMRH). In terms of layers within a firm we do not keep track of the specific occupational categories but simply rank them. Hence a firm with occupational categories 2 and 0 will have 1 layer of management, and its organization will consist of a layer 0 corresponding to some skilled and non-skilled professionals, and a layer 1 corresponding to intermediary executives and supervisors.[15]

Table 1 presents some basic statistics for Sample 1 for the ten years spanned by our data. The data exhibits some clear trends over time. In particular, the number of firms declines and firms tend to become larger. In all our regressions we control for time and industry fixed effects.

Table 1: Firm-level data description by year

		Mean			
Year	Firms	Value Added	Hours	Wage	# of layers
1996	8,061	1,278	102,766	4.37	1.25
1997	8,797	1,227	91,849	4.48	1.20
1998	7,884	1,397	96,463	4.81	1.28
1999	7,053	1,598	105,003	4.93	1.31
2000	4,875	2,326	139,351	5.13	1.62
2002	4,594	2,490	125,392	5.63	1.62
2003	4,539	2,363	124,271	5.65	1.70
2004	4,610	2,389	124,580	5.82	1.74
2005	3,962	2,637	129,868	6.01	1.76

Notes: Value added in 2005 euros. Wage is average hourly wage in 2005 euros.

the ISCO.

[15] One potential concern with this methodology to measure the number of layers is that many firms will have layers with occupations that are not adjacent in the rank. This does not seem to be a large problem. More than 75% of firms have adjacent layers.

4 Portuguese Production Hierarchies: Basic Facts

In this section we reproduce some of the main results in CMRH for France using our data for Portugal in Sample 1. These results underscore our claim that the concept of layers we use is meaningful. We show this by presenting evidence that shows, first, that firms with different numbers of layers are systematically different in a variety of dimensions; second, that firms change layers in a systematic and expected way; third, that the workforce within a layer responds as expected as firms add or subtract layers. This evidence makes us confident that interpreting the adding and dropping of layers in data as a firm reorganization is warranted by the evidence.

Table 2: Firm-level data description by number of layers

# of layers	Firm-years	Mean			Median
		Value added	Hours	Wage	Wage
0	14,594	267.2	12,120.7	3.55	3.16
1	14,619	648.4	31,532.0	4.03	3.64
2	12,144	2,022.7	96,605.2	4.51	4.11
3	13,018	10,286.2	327,166.8	5.73	5.20

Notes: Value added in 000s of 2005 euros. Wage is either average or median hourly wage in 2005 euros. Hours are yearly.

Table 2 presents the number of firm-year observations by number of management layers as well as average value added, hours, and wages. It also presents the median wage given that the wage distribution can be sometimes very skewed. The evidence clearly shows that firms with more layers are larger in terms of value added and hours. It also shows that firms with more layers pay on average higher wages.

Figures 5 to 7 present the distributions of value added, employment and the hourly wage by layer. The distributions are clearly ordered. The distributions for firms with more layers are shifted to the right and exhibit higher variance. In Figure 6 the modes in the distribution of hours corresponds to the number of hours of one full-time employee, two full-time employees, etc. The figures show that firms with different numbers of layers are in fact very different. The notion of layers seems to be capturing a stark distinction among firms.

Our definition of layers of management is supposed to capture the hierarchical structure of the firm. So it is important to verify that the implied hierarchies are pyramidal in the sense that lower layers employ more hours and pay lower hourly wages. Table 3 shows that the implied hierarchical structure of firms is hierarchical in the majority of cases. Furthermore, the implied ranking holds for 76% of the cases when comparing any individual pair of layers. Similarly, Table 4 shows that lower layers command lower wages in the vast majority of cases. We conclude that, although perhaps with some imprecision, our definition of layers does a good job in capturing the hierarchical structure of firms.

Our primary goal is to study the endogenous productivity responses of firm that reorganize. So it is

Figure 5: Value Added Density

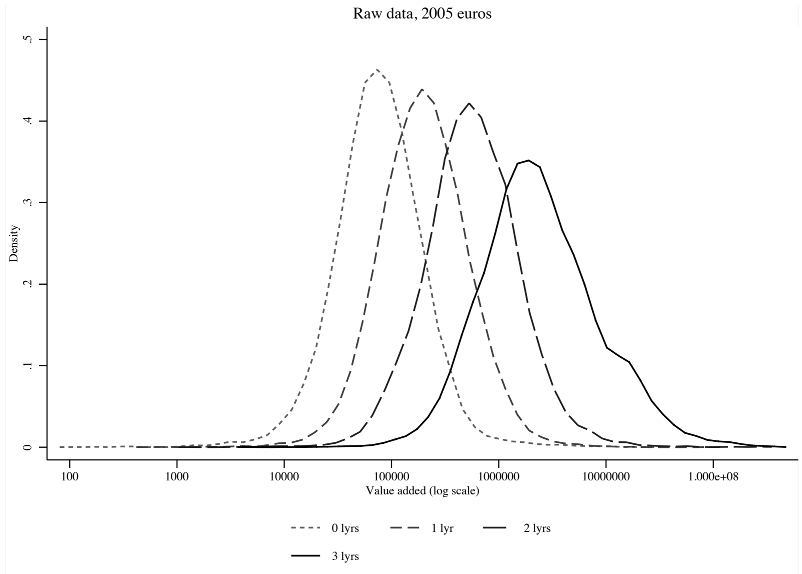
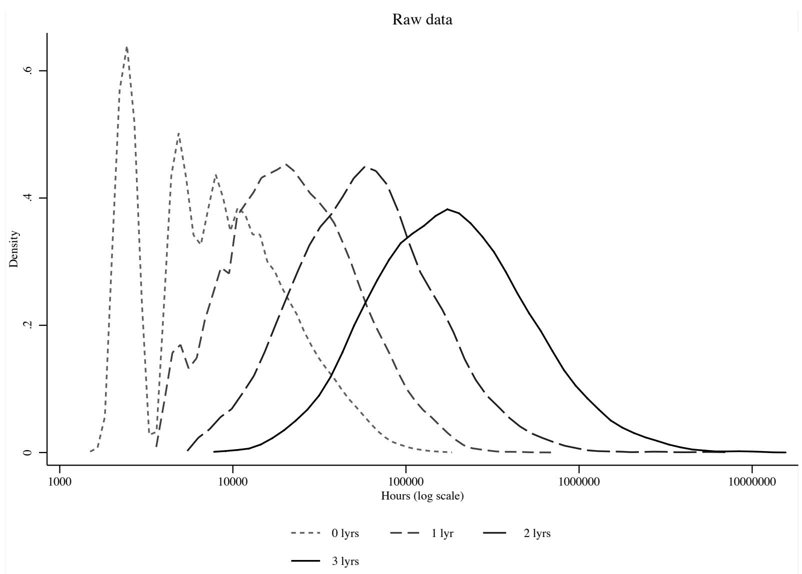


Figure 6: Employment (Hours) Density



important to establish how often they do so. Table 5 presents a transition matrix across layers. In a given year about half the total number of firms keep the same number of layers, with the number increasing to 70% for firms with 4 layers (3 layers of management). Most of the firms that do not reorganize just exit, with the percentage of exiting firms declining with the number of layers. About 12% of firms in a layer reorganize

Figure 7: Hourly Wage Density

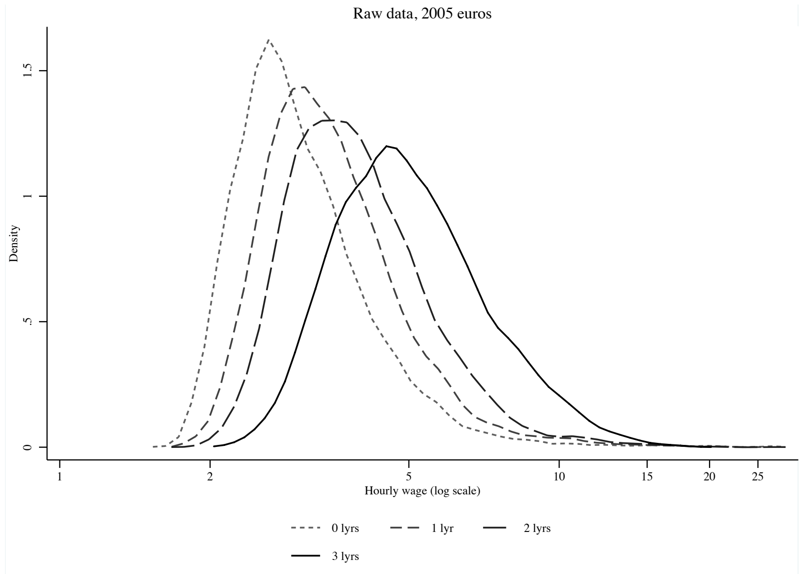


Table 3: Percentage of firms that satisfy a hierarchy in hours

# of layers	$N_L^l \geq N_L^{l+1}$ all l	$N_L^0 \geq N_L^1$	$N_L^1 \geq N_L^2$	$N_L^2 \geq N_L^3$
1	91.64	91.64	–	–
2	69.62	92.07	77.35	–
3	50.51	88.70	74.34	83.65

N_L^l = hours at layer l of a firm with L layers.

by adding a layer, and about the same number downscale and drop one. Overall, as in France, there seem to be many reorganizations in the data. Every year around 20% of firms add and drop occupations, and therefore restructure their labor force (the number is lower for firms with 3 layers of management since, given that the maximum number of management layers is 3, they can only drop layers).

A reorganization is accompanied with many other firm-level changes. In Table 6 we divide firms depending on whether they add, do not change, or drop layers, and present measured changes in the total number

Table 4: Percentage of firms that satisfy a hierarchy in wages

# of layers	$w_L^l \leq w_L^{l+1}$ all l	$w_L^0 \leq w_L^1$	$w_L^1 \leq w_L^2$	$w_L^2 \leq w_L^3$
1	75.87	75.87	–	–
2	65.66	85.21	79.57	–
3	67.11	92.36	84.62	87.82

Table 5: Distribution of layers at $t+1$ conditional on layers at t

		# of layers at $t+1$					
		Exit	0	1	2	3	Total
# of layers at t	0	31.19	54.29	12.54	1.69	0.29	100.00
	1	25.75	10.26	51.12	11.35	1.51	100.00
	2	21.73	1.49	12.06	49.62	15.09	100.00
	3	15.68	0.37	1.46	12.90	69.59	100.00
	New	85.08	5.31	3.77	3.01	2.83	100.00

of hours, number of hours normalized by the number of hours in the top layer, value added, and average wages. For all these measures we present changes after de-trending in order to control for the time trends in the data that we highlighted before. First, note that firms that either expand or contract substantially tend to reorganize. This is the case both in terms of hours or in terms of value added. Furthermore, changes in either hours or value added seem to be symmetric, but with opposite sign, for firms that add and drop layers. Finally, firms that add layers tend to pay higher wages. However, once we de-trend, it is clear that wages in the preexisting layers decline. So average wages increase because the agents in the new layer earn more than the average but workers in preexisting layers earn less as their knowledge is now less useful (as found for France in CMRH).

Table 6: Changes in firm-level outcomes

# of layers	All	Increase L	No Change in L	Decrease L
dln total hours	-0.0068^a	0.2419^a	-0.0080^a	-0.2992^a
- detrended		0.2472^a	-0.0011	-0.2911^a
dln normalized hours	0.0099^b	1.0890^a	-0.0204^a	-1.1043^a
- detrended		1.0761^a	-0.0299^a	-1.1128^a
dlnVA	0.0173^a	0.0509^a	0.0155^a	-0.0126^a
- detrended		0.0323^a	-0.0013	-0.0307^a
dln avg. wage	0.0369^a	0.0683^a	0.0348^a	0.0122^a
- detrended		0.0303^a	-0.0018^c	-0.0253^a
common layers	0.0356^a	0.0068^b	0.0348^a	0.0750^a
- detrended		-0.0295^a	-0.0005	0.0387^a

Notes: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

The results above can be further refined by looking at layer-level outcomes for firms that expand without reorganizing and firms that expand as a result of a reorganization. The theory predicts that firms that expand but keep the same number of layers will increase employment and wages in all layers. In contrast, firms that expand and add layers, will increase employment in all layers but will decrease wages (and according to the theory, knowledge) in all preexisting layers. That is, adding a layer allows the firm to economize on the knowledge of all the preexisting layers. Tables 7 and 8 present the elasticity of normalized hours (hours at each layer relative to the top layer) and wages, respectively, to value added for firms that do not add layers. The first column indicates the number of layers in the firm, and the second the particular layer for which the elasticity is calculated. The theory predicts that all elasticities should be positive. This prediction is confirmed for all elasticities except for one case where the estimate is not significant. Hence, we conclude that firms that grow without reorganizing increase employment and wages in all layers.

Table 7: Elasticity of n_L^ℓ with respect to value added for firms that do not change L

# of layers	Layer	Elasticity	# observations
1	0	0.1155^a	6,351
2	0	0.1146^a	4,998
2	1	-0.0147	4,998
3	0	0.1760^a	7,079
3	1	0.0847^a	7,079
3	2	0.0987^a	7,079

Notes: Robust standard errors in parentheses: [a] p<0.01, [b] p<0.05, [c] p<0.1.

Table 8: Elasticity of w_L^ℓ with respect to value added for firms that do not change L

# of layers	Layer	Elasticity	# observations
0	0	0.0056	6,987
1	0	0.0216^a	6,351
1	1	0.0283^a	6,351
2	0	0.0150^b	4,998
2	1	0.0229^b	4,998
2	2	0.0303^b	4,998
3	0	0.0225^a	7,079
3	1	0.0201^a	7,079
3	2	0.0298^a	7,079
3	3	0.0199^b	7,079

Notes: Robust standard errors in parentheses: [a] p<0.01, [b] p<0.05, [c] p<0.1.

Tables 9 and 10 show changes in normalized hours and wages when firms reorganize. The tables show the total number of layers before and after the reorganization, as well as the layer for which the log-change is computed. As emphasized before, adding layers should lead to increases in employment but declines in wages in all preexisting layers. These implications are verified for all transitions in all layers except for two non-significant results for firms that start with zero layers of management. Similar to the results in CMRH for France, our estimates for Portugal show that firms that add layers in fact concentrate workers' knowledge, as proxied by their wages, on the top layers. This is one of the consequences of a firm reorganization and supports empirically the underlying mechanism that, we hypothesize, leads to an increase (decrease) in quantity-based productivity as a result of a reorganization that adds (drops) layers.

Table 9: $d \ln n^{\ell}_{Lit}$ for firms that transition

# of layers before	# of layers after	Layer	$d \ln n^{\ell}_{Lit}$	# observations
0	1	0	1.2777^a	1,614
0	2	0	1.6705^a	218
0	3	0	2.3055^a	37
1	0	0	-1.2304^a	1,275
1	2	0	0.5178^a	1,410
1	2	1	0.4920^a	1,410
1	3	0	0.9402^a	188
1	3	1	0.8367^a	188
2	0	0	-1.6449^a	150
2	1	0	-0.5645^a	1,215
2	1	1	-0.5060^a	1,215
2	3	0	0.6806^a	1,520
2	3	1	0.7098^a	1,520
2	3	2	0.6340^a	1,520
3	0	0	-2.5187^a	38
3	1	0	-0.9772^a	149
3	1	1	-0.8636^a	149
3	2	0	-0.7977^a	1,312
3	2	1	-0.7532^a	1,312
3	2	2	-0.6465^a	1,312

Notes: Robust standard errors in parentheses: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

5 Changes in Productivity

We now present our methodology to measure changes in revenue-based and quantity-based productivity. The measurement of revenue productivity has received a lot of attention in the industrial organization literature, and so we expand standard methodologies to account for the role of layers. Measuring quantity-based productivity is more involved and requires more detailed data. We address each measurement exercise in turn.[16]

We use two complementary approaches to select comparable groups of firms. The first approach uses sequences of firm-years with either one or zero changes in layers. We define a sequence of type $L - L'$ as the

[16] The main insight in CRH is that reorganization increases quantity-based-productivity via a reduction in marginal costs. This in turn translates into a reduction in prices that lowers revenue-based productivity. Therefore, with information on prices one might be tempted to look at how firm level prices change as firms change their organization. However, prices change as a consequence of supply side shifters, like costs, as well as other demand side shifters, like markups and taste shocks. As a result, prices might be a noisy measure of firms' performance. We instead focus on measuring changes in quanty-based productivity and use the methodology in MULAMA to account for different demand and supply side shocks.

Table 10: $d \ln w^{\ell}_{Lit}$ for firms that transition

# of layers before	# of layers after	Layer	$d \ln w^{\ell}_{Lit}$	# observations
0	1	0	0.0062	1,614
0	2	0	0.0207	218
0	3	0	-.1878^{a}	37
1	0	0	0.0557^{a}	1,275
1	2	0	0.0038	1,410
1	2	1	-0.0624^{a}	1,410
1	3	0	-0.0230^{a}	188
1	3	1	-0.1710^{a}	188
2	0	0	0.0692^{a}	150
2	1	0	0.0373^{a}	1,215
2	1	1	0.1192^{a}	1,215
2	3	0	-0.0015	1,520
2	3	1	-0.0113^{b}	1,520
2	3	2	-0.0676^{a}	1,520
3	0	0	0.2673^{a}	38
3	1	0	0.0691^{a}	149
3	1	1	0.1672^{a}	149
3	2	0	0.0313^{a}	1,312
3	2	1	0.0467^{a}	1,312
3	2	2	0.1114^{a}	1,312

Notes: Robust standard errors in parentheses: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

series of years in which a firm has the same consecutively observed number of management layers L plus the adjacent series of years in which a firm has the same consecutively observed number of management layers L'. For example, a firm that we observed all years between 1996 and 2000 and that has zero layers in 1996, 1997, and 2000 and one layer in 1998 and 1999 would have two sequences: A 0-1 sequence (1996 to 1999) as well as a 1-0 sequence (1998 to 2000). Firms that never change layers in our sample form a constant-layer sequence. We then calculate our results using firm-product-sequence fixed effects for Sample 2 and firm-sequence fixed-effects for Sample 1.

The second approach groups firms according to the number of layers (0, 1, 2 or 3) the firm has in the year of the first observed reorganization. We then present results for each of these groups using firm-product fixed effects. On top of these fixed effects, in both approaches we use a battery of industry or product group dummies as well as time dummies. Throughout, standard errors are clustered at the firm-level. Bootstrapped standard errors are virtually identical to firm-clustered ones.

5.1 Revenue-based Productivity

The Cobb-Douglas production function for firm i in period t introduced in equation (1) can be expressed in logs as

$$q_{it} = a_{it} + \alpha_O o_{it} + \alpha_M m_{it} + (\gamma - \alpha_M - \alpha_O) k_{it} \tag{6}$$

where a_{it} denotes productivity, o_{it} the log of the labor input, m_{it} the log of materials, and k_{it} the log of capital.

The labor input O_{it} is not directly observable, but we can use the fact that

$$O_{it} = \frac{C\left(O_{it}; w\right)}{C\left(O_{it}; w\right)} O_{it} = \frac{C\left(O_{it}; w\right)}{AC\left(O_{it}; w\right)}. \tag{7}$$

The numerator of this expression, $C\left(O_{it}; w\right)$, is the total expenditure on the labor input, i.e., the total wage bill of the firm (which is observable in standard data) while the denominator, $AC\left(O_{it}; w\right) = C\left(O_{it}; w\right) / O_{it}$, is the unit cost of the labor input (which is, by contrast, unobservable). Substituting this into the production function and multiplying by the price leads to an equation for revenue given by

$$r_{it} = \bar{a}_{it} + \alpha_O \ln C\left(O_{it}; w\right) + \alpha_M m_{it} + (\gamma - \alpha_M - \alpha_O) k_{it}, \tag{8}$$

where $\bar{a}_{it} \equiv p_{it} + a_{it} - \alpha_O \ln AC\left(O_{it}; w\right) = p_{it} + a_{it} + \beta L_{it}$ denotes revenue-based productivity and $C\left(O_{it}; w\right)$ is the total wage bill of the firm. Note that $-\alpha_O \ln AC\left(O_{it}; w\right) = \beta L_{it}$ is what is implied by the CRH model if we substitute the constraint (5) in the organizational problem with $n_L^L \geq \epsilon$, for small enough $\epsilon > 0$. Therefore, we assume that revenue-based productivity follows the process

$$\bar{a}_{it} = \phi_a \bar{a}_{it-1} + \beta L_{it} + \nu_{ait}, \tag{9}$$

where ν_{ait} is a productivity shock that is i.i.d. across firms and time, uncorrelated with all past values of $\bar{a}_{it}$ and L_{it} but correlated with the number of layers in t. Indeed, a firm will optimally choose the number of layers in t depending on, among others, the realization of ν_{ait}.

We also assume that capital is predetermined in t and that firms optimally choose materials and the labor input in order to minimize short-run costs. The cost of materials is common across firms but can vary over time while the unit cost of the labor input $AC\left(O_{it}; w\right)$ varies across firms and time. From first-order cost minimization conditions we have that materials' choice m_{it} is a function of k_{it} and $\bar{a}_{it}$. After inverting the first-order conditions of the firm we can express $\bar{a}_{it}$ as a function of capital and materials, namely, $\bar{a}_{it} = g(k_{it}, m_{it})$.

Using revenue (or value added) data we can then estimate the revenue-based productivity process, $\bar{a}_{it}$, which should decline with the number of layers, as described above. The literature has proposed several ways to estimate revenue productivity. Below we use simple measures as revenue per worker as well as the methodology of Wooldridge (2009) to deal with the endogeneity in input use. Similar results can be

Table 11: Revenue Labor Productivity. Firm-product-sequence FE

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.073^b	-0.016	0.090^c	0.055^b
	(0.035)	(0.059)	(0.053)	(0.027)
Number of management layers t	-0.090^a	-0.258^a		-0.151^a
	(0.017)	(0.040)		(0.018)
Constant	10.795^a	11.924^a	10.527^a	11.109^a
	(0.393)	(0.659)	(0.617)	(0.303)
Observations	4,206	2,750	3,090	10,046
Number of fixed effects	1,687	1,289	1,310	4,286
Adjusted R^2	0.029	0.151	0.026	0.060

Firm-level clustered standard errors in parentheses

^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

obtained using either value added per worker, OLS Revenue TFP or the methodology in Olley and Pakes (1996), Levinsohn and Petrin (2003), and DeLoecker and Warzynski (2012). See Appendix A for further details.

The first set of results is presented in Table 11. The measure of revenue-based productivity is revenue per worker and we present results for sequences in which firms increase, decrease, or keep constant the number of layers. We also present results for the whole sample. The number of layers is significantly related to lower revenue-based productivity. Revenue-based productivity decreases by about 9% when firms add a layer and increases by about 25% when they drop one. Note that past revenue-based productivity has a small effect on current revenue-based productivity given that we are using firm-product-sequence fixed effects.[17]

Table 12 presents the results for the alternative grouping. It presents the effect of layers on productivity as a function of the number of layers after we see the firm switching for the first time. The coefficient on layers is again negative and highly significant for firms with any initial number of layers. Revenue productivity declines by about 12% when firms add a layer, although the number increases to 18% for the largest firms with 3 layers (this last coefficient is identified only from firms that drop layers since 3 layers is the maximum number of layers in our measure).

Of course, there are many potential objections to using revenue per worker as the relevant measure of revenue productivity. In particular, as Olley and Pakes (1996) famously argued, there are other inputs and their use is endogenous to the level of productivity of the firm. Levinsohn and Petrin (2003) propose a methodology to deal with this problem using the first-order conditions of the firm input choice problem with respect to the variable inputs.

We use the variant of this methodology proposed by Wooldridge (2009).[18] If we subtract the value

[17] This result also holds for all subsequent Tables. When not using fixed effects we find standard auto-regressive coefficients of around 0.8 and R^2 of about 0.7.

[18] Wooldridge (2009) builds on Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2007) and shows

Table 12: Revenue Labor Productivity. Firm-product FE

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	-0.070	0.171^a	0.107^b	0.016
	(0.059)	(0.047)	(0.048)	(0.067)
Number of management layers t	-0.103^b	-0.121^a	-0.115^a	-0.187^a
	(0.045)	(0.023)	(0.021)	(0.037)
Constant	12.017^a	9.587^a	10.446^a	11.891^a
	(0.674)	(0.537)	(0.551)	(0.797)
Observations	533	1,665	3,690	3,569
Number of fixed effects	164	552	1,157	1,161
Adjusted R^2	0.068	0.087	0.061	0.093

Firm-level clustered standard errors in parentheses

^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

of materials from revenue to obtain value added, va_{it}, we can estimate a value added-based production function featuring two inputs (the labor input and capital) and one proxy variable (materials). Building on the production function in (8) and given (9) as well as the inverted input demand equation of the firm, $\bar{a}_{it-1} = g(k_{it-1}, m_{it-1})$, we obtain

$$va_{it} = \alpha_O \ln C\left(O_{it}; w\right) + (\gamma - \alpha_M - \alpha_O)k_{it} + \phi_a g(k_{it-1}, m_{it-1}) + \beta L_{it} + \nu_{ait}. \tag{10}$$

Under our assumptions, the error term ν_{ait} in (9) is uncorrelated with k_{it} and m_{it-1}. Hence, $g(k_{it-1}, m_{it-1})$ is also uncorrelated with ν_{ait}. The wage bill $C\left(O_{it}; w\right)$ and the number of layers at t, L_{it}, are instead endogenous and we instrument them with the wage bill at $t-1$ and the number of layers in $t-1$. As for the term $\phi_a g(k_{it-1}, m_{it-1})$ we use a second order polynomial approximation in k_{it-1} and m_{it-1}. We finally estimate (10) by IV and ultimately get an estimate of revenue TFP as

$$\widehat{\bar{a}} = va_{it} - \hat{\alpha}_O \ln C\left(O_{it}; w\right) - \hat{\alpha}_K k_{it}.$$

Table 13 shows that adding a layer reduces revenue-based productivity by about 5%, while dropping one increases it slightly less than 14%. The overall effect of changing a layer is 8% and highly significant. These results are qualitatively the same, but quantitatively somewhat smaller than the ones we obtained for revenue labor productivity.

When grouping according to the initial number of layers, Table 14 shows that the effect of an extra layer decreases revenue-based productivity by between 4 and 9% with the number increasing for higher layers (which is probably the result of downward transitions having larger effects than upward transitions). The

how to obtain consistent estimates of input elasticities with a one-step GMM procedure. The results are qualitatively and quantitatively very similar if we simply implement the methodology in Olley and Pakes (1996).

Table 13: Wooldridge Revenue TFP. Firm-product-sequence FE

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.092[b]	-0.077[c]	0.070	0.044[c]
	(0.041)	(0.046)	(0.044)	(0.026)
Number of management layers t	-0.049[a]	-0.138[a]		-0.082[a]
	(0.014)	(0.022)		(0.011)
Constant	0.056	0.137[a]	-0.025	0.261[a]
	(0.041)	(0.039)	(0.016)	(0.029)
Observations	4,127	2,708	2,989	9,824
Number of fixed effects	1,655	1,268	1,276	4,199
Adjusted R^2	0.057	0.063	0.032	0.042

Firm-level clustered standard errors in parentheses

[a] $p<0.01$, [b] $p<0.05$, [c] $p<0.1$

Table 14: Wooldridge Revenue TFP. Firm-product FE

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	-0.041	0.011	0.123[a]	0.139[a]
	(0.061)	(0.060)	(0.039)	(0.046)
Number of management layers t	-0.063[a]	-0.039[c]	-0.067[a]	-0.093[a]
	(0.022)	(0.021)	(0.013)	(0.024)
Constant	0.096[b]	0.023	0.108[a]	0.319[a]
	(0.040)	(0.055)	(0.032)	(0.069)
Observations	528	1,630	3,645	3,446
Number of fixed effects	162	543	1,138	1,122
Adjusted R^2	0.044	0.026	0.060	0.057

Firm-level clustered standard errors in parentheses

[a] $p<0.01$, [b] $p<0.05$, [c] $p<0.1$

results are again consistent with our findings for revenue per worker, although clearly smaller in magnitude.

Overall these results paint a very consistent picture. Revenue-based productivity decreases with a reorganization that adds layers by somewhat between 4 and 9%. The result varies somewhat by layer and depending on whether firms add or drop layers and their initial number of layers. Furthermore, taking care of multiple inputs and adjusting for the endogenous choice of materials and labor is important as well, and reduces the absolute magnitude of the estimated effect of a reorganization. Nevertheless, the main result that revenue productivity jumps in the opposite direction as the number of layers is very robust across specifications and exercises. Clearly, revenue productivity can jump down either because firms reduce their prices or, perhaps, because quantity-base productivity goes down (a result that would contradict our hypothesis). Thus, we now proceed to estimate the effect of a reorganization on quantity-based productivity.

5.2 Quantity-based Productivity

Measuring quantity-based productivity well is hard partly because we need information about prices and quantities. And partly because we need to account properly for other types of shocks that firms face like demand and markup shocks that will affect its pricing and production decisions. The empirical implications derived from the theory that we outlined above crucially depends on accounting well for heterogeneity and shocks in demand and their potential correlation with exogenous productivity and other shocks. So it is important to choose a methodology that allows for correlation between demand and productivity disturbances. To do so we follow Forlani et al., (2014) that impose sufficient structure to compute quantity-based productivity in the presence of potentially correlated productivity, demand and markup shocks. We start with a description of the baseline methodology and subsequently expand on its application by explicitly considering the role of organization in producing the labor input.

5.2.1 Baseline MULAMA

Following Forlani et al. (2014) we use a two-stage estimation procedure to obtain quantity-based productivity. Our approach allows us to explicitly take into account the presence of demand shocks and markup heterogeneity across firms. We do this by both assuming costs minimization, which provides a useful way of computing markups as in De Loecker and Warzynski (2011), and by imposing some restrictions on the way demand shocks enter utility. In our case we simply have that log revenue is given by $r_{it} = p_{it} + q_{it} = \frac{1}{\mu_{it}}(\lambda_{it} + q_{it})$, where λ_{it} is the demand shock to firm i at time t. That is, demand shocks enter log-linearly in the revenue equation, along with quantity, and both multiply the inverse of the markup. Forlani et al. (2014) show how this holds as a first-order linear approximation in a variety of circumstances.

Recall the production function in equation (6) and assume the following quantity-based productivity process: $a_{it} = \phi_a a_{it-1} + \nu_{ait}$. Furthermore, assume that demand for the product of firm i at time t, in logs, is given by

$$p_{it} = \left(1 - \frac{1}{\eta_{it}}\right)\lambda_{it} - \frac{1}{\eta_{it}}q_{it}$$

where λ_{it} denotes the level of demand and η_{it} the elasticity of demand. We assume that

$$\lambda_{it} = \phi_\lambda \lambda_{it-1} + \nu_{\lambda it},$$

where $\nu_{\lambda it}$ is an idiosyncratic demand shock that can be correlated with the productivity shock ν_{ait}. The firm operates in a monopolistically competitive market, so its markup is given by

$$\mu_{it} = \frac{\eta_{it}}{\eta_{it} - 1}.$$

Thus, heterogeneity or shocks in the elasticity of demand will result in variation in markups across firms and time.

Cost minimization implies, for the flexible inputs (which in our model are the labor input and materials), that

$$\frac{\alpha_O}{\mu_{it}} = \frac{O_{it}MC\left(O_{it};w\right)}{P_{it}Q_{it}},$$

and

$$\frac{\alpha_M}{\mu_{it}} = \frac{M_{it}P_{Mt}}{P_{it}Q_{it}},$$

where $MC\left(O_{it};w\right)$ is the marginal cost of the labor input and P_{Mt} denotes the price of materials. Note that $M_{it}P_{Mt}$ is the total expenditure on materials, and $O_{it}MC\left(O_{it};w\right)$ is the total expenditure on the labor input when we allow $n_L^L \geq \epsilon$, for $\epsilon > 0$ (since the unit and marginal cost coincide at the MES). Given that we observe the total expenditure on the labor input, this is the case that we consider throughout.

With this structure in hand, we can proceed to measure quantity-based productivity using data on quantities, revenue, labor, capital, material and expenditure shares.

Denote

$$LHS_{it} = \frac{r_{it} - s_{Oit}\left(o_{it} - k_{it}\right) - s_{Mit}\left(m_{it} - k_{it}\right)}{s_{Mit}},$$

where s_{xit} is the share in expenditure of input x. After some manipulations of the revenue equation we can obtain an expression for LHS_{it} that we can estimate in the first stage. Namely,

$$LHS_{it} = b_1 z_{1it} + b_2 z_{2it} + b_3 z_{3it} + b_4 z_{4it} + b_5 z_{5it} + u_{it}, \tag{11}$$

where $z_{1it} = k_{it}$, $z_{2it} = LHS_{it-1}$, $z_{3it} = k_{it-1}$, $z_{4it} = \frac{r_{it-1}}{s_{Mit-1}}$, $z_{5it} = q_{it-1}$, $u_{it} = \left(\nu_{ait} + \nu_{\lambda it}\right)/\alpha_M$. Appendix B presents a detailed derivation of equation (11). Note that we can simply use OLS to estimate equation (11) since u_{it} is not correlated with the covariates. This equation allows us to identify several of the model's parameters. From the estimates of this equation we can identify all the parameters since $\hat{b}_1 = \frac{\gamma}{\alpha_M}$, $\hat{b}_2 = \phi_a$, $\hat{b}_3 = -\phi_a \frac{\gamma}{\alpha_M}$, $\hat{b}_4 = \phi_\lambda - \phi_a$, and $\hat{b}_5 = \frac{-\phi_\lambda + \phi_a}{\alpha_M}$.

Using $\hat{b}_1$ and $\hat{b}_2$ we can implement a second stage to separately identify γ where we use the productivity process and the production function to obtain

$$q_{it} - \hat{b}_2 q_{it-1} = b_6 z_{6it} + \nu_{ait}$$

where

$$z_{6it} = \frac{o_{it} - k_{it}}{\hat{b}_1}\frac{s_{Oit}}{s_{Mit}} + \frac{m_{it} - k_{it}}{\hat{b}_1} + k_{it} + \frac{\hat{b}_2}{\hat{b}_1}LHS_{it-1} - \hat{b}_2 k_{it-1} - \frac{r_{it-1}\hat{b}_2}{\hat{b}_1 s_{Mit-1}},$$

with $b_6 = \gamma$.[19] Note that since k_{it} is predetermined in t we can instrument for z_{6it} with k_{it}. This is what

[19]Note that, alternatively, we could have used a combination of the estimates in the first stage to obtain an estimate of γ, instead of using a second stage. This alternative methodology is in general not as robust and precise, since it involves

we do using the instrumental variables (IV) estimator.

Then, our estimate of productivity is given by

$$\hat{a}_{it} = q_{it} - \frac{\hat{b}_6}{\hat{b}_1}\frac{s_{Oit}}{s_{Mit}}(o_{it} - k_{it}) - \frac{\hat{b}_6}{\hat{b}_1}(m_{it} - k_{it}) - \hat{b}_6 k_{it},$$

our estimate of demand shocks by

$$\hat{\lambda}_{it} = \frac{\hat{b}_6}{\hat{b}_1 s_{Mit}} r_{it} - q_{it},$$

and our estimate of markups by

$$\hat{\mu}_{it} = \frac{\hat{b}_6}{\hat{b}_1 s_{Mit}}.$$

The basic estimation methodology that we just described is amenable to various generalizations. In particular we can allow for a translog production function and can allow for a quadratic rather than the linear dependence of the productivity process on past productivity.

5.2.2 Changes in layers in MULAMA

The methodology to estimate quantity productivity introduced above does not incorporate the effect of changes in organization in the labor input. To do so, we parallel what we did in Section 5.1, but use two distinct variants. In Case 1, we proxy for the organizational part of productivity using quantity. In Case 2, we do exactly what we did in the case of revenue-based productivity and use the number of layers.

Case 1 The labor input corresponds to the output of the knowledge-based hierarchy, namely $F\left(Z_{Lit}^{L}\right) n_{Lit}^{0}$. So,

$$q_{it} = \tilde{a}_{it} + \alpha_O \ln n_{Lit}^{0} + \alpha_M m_{it} + (\gamma - \alpha_M - \alpha_O) k_{it},$$

where $\tilde{a}_{it} = a_{it} + \alpha_O \ln F\left(Z_{Lit}^{L}\right)$ denotes quantity-based productivity. Note that quantity -based productivity $\tilde{a}_{it}$ now incorporates the effect of changes in organization in the labor input via $\ln F\left(Z_{Lit}^{L}\right)$. Furthermore, since every value of $\ln F\left(Z_{Lit}^{L}\right)$ corresponds to a value of q_{it} we account for this by assuming that the autoregressive process for $\tilde{a}_{it}$ is given by

$$\tilde{a}_{it} = \phi_a \tilde{a}_{it-1} + \beta q_{it} + \nu_{ait}, \tag{12}$$

and we use the number of layer-zero employees, n_{Lit}^{0}, as our measure of the labor input. We adjust the estimation of quantity-based productivity described above to take into account the dependence of the process of productivity on the quantity produced.[20]

multiplications and divisions of estimated coefficients as well as a difference between ϕ_λ and ϕ_a that is significantly different from zero.

[20] See Appendix B for further details.

Table 15: MULAMA Quantity TFP: Case 1. Firm-product-sequence FE

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.398^b	-0.120	0.519^a	0.284^b
	(0.155)	(0.097)	(0.102)	(0.132)
Quantity t	0.527^a	0.478^a	0.301	0.445^a
	(0.124)	(0.128)	(0.187)	(0.107)
Constant	-7.531^a	-6.966^a	-4.459	-6.345^a
	(1.761)	(1.840)	(2.757)	(1.528)
Observations	4,171	2,840	3,055	10,066
Number of fixed effects	1,673	1,280	1,298	4,251
Adjusted R^2	0.159	0.039	0.338	0.108

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

A note of caution is in order. There are many potential reasons for why the quantity produced can affect productivity. For example, standard learning-by-doing theories would have very similar implications. Increasing returns, in contrast, is not a potential explanation given that we account for them in our estimation strategy. Our aim here is to first measure the effect of physical quantity on productivity and assess whether this dependence is related to a firm's reorganization. If the reason quantity-based productivity depends on quantity is not related to reorganization, conditioning on the changes in the number of layers should not affect the results. If, conversely, the effect of quantity on productivity is significantly different depending on whether the firm reorganizes or not, we can claim that part of this relationship is related to our mechanism.

Table 15 presents the results using firm-product-sequence fixed effects. As before, it divides the observations in sequences in which the number of layers increased, decreased, or stayed the same. It also presents results for the whole sample. The results are quite stark. Quantity increases quantity-based productivity significantly for firms that add or drop layers, but is not significantly related to quantity-based productivity for firms that do not reorganize. The magnitude of the overall effect implies that a 10% change in quantity increases productivity by about 4.5%. In firms that add layers a 10% increase in quantity increases productivity by 5.3%, while in firms that drop them only by 4.8%. Note that the coefficient on quantity for firms that do not reorganize, although not significant, is positive. This is consistent with alternative explanations of the link between quantity and quantity-based productivity. Although these alternative explanations do not receive significant support in the data once we control for changes in layers. Of course, they could still be potentially important in determining the timing of a firm's reorganization. An argument we return to when we instrument for changes in organization below.

Table 16 presents the results for the alternative grouping based on the initial number of layers. The results are, again, encouraging. Quantity affects positively and significantly quantity-based productivity for firms with any initial number of layers. Note also that the effect seems to decline with the number of layers.

Table 16: MULAMA Quantity TFP: Case 1. Firm-product FE

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	0.279^a	0.056	0.221^b	0.526^a
	(0.096)	(0.233)	(0.108)	(0.088)
Quantity t	1.603^a	0.544^a	0.434^a	0.315^b
	(0.591)	(0.169)	(0.101)	(0.129)
Constant	-22.297^a	-8.034^a	-5.966^a	-4.581^b
	(7.494)	(2.319)	(1.460)	(2.001)
Observations	532	1,649	3,674	3,523
Number of fixed effects	163	549	1,149	1,148
Adjusted R^2	0.127	0.017	0.086	0.309

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

This means that reorganizations affect quantity productivity more for firms with smaller numbers of layers. The effect is particularly large (although also more noisy given the smaller number of observations) for firms with zero layers. For these firms a 1% increase in quantity leads to a 1.6% increase in quantity-based productivity.

Case 2 One of the main difficulties in interpreting the results for Case 1, that we just described, is that the mapping between reorganization and quantity is not measured directly. Hence, we do not know the magnitude of the change in quantity-based productivity that results from a reorganization. Furthermore, the previous method might confound other effects entering through q_{it}, although we find no significant evidence that this is the case. To address these potential concerns in this case we substitute O_{it} in the production function for $C\left(O_{it};w\right)/AC\left(O_{it};w\right)$ (as we did for the case of revenue-based productivity, see Equation 7). We now assume that the process for quantity-based productivity is given by

$$\tilde{a}_{it} = \phi_a \tilde{a}_{it-1} + \beta L_{it} + \nu_{ait}, \tag{13}$$

where, as before, we adjust the process for quantity-based productivity to take into account the dependence on layers, see Appendix B for further details. As we mention before, this is the implication of the theory if we replace constraint (5) by $n_L^L \geq \epsilon > 0$, since in this case the average cost function is a step function where the steps correspond to changes in layers.

Table 17 confirms, for the case of Case 2, our findings for Case 1. Adding a layer is associated with an increase of around 4% in quantity-based productivity. The effect is also positive for firms that drop layers (so their quantity-based productivity declines), although not significant at the 10% level. Note the high bar that we are setting for our empirical estimation. We are including a large number of fixed effects and dummies in the estimation, so everything is estimated out of changes in the number of layers for a given

Table 17: MULAMA Quantity TFP: Case 2. Firm-product-sequence FE

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.316^a	0.348^b	0.359^a	0.341^a
	(0.056)	(0.170)	(0.092)	(0.067)
Number of management layers t	0.040^c	0.013		0.032^b
	(0.022)	(0.022)		(0.014)
Constant	-0.113^c	-0.008	-0.001	-0.034
	(0.062)	(0.035)	(0.023)	(0.041)
Observations	4,171	2,840	3,055	10,066
Number of fixed effects	1,673	1,280	1,298	4,251
Adjusted R^2	0.076	0.111	0.101	0.094

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Table 18: MULAMA Quantity TFP: Case 2. Firm-product FE

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	0.240^a	0.276^a	0.408^a	0.480^a
	(0.066)	(0.056)	(0.061)	(0.099)
Number of management layers t	0.055^b	-0.013	0.029^c	0.049^a
	(0.028)	(0.023)	(0.017)	(0.018)
Constant	0.110^c	0.063	-0.069	-0.086
	(0.065)	(0.048)	(0.044)	(0.053)
Observations	532	1,649	3,674	3,523
Number of fixed effects	163	549	1,149	1,148
Adjusted R^2	0.053	0.079	0.140	0.186

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

firm-product-sequence. Furthermore given that we are adding time dummies, we are controlling for any time trend in the data.

In Table 18 we present results grouped by the initial layer. Here the estimates show an overall effect of layers on quantity-based productivity of between 3 and 5%. Note that for the case of firms with 1 layer our results are negative, although not significant. In sum, all the significant results we have shown in this section indicate that a reorganization that adds layers has a positive effect on quantity-based productivity, while a reorganization that drops layers and shrinks the firm has a negative effect on quantity-based productivity.

Table 19: Wooldridge Revenue TFP: Contemporaneous Firm-product-sequence FE

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.091^b	-0.064^c	0.065^b	0.053^b
	(0.038)	(0.038)	(0.041)	(0.024)
Change in quantity	-0.033^a	-0.074^a		-0.048^a
	(0.009)	(0.014)		(0.008)
Constant	-0.084^a	-0.022	0.050^b	0.075^a
	(0.013)	(0.027)	(0.020)	(0.019)
Observations	4,057	2,686	2,934	9,677
Number of fixed effects	1,630	1,258	1,248	4,136
Adjusted R^2	0.053	0.042	0.031	0.027

Firm-level clustered standard errors in parentheses

^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

5.3 Contemporaneous Effects

The previous formulations allow for a slow adjustment of productivity to the number of layers. Namely, the actual number of layers affects productivity and so a change in the number of layers affects the level of productivity, conditional on past productivity, in every period afterwards. In the model the effect is not necessarily a permanent effect, but rather just a level effect in the period of the switch. Hence, we consider an alternative version for the process of revenue-based and quantity-based productivity where we only incorporate the change in the number of layers and not their level. That is, in this alternative specifications the change in layers affects the level of productivity in just one period. Namely, the additional layer has only a contemporaneous level effect. Of course, a potential problem with this specification is that it might fail to capture protracted effects of changes in layers on productivity. This is why we started with the permanent case in the previous sub-sections.

Consider first the case of revenue-based productivity. The process for revenue-based productivity in equation (9), when we only consider contemporaneous effects, becomes

$$\bar{a}_{it} = \phi_a \; \bar{a}_{it-1} + \beta \Delta L_{it} + \nu_{ait}. \tag{14}$$

We can use the process (14) and the methodology of Wooldridge (2009) to recompute the revenue-based productivity process. The results are presented in Table 19 and 20. The findings in Table 19 are consistent with our hypothesis and all highly significant. The effect of a change in layers on revenue-based productivity is negative for firms that either increase or drop layers. Note also that the effects are somewhat smaller than in Table 13. This is natural given that firms change the number of layers in most cases by only one layer, while the average number of layers is between 1.20 and 1.76 as we saw in Table 1. Once we take this effect into account the contemporaneous effect is similar, although still bit smaller than the permanent ones.

Table 20: Wooldridge Revenue TFP: Contemporaneous. Firm-product FE

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	0.001	0.015	0.185^{a}	0.126^{a}
	(0.044)	(0.048)	(0.042)	(0.039)
Change in quantity	-0.047^{a}	-0.021	-0.038^{a}	-0.055^{a}
	(0.018)	(0.015)	(0.008)	(0.012)
Constant	0.004	-0.035^{a}	-0.057^{a}	0.048^{a}
	(0.032)	(0.027)	(0.012)	(0.015)
Observations	525	1,602	3,591	3,385
Number of fixed effects	159	535	1,119	1,096
Adjusted R^2	0.049	0.011	0.059	0.040

Firm-level clustered standard errors in parentheses

^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Table 20 presents the results ordered by layer. The results are, again, all negative. Note that for firms that start with zero, two, or three layers, the coefficients are all significant, while the estimate for firms with one layer is not significant. The imprecision in the estimate of the contemporaneous effect might be the result of the relatively smaller sample size compared to groups of firms with 2 and 3 layers.

Consider now a similar modification in our estimation of quantity-based productivity. When we only consider contemporaneous effects, the process for quantity-based productivity in equation (13) becomes

$$\tilde{a}_{it} = \phi_a \tilde{a}_{it-1} + \beta \Delta L_{it} + \nu_{ait}. \tag{15}$$

We can use the process in (15) and recompute all the shocks in the model using the MULAMA methodology described above.[21] Tables 21 and 22 present the two sets of results. The effect in Table 21 is again somewhat smaller than what we found for the permanent case in Table 17, but positive for the whole sample of firms as well as for firms that add or drop layers. It is not significant for the firms that reorganize and add layers. The overall estimated magnitude of the contemporaneous effect of an extra layer is around 1.5%.

The results in Table 22 are somewhat mixed. Again all the significant results are positive, although the effect of changes in layers is not significant for firms that start with zero or 1 management layers. As before, this is probably the result of the large number of fixed effects that we are using, combined with the smaller samples for firms with zero or one layer. Still, our estimates indicate that a reorganization that adds layers, whenever we can measure it somewhat precisely, always leads to an increase in quantity-based productivity.

[21] In addition, Appendix C presents the results for the contemporaneous Case 1, where

$$\tilde{a}_{it} = \phi_a \; \tilde{a}_{it-1} + \beta \Delta q_{it} + \nu_{ait}.$$

The results for this case are all aligned with the theory. Yet, as discussed before, the results for quantity are necessary, but not sufficient, to claim that quantity-based productivity responds positively to a reorganization that adds layers. Therefore, for brevity, we decided to present Case 2 in the main text, and relegated Case 1 to the Appendix.

Table 21: MULAMA Quantity TFP: Case 2 Cont. Firm-product-sequence FE

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.207^a	0.103	0.162^a	0.163^a
	(0.043)	(0.070)	(0.055)	(0.034)
Change in layers	0.007	0.021^c		0.014^c
	(0.011)	(0.011)		(0.008)
Constant	0.001	0.041	0.032	0.040
	(0.016)	(0.037)	(0.019)	(0.025)
Observations	4,171	2,840	3,055	10,066
Number of fixed effects	1,673	1,280	1,298	4,251
Adjusted R^2	0.041	0.012	0.030	0.027

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Table 22: MULAMA Quantity TFP: Case 2 Contemporaneous. Firm-product FE

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	0.154^b	0.226^a	0.263^a	0.215^a
	(0.067)	(0.062)	(0.039)	(0.059)
Change in layers	-0.003	0.004	0.023^a	0.027^b
	(0.016)	(0.013)	(0.009)	(0.011)
Constant	0.209^a	0.080^a	-0.015	0.041^c
	(0.049)	(0.023)	(0.015)	(0.022)
Observations	532	1,649	3,674	3,523
Number of fixed effects	163	549	1,149	1,148
Adjusted R^2	0.017	0.062	0.069	0.051

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

5.4 Instrumental Variables

The results so far have shown, we believe, that a set of detailed and specific predictions on revenue-based and quantity-based productivity changes as a result of a firm's reorganization are robustly present in the data. The fact that when we see firms adding layers revenue-based productivity declines but quantity-based productivity increases, and that this is significantly the case after including a large battery of fixed effects, lends credibility to the causal interpretation that our theory provides for these facts. Still, we cannot rule out the possibility that a positive shock to quantity-based productivity leads to an increase in layers (although it would still be hard to explain why revenue-productivity falls independent of the reorganization). More specifically, it can be the case that the reorganization of a firm (as measured by changes in quantity in Case 1 and the number of layers in Case 2) is the result of the innovations, ν_{ait}, in equations (9) and (12) to (15).

Table 23: Wooldridge Revenue TFP: Contemporaneous with IV

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.775[a]	0.781[a]	0.794[a]	0.792[a]
	(0.023)	(0.033)	(0.022)	(0.016)
Change in layers	-0.047[b]	-0.083[a]		-0.054[a]
	(0.014)	(0.019)		(0.013)
Observations	3,653	2,365	2,020	8,038
Adjusted R^2	0.609	0.490	0.581	0.589

Firm-level clustered standard errors in parentheses
[a] $p<0.01$, [b] $p<0.05$, [c] $p<0.1$

The structural assumptions we have used above provide us with a set of instrumental variables for a reorganization. In particular, for the contemporaneous effects in equations (14) and (15) we can use: the number of layers in $t-1$; quantity, revenue, demand shocks and markups at time $t-1$; capital at time t; and productivity at time $t-2$. We can also use all these variables lagged to the first year available. Note that it is important that we use the productivity process in (14) and (15) where the change in layers, not the number of layers, has an effect on productivity.[22] In these specifications layer changes have a direct effect on productivity only in period t and so we can use the past number of layers to instrument for a change in layers. Otherwise the effect of a change in layers could still be endogenous to current innovations and so all these variables (except perhaps the ones for the initial period) are not valid instruments.

Note also that since we are using this battery of instrumental variables as well as past productivity, the set of fixed effects we used in the previous regression are not obviously necessary. They also reduce the precision of our estimates substantially. So for all results using instrumental variables we drop the set of firm-level fixed effects, although we keep product group and time fixed effects in all regressions.[23]

Table 23 presents the results for the specification of the process of revenue-based productivity in equation (14), but when we instrument for the change in layers using the variables described above. The estimates using instrumental variables still deliver a negative relationship between changes in layers and revenue-based productivity. These results can now be interpreted as causal. So, an extra layer reduces revenue-based productivity by 4.7%. The results are even more negative and significant for firms that drop layers. The average effect across all firms is that a change in layers accounts for a reduction in revenue-based productivity of 5.4%.

Note also that compared to the previous tables, past revenue-based productivity now plays a much more significant role (in line with past estimates of the autoregressive component in the literature) and the persistence coefficient is around .8. The reason is, of course, that we have dropped the set of firm-product-

[22] In Appendix C we also consider Case 1 and instrument for changes in quantity.

[23] In all the tables that use instrumental variables we do not present the value of constants since we de-mean all variables.

Table 24: Wooldridge Revenue TFP: Contemporaneous with IV

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	0.430^{a}	0.612^{a}	0.812^{a}	0.814^{a}
	(0.079)	(0.034)	(0.023)	(0.016)
Change in layers	-0.015	-0.062^{c}	-0.052^{a}	-0.082^{a}
	(0.031)	(0.037)	(0.020)	(0.023)
Observations	525	1,602	3,591	3,385
Adjusted R^2	0.238	0.382	0.584	0.616

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Table 25: MULAMA Quantity TFP: Case 2 Contemporaneous with IV

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.907^{a}	0.880^{a}	0.923^{a}	0.908^{a}
	(0.011)	(0.018)	(0.015)	(0.009)
Change in layers	0.044^{a}	0.060^{b}		0.066^{a}
	(0.017)	(0.026)		(0.017)
Observations	3,748	2,181	2,098	8,027
Adjusted R^2	0.793	0.723	0.860	0.791

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

sequence fixed effects.

Table 24 presents the results when we group by initial layer. All the estimates of the effects of changes in layers are negative, once more indicating that there is a causal effect of changes in layers on revenue-based productivity. However, the estimates for firms with zero layers are again not significant. As before, this might be the result of the relatively smaller number of observations.

Perhaps more revealing are the results when we use the process for quantity-based productivity in (15). These results, presented in Table 25 are very supportive of the theory and show that the causal effects of a change in layers on productivity is between 4 and 6%. These numbers are larger than the ones we found in Table 22, indicating perhaps that the fixed effects in those results where capturing some of the effect of the change in layers. In this case, the autorregressive coefficient is very significant and around .9 for all cases.

Our last set of results is presented in Table 26 where we group firms by their initial number of layers but use the contemporaneous effects version of Case 2. As we have in a series of tables with groupings by layer, the small number of observations imply that the results for zero and one layers are not significant. However, the results for firms with 2 and 3 layers of management are positive and significant.

Table 26: MULAMA Quantity TFP: Case 2 Contemporaneous with IV

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	0.846^{a}	0.905^{a}	0.884^{a}	0.892^{a}
	(0.048)	(0.020)	(0.013)	(0.014)
Change in layers	-0.029	0.030	0.066^{a}	0.046^{c}
	(0.039)	(0.040)	(0.022)	(0.027)
Observations	532	1,649	3,674	3,523
Adjusted R^2	0.689	0.780	0.757	0.776

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Overall, throughout our investigation we did not find any significant evidence to falsify the hypothesis proposed by the hierarchy model. All the significant evidence was in line with the main implications. Hence, we conclude that when firms receive an exogenous shock that makes them reorganize, quantity-based productivity increases significantly. An extra layer increases productivity by around 5%, and a drop in layers decreases productivity by a similar percentage. As firms reorganize they also expand, move down their demand curve and reduce prices. This countervailing force results in a decrease in revenue-base productivity of around 4 percentage points.

5.5 Aggregate Productivity Effects from Reorganization

The results in the previous section indicate that reorganizations lead to large changes in quantity-based productivity conditional on a firm's past productivity, its number of layers, and a variety of fixed effects. If we want to gauge the importance of organization for aggregate productivity dynamics, we need to understand how important is the effect of reoganizations for the average firm that reorganizes. So, for the firms that reorganize we want to ask how important is the change in productivity that resulted from the reorganization, compared to changes in productivity due to idiosyncratic shocks, or the mean reversion implied by the process in (15).

Consider a firm-product i that we observe from $t-T$ to t. Iterating over equation (15) we obtain that

$$\tilde{a}_{it} - \tilde{a}_{it-T} = \left(\phi_a^T - 1\right)\tilde{a}_{it-T} + \beta \sum_{v=0}^{T-1} \phi_a^v \Delta L_{it-v} + \sum_{v=0}^{T-1} \phi_a^v \nu_{ait-v}.$$

Hence, the overall change in productivity for a firm, given by $\tilde{a}_{it} - \tilde{a}_{it-T}$, can be decomposed into three components. The first term is a mean reversion component that is negative when $\tilde{a}_{it-T}$ is positive since $\phi_a < 1$. Namely, productivity tends to revert to its long term mean given a number of layers. The cumulative change in productivity due to a reorganization, is given by the second term, namely, $\beta \sum_{v=0}^{T-1} \phi_a^v \Delta L_{it-v}$. The third term is just the accumulated effect of past shocks.

Table 27: Change in Quantity TFP due to Reorganization

	Firms that increase layers		Firms that reduce layers	
Percentiles	Overall change	Due to reorganization	Overall change	Due to reorganization
10%	-.49	.05	-.53	-.09
25%	-.18	.05	-.27	-.06
50%	.05	.06	-.04	-.06
75%	.32	.07	.20	-.05
90%	.67	.11	.52	-.05
Mean	0.06	0.07	-0.02	-0.07
Observations	817	817	466	466

We calculate these terms for firms that increase and decrease the number of layers between $t-T$ and t. Using our results for β and ϕ_a from the MULAMA Quantity TFP Case 2 Contemporaneous with Instrumental Variables in column 4 of Table 25 we calculate each of these terms for the whole distribution of firms. Clearly, the actual change in productivity across firms is very heterogeneous. Some firms that add layers experiment a large decline in productivity, while some experiment a very large increase. Hence, we order firms by their overall change in productivity and in Table 27 present the distribution of the overall changes in productivity and the change in productivity due to changes in layers.[24] Columns two and three present the results for firms that increase layers, while columns four and five present the results for firms that drop layers.

The results are stark. On average, or for the median firm, the increase in productivity due to reorganization explains essentially all of the increase in overall productivity. This is clearly not the case for all firms, some of them receive large positive or negative productivity shocks that account for most of the changes in productivity, but on average those shocks (and the associated reversion to the mean) roughly cancel out across firms. The result is that reorganization can account for an increase in quantity-based productivity, when firms reorganize, of about 7% while the average increase in productivity for these firms was about 6%. Similarly, when firms reduce the number of layers, reorganization accounts for a 7% decrease in quantity-based productivity while the average decrease in productivity for these firms was about 2%. Reorganization amounts to more than 100% of the change in productivity! The results underscores the importance of the reorganization of firms as a source of aggregate productivity gains in the economy.

6 Conclusion

Large firm expansions involve lumpy reorganizations that affect firm productivity. Firms that reorganize and add a layer increase hours of work by 25% and value added by slightly more than 3%, while firms that do not reorganize decrease hours slightly and value added by only 0.1%. Reorganization therefore accompanies firms' expansions. A reorganization that adds layers allows the firm to operate at a larger scale. We have

[24] The unit of observation is a firm-product and we allow $t-T$ and T to vary across firm-product pairs.

shown that such a reorganization leads to increases in quantity-based productivity of about 4%. Even though the productive efficiency of the firm is enhanced by adding layers, its revenue-based productivity declines by more than 4%. The new organizational structure lowers the marginal cost of the firm and it allows it to increase its scale. This makes firms expand their quantity and move down their demand curves, thereby lowering prices and revenue-based productivity.

We use a detailed data set of Portuguese firms to show that these facts are very robustly present in the data. Our data set is somewhat special in that it not only includes employer-employee matched data, necessary to built a firm's hierarchy, but it also includes information on quantity produced. This allows us to contrast the effect of reorganization, using fairly flexible methodologies, to calculate quantity and revenue-based productivity. Furthermore, given that we have a relatively long panel, we show that the results hold using a large number of firm-product-sequence fixed effects on top of time and industry dummies. We do not find any case in which the evidence significantly falsifies the main hypothesis of the effect of a reorganization on both types of firm productivity. In contrast, we present significant evidence of a causal effect of an increase in layers on quantity-based productivity.

Our findings underscore the role that organizational decisions play in determining firm productivity. Our results, however, can be viewed more broadly as measuring the impact of lumpy firm level changes on the endogenous component of firm productivity. Many changes that increase the capacity of the firm to grow (like building a new plant or production line, or creating a new export link with a foreign partner) will probably result in similar effects on quantity and revenue-based productivity. In our view, the advantage of looking at reorganizations using a firm's management layers, as defined by occupational classifications, is that firms change them often and in a very systematic way. Furthermore, this high frequency implies that many of the observed fluctuations in both quantity-based and revenue-based productivity result from these endogenous firm decisions and should not be treated as exogenous shocks to the firm.

Recognizing that part of a firm's productivity changes are endogenous is relevant because the ability of firms to change their organization might depend on the economic environment in which they operate. We have shown that changing the number of management layers is important for firms to realize large productivity gains when they grow. Environments in which building larger hierarchies is hard or costly due, for example, to the inability to monitor managers or to enforce detailed labor contracts prevent firms from obtaining these productivity gains.[25] This, among other factors, could explain why firms in developing countries tend to grow less rapidly (Hsieh and Klenow, 2014).

[25] See Bloom et al. (2013) for some evidence on potential impediments in India.

References

[1] Ackerberg, D., Benkard, C.L., Berry, S. and Pakes, A. 2007. “Econometric tools for analyzing market outcomes.” In J. Heckman and E. Leamer (eds), Amsterdam: North-Holland, *Handbook of Econometrics,* Vol. 6:1, 4171–4276.

[2] Berry, S. J. Levinsohn, and A. Pakes, 1995. “Automobile Prices in Market Equilibrium.” *Econometrica*, 63:4, 841-890.

[3] Blanchard, O. and Portugal, P., 2001. “What Hides Behind an Unemployment Rate: Comparing Portuguese and U.S. Labor Markets.” *American Economic Review*, 91:1, 187-207.

[4] Bloom, N., B. Eifert, D. McKenzie, A. Mahajan, and J. Roberts, 2013. “Does Management Matter: Evidence from India.” *Quarterly Journal of Economics*, 128:1, 1-51.

[5] Cabral, L. and Mata, J., 2003. “On the Evolution of the Firm Size Distribution: Facts and Theory.” *American Economic Review*, 93:4, 1075-1090.

[6] Caliendo, L., Monte, F., and Rossi-Hansberg, E., 2012. “The Anatomy of French Production Hierarchies.” *Journal of Political Economy*, forthcoming.

[7] Caliendo, L. and E. Rossi-Hansberg, 2012. “The Impact of Trade on Organization and Productivity.” *Quarterly Journal of Economics*, 127:3, 1393-1467.

[8] De Loecker, J. and F. Warzynski, 2012. “Markups and Firm-level Export Status.” *American Economic Review*, 102:6, 2437-2471.

[9] Friedrich, B.U. 2015. “Trade Shocks, Firm Hierarchies and Wage Inequality.” Yale University, mimeo.

[10] Forlani, E., Martin, R., Mion, G. and Muuls, M. “Unraveling Firms: Demand, Productivity and Markups Heterogeneity.” mimeo.

[11] Garicano, L., 2000. “Hierarchies and the Organization of Knowledge in Production.” *Journal of Political Economy*, 108:5, 874-904.

[12] Garicano, L. and E. Rossi-Hansberg, 2006. “Organization and Inequality in a Knowledge Economy.” *Quarterly Journal of Economics*, 121:4, 1383-1435.

[13] Garicano, L. and E. Rossi-Hansberg, 2015. “Knowledge-based Hierarchies: Using Organizations to Understand the Economy.” *Annual Reviews of Economics*, 7, 1-30.

[14] Garcia-Marin, A. and N. Voigtländer, 2014. “Exporting and Plant-Level Efficiency Gains: It’s in the Measure.” NBER Working Paper No. 19033.

[15] Hsieh, C. and P. Klenow, 2014. "The Life Cycle of Plants in India and Mexico." *Quarterly Journal of Economics*, 129:3, 1335-1384..

[16] Klette, T. J. and Griliches, Z., 1996. "The inconsistency of common scale estimators when output prices are unobserved and endogenous." *Journal of Applied Econometrics* 11, 343-361.

[17] Levinsohn, J. and Petrin, A., 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies*, 70:2, 317-341.

[18] Marschak, J. and Andrews, W., 1944. "Random simultaneous equations and the theory of production." *Econometrica* 12, 143-205.

[19] Mion, G. and Opromolla, L., 2014. "Managers' Mobility, Trade Performance, and Wages." *Journal of International Economics*, 94:1, 85-101.

[20] Olley, S. and A. Pakes, 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica,* 64:6, 1263-97.

[21] Rosen, S., 1982. "Authority, Control, and the Distribution of Earnings." *The Bell Journal of Economics*, 13:2, 311-323.

[22] Tåg, J. 2013. "Production Hierarchies in Sweden." *Economics Letters* 121:2, 210-213.

[23] Wooldridge, J.M. 2009. "On estimating firm-level production functions using proxy variables to control for unobservables." *Economics Letters* 104:3, 112–114.

Appendix A

We start with the matched employer-employee data set, keeping only firms in the manufacturing sector located in mainland Portugal and dropping firms with non-positive sales. Information for the year 2001 for the matched employer-employee data set was only collected at the firm-level. Given that worker-level variables are missing in 2001 we have to drop all firm-level observations for 2001. There are in total 353,311 firm-year observations. We then focus on the worker-level information and drop a minority of workers with an invalid social security number and with multiple jobs in the same year. We further drop worker-year pairs whenever (i) their monthly normal or overtime hours are non-positive or above 480; (ii) the sum of weekly normal and overtime hours is below 25 and above 80; (iii) their age is below 16 and above 65 years; (iv) they are not full-time employees of the firm. Based on the resulting sample, we trim worker-year pairs whose monthly wage is outside a range defined by the corresponding year bottom and top 0.5 percentiles. This leaves us with 321,719 firm-year and 5,174,324 worker-year observations. In the analysis, we focus on manufacturing firms belonging to industries (NACE rev.1 2-digits between 15 and 37) excluding 16 "Manufacture of tobacco products", 23 "Manufacture of coke, refined petroleum products and nuclear fuel", 30 "Manufacture of office machinery and computers", and 37 "Recycling", due to confidentiality reasons.

We then turn to the balance sheet data set and recover information on firms' operating revenues, material assets, costs of materials, and third-party supplies and services. We compute value-added as operating revenues minus costs of materials and third-party supplies and services. We drop firm-year pairs with non-positive value-added, material assets, cost of materials, and size. This reduces the size of the overall sample to 61,872 firm-year observations and 2,849,363 worker-year observations.

Finally, we turn to the production data set and recover information on firm-product sales and volume for each firm-product-year triple in the data set. In the production data set a product is identified by a 10-digits code, plus an extra 2-digits that are used to define different variants of the variable.[26] The first 8 digits correspond to the European PRODCOM classification while the additional two have been added by INE to further refine PRODCOM. The volume is recorded in units of measurement (number of items, kilograms, litres) that are product-specific while the value is recorded in current euros. We drop observations where the quantity produced, quantity sold, and sales are all zero. For each product-firm-year combination, we are able to compute a unit value. We adjust the quantity sold, for each firm-year-product, by multiplying it by the average (across firms) product-year unit value. We then construct a more aggregate partition of products based on the first 2-digits as well as on the unit of measurement. More specifically, we assign 10-digits products sharing the same first 2 digits and unit of measurement to the same aggregate product. We keep only manufacturing products, and aggregate quantity sold and sales at the firm-year-product level following the new definition of a product. We restrict the analysis to products with at least 50 firm-year

[26] From the raw data it is possible to construct different measures of the volume and value of a firm's' production. For the sake of this project we use the volume and value corresponding to a firms' sales of its products. This means we exclude products produced internally and to be used in other production processes within the firm as well as products produced for other firms, using inputs provided by these other firms. The advantage of using this definition is that it nicely corresponds to the cost of materials coming from the balance sheet data.

observations. Finally, we merge the production data with the matched employer-employee and firm balance sheet data.

Given that we restricted the set of products considered in the analysis, we compute the ratio between total firm-year sales in the sample coming from the production data set and firm-year sales in the firm balance sheet sample and drop firm-year pairs we extreme values of the ratio (below 25 percent and above 105 percent). We then adjust firm sales (from the balance sheet data), cost of materials, material assets, wage bill, size, value-added, wage bill of layer zero, and number of employees in layer zero using the above sales ratio. We then split the same set of variables into parts associated with each product, using the product sales in the production data set. We trim firm-year-product triples that do not satisfy one or more of the following constraints: the sum of cost of materials and wages, as a share of sales, below one; unit value between the 1st and 99th percentiles; cost of materials as a share of sales between the 1st and 99th percentiles; ratio of material assets to size between the 1st and 99th percentiles. The size of the sample is now 19,031 firm-year observations and 1,593,294 worker-year observations.

Table A.1: Classification of Workers According to Hierarchical Levels

Level	**Tasks**	**Skills**
1. Top executives (top management)	Definition of the firm general policy or consulting on the organization of the firm; strategic planning; creation or adaptation of technical, scientific and administrative methods or processes	Knowledge of management and coordination of firms fundamental activities; knowledge of management and coordination of the fundamental activities in the field to which the individual is assigned and that requires the study and research of high responsibility and technical level problems
2. Intermediary executives (middle management)	Organization and adaptation of the guidelines established by the superiors and directly linked with the executive work	Technical and professional qualifications directed to executive, research, and management work
3. Supervisors, team leaders	Orientation of teams, as directed by the superiors, but requiring the knowledge of action processes	Complete professional qualification with a specialization
4. Higher-skilled professionals	Tasks requiring a high technical value and defined in general terms by the superiors	Complete professional qualification with a specialization adding to theoretical and applied knowledge
5. Skilled professionals	Complex or delicate tasks, usually not repetitive, and defined by the superiors	Complete professional qualification implying theoretical and applied knowledge
6. Semi-skilled professionals	Well defined tasks, mainly manual or mechanical (no intellectual work) with low complexity, usually routine and sometimes repetitive	Professional qualification in a limited field or practical and elementary professional knowledge
7. Non-skilled professionals	Simple tasks and totally determined	Practical knowledge and easily acquired in a short time
8. Apprentices, interns, trainees	Apprenticeship	

Notes: Hierarchical levels defined according to Decreto Lei 121/78 of July 2nd (Lima and Pereira, 2003).

All monetary values are deflated to 2005 euros using the monthly (aggregated to annual) Consumer Price Index (CPI - Base 2008) by Special Aggregates from Statistics Portugal. Monthly wages are converted to annual by multiplying by 14.

Some concepts are recurring in the explanation of a majority of the tables and figures. We define them here and consider them understood main text:

- **Layer number.** In the matched employer-employee data set, each worker, in each year, has to be assigned to a category following a (compulsory) classification of workers defined by the Portuguese law (see Table A.1 and Mion and Opromolla, 2014). Classification is based on the tasks performed and skill requirements, and each category can be considered as a level in a hierarchy defined in terms of increasing responsibility and task complexity. On the basis of the hierarchical classification and taking into consideration the actual wage distribution, we partition the available categories into occupations. We assign "Top executives (top management)" to occupation 3; "Intermediary executives (middle management)" and "Supervisors, team leaders" to occupation 2; "Higher-skilled professionals" and some "Skilled professionals" to occupation 1; and the remaining employees, including "Skilled professionals", "Semi-skilled professionals", "Non-skilled professionals", and "Apprenticeship" to occupation 0. The position of the workers in the hierarchy of the firm, starting from 0 (lowest layer, present in all firms) to 3 (highest layer, only present in firms with 3 layers of management).

- **Number of layers of management.** A firm reporting c occupational categories will be said to have $L = c - 1$ layers of management: hence, in our data we will have firms spanning from 0 to 3 layers of management (as in CMRH). In terms of layers within a firm we do not keep track of the specific occupational categories but simply rank them. Hence a firm with occupational categories 2 and 0 will have 1 layer of management, and its organization will consist of a layer 0 corresponding to some skilled and non-skilled professionals, and a layer 1 corresponding to intermediary executives and supervisors.

- **Reorganization in year t.** A firm reorganizes in year t when it changes the number of management layers with respect to those observed in the most recent prior available year (year $t-1$ in most cases).

- **Year of the first observed reorganization for a firm.** The earliest reorganization year observed (for those firms first appearing in the data prior to 1997) or the first year in which a firm appears in the data (for those firms first appearing in the data in 1997 or later).

- **Firm industry.** The industry of the firm is measured according to the NACE rev.1 2-digits disaggregation. This includes 19 divisions, from division 15 (Manufacture of food products and beverages) to division 36 (Manufacture of furniture; manufacturing n.e.c.). We drop division 16 (Manufacture of tobacco products), 23 (Manufacture of coke, refined petroleum products and nuclear fuel), and 30 (Manufacture of office machinery and computers) because they comprise very few observations.

- **Wage bill.** A worker annual wage is computed adding the monthly base and overtime wages plus regular benefits and multiplying by 14. We apply a trimming of the top and bottom 0.5 per cent within each year. A firm wage bill is the sum of the annual wages of all its workers that satisfy the criteria listed above.

- **Value added.** Value added is computed, from the balance sheet data set, as operating revenues minus costs of materials and third-party supplies and services.

- **Revenue productivity.** The log of the ratio between firm sales and employment.
- **Value added productivity.** The log of the ratio between firm value added and employment.
- **OLS TFP.** Log total factor productivity computed from a standard three factors (labour, capital and materials) Cobb-Douglas production function model where output is measured by firm sales and the model is estimated via OLS. Separate estimations have been carried for each industry.
- **Olley and Pakes revenue-based TFP.** Log total factor productivity computed from a standard two factors (labour and capital) Cobb-Douglas production function model where output is measured by firm value-added. Productivity shocks are modeled as in Olley and Pakes (1996) while being further enriched with layers along the lines presented in Section 5.1. We use the lagged number of management layers to instrument L_{it} in the second stage of the OP procedure. Separate estimations have been carried for each industry.
- **Wooldridge revenue-based productivity.** Log total factor productivity computed from a standard two factors (labour and capital) Cobb-Douglas production function model where output is measured by firm value-added. Productivity shocks are modeled as in Wooldridge (2009) while being further enriched with layers as described in Section 5.1. Separate estimations have been carried for each industry.
- **Price.** In the PC data set a product is identified by a 10-digits code. The first 8 digits correspond to the European PRODCOM classification while the additional two have been added by INE to further refine PRODCOM. The volume is recorded in units of measurement (number of items, kilograms, litres) that are product-specific while the value is recorded in current euros. Therefore, for each product-firm-year combination, we are able to compute a price.
- **Quantity-based TFP.** We run separate quantity-based productivity estimations for each of the aggregate products using variations of the MULAMA methodology. See the next appendix for a detailed explanation of the estimation methodologies.

Table Descriptions

Table 1: This table reports, for each year, the number of firms in Sample 1 and corresponding averages across all firms for selected variables. Value added, hours, and wage are defined above. Value added is in 2005 euros. Wage is average hourly wage in 2005 euros. Hours are yearly. # of layers is the average number of layers of management across firms in each year.

Table 2: Table 2 reports summary statistics on firm-level outcomes, grouping firm-year observations according to the number of layers of management reported (# of layers). Firm-years is the number of firm-years observations in the data with the given number of layers of management. Value added, hours,

and wage are defined above. Value added in 000s of 2005 euros. Wage is either average or median hourly wage in 2005 euros. Both value added and wages are detrended. Hours are yearly.

Table 3 and 4: Table 3 reports the fraction of firms that satisfy a hierarchy in hours, grouping firms by their number of layers of management (# number of layers). Hours N_L^l is the number of hours reported in layer l in an L layers of management firm. For $L = 1, 2, 3$, and $l = 0, ..., L-1$ we say that a firm satisfies a hierarchy in hours between layers number l and $l+1$ in a given year if $N_L^l \geq N_L^{l+1}$, i.e. if the number of hours worked in layer l is at least as large as the number of hours worked in layer $l+1$; moreover, we say that a firm satisfies a hierarchy at all layers if $N_L^l \geq N_L^{l+1} \quad \forall l = 0, ..., L-1$, i.e. if the number of hours worked in layer l is at least as large as the number of hours in layer $l+1$, for all layers in the firm. Following these definitions, the top panel reports, among all firms with $L = 1, 2, 3$ layers of management, the fraction of those that satisfy a hierarchy in hours at all layers (first column), and the fraction of those that satisfy a hierarchy in hours between layer l and $l+1$, with $l = 0, ..., L-1$ (second to fourth column).

Table 4 is the same as Table 3 for the case of wages, where w_L^l is the average hourly wage in layer l in an L layers of management firm.

Table 5: Table 5 reports the distribution of the number of layers of management at time $t+1$, grouping firms according to the number of layers of management at time t. Among all firms with L layers of management ($L = 0, ..., 3$) in any year from 1996 to 2004, the columns report the fraction of firms that have layers $0, ..., 3$ the following year (from 1997 to 2005), or are not present in the dataset, Exit. The table also reports, in the bottom row, the distribution of the new firms by their initial number of layers. The elements in the table sum to 100% by row.

Table 6: This table shows changes in firm-level outcomes between adjacent years for all firms (All), and for the subsets of those that increase (Increase L), don't change (No change in L) and decrease (Decrease L) layers. It reports changes in log hours, log normalized hours, log value added, log average wage, and log average wage in common layers for the whole sample. The change in average wage for common layers in a firm that transitions from L to L' layers is the change in the average wage computed using only the $min\{L, L'\}$ layers before and after the transition. To detrend a variable, we subtract from all the log changes in a given year the average change during the year across all firms.

Table 7: This table reports the results of regressions of log change in normalized hours by layer on log change in value added for firms that do not change their number of layers of management L across two adjacent periods. Specifically, we run a regression of log change in normalized hours at layer l (layer) in a firm with L (# of layers in the firm) layers of management on a constant and log change in value added across all the firms that stay at L layers of management across two adjacent years. Robust standard errors are in parentheses.

Table 8: This table reports the results of regressions of log change in hourly wage by layer on log change in value added for firms that do not change their number of layers of management L across two adjacent periods. Specifically, we run a regression of log change in average hourly wage at layer l (layer) in a firm

with L (# of layers in the firm) layers of management on a constant and log change in value added across all the firms that stay at L layers of management across two adjacent years. Robust standard errors are in parentheses.

Table 9: This table shows estimates of the average log change in normalized hours at each layer l (Layer) among firms that transition from L (# of layers before) to L' layers (# of layers after), with $L \neq L'$: for a transition from L to L', we can only evaluate changes for layer number $l = 0, ..., \min\{L, L'\}$. $d \ln n^l_{Lit}$ is the average log change in the transition, estimated as a regression of the log change in the number of normalized hours in layer l in two adjacent years on a constant. Robust standard errors are in parentheses.

Table 10: This table shows estimates of the average log change in hourly wage at each layer l (Layer) among firms that transition from L (# of layers before) to L' layers (# of layers after), with $L \neq L'$: for a transition from L to L', we can only evaluate changes for layer number $l = 0, ..., \min\{L, L'\}$. $d \ln w^l_{Lit}$ is the average log change in the transition, estimated as a regression of the log change in the average hourly wage in layer l in two adjacent years on a constant. Robust standard errors are in parentheses.

Table 11: The data underlying Table 11 is composed of sequences of firm-product-years with either one or zero changes in layers. For a given product, we define a firm sequence of type $L - L'$ as the series of years in which a firm sells the corresponding product and has the same consecutively observed number of management layers L plus the adjacent series of years in which a firm sells the product and has the same consecutively observed number of management layers L'. For example, a firm that we observed selling the product all years between 1996 and 2000 and that has zero layers in 1996, 1997, and 2000 and one layer in 1998 and 1999 would have two sequences: An (increasing) 0-1 sequence (1996 to 1999) as well as a (decreasing) 1-0 sequence (1998 to 2000). Firms that never change layers in our sample form a constant-layer sequence. We group firm-product sequences into "Increasing", "Decreasing", and "Constant" sequence types.

For each type of sequence, Table 11 shows estimates of regressions where the dependent variable is (log) revenue labor productivity in a given year. The key regressor is the number of management layers in the firm in the same year. We control for the (log) revenue labor productivity in the previous year, and include firm-product-sequence fixed effects. We also include a set of year dummies. Firm-level clustered standard errors are in parentheses. The last column of Table 11 shows estimates of a regression that pools all types of sequences.

Table 12: The data underlying Table 12 is at the firm-product-year level. We groups firms according to the number of layers ($0, 1, 2$ or 3) the firm has in the year of the first observed reorganization (see definition above). For each of these groups of firm-product-years Table 12 shows estimates of regressions where the dependent variable is (log) revenue labor productivity in a given year. The key regressor is the number of management layers in the firm in the same year. We control for the (log) revenue labor productivity in the previous year, and include firm-product fixed effects. We also include a set of year dummies. Firm-level clustered standard errors are in parentheses.

Table 13 and 14: Tables 13 and 14 show estimates of the same type of regressions described for Table 11 and 12. The only difference being that the dependent variable is (log) revenue total factor productivity computed according to the Wooldridge methodology.

Table 15 to 22, and C1 and C2 : This set of tables show estimates of the same type of regressions described for Table 11 and 12.

In Table 15 and 16 the dependent variable is (log) quantity-based productivity computed according to the MULAMA methodology extended for incorporating changes in the organization of the labor input as described in Case 1 of Section 5.2.2. In this case the key regressor is the quantity sold by the firm in t, which is allowed to have permanent effects on the firm productivity.

In Table 17 and 18 the dependent variable is (log) quantity-based productivity computed according to the MULAMA methodology extended for incorporating changes in the organization of the labor input as described in Case 2 of Section 5.2.2. In this case the key regressor is the number of management layers of the firm in t, which is allowed to have permanent effects on the firm productivity.

In Table 19 and 20 the dependent variable is (log) revenue-based productivity computed according to the Wooldridge methodology extended for incorporating changes in the organization of the labor input as described in Case 2 of Section 5.2.2. In this case the key regressor is the number of management layers of the firm in t, which is allowed to have only a contemporaneous effect on the firm productivity.

In Table 21 and 22 the dependent variable is (log) quantity-based productivity computed according to the MULAMA methodology extended for incorporating changes in the organization of the labor input as described in Case 2 of Section 5.2.2. In this case the key regressor is the number of management layers of the firm in t, which is allowed to have only a contemporaneous effect on the firm productivity.

In Table C1 and C2 the dependent variable is (log) quantity-based productivity computed according to the MULAMA methodology extended for incorporating changes in the organization of the labor input as described in Case 1 of Appendix C. In this case the key regressor is the quantity sold by the firm in t, which is allowed to have only a contemporaneous effect on the firm productivity.

Table 23 to 26, and C3 and C4: This set of tables show estimates of the same type of regressions described for Table 19 to 22, but we instrument for the change in quantity or for the change in the number of layers using the number of layers in $t-1$, quantity, revenue, demand shocks and markups at time $t-1$, capital at time t, and productivity at time $t-2$. Also, we replace the firm-product or firm-product-sequence fixed effects with a set of product dummies and year dummies.

Figure 5, 6, and 7: These figures report kernel density estimates of the distribution of log value added (Figure 5), log hours worked (Figure 6) and log hourly wage (Figure 7) by number of layers in the firm. One density is estimated for each group of firms with the same number of layers.

Appendix B

Baseline MULAMA

In this appendix we show how to derive the first stage estimating equation (11), for the baseline MULAMA

Demand in log is given by

$$p_{it} = \left(1 - \frac{1}{\eta_{it}}\right)\lambda_{it} - \frac{1}{\eta_{it}} q_{it}, \tag{16}$$

the markup then is

$$\mu_{it} = \frac{\eta_{it}}{\eta_{it} - 1}. \tag{17}$$

The production function in log is

$$q_{it} = a_{it} + \alpha_O o_{it} + \alpha_M m_{it} + (\gamma - \alpha_M - \alpha_O) k_{it}, \tag{18}$$

and the assumptions over Markov process for productivity and demand are given by

$$a_{it} = \phi_a \; a_{it-1} + \nu_{ait}, \tag{19}$$

$$\lambda_{it} = \phi_\lambda \lambda_{it-1} + \nu_{\lambda it}. \tag{20}$$

From cost minimization one obtains that the marginal cost is given by

$$\frac{\partial C_i}{\partial Q_i} = \frac{1}{\alpha_O + \alpha_M} \frac{C_i}{Q_i} \tag{21}$$

Using (16) and (17), revenue can be expressed in the following way

$$r_{it} = \frac{1}{\mu_{it}} \left(q_{it} + \lambda_{it}\right). \tag{22}$$

Using (21) note that expenditure shares are related to markups in the following way

$$\frac{1}{\mu_{it}} = \frac{s_{Oit}}{\alpha_O} = \frac{s_{Mit}}{\alpha_L}. \tag{23}$$

First stage

In order to derive the estimating equation start from (18) and (19) to obtain

$$q_{it} = \alpha_O \left(o_{it} - k_{it}\right) + \alpha_M \left(m_{it} - k_{it}\right) + \gamma k_{it} + \phi_a \; a_{it-1} + \nu_{ait}, \tag{24}$$

then substituting this expression into (22)

$$r_{it} = \frac{\alpha_O}{\mu_{it}}\left(o_{it} - k_{it}\right) + \frac{\alpha_M}{\mu_{it}}\left(m_{it} - k_{it}\right) + \frac{\gamma}{\mu_{it}} k_{it} + \frac{\phi_a}{\mu_{it}}\ a_{it-1} + \frac{1}{\mu_{it}}\nu_{ait} + \frac{1}{\mu_{it}}\lambda_{it},$$

rearranging and using (23) we define as in the main body of the text

$$LHS_{it} = \frac{r_{it} - s_{Oit}\left(o_{it} - k_{it}\right) - s_{Mit}\left(m_{it} - k_{it}\right)}{s_{Mit}}, \tag{25}$$

or written differently

$$LHS_{it} = \frac{\gamma}{\alpha_M} k_{it} + \frac{\phi_a}{\alpha_M} a_{it-1} + \frac{1}{\alpha_M}\lambda_{it} + \frac{1}{\alpha_M}\nu_{ait}.$$

We need to find expressions for a_{it-2} and λ_{it-1}.

From (22) note that

$$\lambda_{it-1} = \mu_{it-1} r_{it-1} - q_{it-1}, \tag{26}$$

then using this expression into (20) we obtain

$$\lambda_{it} = \phi_\lambda\left(\mu_{it-1} r_{it-1} - q_{it-1}\right) + \nu_{\lambda it}. \tag{27}$$

Now from (25) we can obtain

$$a_{it-2} = \frac{\alpha_M}{\phi_a} LHS_{it-1} - \frac{\gamma}{\phi_a} k_{it-1} - \frac{1}{\phi_a}\lambda_{it-1} - \frac{1}{\phi_a}\nu_{ait-1},$$

using (26) we get

$$a_{it-2} = \frac{\alpha_M}{\phi_a} LHS_{it-1} - \frac{\gamma}{\phi_a} k_{it-1} - \frac{1}{\phi_a}\left(\mu_{it-1} r_{it-1} - q_{it-1}\right) - \frac{1}{\phi_a}\nu_{ait-1},$$

and after substituting this expression in (19),

$$a_{it-1} = \alpha_M LHS_{it-1} - \gamma k_{it-1} - \left(\mu_{it-1} r_{it-1} - q_{it-1}\right). \tag{28}$$

Using (28) and (27) into (25) we obtain,

$$\begin{aligned}
LHS_{it} \quad = \quad & \frac{\gamma}{\alpha_M} k_{it} + \phi_a LHS_{it-1} - \gamma \frac{\phi_a}{\alpha_M} k_{it-1} \\
& + \left(\phi_\lambda - \phi_a\right) \frac{r_{it-1}}{s_{Mit-1}} \\
& + \left(\phi_a - \phi_\lambda\right) \frac{1}{\alpha_M} q_{it-1} \\
& + \frac{1}{\alpha_M}\left(\nu_{ait} + \nu_{\lambda it}\right).
\end{aligned}$$

More compactly, we end up with (11)

$$LHS_{it} = b_1 z_{1it} + b_2 z_{2it} + b_3 z_{3it} + b_4 z_{4it} + b_5 z_{5it} + u_{it}.$$

Second stage

Using (23) firm log output q_{it} can be written as

$$q_{it} = \mu_{it} s_{Oit} (o_{it} - k_{it}) + \mu_{it} s_{Mit} (m_{it} - k_{it}) + \gamma k_{it} + a_{it}.$$

Further exploiting (19) and (28), as well as $\hat{b}_1 = \frac{\gamma}{\alpha_M} = \frac{\gamma}{s_{Mit}\mu_{it}}$ and $\hat{b}_2 = \phi_a$, we get

$$q_{it} = \gamma \frac{o_{it} - k_{it}}{\hat{b}_1} \frac{s_{Oit}}{s_{Mit}} + \gamma \frac{m_{it} - k_{it}}{\hat{b}_1} + \gamma k_{it} + \gamma \frac{\hat{b}_2}{\hat{b}_1} LHS_{it-1} - \gamma \hat{b}_2 k_{it-1} - \gamma \frac{r_{it-1}\hat{b}_2}{\hat{b}_1 s_{Mit-1}} + \hat{b}_2 q_{it-1} + \nu_{ait}$$

and so

$$q_{it} - \hat{b}_2 q_{it-1} = b_6 z_{6it} + \nu_{ait}$$

where

$$z_{6it} = \frac{o_{it} - k_{it}}{\hat{b}_1} \frac{s_{Oit}}{s_{Mit}} + \frac{m_{it} - k_{it}}{\hat{b}_1} + k_{it} + \frac{\hat{b}_2}{\hat{b}_1} LHS_{it-1} - \hat{b}_2 k_{it-1} - \frac{r_{it-1}\hat{b}_2}{\hat{b}_1 s_{Mit-1}},$$

and $b_6 = \gamma$.

Derivations for Case 1

In this case, the production function is given by

$$q_{it} = \tilde{a}_{it} + \alpha_O \ln n^0_{iL,t} + \alpha_M m_{it} + (\gamma - \alpha_M - \alpha_O) k_{it}, \tag{29}$$

and the Markov process for productivity and demand shocks

$$\begin{aligned} \tilde{a}_{it} &= \phi_a \, \tilde{a}_{it-1} + \phi_q \, q_{it} + \nu_{ait}, & (30) \\ \lambda_{it} &= \phi_\lambda \lambda_{it-1} + \nu_{\lambda it}, & (31) \end{aligned}$$

where ν_{ait} and $\nu_{\lambda it}$ can be correlated with each other.

At any given point in time firms minimize costs for flexible inputs (number of layer zero workers $n^0_{iL,t}$ and materials M_{it}) considering capital, as well as $\tilde{a}_{it-1}$, ν_{ait}, λ_{it} and the price of materials and knowledge as given. Short-run marginal cost thus satisfies

$$\frac{\partial C_{it}}{\partial Q_{it}} = \frac{1 - \phi_q}{\alpha_O + \alpha_M} \frac{C_{it}}{Q_{it}}. \tag{32}$$

Using (16) and (17), log revenue can be expressed in the following way:

$$r_{it} = \frac{1}{\mu_{it}} \left(q_{it} + \lambda_{it} \right) \tag{33}$$

Using (32) note that expenditure shares are related to markups

$$\frac{1}{\mu_{it} \left(1 - \phi_q \right)} = \frac{s_{Mit}}{\alpha_M} = \frac{s_{Oit}}{\alpha_O}, \tag{34}$$

where s_{Oit} here represents the share of layer zero workers expenditure in total revenue.

Using (29) and (30) one obtains:

$$q_{it} = \frac{\alpha_O}{1-\phi_q} \left(\ln n_{iL,t}^0 - k_{it} \right) + \frac{\alpha_M}{1-\phi_q} \left(m_{it} - k_{it} \right) + \frac{\gamma}{1-\phi_q} k_{it} + \frac{\phi_a}{1-\phi_q} \tilde{a}_{it-1} + \frac{1}{1-\phi_q} \nu_{ait}, \tag{35}$$

then substituting this expression into (33)

$$\begin{aligned} r_{it} \; = \; & \frac{\alpha_O}{\left(1-\phi_q\right)\mu_{it}} \left(\ln n_{iL,t}^0 - k_{it} \right) + \frac{\alpha_M}{\left(1-\phi_q\right)\mu_{it}} \left(m_{it} - k_{it} \right) \\ & + \frac{\gamma}{\left(1-\phi_q\right)\mu_{it}} k_{it} + \frac{\phi_a}{\left(1-\phi_q\right)\mu_{it}} \tilde{a}_{it-1} + \frac{1}{\left(1-\phi_q\right)\mu_{it}} \nu_{ait} + \frac{1}{\mu_{it}} \lambda_{it}. \end{aligned}$$

First stage

Rearranging and using (34) we define LHS_{it} and get:

$$\begin{aligned} LHS_{it} \; \equiv \; & \frac{r_{it} - s_{Oit} \left(\ln n_{iL,t}^0 - k_{it} \right) - s_{Mit} \left(m_{it} - k_{it} \right)}{s_{Mit}} \\ = \; & \frac{\gamma}{\alpha_M} k_{it} + \frac{\phi_a}{\alpha_M} \tilde{a}_{it-1} + \frac{1-\phi_q}{\alpha_M} \lambda_{it} + \frac{1}{\alpha_M} \nu_{ait}. \end{aligned} \tag{36}$$

We need to find expressions for $\tilde{a}_{it-2}$ and λ_{it-1}. From (33) note that:

$$\lambda_{it-1} = \mu_{it-1} r_{it-1} - q_{it-1}, \tag{37}$$

then using this expression into (31) we obtain:

$$\lambda_{it} = \phi_\lambda \left(\mu_{it-1} r_{it-1} - q_{it-1} \right) + \nu_{\lambda it}. \tag{38}$$

Now from (36) we can obtain

$$\tilde{a}_{it-2} = \frac{\alpha_M}{\phi_a} LHS_{it-1} - \frac{\gamma}{\phi_a} k_{it-1} - \frac{1-\phi_q}{\phi_a} \lambda_{it-1} - \frac{1}{\phi_a} \nu_{ait-1},$$

while using (37) we get

$$\tilde{a}_{it-2} = \frac{\alpha_M}{\phi_a} LHS_{it-1} - \frac{\gamma}{\phi_a} k_{it-1} - \frac{1-\phi_q}{\phi_a} \left(\mu_{it-1} r_{it-1} - q_{it-1}\right) - \frac{1}{\phi_a} \nu_{ait-1},$$

and after substituting this expression in (30)

$$\tilde{a}_{it-1} = \alpha_M LHS_{it-1} - \gamma k_{it-1} - \left(1-\phi_q\right) \left(\mu_{it-1} r_{it-1} - q_{it-1}\right) + \phi_q \; q_{it-1}. \tag{39}$$

Using (39) and (38) into (36) we obtain

$$\begin{aligned} LHS_{it} &= \frac{\gamma}{\alpha_M} k_{it} + \phi_a LHS_{it-1} - \gamma \frac{\phi_a}{\alpha_M} k_{it-1} \\ &\quad + \left(\phi_\lambda - \phi_a\right) \frac{r_{it-1}}{s_{Mit-1}} \\ &\quad + \left(\phi_a - \phi_\lambda + \phi_q \phi_\lambda\right) \frac{1}{\alpha_M} q_{it-1} \\ &\quad + \frac{1}{\alpha_M} \left(\nu_{ait} + \left(1-\phi_q\right) \nu_{\lambda it}\right), \end{aligned} \tag{40}$$

that we can rewrite as:

$$LHS_{it} = b_1 z_{1it} + b_2 z_{2it} + b_3 z_{3it} + b_4 z_{4it} + b_5 z_{5it} + u_{it}, \tag{41}$$

where z_{1it}=k_{it}, z_{2it}=LHS_{it-1}, z_{3it}=k_{it-1}, z_{4it}=$\frac{r_{it-1}}{s_{Mit-1}}$, z_{5it}=q_{it-1}, u_{it}=$\frac{1}{\alpha_M}\left(\nu_{ait} + \left(1-\phi_q\right)\nu_{\lambda it}\right)$ as well as b_1=$\frac{\gamma}{\alpha_M}$, b_2=ϕ_a, b_3=$-\gamma\frac{\phi_a}{\alpha_M}$, b_4=$(\phi_\lambda - \phi_a)$ and b_5=$\frac{1}{\alpha_M}\left(\phi_a - \phi_\lambda + \phi_q\phi_\lambda\right)$. Given our assumptions the error term u_{it} in is uncorrelated with all of the regressors. Therefore (41) can be estimated via simple OLS. After doing this we set $\hat{\beta}$=$\hat{b}_1$ and $\hat{\phi}_a$=$\hat{b}_2$ and do not exploit parameters' constraints in the estimation.

Second stage

From (35) and (39) we have that log output is given by:

$$\begin{aligned} q_{it} &= \frac{\alpha_O}{1-\phi_q}\left(\ln n^0_{iL,t} - k_{it}\right) + \frac{\alpha_M}{1-\phi_q}\left(m_{it} - k_{it}\right) + \frac{\gamma}{1-\phi_q} k_{it} \\ &\quad + \frac{\phi_a}{1-\phi_q} \alpha_M LHS_{it-1} - \frac{\phi_a}{1-\phi_q} \gamma k_{it-1} - \frac{\phi_a}{1-\phi_q}\left(\frac{\alpha_M}{s_{Mit-1}} r_{it-1} - q_{it-1}\right) + \frac{1}{1-\phi_q} \nu_{ait}. \end{aligned}$$

Substituting (34) and known parameters from the first stage, we obtain

$$\begin{aligned} q_{it} &= \frac{\gamma}{1-\phi_q} \frac{1}{\hat{\beta}} \frac{s_{Oit}}{s_{Mit}} \left(\ln n^0_{iL,t} - k_{it}\right) + \frac{\gamma}{1-\phi_q} \frac{1}{\hat{\beta}} \left(m_{it} - k_{it}\right) + \frac{\gamma}{1-\phi_q} k_{it} \\ &\quad + \frac{\gamma}{1-\phi_q} \frac{\hat{\phi}_a}{\hat{\beta}} LHS_{it-1} - \frac{\gamma}{1-\phi_q} \hat{\phi}_a k_{it-1} - \frac{\gamma}{1-\phi_q} \frac{1}{\hat{\beta}} \frac{\hat{\phi}_a}{s_{Mit-1}} r_{it-1} + \frac{\hat{\phi}_a}{1-\phi_q} q_{it-1} + \frac{1}{1-\phi_q} \nu_{ait.} \end{aligned}$$

Note that the only parameters to estimate are γ and ϕ_q. These are identified using a linear model:

$$q_{it} = b_6 z_{6it} + b_7 z_{7it} + \varepsilon_{it}, \tag{42}$$

where:

$$\begin{aligned} z_{6it} &= \frac{1}{\hat{\beta}} \frac{s_{Oit}}{s_{Mit}} \left(\ln n^0_{iL,t} - k_{it} \right) + \frac{1}{\hat{\beta}} (m_{it} - k_{it}) + k_{it} + \frac{\hat{\phi}_a}{\hat{\beta}} LHS_{it-1} - \hat{\phi}_a k_{it-1} - \frac{1}{\hat{\beta}} \frac{\hat{\phi}_a}{s_{Mit-1}} r_{it-1}, \\ z_{7it} &= \hat{\phi}_a q_{it-1}, \\ \varepsilon_{it} &= \frac{1}{1-\phi_q} \nu_{ait}, \end{aligned}$$

as well as $b_6 = \frac{\gamma}{1-\phi_q}$ and $b_7 = \frac{1}{1-\phi_q}$. Note that z_{6it} is endogenous and we instrument it with k_{it}. We then set $\hat{\gamma} = \hat{b}_6 / \hat{b}_7$ and $\hat{\phi}_q = \frac{\hat{b}_7 - 1}{\hat{b}_7}$ and obtain:

$$\hat{a}_{it} = q_{it} - \frac{\hat{\gamma}}{\hat{\beta}} \frac{s_{Oit}}{s_{Mit}} \left(\ln n^0_{it} - k_{it} \right) - \frac{\hat{\gamma}}{\hat{\beta}} (m_{it} - k_{it}) - \hat{\gamma} k_{it}$$

$$\hat{\mu}_{it} = \frac{\hat{\gamma}}{\hat{\beta} s_{Mit} (1 - \hat{\phi}_q)}$$

$$\hat{\lambda}_{it} = \frac{\hat{\gamma}}{\hat{\beta} s_{Mit} (1 - \hat{\phi}_q)} r_{it} - q_{it}.$$

Derivations for Case 2

In this case, the production function in logs can be written as

$$q_{it} = \tilde{a}_{it} + \alpha_O \ln WB_{it} + \alpha_M m_{it} + (\gamma - \alpha_M - \alpha_O) k_{it}, \tag{43}$$

where WB_{it} is the wage bill at time t of firm i.

The Markov process for productivity and demand shocks are

$$\begin{aligned} \tilde{a}_{it} &= \phi_a \, \tilde{a}_{it-1} + \phi_L L_{it} + \nu_{ait}, & (44) \\ \lambda_{it} &= \phi_\lambda \lambda_{it-1} + \nu_{\lambda it}, & (45) \end{aligned}$$

where ν_{ait} and $\nu_{\lambda it}$ can be correlated with each other.

Following the same steps as we did for Case 1, the short-run marginal cost satisfies

$$\frac{\partial C_{it}}{\partial Q_{it}} = \frac{1}{\alpha_O + \alpha_M} \frac{C_{it}}{Q_{it}}. \tag{46}$$

Using (16) and (17), log revenue can be expressed in the following way:

$$r_{it} = \frac{1}{\mu_{it}} \left(q_{it} + \lambda_{it} \right). \tag{47}$$

Using (21) note that expenditure shares are related to markups in the following way

$$\frac{1}{\mu_{it}} = \frac{s_{Mit}}{\alpha_M} = \frac{s_{Oit}}{\alpha_O}. \tag{48}$$

where s_{Oit} here represents the share of total labor expenditure in total revenue.

Using (43) and (44) one obtains:

$$q_{it} = \alpha_O \left(\ln WB_{it} - k_{it} \right) + \alpha_M \left(m_{it} - k_{it} \right) + \gamma k_{it} + \phi_a \tilde{a}_{it-1} + \phi_L L_{it} + \nu_{ait}, \tag{49}$$

then substituting this expression into (47):

$$\begin{aligned} r_{it} &= \frac{\alpha_O}{\mu_{it}} \left(\ln WB_{it} - k_{it} \right) + \frac{\alpha_M}{\mu_{it}} \left(m_{it} - k_{it} \right) \\ &\quad + \frac{\gamma}{\mu_{it}} k_{it} + \frac{\phi_a}{\mu_{it}} \tilde{a}_{it-1} + \frac{\phi_L}{\mu_{it}} L_{it} + \frac{1}{\mu_{it}} \nu_{ait} + \frac{1}{\mu_{it}} \lambda_{it}. \end{aligned}$$

First stage

Rearranging and using (48) we define LHS_{it} and get:

$$\begin{aligned} LHS_{it} &\equiv \frac{r_{it} - s_{Oit} \left(\ln WB_{it} - k_{it} \right) - s_{Mit} \left(m_{it} - k_{it} \right)}{s_{Mit}} \\ &= \frac{\gamma}{\alpha_M} k_{it} + \frac{\phi_a}{\alpha_M} \tilde{a}_{it-1} + \frac{1}{\alpha_M} \lambda_{it} + \frac{\phi_L}{\alpha_M} L_{it} + \frac{1}{\alpha_M} \nu_{ait}. \end{aligned} \tag{50}$$

We need to find expressions for $\tilde{a}_{it-2}$ and λ_{it-1}. From (22) note that:

$$\lambda_{it-1} = \mu_{it-1} r_{it-1} - q_{it-1}, \tag{51}$$

then using this expression into (45) we obtain:

$$\lambda_{it} = \phi_\lambda \left(\mu_{it-1} r_{it-1} - q_{it-1} \right) + \nu_{\lambda it}. \tag{52}$$

Now from (50) we can obtain:

$$\tilde{a}_{it-2} = \frac{\alpha_M}{\phi_a} LHS_{it-1} - \frac{\gamma}{\phi_a} k_{it-1} - \frac{1}{\phi_a} \lambda_{it-1} - \frac{\phi_L}{\phi_a} L_{it-1} - \frac{1}{\phi_a} \nu_{ait-1},$$

while using (51) we get:

$$\tilde{a}_{it-2} = \frac{\alpha_M}{\phi_a} LHS_{it-1} - \frac{\gamma}{\phi_a} k_{it-1} - \frac{1}{\phi_a} \left(\mu_{it-1} r_{it-1} - q_{it-1}\right) - \frac{\phi_L}{\phi_a} L_{it-1} - \frac{1}{\phi_a} \nu_{ait-1},$$

and after substituting this expression in (44)

$$\tilde{a}_{it-1} = \alpha_M LHS_{it-1} - \gamma k_{it-1} - \left(\mu_{it-1} r_{it-1} - q_{it-1}\right). \tag{53}$$

Using (53) and (52) into (50) we obtain:

$$\begin{aligned} LHS_{it} = {} & \frac{\gamma}{\alpha_M} k_{it} + \phi_a LHS_{it-1} - \gamma \frac{\phi_a}{\alpha_M} k_{it-1} \\ & + (\phi_\lambda - \phi_a) \frac{r_{it-1}}{s_{Mit-1}} \\ & + (\phi_a - \phi_\lambda) \frac{1}{\alpha_M} q_{it-1} \\ & + \frac{\phi_L}{\alpha_M} L_{it} + \frac{1}{\alpha_M} (\nu_{ait} + \nu_{\lambda it}), \end{aligned} \tag{54}$$

that we can rewrite as

$$LHS_{it} = b_1 z_{1it} + b_2 z_{2it} + b_3 z_{3it} + b_4 z_{4it} + b_5 z_{5it} + b_6 z_{6it} + u_{it}, \tag{55}$$

where $z_{1it}{=}k_{it}$, $z_{2it}{=}LHS_{it-1}$, $z_{3it}{=}k_{it-1}$, $z_{4it}{=}\frac{r_{it-1}}{s_{Mit-1}}$, $z_{5it}{=}q_{it-1}$, $z_{6it}{=}L_{it}$, $u_{it}{=}\frac{1}{\alpha_M}(\nu_{ait} + \nu_{\lambda it})$ as well as $b_1{=}\frac{\gamma}{\alpha_M}$, $b_2{=}\phi_a$, $b_3{=}-\gamma\frac{\phi_a}{\alpha_M}$, $b_4{=}(\phi_\lambda - \phi_a)$, $b_5{=}\frac{1}{\alpha_M}(\phi_a - \phi_\lambda)$ and $b_6{=}\frac{\phi_L}{\alpha_M}$. Given our assumptions the error term u_{it} in is uncorrelated with all of the regressors but $z_{6it}{=}L_{it}$. Coherently with our assumptions we instrument L_{it} with L_{it-1}. After doing this we set $\hat{\beta}{=}\hat{b}_1$ and $\hat{\phi}_a{=}\hat{b}_2$ and do not exploit parameters' constraints in the estimation.

Second stage

From (49) and (53) we have that log output is given by

$$\begin{aligned} q_{it} = {} & \alpha_O (\ln WB_{it} - k_{it}) + \alpha_M (m_{it} - k_{it}) + \gamma k_{it} \\ & + \phi_a \alpha_M LHS_{it-1} - \phi_a \gamma k_{it-1} - \phi_a \left(\frac{\alpha_M}{s_{Mit-1}} r_{it-1} - q_{it-1}\right) + \phi_L L_{it} + \nu_{ait}. \end{aligned}$$

Substituting (48) and known parameters from the first stage, we obtain

$$\begin{aligned} q_{it} = {} & \gamma \frac{1}{\hat{\beta}} \frac{s_{Oit}}{s_{Mit}} (\ln WB_{it} - k_{it}) + \gamma \frac{1}{\hat{\beta}} (m_{it} - k_{it}) + \gamma k_{it} \\ & + \gamma \frac{\hat{\phi}_a}{\hat{\beta}} LHS_{it-1} - \gamma \hat{\phi}_a k_{it-1} - \gamma \frac{1}{\hat{\beta}} \frac{\hat{\phi}_a}{s_{Mit-1}} r_{it-1} + \hat{\phi}_a q_{it-1} + \phi_L L_{it} + \nu_{ait.} \end{aligned}$$

Note that the only parameters to estimate are γ and ϕ_L. These are identified using a linear model:

$$q_{it} - \hat{\phi}_a q_{it-1} = b_7 z_{7it} + b_8 z_{8it} + \varepsilon_{it}, \tag{56}$$

where:

$$\begin{aligned} z_{7it} &= \frac{1}{\hat{\beta}} \frac{s_{Oit}}{s_{Mit}} (\ln WB_{it} - k_{it}) + \frac{1}{\hat{\beta}} (m_{it} - k_{it}) + k_{it} + \frac{\hat{\phi}_a}{\hat{\beta}} LHS_{it-1} - \hat{\phi}_a k_{it-1} - \frac{1}{\hat{\beta}} \frac{\hat{\phi}_a}{s_{Mit-1}} r_{it-1}, \\ z_{8it} &= L_{it}, \\ \varepsilon_{it} &= \nu_{ait}, \end{aligned}$$

as well as $b_7 = \gamma$ and $b_8 = \phi_L$. Note that z_{7it} and z_{8it} are endogenous and we instrument them with k_{it} and L_{it-1}. We then set $\hat{\gamma} = \hat{b}_7$ and obtain

$$\hat{a}_{it} = q_{it} - \frac{\hat{\gamma}}{\hat{\beta}} \frac{s_{Oit}}{s_{Mit}} (\ln WB_{it} - k_{it}) - \frac{\hat{\gamma}}{\hat{\beta}} (m_{it} - k_{it}) - \hat{\gamma} k_{it}$$

$$\hat{\mu}_{it} = \frac{\hat{\gamma}}{\hat{\beta} s_{Mit}}$$

$$\hat{\lambda}_{it} = \frac{\hat{\gamma}}{\hat{\beta} s_{Mit}} r_{it} - q_{it}.$$

7 Appendix C

In this Appendix we present the results for Case 1 without and with IV when we allow only for contemporaneous effects quantity-based productivity.

The process for quantity-based productivity in equation (12), when we only consider contemporaneous effects, becomes

$$\tilde{a}_{it} = \phi_a \ \tilde{a}_{it-1} + \beta \Delta q_{it} + \nu_{ait}. \tag{57}$$

We use the process in (57) and recompute all the shocks in the model using the MULAMA methodology. The results are presented in Table C.1 and C.2. The results in Table C.1 are consistent with our hypothesis and all highly significant. The effect of quantity on productivity is positive and significant for firms that either increase or drop layers, and it is larger for these firms than for firms that do not change layers. Note also that the effects are much smaller than in Table 15 which indicates that the average level effect over time might be larger than the immediate contemporaneous effects.

Table C.2 presents the results ordered by layer. The results for firms that start with 1, 2 or 3 layers are all positive and significant, although the estimate for firms with zero layers is negative and not significant.

Table C.3 presents the results for the specification of the process of quantity-based productivity in

Table C.1: MULAMA Quantity TFP Case 1 Cont. Firm-product-sequence FE

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.141^c	0.061	0.248^b	0.149^a
	(0.086)	(0.050)	(0.111)	(0.054)
Change in quantity	0.233^a	0.205^a	0.166^a	0.194^a
	(0.081)	(0.065)	(0.058)	(0.041)
Constant	0.074^c	0.167^a	0.049	-0.048
	(0.041)	(0.061)	(0.034)	(0.052)
Observations	4,171	2,840	3,055	10,066
Number of fixed effects	1,673	1,280	1,298	4,251
Adjusted R^2	0.038	0.023	0.069	0.035

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Table C.2: MULAMA Quantity TFP: Case 1 Contemporaneous. Firm-product FE

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	-0.179	0.201	0.370^a	0.231^a
	(0.146)	(0.133)	(0.060)	(0.048)
Change in quantity	-0.245	0.182^b	0.332^a	0.167^a
	(0.237)	(0.084)	(0.078)	(0.048)
Constant	0.818^a	0.251^a	0.014	-0.026
	(0.205)	(0.061)	(0.029)	(0.036)
Observations	532	1,649	3,674	3,523
Number of fixed effects	163	549	1,149	1,148
Adjusted R^2	0.037	0.067	0.141	0.073

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

equation (57), but when we instrument for the change in quantity using the variables described in the main text. As we can see, for a firm that adds layers, a 10% increase in quantity leads to a 2.5% increase in productivity. The results are much weaker (and not significant) for firms that drop layers. They are also positive and significant for firms with constant numbers of layers, which indicates that quantity and quantity-based productivity are probably related through other channels, as we had found before. Still, for the significant estimates, we find that changes in organization lead to a larger change in real-based productivity than when we do not observe a reorganization. The persistence coefficient in this case is around .9. As before, the reason is that we have dropped the set of firm-product-sequence fixed effects.

Table C.4 presents the results when we group by initial layer. All the estimates of the effects of changes in quantity are positive, although as in previous cases, not significant for firms that start with either zero

Table C.3: MULAMA Quantity TFP: Case 1 Contemporaneous with IV

VARIABLES	(1) Increasing	(2) Decreasing	(3) Constant	(4) All
Productivity t-1	0.872^a	0.856^a	0.920^a	0.884^a
	(0.024)	(0.024)	(0.026)	(0.016)
Change in quantity	0.246^b	0.095	0.148^b	0.170^a
	(0.100)	(0.138)	(0.073)	(0.062)
Observations	3,748	2,181	2,098	8,027
Adjusted R^2	0.720	0.715	0.800	0.750

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Table C.4: MULAMA Quantity TFP: Case 1 Contemporaneous with IV

VARIABLES	(1) 0 layers	(2) 1 layer	(3) 2 layers	(4) 3 layers
Productivity t-1	0.802^a	0.866^a	0.866^a	0.883^a
	(0.074)	(0.040)	(0.020)	(0.031)
Change in quantity	0.317	0.085	0.331^a	0.161^a
	(0.372)	(0.150)	(0.105)	(0.061)
Observations	532	1,649	3,674	3,523
Adjusted R^2	0.500	0.677	0.737	0.746

Firm-level clustered standard errors in parentheses
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

or one layer.