The Anatomy of French Production Hierarchies

Lorenzo Caliendo

Yale University

Ferdinando Monte

Johns Hopkins University

Esteban Rossi-Hansberg

Princeton University

We study the internal organization of French manufacturing firms. We divide the employees of each firm into "layers" using occupational categories. Layers are hierarchical in that the typical worker in a higher layer earns more, and the typical firm occupies less of them. The probability of adding/dropping a layer is positively/negatively correlated with value added. Reorganization, through changes in layers, is essential to understanding how firms grow. Firms that expand substantially add layers and pay lower average wages in all preexisting layers. In contrast, firms that expand little and do not reorganize pay higher average wages in all preexisting layers.

I. Introduction

Labor is not a homogeneous input. Employees are distinct in their levels of skill, knowledge, experience, and a vast variety of other dimensions.

We thank Ariel Burstein, Pierangelo De Pace, Luis Garicano, Bob Gibbons, Gene Grossman, Maria Guadalupe, Gordon Hanson, Oleg Itskhoki, Wolfgang Keller, Sam Kortum, Francis Kramarz, Claire Lelarge, Bob Lucas, Marc Melitz, Alex Monge, Derek Neal, Steve Redding, Peter Schott, Chris Sims, John van Reenen, and Julie Wulf for useful conversations and comments. The computations in this paper were done at a secure data center

Electronically published July 9, 2015
[*Journal of Political Economy*, 2015, vol. 123, no. 4]

So an important decision made by firms is to determine not only the number but also the characteristics of their employees, as well as the role that each of them plays in the firm. We refer to these decisions as the organization of a firm. In this paper we aim to describe empirically the organization of firms and how this organization is related to other firm characteristics. We are particularly interested in understanding if firms actively manage their organization—and therefore the number and characteristics of their employees—and how they do this. Understanding these decisions is important in order to understand the behavior of firms and therefore that of the aggregate economy. As far as we know, this is the first empirical study of the internal organization of firms that uses a comprehensive data set with a large number of firms.

We use a sample of the large majority of French manufacturing firms during the period 2002–7.[1] To organize the data in a practical and meaningful way, we first introduce the concept of a "layer" of employees. The concept of a layer is adopted from the theory of management hierarchies proposed initially by Rosen (1982) and Garicano (2000) and used in the context of heterogeneous firms in Caliendo and Rossi-Hansberg (2012). In this theory a layer is a group of employees, with similar characteristics summarized in their knowledge, who perform similar tasks within the organization. Conceptually, these layers are hierarchical in the sense that higher layers of management are smaller and include more knowledgeable employees who have as subordinates employees in lower layers. Dividing the employees in real firms into layers requires some mapping between these concepts and the data.

The first part of the map involves using wages as a one-dimensional measure of the marketable characteristics of employees, that is, to view wages as a measure of the "knowledge" of workers. In this vein, if one individual is more knowledgeable than another (in terms of practical knowledge used in production), he will obtain a higher wage. Later in the paper we discuss the particular way in which this average knowledge in a layer is modified and how it is created using a combination of education and experience.

The second part of the map is to group workers into layers. To do so we use information on occupational characteristics. Fortunately, the French data we use provide hierarchical occupational categories. The top occupation includes owners who receive a wage. The next one below includes senior staff and top management positions and the next one employees at a supervisory level. The lowest two occupations include clerks and blue-collar workers. We document that they earn similar

located at CREST, Paris. Caliendo acknowledges the support from the Yale MacMillan Center. Data are provided as supplementary material online.

[1] A detailed description of the data is relegated to online app. B.

wages, and so we pull them together into one layer. This gives us a maximum of four hierarchical layers of employees. Of course, many firms (in fact, most of them) do not employ agents in all of these layers, something we exploit extensively in this paper.

We then investigate whether this division of the employees in a firm into layers is an economically meaningful classification. We cut the data in a variety of ways that indicate that it is. Firms with more layers are larger in terms of value added and employment and, in general, pay higher wages. Around 50 percent of the variation in wages within firms is variation across layers. A large majority of firms have consecutively ordered layers that start at the bottom. When they add or drop a layer, it is mostly a consecutive layer, and they add or drop only one. Layers within firms are also different from each other. Lower layers are larger in the number of hours of work and employ agents who earn lower wages. Thus, most firms are hierarchical in their layers both in terms of wages and in terms of time employed at each layer. In addition, the probability of adding a layer is increasing in value added, and the probability of dropping one is decreasing in value added. Finally, the firms that we observe adding/dropping a layer in a given period tend to grow/shrink in the previous periods, indicating that they progressively get closer to a size threshold that triggers the change. All these facts are very significant in the data and robust to accounting for industry and time fixed effects. So we conclude that the layers we identify using occupations are not arbitrary names but have an economic meaning in terms of the characteristics of the employees they group and the tasks they perform.

The next step is to understand how firms change their organization—the knowledge and number of employees at each layer—as they grow. It is useful to go back to the theory in order to guide our exploration. We rely on Caliendo and Rossi-Hansberg (2012) as our guide since their general equilibrium theory of production hierarchies allows for firm heterogeneity, which is important in the data.[2] In this theory, firms organize production to economize on their use of knowledge: a costly input. Production requires time and knowledge. Workers in layer 1 work on the production floor. To produce, they need to solve the problems they face in production. Their knowledge allows them to solve some, but not all, of the problems they face. If they can solve a problem, output is immediately realized. Otherwise they can ask agents in higher layers how to solve them. Since they do not know anything about these problems, they first ask the managers in layer 2. These managers spend their time communicating with the workers and understanding their problems.

[2] Throughout, we refer mostly to the theory of Caliendo and Rossi-Hansberg (2012), although some of the arguments we advance can be traced back to Garicano (2000) or Garicano and Rossi-Hansberg (2006).

They in turn solve some of them and pass the rest to the third layer and so on. The problem of the firm is to decide the number of hours of work and the level of knowledge of employees in each layer and how many layers to have in the firm. The number of hours of the top manager is fixed and common across firms.

A firm with higher demand, or higher exogenous idiosyncratic productivity, optimally decides to have more layers. Its larger scale allows it to economize on the total cost of knowledge by having many layers of management with very knowledgeable managers at the top but much less knowledgeable employees in the bottom layers. The theory in Caliendo and Rossi-Hansberg (2012) implies that some firms that expand value added will add layers. However, some others might expand without adding layers since the expansion is not large enough to make the added cost of an extra layer (namely, the wage of the new top manager) worth paying. The trade-off is simple: a lower "marginal" cost from having less knowledgeable employees in the existing layers (because the new top manager can solve the less frequent questions) versus a higher "fixed" cost from having to pay an extra, and large, wage of the top manager. So it is worth paying the fixed cost only if the cumulative expansion since the last change is large enough. Thus, in the theory, firms that expand by adding layers reduce wages and increase the number of hours at all layers, while firms that expand but do not add layers increase both hours and wages at all layers.

This theory has implications on firm- and layer-level outcomes. That is, it has implications for the number of workers and their average knowledge, and therefore the average wage, for each layer of employees. We go to the data guided by the implications of the theory. We look at firms that expand and add layers and firms that contract and drop layers. In particular, we estimate changes in log average wages by layer and changes in log average hours of work by layer, normalized by hours in the top layer, for firms that add or drop layers from one year to the next. We then look at firms that expand but do not add or drop layers.

In the data, firms that expand substantially in a given year tend to reorganize by changing layers. They contribute almost 40 percent of the total change in value added in the manufacturing sector. We find that wages in firms that expand and add layers behave differently than in firms that expand but do not reorganize. If firms expand by adding layers, average wages in preexisting layers fall, while if firms expand without reorganizing, average wages in all layers rise. All these results reverse when we focus on firms that contract and either drop layers or do not reorganize. These results hold for each layer in firms with any number of layers. We do not find any instance in which they are contradicted by the data. Importantly, our findings are not simply the result of regrouping workers with the same wages across layers. We document that the distribution of wages in preexisting layers shifts down for all percentiles when firms grow by adding layers, while it shifts up for all percentiles when firms grow without reorganizing.

The results above document how firms affect the average "knowledge" in each layer when they grow or decline. They do not, however, explain how firms manage to modify the characteristics of their employees to achieve these average changes.[3] The data available to explore these questions are not as well suited as the data we used for the results above; nevertheless, in Section V, we ask two questions pertinent to understanding how firms adjust average knowledge in a layer.

First, do firms affect the average wage in a layer by changing the composition of the workers in a layer or by changing the knowledge and wages of current employees? We find that when firms grow by adding a layer, average wages in a layer are reduced by changing the composition of employees in a layer and not by reducing individual wages. The extensive margin is also dominant when we look at firms that drop layers.

Second, if knowledge can be created with either formal education or experience, which of the two is affected when firms decide to change the average level of knowledge in a layer? Using estimated measures of labor market experience and years of formal education for each employee,[4] we show that firms use these two forms of acquiring knowledge in systematic but distinct ways. Firms that grow without adding layers increase knowledge by hiring workers with more formal education, particularly at the bottom of the hierarchy. In contrast, firms that expand by adding layers tend to reduce knowledge by hiring less experienced workers.[5] The behavior we uncover is consistent with the view that formal years of education provide the knowledge to solve the most common problems in an organization, the tasks handled at the lower layers of the hierarchy. In contrast, labor market experience provides the knowledge required to solve more infrequent problems, the tasks handled at the higher level of the hierarchy.

[3] The frictions in the French labor market, and therefore the implications of firm reorganization on particular individuals, are certainly worth exploring but do not constitute the main focus of our paper. Instead, our aim is to understand if and how firms actively manage and reorganize their labor force. Our focus on firms' decisions and their performance parallels the work of Bloom and van Reenen (2007), Bloom, Sadun, and van Reenen (2012), and Bloom et al. (2013), all of which study the effect of management, not organization, on firm characteristics. Our approach also relates to those of Baker, Gibbs, and Holmstrom (1994) and Baker and Holmstrom (1995), who study how a particular firm organizes its internal labor force.

[4] Direct measures of formal education and experience are not directly available in a way that can be matched to the firm data we are using. So we need to estimate worker-level education and experience using a methodology that we describe in detail in Sec. V.

[5] We also find that the reverse patterns hold when we look at firms that contract with and without changing layers.

Several other papers have studied the internal organization of firms using small samples of producers (a few hundred). For instance, Caroli and van Reenen (2001) use surveys from England and France to find that the wage bill shares of different skill levels change as firms "delayer." These results support our finding that delayering is associated with systematic occupational shifts.[6] Garicano and Hubbard (2007) study the role of hierarchies as a means of organizing production in law firms. They use confidential data from law offices from the 1992 Census of Services. They find that as market size increases, the ratio of associates to partners increases. We document a similar finding for manufacturing firms in France. We find that as firms expand, by either adding layers or not, the number of hours worked by lower-level employees relative to higher-level ones expands. Rajan and Wulf (2006), using a sample of 300 large US firms for the period 1986–98, analyze how hierarchies of top-level managers have changed over time. The study shows that the chief executive officers' span of control has increased, while the number of layers between division heads and CEOs has gone down during the sample period. Thus, they find evidence that such hierarchies have "flattened" over time and have decentralized their decision making. Using a large comprehensive data set for France, we document that firms have also become "flatter" during the period 2002–7.

The objective of our paper is to study the organization of firms and how this organization changes as firms grow or decline. The theory in Caliendo and Rossi-Hansberg (2012) emphasizes the notion of reorganization, a concept that we find helpful when looking at the data. The theory provides moment restrictions about endogenous variables, and we examine the extent to which these restrictions hold in the data. As such, our empirical strategy is not designed to estimate the causal effect of exogenous changes in organization on firm outcomes. We find that conditioning on whether a firm reorganizes or not allows us to uncover what we view as robust characteristics of firm behavior, for example, that they reduce average wages in preexisting layers when they expand and reorganize, but they increase them when they expand without adding layers. Independently of the theory that provides the true explanation for these facts, they show that the concept of reorganization through changes in layers, measured by changes in the occupations employed in a firm, is useful to characterize the behavior of firms.

The rest of the paper is organized as follows. The next section describes in more detail the essential features of the theory in Caliendo

[6] Caroli and van Reenen (2001) have the advantage of using a measure of delayering directly reported by managers rather than indirectly inferred from the occupational structure, like ours. The advantage of our approach is that it relies less on the subjective views of managers and more on their observed actions. Our measure also allows us to use the universe of manufacturing firms instead of specialized surveys.

and Rossi-Hansberg (2012) that guides our empirical exploration. Section III describes the data and our construction of layers and shows the basic characteristics of firms and layers. Section IV presents our findings on organizational changes as a result of changes in layers and expansions in value added. Section V discusses how firms change average wages in a layer, and Section VI presents conclusions. Online appendix A presents a variety of robustness checks and extensions of the results in the main text. Online appendix B describes in detail the data, their manipulation, and our empirical methodology.

II. A Theory of Organization with Heterogeneous Firms

In this section we discuss briefly the framework in Caliendo and Rossi-Hansberg (2012). Given that the purpose of the current paper is to describe and understand the data, we present the theory in its simplest form and do not discuss all the details fully. The interested reader is directed to Caliendo and Rossi-Hansberg's study for the more technical discussions and all proofs of the results.

We consider an economy with $\tilde{N}$ identical agents with preferences that lead to a demand for variety α given by $x(p, \alpha; R, P)$, where p denotes the price of a variety, R the aggregate revenue of the economy, and P the aggregate price index. We assume that agents like varieties with higher α better, so $\partial x(p, \alpha; R, P)/\partial\alpha > 0$ and, as usual, $\partial x(p, \alpha; R, P)/\partial p < 0$. Agents are endowed with one unit of time that they supply inelastically and obtain an equilibrium wage $\bar{w}$ for their unit of time. Agents acquire knowledge in order to solve the problems they encounter during production. Learning how to solve problems in an interval of knowledge of length z costs $\bar{w}cz$ (c teachers per unit of knowledge at cost $\bar{w}$ per teacher). Since the cost of knowledge is linear, agents receive it back as compensation for their work. Hence, the total wage of an employee with knowledge z is given by $w = \bar{w}[cz + 1]$.

We refer to agents that want to start a new firm as entrepreneurs. An entrepreneur pays a fixed entry cost f^E in units of labor to design her product. After doing so, she obtains a demand draw α from a known distribution $G(\alpha)$. The draw α determines the level of demand of the firm. Firms compete with each other monopolistically, and the only exogenous and heterogeneous characteristic of a firm is the level of demand for its product α.[7]

If the entrepreneur decides to produce, she pays a fixed cost f in units of labor. Production requires labor and knowledge. Agents employed in a firm act as production workers (layer $\ell = 1$) or managers (layers $\ell \geq 2$).

[7] It is important to note that any change that shifts revenue proportionally, like an exogenous firm-level productivity shock, will be isomorphic to a change in α. In this sense, α could be equivalently modeled as an exogenous productivity level as in Melitz (2003).

We denote by n_L^ℓ, z_L^ℓ, and w_L^ℓ the number, knowledge, and total wage of employees at layer $\ell = 1, 2, 3, \ldots$ of an organization with L layers.[8] Production workers use their unit of time to generate a production possibility that can yield one unit of output. For output to be realized, the worker needs to solve a problem drawn from a distribution $F(z)$ with $F''(z) < 0$. Production workers learn how to solve the most frequent problems: the ones in the interval $[0, z_L^1]$. If the problem they face falls in $[0, z_L^1]$, production is realized; otherwise, they can ask a manager one layer above how to solve the problem. Managers spend h units of their time on each problem that gets to them. A manager at layer $\ell = 2$ tries to solve the problems workers could not solve. Hence, they learn how to solve problems in $[z_L^1, z_L^1 + z_L^2]$. In general, the firm needs $n_L^\ell = hn_L^1[1 - F(Z_L^{\ell-1})]$ managers of layer ℓ, where $Z_L^\ell = \sum_{\ell=1}^{\ell} z_L^\ell$.[9]

We characterize the problem using the variable cost function. Let $C(q; w)$ denote the minimum variable cost of producing q units and $C_L(q; w)$ the same cost if we restrict the organization to producing with L layers of management. Then,

$$C(q;w) = \min_{L\geq 1}\{C_L(q;w)\} = \min_{L\geq 1, \{n_L^\ell, z_L^\ell\}_{\ell=1}^L \geq 0} \sum_{\ell=1}^{L} n_L^\ell w_L^\ell \tag{1}$$

subject to

$$q \leq F(Z_L^L) n_L^1, \tag{2}$$

$$w_L^\ell = \bar{w}[cz_L^\ell + 1] \quad \text{for all } \ell \leq L, \tag{3}$$

$$n_L^\ell = hn_L^1[1 - F(Z_L^{\ell-1})] \quad \text{for } L \geq \ell > 1, \tag{4}$$

$$n_L^L = 1. \tag{5}$$

So one entrepreneur, $n_L^L = 1$, chooses the number of layers, L, employees at each layer, n_L^ℓ, and the interval of knowledge that they acquire, z_L^ℓ, subject to the output constraint and the time constraints of employees at each layer. Figure 1 illustrates the resulting average cost function $C(q; w)/q$ as a function of q. It is the lower envelope of the average cost functions restricted to have a given number of layers, $C_L(q; w)/q$. The minimum of these average cost functions decreases with the number of layers and is reached for higher output levels the higher the layer. Each

[8] Note that we label the lowest layer of the organization, the layer of production workers, as layer 1, while in Garicano (2000), Garicano and Rossi-Hansberg (2006), and Caliendo and Rossi-Hansberg (2012), the lowest layer is denoted by layer 0.

[9] To derive some of the implications of the theory, Caliendo and Rossi-Hansberg (2012) specify the distribution of problems as an exponential, so $F(z) = 1 - e^{-\lambda z}$.

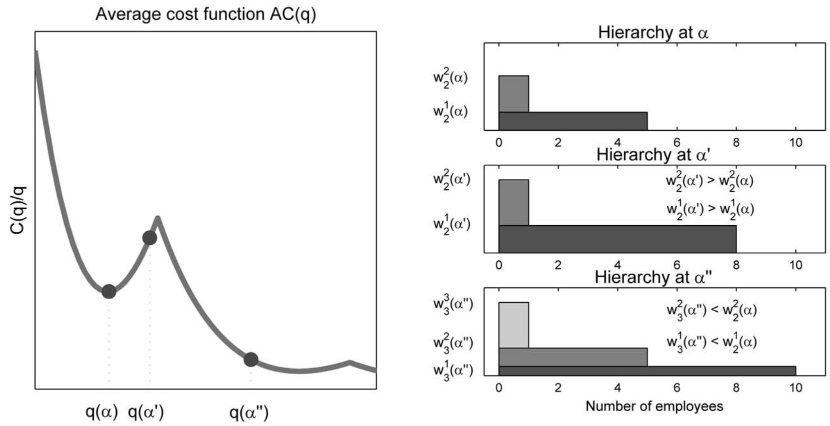

FIG. 1.—The average cost function $C(q; w/q)$ as a function of q

point in that curve is associated with a particular organization. Namely, it is associated with a number of layers and a number of employees and their knowledge at each layer.

Importantly, as proven in Caliendo and Rossi-Hansberg (2012), as firms increase the number of layers by one in order to produce more, the number of agents in each layer increases and the knowledge in all preexisting layers, and therefore the wage, decreases. In figure 1 this is illustrated as a change from $q(\alpha)$ to $q(\alpha'')$. The logic is straightforward. Firms add layers to economize on the knowledge of their workers. So when they add a new top layer, they make the new manager deal with the rare problems and make lower-level employees know less. The lower knowledge in all preexisting layers reduces, by equation (4), the span of control of each manager in the organization (namely, the number of subordinates per manager). However, the number of employees in all layers still goes up since the span of control of the new top manager is larger than one.

Figure 1 also illustrates how firms grow when they do not add layers (a change from $q(\alpha)$ to $q(\alpha')$ in the figure). In this case we see that the number of workers in all layers increases, as do the knowledge and wages of all workers. This is the only way the firm can expand given the number of layers. Because the span of managers is given by the knowledge of their subordinates, the only way the firm has to increase output is to increase the knowledge of its employees. Since the knowledge of agents at different layers is complementary, the firm does so at all layers. So the implications of the model on how firms grow and how the wages and knowledge of their employees vary depend crucially on whether firms add layers or not. Clearly, firms that want to expand substantially change the number of layers, while firms that do not want to expand that much, in general, keep the number of layers fixed. We investigate all these implications in our empirical analysis below.

So far we have not said anything about how the quantity produced is determined. To do so, we need to turn to the profit maximization and entry decision of the firm. Caliendo and Rossi-Hansberg (2012) embed the cost function discussed above into a standard Melitz (2003) type framework with heterogeneity in demand. Given that we exploit the general equilibrium of the model only in a limited way, we direct the reader to Caliendo and Rossi-Hansberg's study for details. Here we state only that the model yields an optimal quantity produced, which is increasing as a function of the demand draw α. So a higher-demand draw leads to a higher quantity and, as described above, a new organization.

To sum up, the model has the following implications:

1. Firms are hierarchical, $n_L^1 \geq \cdots \geq n_L^\ell \geq \cdots \geq n_L^L$ for all L.
2. Layers, L, sales, pq, and total number of employees, $\sum_{\ell=1}^{L} n_L^\ell$, increase with α.
3. Given L, w_L^ℓ and n_L^ℓ increase with α at all ℓ.
4. Given α, w_L^ℓ decreases and n_L^ℓ increases with an increase in L at all ℓ.

Armed with these implications and the way of organizing the data dictated by the theory, we now turn to our empirical analysis of the anatomy of French production hierarchies.

III. The Data

We use confidential data collected by the French National Statistical Institute (INSEE) for the period 2002–7. We do not use the data before 2002 because the occupational categories, which we use to determine layers below, changed that year. To construct our unique data set, we merge two different sources of mandatory reports. First is the Bénéfice Réel Normal (BRN) data set, which includes the balance sheet data of private firms. It includes 553,125 firm-year observations in the manufacturing sector. Second is the Déclarations Annuel des Données (DADS) data set, which includes occupation, hours of work, and earning reports of salaried employees. In matching the data sets, we lose 6.9 percent of the observations, and we lose another 11.5 percent from cleaning the data. The resulting sample covers, on average over time, 90.3 percent of total value added in manufacturing. We should note that small firms can choose not to report in the BRN. However, firms that choose not to report add up to a small share of value added since they are included in the 9.7 percent of value added not included in our sample. A detailed description of the construction and characteristics of the data set is included in online appendix B.

In order to dissect this large data set in a way that we can understand and analyze through the lens of the theory described in the previous section, we first need to determine what constitutes a layer of workers or management in the data. To do so, we use the PCS-ESE classification codes for workers in the manufacturing sector. Remember that the notion of a layer is a group of employees who have similar knowledge levels and wages and who perform tasks at a similar level of authority. That is, our purpose is not to separate employees in a firm according to the functional characteristics of the tasks they perform (e.g., whether they are accountants or lawyers) but rather on the basis of their hierarchical level in the organization, that is, on the basis of the number of layers of subordinates that they have below them. The PCS-ESE is, we believe, ideal for this purpose. For manufacturing it includes five occupational categories classified with numbers from 2 to 6.[10] The classes we use, together with their class number, are as follows:

2. firm owners receiving a wage (which includes the CEO or firm directors);
3. senior staff or top management positions (which includes chief financial officers, heads of human resources, and logistics and purchasing managers);
4. employees at the supervisor level (which includes quality control technicians, technical, accounting, and sales supervisors);
5. qualified and nonqualified clerical employees (secretaries, human resources or accounting employees, telephone operators, and sales employees);
6. blue-collar qualified and nonqualified workers (welders, assemblers, machine operators, and maintenance workers).

Throughout the paper we merge classes 5 and 6 since the distribution of wages of workers in these two classes is extremely similar, indicating similar levels of knowledge. Table 1 shows percentiles of the distribution of wages in the different classes of workers, all expressed in 2005 euros.[11]

The distributions are clearly ranked. CEOs make the most money, and wages decrease as we reach classes 5 and 6, which are practically identical. In order to match the numbers of these occupational classes with the theory, we order them from the bottom up. So classes 5 and 6 will form the layer of production workers, namely, layer 1. Class 4 of super-

[10] Class 1 is used only for farmers and so is never present in our data.
[11] Throughout the paper all nominal variables are expressed in 2005 euros.

TABLE 1
Distribution of Average Hourly Wage by Occupation in 2005 Euros

	CEO, Directors	Senior Staff	Supervisors	Clerks	Blue-Collar
Mean	81.39	47.83	26.58	19.01	20.70
p5	23.68	21.45	14.35	10.63	10.64
p10	28.60	25.01	16.21	11.79	11.82
p25	41.51	31.00	19.36	13.84	13.65
p50	58.06	38.28	23.11	16.49	15.97
p75	80.48	47.26	27.76	19.95	19.07
p90	114.51	59.91	34.15	24.66	23.40
p95	142.29	72.08	40.45	29.37	27.87

Note.—This table reports, for each one-digit occupational code present in the PCS-ESE 2003, mean and percentiles of the hourly wage distribution across all firms and years in the data. One observation in an occupation is the average hourly wage in a given firm-year from the BRN source, conditional on the firm reporting the occupation. The average hourly wage is the total labor cost from the BRN data set for an occupation, divided by the number of hours reported in this occupation. The total labor cost for an occupation is computed multiplying the total labor cost from the firm balance sheet times the share of wages paid to the occupation as resulting from the DADS source. Occupation 6 excludes one outlier, which would have driven its mean to 28.89.

visors will form the second layer, layer 2. Senior staff will be included in layer 3, and CEOs and firm directors will form layer 4. The number of layers in the firm is then the number of occupational classes reported in a year (i.e., a layer exists as long as there is at least 1 hour of work employed in it, and firms add or drop layers as the number of occupational classes changes). Firms can have a maximum of four layers, starting with layer 1 and moving all the way up to layer 4.

A. *Firms with a Different Number of Layers Are Different*

We aim to establish that this classification of employees into layers is a meaningful economic classification. Of course, this occupational classification could just constitute some arbitrary names given to particular workers in an organization that are not systematic across firms. The evidence in table 1 suggests otherwise. Wages across these occupations are evidently ranked. Clearly, much more is needed. We dedicate the rest of this section to convincing the reader that this classification is useful. In tables 2 and 3 we present some basic statistics of our data set.

Table 2 presents the number of firms by year as well as average value added, hours of work, wages, and layers. There is little variation by year in the data, as is evident from the table. We classify a firm as having a particular layer if it reports employing a positive number of hours in that layer. On average, firms in our sample employ a positive number of hours in about two and a half layers. It is important for our purposes that firms do not tend to employ workers in all layers, since we will analyze

TABLE 2
DATA DESCRIPTION BY YEAR

		AVERAGE			
YEAR	FIRMS	Value Added	Hours	Wage	Number of Layers
2002	78,494	2,929	78,775	22.46	2.60
2003	76,927	2,922	77,813	22.69	2.58
2004	75,555	2,957	76,574	23.47	2.59
2005	74,806	2,799	73,078	23.64	2.55
2006	73,834	2,847	72,770	23.50	2.53
2007	71,859	2,709	68,908	24.10	2.51

NOTE.—This table reports, for each year, the number of firms in the data set and corresponding averages across all firms for selected variables. Value added pertains to the firm's balance sheet. Hours is the average number of hours from the DADS source. Wage is the average hourly wage from the BRN in 2005 euros. Number of layers is the average number of layers across firms in each year. Value added is in thousands of 2005 euros.

TABLE 3
DATA DESCRIPTION BY NUMBER OF LAYERS IN THE FIRM

		AVERAGE			
NUMBER OF LAYERS	FIRM-YEARS	Value Added	Hours	Wage	MEDIAN WAGE
1	80,326	201	7,656	26.90	17.50
2	124,448	401	15,706	21.82	18.64
3	160,030	2,834	80,488	22.31	20.41
4	86,671	8,916	211,098	23.89	22.04

NOTE.—This table reports summary statistics on firm-level outcomes, grouping firm-year observations according to the number of layers reported. Firm-years is the number of firm-year observations in the data with the given number of layers. Value added is the average from the firm's balance sheet. Hours is the average number of total hours from the DADS source. Wage is the average hourly wage from the BRN in 2005 euros. Median wage is the median across all firms in the cell of the average hourly wage from the BRN source in 2005 euros. Value added is in thousands of 2005 euros.

how firms change as they add or drop layers of management. Given that, on average, firms have only slightly more than two and a half layers and that they can have a maximum of four layers, there are ample opportunities for firms to add new layers.

Table 3 presents the average characteristics of firms across layers. One-layer firms, self-employed workers, or firms in which everybody is engaged in the same production activities (i.e., have the same occupation) employ about four people on average.[12] Clearly, firms with more layers

[12] Note that since class 2 includes only owners that receive a wage, some of these firms might have an owner (acting as top manager) who is not being compensated directly with a wage. Even though one might regard one-layer firms as nonstandard, we keep them in the analysis since their behavior is very much in line with the behavior of firms with more layers.

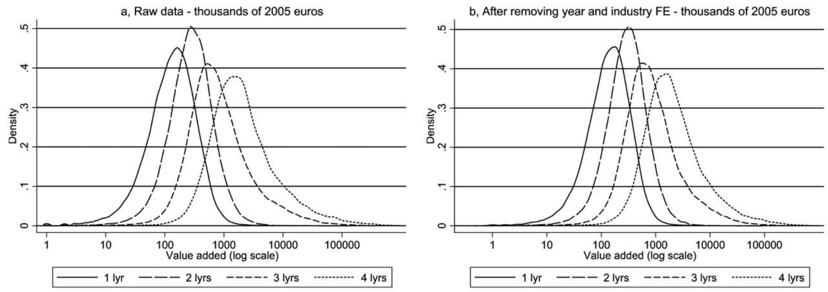

FIG. 2.—Value-added distribution by number of layers. These figures report kernel density estimates of the distribution of log value added by number of layers in the firm. The left panel reports the kernel density estimate of the distribution of log value added on the raw data: one density is estimated for each group of firms with the same number of layers. The right panel shows the same density estimated after removing year and industry fixed effects. To remove these effects, we run a regression of the form $\log v_i = \alpha + \sum_j \beta_j \text{layers}_{ij} + \sum_i \delta_i \text{industry}_{ij} + \sum \gamma_i \text{year}_{ij} + \varepsilon_i$, where v_i is value added for firm-year i, and $\text{layers}_{i\cdot}$, $\text{industry}_{i\cdot}$, and $\text{year}_{i\cdot}$ are a set of layers, two-digit industry, and year dummies, respectively. The omitted dummy for layers is for firms with zero layers of management. The log value added for firm-year i without year and industry fixed effects is then $\log \hat{v}_i = \log m + \sum_j \hat{\beta}_j \text{layers}_{ij} + \hat{\varepsilon}_i$, where we set m to the median value added in 2002 for firms with zero layers of management. We then compute four kernel density estimates of the distribution of $\log \hat{v}_i$, grouping firms according to their number of layers.

are larger in terms of value added and hours. They also tend to pay higher wages. The last fact is more evident when we look at the median than at the mean given that there are some outliers for firms with only layer 1. Figures 2, 3, and 4 document the same facts using the whole distribution.

The figures also show the distributions after we control for time and industry fixed effects (which as can be seen in the figure do very little). The distributions provide a picture very similar to the one we obtained from just looking at the means. Firms with more layers are larger and tend to pay higher average wages. As a robustness check, table A1 and

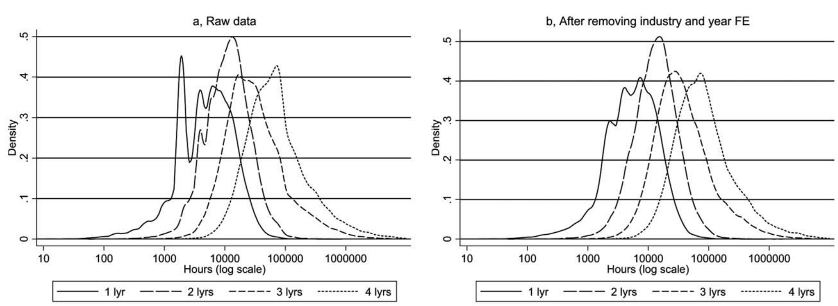

FIG. 3.—Hours distribution by number of layers. These figures report kernel density estimates of the distribution of log hours worked by number of layers in the firm. See the note to figure 2 for a description of how the densities are computed.

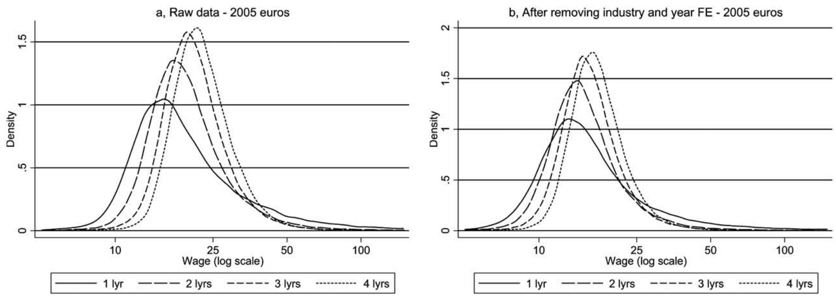

FIG. 4.—Firm average hourly wage distribution by number of layers. These figures report kernel density estimates of the distribution of log hourly wage by number of layers in the firm. See the note to figure 2 for a description of how the densities are computed.

figure A1 in online appendix A show the same findings using an alternative measure of wages (from DADS) that does not include payroll taxes and other expenses. A description of the data is presented in online appendix B. The evidence in this section has documented not only that layers do not simply group workers in arbitrary ways but that firms with different numbers of layers are different in economically meaningful ways.

B. *Firms Have Consecutively Ordered Layers and Form Hierarchies*

So far we have studied firms with different total numbers of layers but not which layers firms actually do include in their organization. Table 4 shows that the vast majority of firms have consecutively ordered layers starting from layer 1.[13] That is, 87.42 percent of the firms in our sample that have one layer actually have occupations 5 and 6. These are also the largest firms in terms of employment, as they account for 99.17 percent of the total employment of firms with one layer. As another example, consider the firms with three layers. About 80 percent of them include occupations from 3 to 5 and 6. The remaining 20 percent are missing one of these occupations and include occupation 2. Note that, weighted by employment or value added, these firms account for less than 7 percent. So the producers that do not have consecutively ordered layers starting from layer 1 are the smallest firms in the sample. Clearly, all firms with four layers have consecutively ordered layers starting from one since they have all layers. Overall, 81.69 percent of firms have consecutively ordered layers starting at one and they account for 96.73 percent of value added. The results that follow in general do not depend

[13] We define a firm as having "consecutively ordered layers" if it has only layer 1 (occupation 5 or 6), layers 1 and 2 (occupations 4–6), layers 1–3 (occupations 3–6), or all layers (occupations 2–6).

TABLE 4
PERCENTAGE OF FIRMS THAT HAVE CONSECUTIVELY ORDERED LAYERS

	Among firms with				
	1 Layer	2 Layers	3 Layers	4 Layers	All Firms
Unweighted	87.42	67.39	80.01	100	81.69
Weighted by value added	87.69	68.40	94.60	100	96.73
Weighted by hours	99.17	72.56	93.07	100	95.69

NOTE.—This table reports the fraction of firms with consecutively ordered layers conditioning on the number of layers in the firm (first four columns) and overall (fifth column). The first row reports the simple fraction of firms; the second and third rows assign a weight that is proportional to the total value added in the balance sheet and to the total hours in the DADS, respectively.

on whether we restrict the sample to firms with consecutively ordered layers (what we label "selected sample"). These robustness checks are included in online appendix A.

Not only do firms have consecutively ordered layers but they form hierarchies. That is, the number of hours employed in the lowest layer is, in most of them, larger than in the second layer, which is larger than the third layer, which is larger than the top layer (if the firm has all these layers). Table 5 presents the fraction of firms that satisfy this hierarchical criterion for hours in all layers and in each of them individually. Table A33 in online appendix A presents the averages weighted by value added.

A particular ranking of layers is hierarchical in the sense that the upper layer is smaller than the lower one in at least 74 percent of cases in the data. Almost all firms with two layers satisfy the ranking. However, only slightly more than half of the firms with all layers satisfy the ranking

TABLE 5
FIRMS THAT SATISFY A HIERARCHY IN HOURS

Number of Layers	$N_L^\ell \geq N_L^{\ell+1}$ All ℓ	$N_L^1 \geq N_L^2$	$N_L^2 \geq N_L^3$	$N_L^3 \geq N_L^4$
2	85.6	85.6	. . .	. . .
3	63.4	85.9	74.8	. . .
4	56.5	86.9	77.5	86.9

NOTE.—This table reports, among all firms with $L = 2, 3, 4$ layers, the fraction of firms 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 $\ell + 1$, with $\ell = 1, \ldots, L - 1$ (second to fourth columns). A firm satisfies a hierarchy in hours between layers number ℓ and $\ell + 1$ in a given year if the number of hours worked in layer ℓ is at least as large as the number of hours worked in layer $\ell + 1$; moreover, a firm satisfies a hierarchy at all layers if the number of hours worked in layer ℓ is at least as large as the number of hours in layer $\ell + 1$, for all layers in the firm. The term N_L^ℓ is the number of hours reported in layer ℓ in a firm with L layers from the DADS source.

in all layers. So most firms are hierarchical in terms of hours, but we have a relatively large number of exceptions in at least one layer. In contrast, when we look at the hierarchy in wages—namely, whether workers in higher layers earn more than workers in lower layers—the hierarchy is satisfied in the vast majority of cases. We present this evidence in table 6. All individual rankings are hierarchical in more than 87 percent of cases, and even firms with four layers are hierarchical in all the rankings in about 80 percent of cases. Table A34 in online appendix A presents the averages weighted by value added.

In sum, we conclude from this evidence that it is accurate to think of the representative firm as hierarchical, with more hours of work in lower layers but workers that are paid less.

Figure 5 represents graphically the firms in our sample. Each panel in the graph represents firms with different numbers of layers. Each layer is represented using a rectangle. The length of the rectangle represents the average number of hours employed in the layer by firms with a given number of layers. The height of the rectangle represents the average hourly wage of employees in that layer (so the area is the total wage bill of the layer). The hierarchical organization of labor is evident. Also evident is the way in which firms with more layers organize differently. In the next section we study the particular changes in wages and hours by layer as firms expand. In the graph we normalize the number of hours of each layer by the number of hours in the top layer. Our model keeps the number of hours at the top layer fixed, so this normalization is desirable when we contrast the implications of the theory with our data. All the characteristics of the representative hierarchies that we discussed in figure 5 are also present if we do not normalize the top layer.

The theory in Caliendo and Rossi-Hansberg (2012) as well as our empirical analysis underscores average wages at each layer. Clearly this is a simplification since workers within a layer are bound to be heterogeneous in their knowledge, for example, because of the individual histories of workers and the frictions faced by firms to hire and fire employees with particular levels of knowledge. Still, we have not said anything about

TABLE 6
Firms That Satisfy a Hierarchy in Wages

Number of Layers	$w_L^{\ell+1} \geq w_L^{\ell}$ All ℓ	$w_L^2 \geq w_L^1$	$w_L^3 \geq w_L^2$	$w_L^4 \geq w_L^3$
2	85.60	85.60	. . .	. . .
3	63.40	85.90	74.80	. . .
4	56.50	86.90	77.50	86.90

Note.—This table is the same as table 5 for the case of wages, where w_L^{ℓ} is the average hourly wage in layer ℓ from the BRN in an L-layer firm.

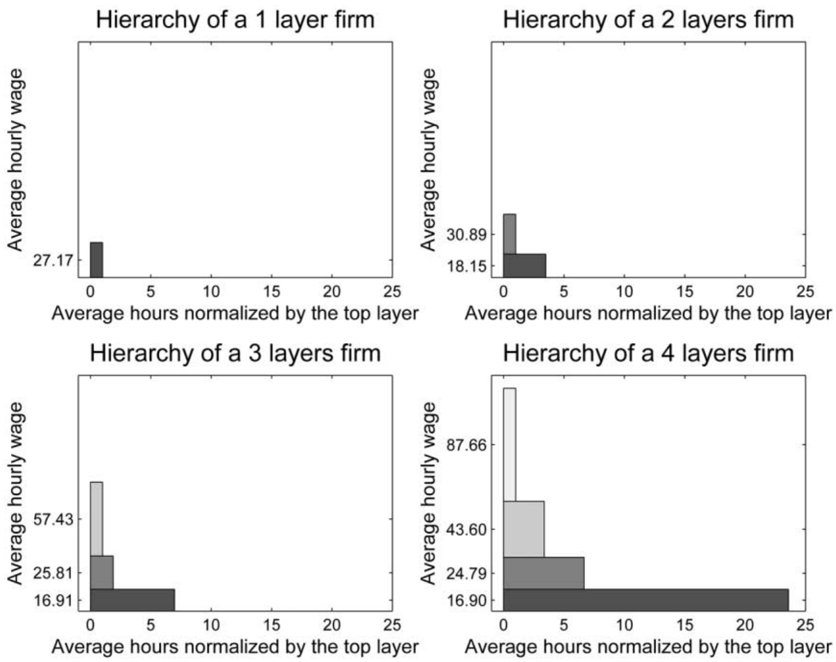

FIG. 5.—Representative hierarchies normalized by hours in the top layer. These figures portray the representative hierarchy of an L-layers firm. We first consider all firms with a given number of layers $L = 1, 2, 3, 4$ and focus on the middle tercile according to total value added. For each layer $\ell = 1, \ldots, L$, we then compute the average number of hours and the average hourly wage in the layer from the BRN in this group. In a figure portraying a firm with L layers, L rectangles are shown, one for each layer. The horizontal length of each rectangle is proportional to the number of normalized hours in the layer (hours are reported along the x-axis) while its height is proportional to the average hourly wage (the value of the wage is reported to the left of each rectangle).

how much of the variation in wages within the firm is explained by variation across layers rather than by variation within layers. This is relevant since if the fraction explained by cross-layer variation was negligible, our focus on layers would be clearly misguided, at least when it comes to analyzing the distribution of wages within firms. Table 7 shows that this is not the case.

The mean share of variation in log wages explained by variation in layers for all firms is about half, independently of how we weight firms. The share is zero for firms with one layer (since, by definition, for these firms there is no cross-layer variation in log wages) and grows to 66 percent for firms with four layers. The table reassures us that variation in wages across layers is essential to understanding the distribution of wages within firms.

TABLE 7
MEAN SHARE OF VARIATION IN WAGES EXPLAINED BY CROSS-LAYER VARIATION

			Weighted by	
	Firm-Years	Unweighted	Hours	Value Added
All firms	434,872	.50	.51	.49
Firms with more than 1 layer	370,997	.59	.51	.50
Firms with 1 layer	63,875	.00	.00	.00
Firms with 2 layers	124,299	.50	.41	.43
Firms with 3 layers	160,028	.62	.51	.50
Firms with 4 layers	86,670	.66	.53	.50

NOTE.—For this table we compute the R^2 of a regression of log hourly wages of workers within a firm on a constant and dummies for layers (all except one), weighted by the number of hours each worker provides to the firm. For each row, the column unweighted reports the average R^2 across all firm-years, while the remaining two columns to the right report the same average when weighting firms by their total number of hours or total value added. The column firm-years reports the number of firm-years used to compute the statistics in the corresponding row. Note that for some firms—e.g., firms with only one worker—the R^2 cannot be computed, and hence the total number of firm-years in the data set does not correspond to the total number of firm-years used. Each row differs from the others according to the subsample of firm-years used in computing the average.

C. *Layer Transitions Depend on Size and Firms Add or Drop Consecutive Layers*

Let us now investigate how many producers add or drop layers in a given period. Table 8 shows that between 60 and 70 percent of firms in a given period maintain their number of layers. From the remaining, some firms exit, with the exit rate decreasing with the number of layers. Clearly, of the firms that change layers, the majority add or drop only one of them. In fact, out of the firms with consecutively ordered layers, most of the firms that add one add a consecutive layer (75.5 percent for one-layer

TABLE 8
DISTRIBUTION OF LAYERS AT $t + 1$ CONDITIONAL ON LAYERS AT t

Number of Layers at t	Number of Layers at $t + 1$					
	Exit	1	2	3	4	Total
1	15.3	67.5	15.2	1.9	.2	100
2	9.8	10.7	62.2	16.2	1.1	100
3	7.7	1.2	13.1	67.6	10.5	100
4	6.2	.2	2.0	20.5	71.3	100

NOTE.—This table reports the distribution of the number of layers at time $t + 1$, grouping firms according to the number of layers at time t. Among all firms with L layers ($L = 1, \dots, 4$) in any year from 2002 to 2006, the columns report the fraction of firms that have layers $1, \dots, 4$ the following year (from 2003 to 2007) or are not present in the data set, exit. The elements in the table sum to 100 percent by row.

firms and 82.3 percent for firms with two layers; see table A2 in online app. A). Hence, when firms add or drop layers, they tend to drop or add a consecutive layer, and only one of them. This is all consistent with the view, provided by the theory, that firms add layers to expand and drop layers to contract, and do so in a systematic way. Since very large expansions are rare, we see few transitions that add or drop more than one layer. Table A3 in online appendix A shows that the same pattern as in table 8 is observed even if we weight the firms by their value added.

We can also study the probability of adding or dropping one or more layers as a function of the size of the firm in terms of its value added. Figure 6 shows a lowess smoothing interpolation of the probability of changing the number of layers to any count, as a function of the value added of the firm, for firms with different initial numbers of layers. If, for example, firms receive shocks to their demand parameter α over time and these shocks are drawn from a common arbitrary distribution, the model predicts that the probability of adding a layer should

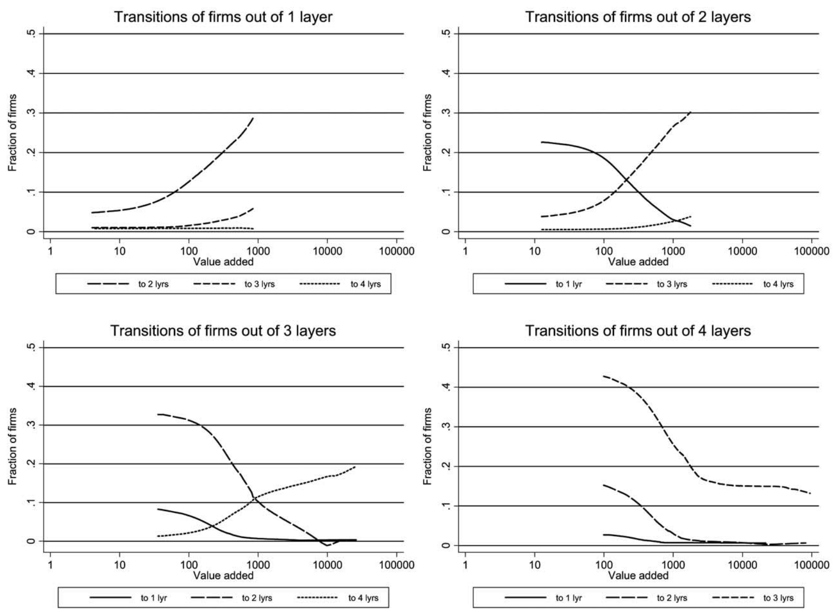

FIG. 6.—Transitions across layers depend on value added. These figures show the probability of transition away from the current layer as a function of the initial value added of the firm. Each panel reports transition probabilities starting from a different initial number of layers. To produce the panel of transitions out of layer $L = 1, \ldots, 4$, we take for each year (from 2002 to 2006) all the firms with L layers and group them into 100 bins according to their value added; for each bin, we compute the fraction of firms that will have any number of layers (or exit the data set) in the following period and plot the average value added in the bin against this fraction. For each transition series we then apply a lowess smoothing for all the probabilities estimated from the first to the ninety-ninth bin.

increase with value added. In contrast, the probability of dropping a layer should decrease in value added. Furthermore, the probability of adding one layer should be larger than the probability of adding two, which should also be increasing in value added. This is exactly what happens in figure 6.

The probability of adding layers is always increasing in value added and of dropping is always decreasing, and the ranking of probabilities is always consistent with the predictions of our theory, augmented with some simple stochastic process for the fundamentals. Figure 6 does not include confidence bands in order to enhance the visibility of the curves. However, in figure A2 in online appendix A, we present a graph with all the individual observations and show that they line up fairly tightly around the interpolation estimates. Online appendix B describes all the details to construct figure 6.

D. *Trends before Adding or Dropping Layers*

The model we sketched in Section II suggests that the firm's decision of how many layers to have is characterized by a series of size thresholds. The model is static, but it lends itself to a trivial frictionless dynamic extension in which, conditional on a distribution of shocks to a firm's revenues (e.g., a demand or a productivity shock), a firm will add a new layer if the cumulative set of shocks since its last change is large enough.[14] This logic implies that, on average and conditional on a number of layers, firms that add layers at t should grow faster in the previous couple of years than other firms with the same number of layers. Similarly, firms that drop layers at t should tend to grow slower in $t-1$ and $t-2$ than other firms with the same number of layers. Figure 7 presents evidence on this hypothesis.

We take all firms i in our data set that have the sequence of layers (L, L, L, L') over time for any L and $L' = \{1, 2, 3, 4\}$. We then estimate, for $k = 0, 1, 2,$

$$d\ln \widetilde{VA}_{it-k} = \sum_{L'=1}^{4} \gamma^{1}_{LL't-k} D_{LL'} + \gamma^{2}_{Lt-k} \ln VA_{it-k} + \epsilon_{it-k},$$

where $d\ln \widetilde{VA}_{it-k}$ is the detrended log change in value added from period $t-k$ for firm i, and $D_{LL'}$ is a dummy that takes the value of one if a firm with L layers at the end of the sequence has L' layers.[15] Figure 7 presents

[14] In moving from a static to a dynamic interpretation, we are assuming that these thresholds are firm specific and that firms can quickly react to revenue shocks with organizational changes when the relevant thresholds are crossed.

[15] Here, and in all the subsequent empirical analysis, we study only log changes, although the theory refers to changes in levels. All the theoretical implications can alternatively be expressed in log changes. For our purposes, this distinction is irrelevant since we are almost

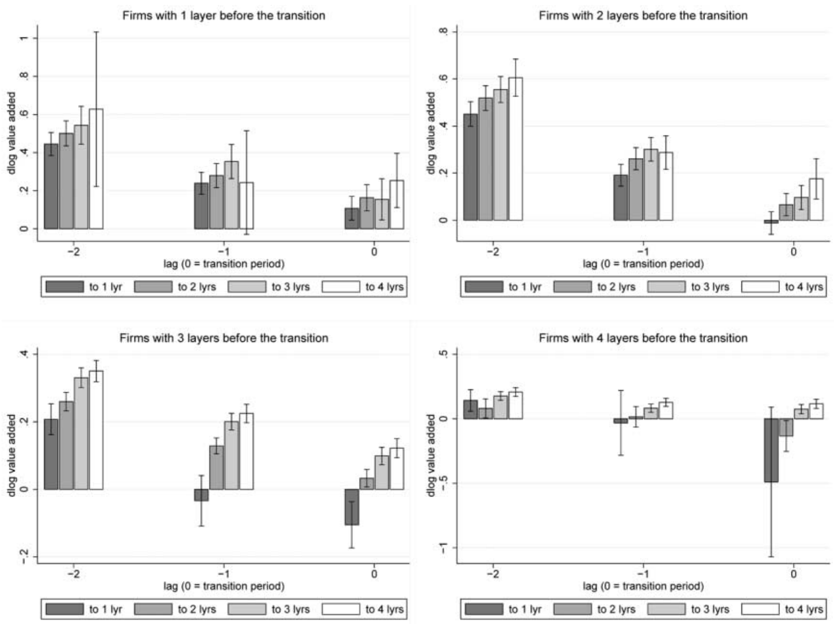

FIG. 7.—Changes in value added before and during transition. These figures show the fixed effect of a given transition (from L to L' number of layers, with L, $L' = 1, 2, 3, 4$) on the average firm growth in a period, conditional on initial size. We use the specification presented in the text. The figure presents the value of $\gamma^1_{LL't-k}$ and its 95 percent confidence interval, grouping in each panel firms with a given number of layers before the transition (i.e., the top-left panel shows $\gamma^1_{1L't-k}$, the top-right $\gamma^1_{2L't-k}$, and so on). Within each panel, the dummies are grouped according to the time lag before the transition, $k = 2, 1, 0$, and, within each group, ordered by the number of layers L' after the transition.

the value of $\gamma^1_{LL't-k}$, namely, the effect of the LL' transition on the mean percentage growth rate in detrended value added, conditional on firm size, two periods before, one period before, and at the moment of the transition to other layers. We also present the 95 percent confidence interval.

It is clear from figure 7 that firms that add layers in period t grow faster in $t-2$, $t-1$, and t than firms that keep or reduce the number of layers. Furthermore, the more layers they add or drop in t, the faster they grow or shrink (although transitions of several layers are estimated less precisely because of small sample sizes). Note that these growth rates are conditional on the size of the firm, and so they are consistent with the view that firms that grow/shrink faster since the last transition are closer to the firm-specific threshold that makes them add/drop a layer.

exclusively interested in sign restrictions. Log changes are scale free, which makes their use convenient in the empirical analysis.

The evidence in this section has documented that layers do not just group workers in arbitrary ways but that firms with different numbers of layers are different in economically meaningful ways. In addition, we have documented that changes in the number of layers are also systematic and are determined by the size and growth of the firm. We now turn to analyze how firms change their organization, layer by layer, when they decide to expand or contract.

IV. How Do Firms Expand?

In this section we analyze how firms change when they expand or contract. The main body of evidence we present tracks firms over time and, therefore, controls for a variety of individual firm characteristics, such as industry. This, we believe, is the ideal way of analyzing the predictions of our theory in the data, and we do so below.

We first study the relationship between firm real value added (VA) and normalized hours n_L^ℓ in a given layer, ℓ, for firms with L layers. That is, we study how n_L^ℓ changes as the firm increases its value added. We first detrend all variables using aggregate trends. Namely, if i refers to a particular firm, $\tilde{n}_{Lit}^\ell = n_{Lit}^\ell/\bar{n}_t$ and $\widetilde{VA}_{it} = VA_{it}/\overline{VA}_t$, where $\bar{n}_t$ and $\overline{VA}_t$ are the average normalized hours and value added. We then estimate a regression of the form

$$d\ln \tilde{n}_{Lit}^\ell = \beta_L^\ell d\ln \widetilde{VA}_{it} + \varepsilon_{it}, \tag{6}$$

where d denotes a yearly time difference (e.g., $d\ln n_{Lit}^\ell = \ln n_{Lit+1}^\ell - \ln n_{Lit}^\ell$) among all firms that stay at L layers for 2 consecutive years.[16]

Table 9 presents the estimates of β_L^ℓ for all $\ell \in 1, 2, \ldots, L$ and every L. Note that since we are normalizing hours by the number of hours in the top layer, we can look at the value of β_L^ℓ for $\ell = 1, \ldots, L-1$ only. First note that as predicted by implication 3 of the theory in Section II, given L, firms grow by increasing the number of hours at all layers. Furthermore, the ranking of the values of β_L^ℓ always satisfies that $\beta_L^\ell > \beta_L^{\ell'}$ for $\ell < \ell'$, although in one instance the difference is not significant.[17] Hence, these results show that once we control for firm fixed effects, the predictions of the theory are in line with our findings. As a robustness check we present in table A35 the results with the selected sample. We conclude that when firms grow but keep the same number of layers, they employ

[16] In the main text we use a specification that controls only for aggregate trends in normalized hours. However, all our results are robust to adding layer-specific time trends.

[17] The theory in fact has a more subtle prediction. Namely, the slope of the relationship between the log of value added and normalized hours should decrease as we consider higher layers. The reason is that a larger firm with the same L has more knowledgeable workers in all layers, as discussed in Sec. II, and so larger spans of control at all layers.

TABLE 9
Elasticity of Hours with Value Added for Firms That Do Not Change L

Number of Layers	Layer	β_L^ℓ	Standard Error	p-Value	Observations
2	1	.042	.012	.000	64,536
3	1	.039	.009	.000	91,253
3	2	.013	.010	.200	91,253
4	1	.107	.014	.000	52,799
4	2	.051	.013	.000	52,799
4	3	.037	.013	.000	52,799

Note.—This table reports the results of regressions of detrended log change in normalized hours at layer ℓ in a firm with L layers on its detrended log change in value added, and no constant, selecting all the firms that stay at L layers across 2 consecutive years. The term β_L^ℓ is the coefficient on log change in value added.

more hours of work at all layers but proportionally more in the lower layers. So firms become flatter, with a wider base.

We do the same analysis for wages. Namely, we study the relationship between value added and wages in a given layer for all firms. As we did for hours above, we detrend wages and value added by removing the yearly mean across all layers and firms. So we run

$$d\ln \tilde{w}_{Lit}^\ell = \gamma_L^\ell d\ln \widetilde{VA}_{it} + \varepsilon_{it}, \tag{7}$$

where $d\ln\tilde{w}_{Lit}^\ell$ is the log difference in detrended wages, $\tilde{w}_{Lit}^\ell = w_{Lit}^\ell/\bar{w}_t$, and $\bar{w}_t$ is the mean hourly wage across all firms in year t among all firms that stay at L layers for 2 consecutive years.[18]

The results are presented in table 10 and are all consistent with the theory. Namely, γ_L^ℓ is positive and significant for all $\ell \in 1, 2, \ldots, L$ and every L. Furthermore, $\gamma_L^\ell < \gamma_L^{\ell'}$ for $\ell < \ell'$ in all cases. Hence, when firms grow without changing the number of layers, they increase wages (or knowledge according to the theory) in all layers, but they increase wages proportionally more at the top of the firm as the model predicts. Table A32 in online appendix A presents several robustness checks.

The above analysis paints a familiar picture of the way firms expand. Firms expand by adding more workers of all types, by hiring more knowledgeable workers, and by paying them more. Most models of firm dynamics (Lentz and Mortensen [2008], among many others) share these features with the theory outlined in Section II when firms keep the number of layers constant. The next subsection shows that when a firm's expansion leads to a change in the number of layers—a reorganization of the firm—many of these findings are altered in a significant way, specifically, the one predicted by implication 4 in Section II above.

[18] Again, all our results are robust to adding layer-specific time trends.

TABLE 10
ELASTICITY OF WAGES WITH VALUE ADDED FOR FIRMS THAT DO NOT CHANGE L

Number of Layers	Layer	γ_L^ℓ	Standard Error	p-Value	Observations
1	1	.077	.007	.000	45,045
2	1	.100	.006	.000	64,536
2	2	.118	.006	.000	64,536
3	1	.145	.006	.000	91,253
3	2	.155	.006	.000	91,253
3	3	.170	.006	.000	91,253
4	1	.171	.009	.000	52,799
4	2	.185	.009	.000	52,799
4	3	.186	.010	.000	52,799
4	4	.217	.011	.000	52,799

NOTE.—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 L across two consecutive periods, where both variables are detrended as specified in the main text. Specifically, we run a regression of detrended log change in average hourly wage at layer ℓ in a firm with L layers on the detrended log change in value added across all the firms that stay at L layers across two consecutive years, with no constant. The term γ_L^ℓ is the coefficient on log change in value added. We use hourly wage at layer ℓ from the BRN.

Expansions that add layers.—We first look at how firm-level outcomes change depending on whether firms add or drop layers of management. Table 11 shows the average log changes in total hours, total normalized hours, value added, and average wages (including and excluding the new top manager in the case of adding layers) for all firms, the ones that add layers, the ones that do not change layers, and the ones that drop layers. As one can see in the first column of the table, most of these variables exhibit some trend over time, and so the average log change is significantly different from zero. To account for this, we also present average changes after we control for time trends (see online app. B for details). Clearly, adding layers is related to increasing hours, normalized hours, and value added. In contrast, firms that add layers decrease average wages once we take out the common time trend. Furthermore, if we look at wages in the preexisting layers only, wages fall significantly, by 12.2 percent. The results are reversed when we select only firms that drop layers. Now wages rise by 12.2 percent.

These estimates demonstrate that in firms that expand by adding layers, average wages in preexisting layers fall. This is inconsistent with many conceptualizations of firm dynamics in which firms that expand always increase the wage of all their employees. Note also that since overall wages increase (without detrending), as do wages of firms that do not exhibit changes in layers, the fall in wages cannot be the result of reverse causality in which drops in wages cause expansions. If that were the case, we would see drops in wages associated with expansions in all firms, not only the ones that add layers. Furthermore, we would not ob-

TABLE 11
Change in Firm-Level Outcomes

	All	Increase L	No Change in L	Decrease L
d ln total hours	−.015***	.040***	−.012***	−.081***
Detrended	. . .	.055***	.003***	−.066***
d ln normalized hours	−.011***	1.362***	.012***	−1.404***
Detrended	. . .	1.373***	.023***	−1.392***
d ln VA	−.008***	.032***	−.007***	−.050***
Detrended	. . .	.040***	.001	−.041***
d ln average wage	.019***	.015***	.019***	.025***
Detrended	. . .	−.005***	.000	.006***
Common layers	.021***	−.101***	.019***	.143***
Detrended	. . .	−.122***	−.002***	.122***
% of firms	100	12.65	73.66	13.68
% value added change	100	40.12	65.08	−5.19

Note.—This table reports changes in firm-level outcomes between consecutive years for all firms and for the subsets of those that increase, do not change, and decrease layers. It reports changes in log hours, log normalized hours, log average wage from the BRN, 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 from the BRN computed using only the first $\min\{L, L'\}$ 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. In the last two rows of the table, % of firms is the percentage of firms observed having each type of behavior; % value added change is the fraction of the total change in real value added observed in the data set accounted for by firms making the given transition.

*** Significant at 1 percent.

tain the opposite result when we select only firms that drop layers. Of course, the theory in Section II is exactly consistent with this evidence on wages. Implication 4 says that, as firms add layers, the knowledge and therefore wages at all layers should decrease.

The results in table 11 suggest that wages in firms that add or drop layers behave differently than previously thought. Table 11 also shows that the firms that add or drop layers represent an important fraction of firms in the economy, as well as an important fraction of value added. Firms that add layers represent 12.65 percent of the total. Furthermore, as they are, on average, larger than their counterparts that do not add layers, they contribute 40.12 percent of the total change in value added. Conversely, firms that drop layers represent 13.68 percent of firms and contribute −5.19 percent to the change in value added. Together, the firms that reorganize by changing layers and that therefore change wages in the new way we uncover represent more than a quarter of the firms in the economy and contribute more than 40 percent of the absolute changes in value added. So the firms that change their organization to expand and contract do not represent a fringe of the firms in the economy. They are essential to understanding firm dynamics and the associated labor market outcomes. The theory of organization with heterogeneous firms in Caliendo and Rossi-Hansberg (2012) can rationalize the behavior of these firms.

We now proceed to analyze in more detail firms that change their layers of management. In particular, we are interested in whether firms that add layers add hours of work to all layers and decrease wages in all layers. The results above tell us that this is the case on average, but they do not imply that this happens layer by layer. Table 12 computes average log changes in detrended normalized hours for firms that transition between layers. Each line in the table represents a particular type of transition (e.g., from two to three layers) and a particular layer in firms that undergo that transition. The first column in the table indicates the number of management layers in the initial period and the second column the number of layers in the second period. The third column indicates the layer, ℓ, for which we are calculating the average (over i and t) of $d\ln\tilde{n}^{\ell}_{Lit}$. The fourth column indicates the coefficient of interest. Note first the sign of the average change. It is positive and significant for all firms that increase the number of layers (by one or more layers).

TABLE 12
AVERAGE LOG CHANGE IN HOURS FOR FIRMS THAT TRANSITION

Number of Layers Before	Number of Layers After	Layer	$d\ln\tilde{n}^{\ell}_{Lit}$	Standard Error	p-Value	Observations
1	2	1	1.537	.018	.000	10,177
1	3	1	1.762	.056	.000	1,263
1	4	1	2.266	.212	.000	97
2	1	1	−1.582	.017	.000	11,106
2	3	1	.716	.012	.000	16,800
2	3	2	.539	.012	.000	16,800
2	4	1	1.205	.049	.000	1,129
2	4	2	1.004	.048	.000	1,129
3	1	1	−1.795	.048	.000	1,584
3	2	1	−.682	.012	.000	17,666
3	2	2	−.518	.012	.000	17,666
3	4	1	1.352	.014	.000	14,113
3	4	2	1.289	.016	.000	14,113
3	4	3	1.174	.016	.000	14,113
4	1	1	−2.119	.173	.000	123
4	2	1	−1.059	.041	.000	1,456
4	2	2	−.918	.040	.000	1,456
4	3	1	−1.411	.014	.000	15,160
4	3	2	−1.345	.015	.000	15,160
4	3	3	−1.260	.015	.000	15,160

NOTE.—This table reports estimates of the average detrended log change in normalized hours at each layer among firms that transition from L to L' layers, with $L \neq L'$: for a transition from L to L', we can evaluate only changes for layer number $\ell = 1, \ldots, \min\{L, L'\}$. The detrending is explained in the main text. The term $d\ln\tilde{n}^{\ell}_{Lit}$ is the average detrended log change in the transition, estimated as a regression of the detrended log change in the number of normalized hours in layer ℓ in 2 consecutive years on a constant. The table uses all observed transitions in the sample.

Symmetrically, it is negative in all layers for all firms that drop one or more layers, exactly what we would expect from the theory in Section II.

Table 12 indicates that the firm-level outcomes on normalized hours from table 11 not only hold for the firm as a whole but hold layer by layer too. All our estimates are significant at the 1 percent level.[19] As we did above, we detrend all variables using aggregate trends. Online appendix A presents a variety of robustness checks. In particular, it presents the results when we use only firms with consecutively ordered layers and we condition on firm-level outcomes, such as expansions in hours or value added.

We do the same analysis layer by layer for changes in wages. Namely, we compute the average (over i and t) of $d\ln\tilde{w}^{\ell}_{Lit}$ for firms that add or drop layers. We present the results in table 13. Again, the table confirms that the results we obtained for firm-level outcomes hold layer by layer, as the theory predicts. Firms that add layers reduce wages in all preexisting layers and firms that drop layers increase wages in all the layers of the reorganized firm. Again, these results are robust to conditioning on large firm expansions in value added, normalized hours, or both, as well as to restricting the sample of firms with consecutively ordered layers. Furthermore, we corroborate our results using a different source of wage data. Tables with the robustness checks are presented in online appendix A. The conclusion is that many firms expand by adding layers of management, and these firms reduce the average salary of workers in all preexisting layers (or, accordingly, their knowledge as the theory would suggest).[20] The theory also predicts that, as a firm adds layers, the wages

[19] The theory also predicts that the proportional change in the hours of employees in higher layers should be larger than the proportional change in the hours of employees in lower layers. The reason is that the knowledge of all employees falls and so does the span of each manager. This results in positive but smaller proportional changes in hours at the bottom of the firm, when the firm adds a layer. In table 12 we see, in most cases, exactly the opposite. As in the case of firms that keep the number of layers constant, the lower layers expand proportionally more. Several forces can be responsible for this mismatch between the theory and the data. First, during a year, firms that switch might also have grown without further changes in the number of layers. Since according to the theory the rank of the log changes is different depending on whether layers are added or not, what we see could be the result of one effect dominating the other. Second, there could be frictions in hiring that make lower layers easier to expand than higher ones. This would be the case if hiring more knowledgeable employees is more costly and takes more time. Finally, the theory suggests that changes in communication costs, h, as the firm adds layers could also reverse the implications of the theory on this ranking.

[20] We interpret our results as the effect of organizational changes resulting from expansions in value added as a consequence of technology or demand shocks. However, changes in a firm's organization could also be the result of changes in the cost of acquiring knowledge, c, or in communication technology, h. These alternative sources of shocks could make our interpretations problematic. However, since most changes in information and communication technology are economywide changes, the fact that our results are about changes and hold after detrending at the firm or layer level alleviates, to a large extent, this concern.

TABLE 13
AVERAGE LOG CHANGE IN WAGES FOR FIRMS THAT TRANSITION

Number of Layers Before	Number of Layers After	Layer	$d\ln\tilde{w}^{\ell}_{Lit}$	Standard Error	p-Value	Observations
1	2	1	−.129	.005	.000	10,177
1	3	1	−.332	.020	.000	1,263
1	4	1	−.678	.117	.000	97
2	1	1	.167	.005	.000	11,106
2	3	1	−.050	.002	.000	16,800
2	3	2	−.255	.004	.000	16,800
2	4	1	−.150	.015	.000	1,129
2	4	2	−.409	.019	.000	1,129
3	1	1	.356	.018	.000	1,584
3	2	1	.059	.002	.000	17,666
3	2	2	.249	.004	.000	17,666
3	4	1	−.021	.002	.000	14,113
3	4	2	−.067	.003	.000	14,113
3	4	3	−.199	.004	.000	14,113
4	1	1	.804	.109	.000	123
4	2	1	.139	.012	.000	1,456
4	2	2	.372	.016	.000	1,456
4	3	1	.009	.002	.000	15,160
4	3	2	.040	.003	.000	15,160
4	3	3	.134	.004	.000	15,160

NOTE.—This table reports estimates of the average detrended log change in hourly wage at each layer ℓ among firms that transition from L to L' layers, with $L \neq L'$: for a transition from L to L', we can evaluate only changes for layer number $\ell = 1, \ldots, \min\{L, L'\}$. The term $d\ln\tilde{w}^{\ell}_{Lit}$ is the average detrended log change in the transition, estimated as a regression of the detrended log change in the average hourly wage at layer ℓ in 2 consecutive years on a constant. The table uses all observed transitions in the sample.

of higher-level managers should fall proportionally more than those of lower-level ones (since their knowledge is more substitutable with that of the top manager), a prediction also corroborated by table 13.

The finding is surprising in light of theories that ignore organizational structure, where it is common to assume that firms expand by "cloning" their operations. In those theories, we expect firms to grow and wages of both new hires and existing workers to rise in parallel fashion. In contrast, in the theory of Caliendo and Rossi-Hansberg (2012) and as our results show, firms do not grow by cloning themselves. Often, when firms grow, they add layers and disproportionately expand the number of employees in the lowest layers. This means that overall average wages in the firm may fall even though productivity is rising and workers who have significant tenure in the firm, and have high positions in the hierarchy, are enjoying pay raises.

We can decompose the total log change in average wages in the firm into two parts. The first part is the change in wages of workers in existing layers, which, as we know, is negative by the results discussed above. The second part is the change induced by adding a new agent at the top of the hierarchy. Agents in the new added layer earn more than the average worker in the firm since they are added, in the vast majority of cases, to the top layer. For example, for firms that transition from one to two layers, the new manager in layer 2 makes 50.7 percent more than the average wage in the firm before reorganization. This number can be much higher for larger firms. Firms that go from three to four layers pay the new top manager 338.5 percent more than the average worker in the firm before adding the layer. All these results are presented in table 14. We decompose the detrended average wage in the firm, $d\ln\bar{w}_{Lit}$, as

$$\begin{aligned} d\ln\bar{w}_{Lit} &= \ln\bar{w}_{L'it+1} - \ln\bar{w}_{Lit} \\ &= \ln[(\bar{w}^{\ell\leq L}_{L'it+1}/\bar{w}_{Lit})s + (w^{L'}_{L'it+1}/\bar{w}_{Lit})(1-s)], \end{aligned}$$

TABLE 14
DECOMPOSITION OF TOTAL LOG CHANGE IN AVERAGE WAGES

	$\bar{w}^{\ell\leq L}_{L'it+1}/\bar{w}_{Lit}$				$\bar{w}^{L'}_{L'it+1}/\bar{w}_{Lit}$		
From/To	2	3	4	From/To	2	3	4
1	.963***	.865***	.733***	1	1.507***	1.501***	1.602***
	(10,167)	(1,262)	(96)		(10,166)	(1,263)	(97)
2		.926***	.876***	2		2.040***	2.021***
		(16,783)	(1,128)			(16,783)	(1,129)
3			.958***	3			4.385***
			(14,099)				(14,099)

	s				$d\ln\bar{w}_{Lit}$		
From/To	2	3	4	From/To	2	3	4
1	.741***	.620***	.563***	1	−.007*	−.094***	−.305**
	(10,166)	(1,262)	(97)		(10,166)	(1,263)	(97)
2		.853***	.775***	2		.005**	−.033**
		(16,784)	(1,128)			(16,784)	(1,129)
3			.948***	3			−.001
			(14,099)				(14,098)

NOTE.—This table reports the sources of change in the average hourly wage from the BRN, by type of transition. For a given firm transitioning from L to $L' > L$ layers, write the detrended log change in average wage as

$$d\ln\bar{w}_{Lit} = \ln\bar{w}_{L'it+1} - \ln\bar{w}_{Lit} = \ln[(\bar{w}^{\ell\leq L}_{L'it+1}/\bar{w}_{Lit})s + \bar{w}^{L'}_{L'it+1}/\bar{w}_{Lit}(1-s)].$$

In this notation, $\bar{w}^{\ell\leq L}_{L'it+1}$ is the average hourly wage after the transition in the common layers, $\bar{w}_{Lit}$ is the average hourly wage before the transition, s is the share of hours of the common layers after the transition, and $\bar{w}^{L'}_{L'it+1}$ is the average wage in the layers added after the transition. We report in the cells the average of each of these quantities in the first three panels; the fourth panel shows the overall average log change in hourly wage during the indicated transition. Each cell is computed excluding observations below the 0.05th and above the 99.95th percentile. Numbers of observations are in parentheses.

* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

where $\bar{w}^{\ell \leq L}_{L'it+1}$ is the average wage in all preexisting layers in the reorganized firms with $L' > L$ layers, $w^{L'}_{L'it+1}$ is the wage of the new top manager, and s is the fraction of hours of work done by employees in preexisting layers. Table 14 presents each of these components. The fact that $\bar{w}^{\ell \leq L}_{L'it+1}/\bar{w}_{Lit}$ is below one for all transitions is, for practical purposes, just a reexpression of the results in table 13. The upper-right panel shows the earnings of the top managers as a fraction of the average wage in the firm before transition. Clearly, since workers in preexisting layers earn less but the new top manager makes more, the overall effect of adding a layer is ambiguous and not particularly robust. The relevant finding is that new managers are the only ones in the reorganized firm who earn more after adding layers.

The results in this section suggest that in order to understand the behavior of firms that expand, it is essential to condition on whether the expansion requires a reorganization. On average, the firms that expand the most tend to reorganize. So to understand expansions, we need to understand reorganization. The salient fact in the data is that when firms expand and reorganize, they pay workers on preexisting layers less on average. This is consistent with the view, borrowed from the theory in Caliendo and Rossi-Hansberg (2012), that the firm wants less knowledgeable workers on preexisting layers after the reorganization. Note that this finding does not challenge the many empirical studies (Brown and Medoff 1989; Bernard and Jensen 1997, 1999; Abowd, Kramarz, and Margolis 1999; Oi and Idson 1999; Frías, Kaplan, and Verhoogen 2009) that have found that average firm wages (or the wages of a particular class of workers) increase with firm size or as firms expand. In fact, we find some evidence that this is true in our sample too. What our finding says is that this is not the case when we condition on the firm reorganizing by adding layers. Furthermore, we find that a substantial fraction of expansions are in fact paired with this type of reorganization.

One might be concerned that the results presented above are just driven by a reclassification of agents across occupations. For example, as the firm expands, the oldest and best-paid worker might be now called a floor supervisor, and since she has the highest wage, this relabeling results in a lowering of the average wage of the agents who remain as workers in layer 1. This relabeling of jobs would show up in our results as a firm that added a layer and reduced the average wage in preexisting layers, even though the actual distribution of wages within the firm has not changed. To address this potential concern, we look at the change in the distribution of wages within the firm after a reorganization.

Figure 8 presents the difference in the log wages at each percentile of the distribution for the different one-layer transitions.[21] The figure also shows a bootstrapped 95 percent confidence interval constructed as we

[21] Transitions of more than one layer look similar, and we omit them for brevity.

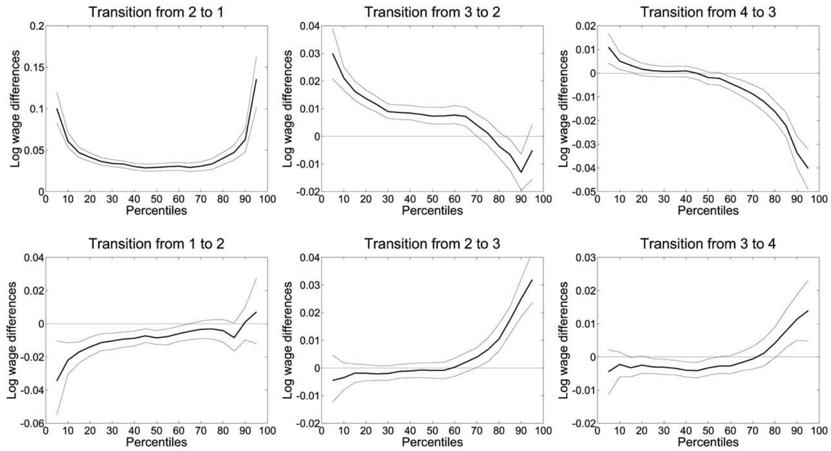

Fig. 8.—Difference in the distribution of wages after minus before the transition. These figures portray, for each transition type, difference in the distribution of wages after minus before the transition. The *y*-axis in each panel reports the log difference between the wage at the *p*th percentile after, less its correspondent wage before the transition. To compute standard errors, we performed 500 bootstrap replications of this process, clustering the sampling at the firm-transition level (i.e., one cluster contains all the employees present either before or after a transition in one firm) to preserve the within-firm and within-transition correlation in wages present in the data, and we report the 5th and 95th percentiles of these replications. To compute these panels, we first construct an employee-level data set that contains log hourly wage of each employee, a firm identifier, and year and current number of layers of the firm, and we remove year and firm fixed effects from the log hourly wage distribution. We then focus on all the hours worked in firm-year observations in which the firm is making a transition from ℓ to ℓ' layers (both before and after the transition takes place) and compute the *p*th percentile (for $p = 5, 10, \ldots, 95$) in the two distributions: the wage distribution after and before the transition. Each distribution of hourly wages is computed making sure that each employee receives a weight proportional to her number of hours in the firm but giving to each firm the same weight, regardless of the total number of hours worked in it (to be consistent with our estimates of changes in log wages in table 13).

explain in online appendix B. First note that the distribution of wages in firms that transition actually changes significantly. This eliminates the concern that our results are just the result of meaningless relabeling of employees. Furthermore, note how wages in the lower part of the distribution fall for all transitions in which firms add layers and they increase when firms drop them, exactly as we have been arguing.

Figure 8 also shows that, in some transitions, at the upper tail of the distribution, wages rise when firms add layers and fall when they drop them. This was expected given the results in table 14, where we show that agents in the newly added layer make more money than the average employee in the firm before the transition. Still, to confirm that this is in fact driven by the agents hired at the new layer, in figure 9 we present a parallel figure without including the hours of work in the new layer (or the dropped layer). The results are very stark and consistent with our

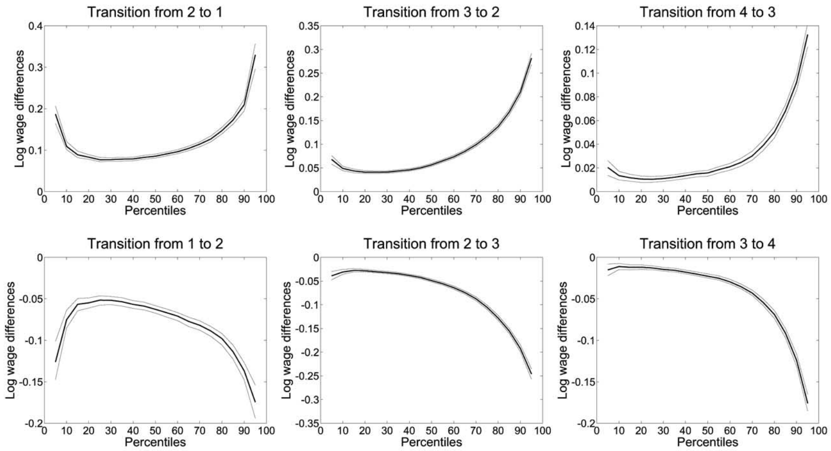

FIG. 9.—Difference in the distribution of wages in common layers after minus before the transition. This figure portrays, for each transition type, the estimated log change (on the *y*-axis) in the percentiles (on the *x*-axis) of the wage distribution in common layers within firms, after the transition versus before the transition, and associated 95 percent bootstrapped confidence intervals. To build it we follow the same process described in figure 8, to which we refer, with the only difference that after removing year and firm fixed effects, we focus on the wage distribution implied by all hours worked in layers that are common before and after the transition.

interpretation. In all transitions in which the firms add (drop) a layer, we observe a shift down (up) in the distribution of wages after the transition. In fact, the largest changes in the distribution are now observed at the upper tail of the distribution. This was expected from the results in table 13, which indicate that average wages in the upper preexisting layers are the ones that fall the most when firms add a layer.

These results are particularly meaningful when contrasted with figure 10, where we present the difference in the log wages at each percentile of the distribution for firms that expand without reorganizing.[22] As before, the distribution of wages changes significantly. However, now the distribution of wages shifts up significantly at all percentiles of the distribution. So, as expected from the layer-level results above, the distribution of wages of firms that grow and reorganize shifts down for preexisting hours of work, while the distribution of wages of firms that grow without reorganizing shifts up.[23] These results underscore how, even if one

[22] Again, bootstrapped 95 percent confidence intervals are included in the figure, and an explanation of how we constructed the figure is presented in online app. B.

[23] In fig. A5 in online app. A, we present the change in the distribution of wages conditioning on firms that contract. In this case the distribution shifts down, as expected from the theory and the layer-level outcomes, for all percentiles in firms with three and four layers and for most percentiles in firms with one and two layers. In the case of one- and two-layer firms, wages at the very top of the distribution increase. This might be the result of some rent extraction by top managers and owners before firms exit.

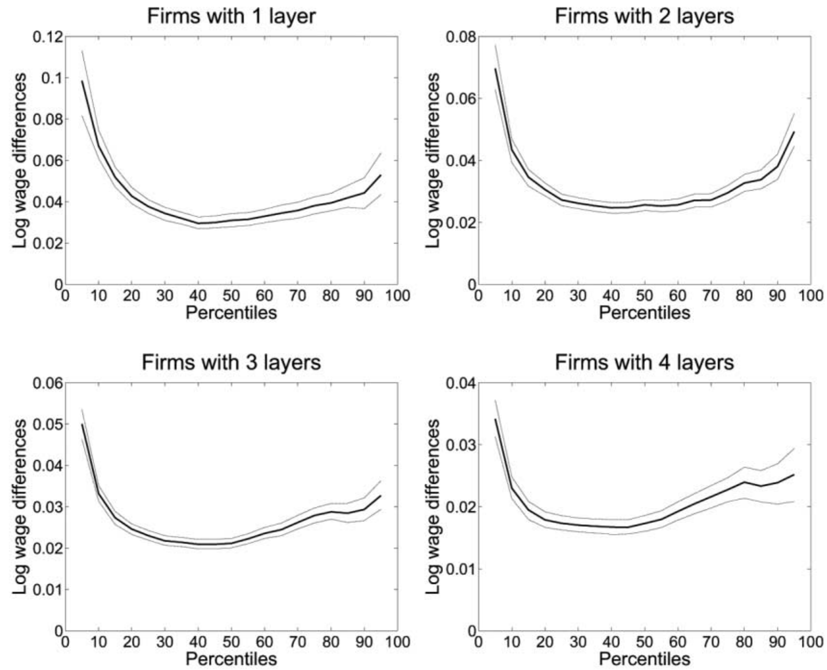

FIG. 10.—Difference in the distribution of wages for firms that do not transition and $d\ln VA > 0$. This figure portrays, for firms staying at a given number of layers in 2 consecutive years and positive change in value added, the estimated log change (on the y-axis) in the percentiles (on the x-axis) of the wage distribution within firms, the year after less the year before, and associated 95 percent bootstrapped confidence intervals. To build it we follow the same process described in figure 8.

is not interested in layer-level outcomes, conditioning on reorganization is essential to understanding firm growth and its implications for wages and the characteristics of a firm's labor force.[24]

We finish this section with a graphical illustration of firm transitions, as well as two examples, of the changes in wages and normalized hours that result from those transitions. Figures 11–13 showcase the representative effects for transitions between one and two, two and three, and three and four layers, respectively. The main characteristics of these changes have been analyzed before and are consistent with the first four implications of the theory in Section II. The figures emphasize the dramatic effects associated with reorganizations in the data. Beyond representative effects, consider the particular example of an individual

[24] We want to stress again that the results of this analysis are never to be understood as causal effects. We are exploring to which extent moment restrictions among endogenous variables from the theory are present in the data.

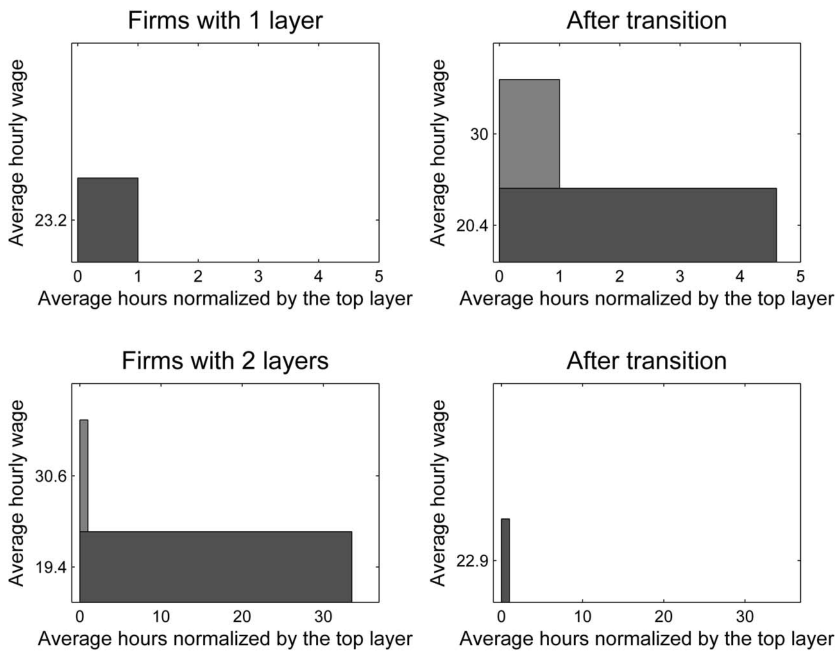

FIG. 11.—Representative transitions between firms with one and two layers. These figures depict hierarchies of firms that transition from L to $L + 1$ layers and from $L + 1$ to L. In each panel, for fixed L, the left column portrays the representative hierarchy with L layers and normalized hours, using only firms that will make the indicated transition. Each representative hierarchy is computed as in figure 5. The right column estimates the hierarchy after the transition using the average log changes in each quantity as resulting from our estimates reported in tables 12 and 13: the wage (normalized hours) in layer ℓ after the transition from L to $L + 1$ is computed as the average hourly wage (normalized hours) before the transition multiplied times $\exp(b)$, where b is the average log change in wages (normalized hours) at layer ℓ in transitions from L to $L + 1$. Normalized hours after a transition are always set to one for the top layer. An analogous procedure is followed for transitions from $L + 1$ to L. For transitions one layer up, the hourly wage for the top layer after the transition is computed as follows: we estimate the average log change in the detrended wage of the top layer and multiply the wage at the top layer before the transition times the exponential of this change.

firm that went from two to three layers. This firm increased its value added by about 12 percent, added about seven workers in layer 1 and four managers in layer 2, and assigned a manager to the new layer 3. As a result of the reorganization, average wages by layer declined approximately 6 percent and 17 percent, respectively. Consider now an example of a firm that dropped a layer to go from three to two. This firm reduced its value added by approximately 20 percent and dropped 10 workers at layer 1 and one manager at layer 2 (in addition to the top layer, of course); average wages increased by 3 percent and 13 percent, respectively.

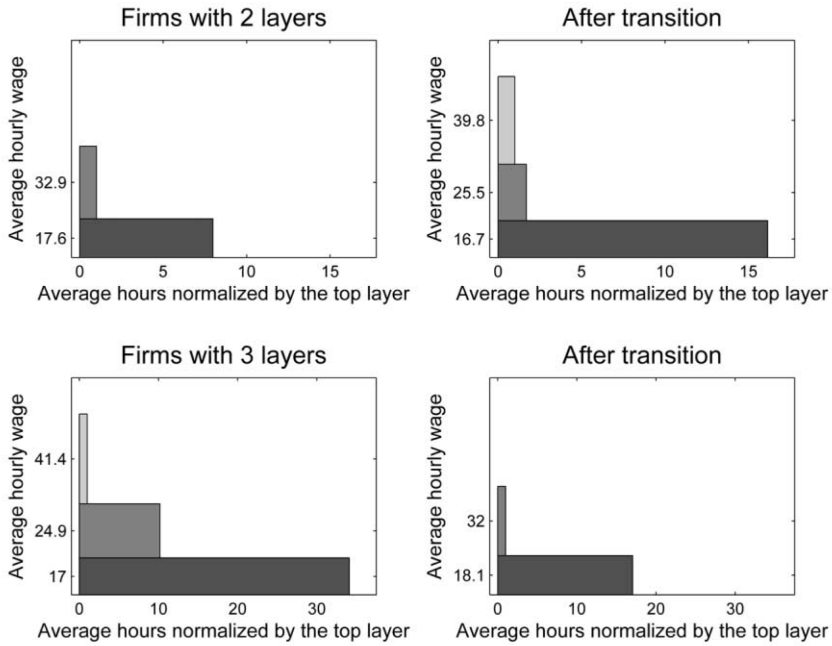

FIG. 12.—Representative transitions between firms with two and three layers. These figures depict hierarchies of firms that transition from two to three layers and from three to two. To build it we follow the same process described in figure 11.

We could also illustrate how firms respond to an expansion of, say, 5 percent in value added. In that case, we can also illustrate what happens with the firms that do not add layers (which we know increase normalized hours and wages). We present these results in figures A3 and A4 in online appendix A.

V. How Do Firms Change the Average Wage in a Layer?

We finish our analysis with an exploration of how firms manage to adjust the average wage in preexisting layers. The answer to this question will be specific to the particularities of the French labor market. As such, it is interesting for learning something about France, but perhaps less so about firms in general. Still, we want to illuminate as much as possible how firms do what we are arguing they are, in fact, doing. Of course, we know well that labor market frictions, institutions, and regulations make reducing the wage of a particular employee complicated, if not impossible. So, for example, how can the prediction that firms that add layers reduce average wages in preexisting layers be so clearly present in our data? We explore two margins of adjustment. First, we study if firms adjust average wages in a layer, which we interpret as the average knowl-

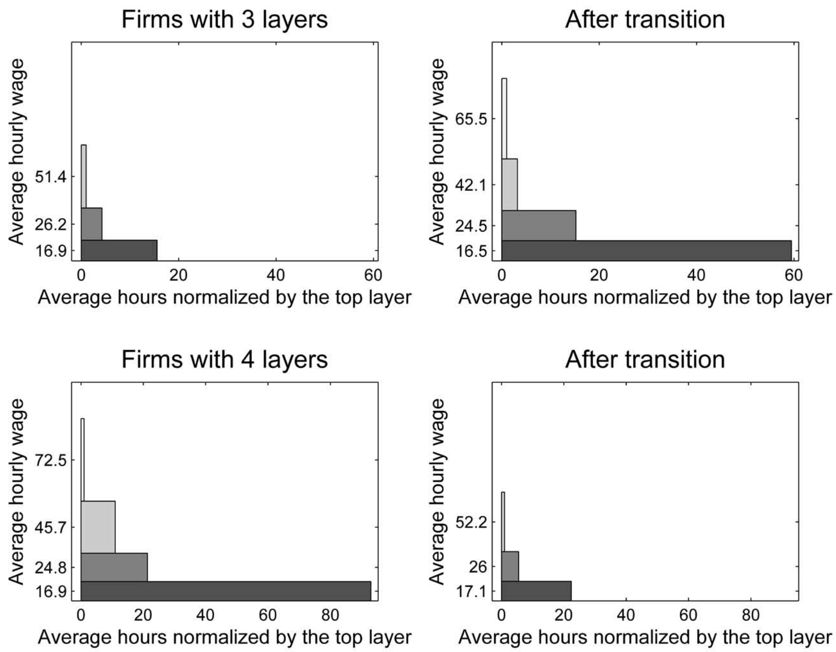

FIG. 13.—Representative transitions between firms with three and four layers. These figures depict hierarchies of firms that transition from three to four layers and from four to three. To build it we follow the same process described in figure 11.

edge in the layer, by adjusting the wages of existing employees or by adjusting their composition. We then explore which changes in the characteristics of employees lead to the changes in wages or knowledge that we observe. More specifically, we study to what extent the changes in wages are the result of changes in formal education or changes in labor market experience.

A. *Extensive versus Intensive Margin*

Our data, and any other data in France, are insufficient to track the universe of employees in a firm over time, as would be required to understand the individual agent effects of changes in organization.[25] The data allow us to study only how reorganization affects layer-level average wages by tracking, across two consecutive periods, hours of work, and

[25] We cannot use a methodology like the one in Abowd et al. (1999). The reason is that their data are just a random subsample of the workforce over time (all individuals employed in French enterprises who were born in October of even-numbered years), and therefore, it is impossible to reconstruct the whole organization of the firm to study individual workers' heterogeneity using such data.

average hourly wages, as employees enter, stay, or leave the layer during a transition.[26] Studying these transitions, we find that firms reduce wages by promoting or firing the highest-paid hours of work in a layer and by hiring new hours of work that are paid less. They keep the salaries of hours of work that stay in the layer essentially unchanged, although in some cases they raise them slightly.[27] This reduces the average wage in the layer. That is, the adjustment in the average wage happens mostly through the extensive margin. Firms adjust the composition of the employees in the layer and not the individual wages that they command.[28]

Note that there is nothing mechanical about this finding. Firms are actively deciding to use the extensive margin to change the average wage in a layer in a particular direction. This is particularly clear when we contrast this finding with the behavior of firms that expand without reorganizing. In those firms, average wages in each layer increase even though the normalized number of hours increases as well. That is, those firms choose to hire new workers in each layer that earn more than the average, not less.

This way of adjusting wages in a layer could be rationalized with the presence of some form of downward wage rigidities. In that case, the reductions we observe in the average wages of preexisting layers in firms that add layers necessarily understate the reductions that we would observe without these frictions. Similar empirical explorations in other, more flexible, labor markets would shed light on the validity of this argument.

B. *Education or Experience to Adjust Knowledge and Wages*

So far we have interpreted changes in the average wage in a layer as a direct measure of changes in the average knowledge of agents employed in that layer. This is consistent with a long tradition in labor economics that interprets wage differentials as compensating for individual characteristics (Rosen 1986). In our case, wage differentials are compensating for the marketable characteristics of individuals that we call knowledge and that individuals use to “solve problems.” Of course, if we had a

[26] See descriptions of tables A36–A39 in online app. A for a detailed account of this exercise and findings.

[27] Tables A36–A39 in online app. A present these results for all transitions. Table A36 shows that the results for hours that stay in the layer are small and not particularly systematic, so they do not dominate the effect on the average wage in a layer. In contrast, table A37 shows that the wages of hours that enter the layer are always significantly lower than the wages of hours that leave the layer (and vice versa for firms that drop layers, except for layer 1 in firms that go from four to three layers of management).

[28] Several studies have documented that average wages at firms increase with firm size or as firms expand: the firm size-wage premium (Brown and Medoff 1989; Bernard and Jensen 1997, 1999; Abowd et al. 1999; Oi and Idson 1999; Frías et al. 2009). Our results are not inconsistent with these findings: our theory does not address how individual-level outcomes change in response to a reorganization, and we cannot study this question with our data.

direct and reliable measure of average knowledge in a layer, we could repeat our study using this measure instead of wages.

We face several difficulties when we embark on this alternative exercise. First, knowledge is a complex concept that is not easily observable in the data. We can observe only a subset of worker characteristics that in reality are combined together to form the "knowledge" of an individual. Perhaps the two most important of these characteristics are formal education and experience. Still, it is not obvious that any of these characteristics on its own is a better measure of knowledge than a market-based measure like wages. In fact, it is likely that the knowledge of individuals is optimally created by combining them in a way that depends on the relative cost of acquiring knowledge in these different ways. Second, there is the issue of data availability. There is no data set in France that provides employer-employee matched measures of formal education, experience, and perhaps other worker characteristics for every individual employee in an organization. So, instead, we use the most comprehensive labor market survey available in France to estimate formal education and experience in our data set.

We estimate the average years of education and the potential labor market experience for each employee in DADS. We start by using the French Labor Force Survey (Enquête Emploi), a survey that contains information on wages, years of education, age, and individual characteristics, among other things. These data allow us to estimate a relationship between log years of education (or log potential labor market experience) and the subset of individual characteristics that we can also find in the DADS data set, namely, the hourly wage, age, industry and time fixed effects, and different measures of occupation.[29] We then use the resulting coefficients to estimate the average education across all hours employed in each layer in DADS (and similarly for labor market experience).[30]

Table 15 presents the elasticity of potential labor market experience and education with respect to value added for firms that do not change layers. Table 16 presents the average change in potential labor market experience and education for firms that change layers. We detrended all measures as we did above for the results in tables 10 and 13. Our findings are robust to several different specifications and forms of detrending the data (see again online app. A).

Consider first the firms that grow without adding layers. As we showed in table 10, these firms increase average wages in all layers. Table 15 shows that these firms increase knowledge by increasing the formal education

[29] Potential labor market experience is defined as age minus the age of the worker at the end of her formal education.

[30] Online app. B provides details about the exact estimating equation we use, as well as other details and robustness checks on our procedure to estimate these formal education and potential experience measures.

TABLE 15
Elasticity of Knowledge with Value Added for Firms That Do Not Change L

Number of Layers	Layer	Experience	p-Value	Education	p-Value	Observations
1	1	.001	.690	.002	.030	45,009
2	1	−.010	.010	.004	.000	64,469
2	2	.009	.030	.003	.000	64,469
3	1	−.010	.000	.004	.000	91,161
3	2	−.001	.970	.003	.000	91,161
3	3	.008	.000	.001	.100	91,161
4	1	−.015	.000	.003	.000	52,730
4	2	−.004	.280	.003	.000	52,730
4	3	.000	.970	.000	.790	52,730
4	4	.007	.020	−.003	.070	52,730

Note.—This table reports the results of regressions of log change in years of potential labor market experience and of education by layer on log change in value added for firms that do not change their number of layers L across two consecutive periods, where both variables are detrended as specified in the main text. Specifically, we run a regression of detrended log change in each of the two variables at layer ℓ in a firm with L layers on its detrended log change in value added, and no constant, across all the firms that stay at L layers across 2 consecutive years. The columns p-value report the respective p-value for each left-hand side using robust standard errors. We use specification 2 to estimate these measures (for a description of the specifications, refer to the online app.).

in the layer. This is particularly true in the lower layers of the organization. At the highest level of the organization, firms also increase average knowledge by hiring more experienced workers. So all of them increase the average wage or knowledge, but they do so using formal education at the bottom of the organization and using also experience in the top layer. This behavior is consistent with the view that formal years of education primarily provide the knowledge to solve the most routine problems in an organization, the problems that are handled at the lower layers of the hierarchy. In contrast, labor market experience provides the knowledge required to solve more infrequent problems, the tasks handled in the higher levels of the hierarchy.

Table 16 presents results for firms that change layers. Consider the case of firms that add layers, the ones that, according to table 13, reduce the average wage, or knowledge, in each preexisting layer. These firms mostly use experience to reduce the average knowledge in a layer and leave average levels of education mostly unchanged. The result is evident in table 16, where the measure of experience declines in all preexisting layers for firms that add layers and rises for all firms that drop layers. These changes are large and precisely estimated. In contrast, for education, the results are very small, not systematic, and in many cases not significant.

These results complete our description of the changes in organization as firms grow. A firm that grows and adds a layer hires more employees at the bottom who have a level of education similar to that of the ones it already employs. It also promotes the more experienced employees to

TABLE 16
Average Change in Knowledge for Firms That Change L

Number of Layers Before	Number of Layers After	Layer	Experience	p-Value	Education	p-Value	Observations
1	2	1	−.108	.000	−.004	.000	10,171
1	3	1	−.184	.000	−.003	.290	1,261
1	4	1	−.330	.000	.025	.030	97
2	1	1	.096	.000	.005	.000	11,088
2	3	1	−.044	.000	.000	.820	16,778
2	3	2	−.181	.000	.002	.010	16,778
2	4	1	−.064	.000	.002	.290	1,124
2	4	2	−.228	.000	.008	.010	1,124
3	1	1	.137	.000	.006	.000	1,584
3	2	1	.044	.000	.002	.530	17,626
3	2	2	.153	.000	.000	.000	17,626
3	4	1	−.011	.000	.001	.100	14,098
3	4	2	−.038	.000	−.001	.000	14,098
3	4	3	−.176	.000	.024	.820	14,098
4	1	1	.197	.000	−.002	.950	123
4	2	1	.073	.000	.000	.120	1,454
4	2	2	.172	.000	−.005	.000	1,454
4	3	1	.013	.000	−.002	.260	15,150
4	3	2	.025	.000	−.001	.000	15,150
4	3	3	.113	.000	−.020	.000	15,150

Note.—This table reports estimates of the average detrended log change in years of potential labor market experience and of education at each layer ℓ among firms that transition from L to L' layers, with $L \neq L'$: for a transition from L to L', we can evaluate only changes for layer number $\ell = 1, \ldots, \min\{L, L'\}$. The detrending is explained in the main text. Each average change is estimated as a regression of the detrended log change in the variable of interest in layer ℓ in 2 consecutive years on a constant. We use specification 2 to estimate these measures (for a description of the specifications, refer to online app. B).

the higher layers, which lowers experience in all layers. Of course, the firm could balance the implied reduction in knowledge by employing more knowledgeable employees at the bottom of the organization. However, in contrast to the case in which they expand without adding layers, they choose not to do so. This is the active decision that leads to a lower average wage and level of knowledge.

VI. Conclusion

This paper provides the first anatomy of organizations using a large-scale data set. Previous studies focused on particular details of the organization and included only a few hundred firms. In contrast, we use virtually all the manufacturing firms in France during the period 2002–7. To study the organization of these producers and how they modify their organization, we use occupational data to classify workers in layers. The concept of a layer of employees and our empirical implementation using hierarchical occupations have proven useful in analyzing these data. First, firms actively manage the number of layers in the firm. Firms expand by adding them and contract by dropping them.

We find that the organization of labor in the firm depends on the number of layers. In our data, studying the behavior of hours of work or wages in a firm without classifying workers in layers and conditioning on changes in these layers results in ambiguous firm-level effects that are hard to understand and not robust. In contrast, once we classify workers in layers, the behavior of wages and hours of work for each layer is easy to understand through the lens of a theory of hierarchical organization, such as the one in Caliendo and Rossi-Hansberg (2012) (which borrows the technology from Rosen [1982] and Garicano [2000]).[31] Firms that expand without adding layers increase hours at all layers as well as average wages at all layers. They essentially behave the way most theories would suggest expanding firms behave. In contrast, firms that add layers expand much more, add many hours of work to all layers, but reduce average wages in all preexisting layers. The fact that the wages of all preexisting layers go down is, as far as we know, a novel empirical finding. Furthermore, it is hard to rationalize with theories that do not have the organization of knowledge at their core.[32] The evidence we provide is extremely robust and we hope will be confronted by the different theories of firm dynamics.

Although somewhat more speculative given the nature of the French data, we also present evidence that the changes in layer-level outcomes are in general achieved by changing the composition of workers in a layer and not by changing the knowledge or wages of continuing employees. Furthermore, these changes are achieved by actively managing the experience and education of the labor force in a layer. Our results suggest that changes in experience are used to reduce and increase knowledge when firms change layers, while education is essentially kept constant. In contrast, when firms expand without adding layers, they hire more educated employees in virtually all layers of the hierarchy but primarily more experienced employees at the top.

It is important to be clear about the way in which our results should be interpreted. Our study has identified patterns in the data and clear robust correlations between organizational change and a variety of firm

[31] We have relied on a dynamic interpretation of a theory that is static in nature. In so doing, we have implicitly assumed that firms can quickly react to revenue-enhancing shocks with organizational changes. Fully fledged dynamic extensions of the theory that incorporate organizational adjustment costs may shed further light on the changes in the organization of forward-looking firms and its implications.

[32] Calvo and Wellisz (1978) and Qian (1994) present theories in which efficiency wages decrease as the span of control is reduced. However, these theories are silent about the way in which growing firms reorganize and, as a result, change the number of employees and wages at each layer.

characteristics, like the distribution of employees across layers and their wages. The patterns we uncover are useful to discriminate between theories of firm growth and their implications for labor market outcomes. They establish a set of facts that theories of firms and wages should relate to in the future. What we have not done in this paper is to identify exogenous shocks that lead to reorganization. So our findings cannot be used to compute, say, the size of the effect of an economywide change in technology on the wages paid by firms or the number and types of the workers they hire. This analysis could easily be done in a country in which such a shock is clearly identifiable.

Moreover, our results have implications for firm- and layer-level outcomes but are silent about how reorganization affects individual workers. This explains our empirical approach and justifies why we do not focus on individual worker characteristics. Understanding the experience of individual workers as firms grow is an interesting research agenda, but not one we embark on in this paper. Using the approach in this paper, such a study could be done in a country with a matched employer-employee data set that tracks the universe of employees over time as they move across layers and firms.

Our empirical analysis of firms' organization can be replicated easily for any country that has data on wages by occupation at the firm level. This opens the possibility of studying how a firm's organization affects a variety of economic phenomena. We hope that our paper will convince others of the importance and relevance of this endeavor.

References

Abowd, J., F. Kramarz, and D. Margolis. 1999. "High-Wage Workers and High-Wage Firms." *Econometrica* 67 (2): 251–333.

Baker, G., M. Gibbs, and B. Holmstrom. 1994. "The Wage Policy of a Firm." *Q.J.E.* 109 (4): 921–55.

Baker, G., and B. Holmstrom. 1995. "Internal Labor Markets: Too Many Theories, Too Little Evidence." *A.E.R.* 85 (2): 255–59.

Bernard, A. B., and J. B. Jensen. 1997. "Exporters, Skill Upgrading and the Wage Gap." *J. Internat. Econ.* 42 (1): 3–31.

———. 1999. "Exceptional Exporter Performance: Cause, Effect, or Both?" *J. Internat. Econ.* 47 (1): 1–25.

Bloom, N., B. Eifert, D. McKenzie, A. Mahajan, and J. Roberts. 2013. "Does Management Matter? Evidence from India." *Q.J.E.* 128 (1): 1–51.

Bloom, N., R. Sadun, and J. van Reenen. 2012. "The Organization of Firms across Countries." *Q.J.E.* 127 (4): 1663–1705.

Bloom, N., and J. van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Q.J.E.* 122 (4): 1351–1408.

Brown, C., and J. Medoff. 1989. "The Employer Size–Wage Effect." *J.P.E.* 97 (5): 1027–59.

Caliendo, L., and E. Rossi-Hansberg. 2012. "The Impact of Trade on Organization and Productivity." *Q.J.E.* 127 (3): 1393–1467.

Calvo, G. A., and S. Wellisz. 1978. "Supervision, Loss of Control, and the Optimum Size of the Firm." *J.P.E.* 86 (5): 943–52.

Caroli, E., and J. van Reenen. 2001. "Skill-Biased Organizational Change? Evidence from a Panel of British and French Establishments." *Q.J.E.* 116 (4): 1449–92.

Frías, J., D. S. Kaplan, and E. A. Verhoogen. 2009. "Exports and Wage Premia: Evidence from Mexican Employer-Employee Data." Working paper, Columbia Univ.

Garicano, L. 2000. "Hierarchies and the Organization of Knowledge in Production." *J.P.E.* 108 (5): 874–904.

Garicano, L., and T. N. Hubbard. 2007. "Managerial Leverage Is Limited by the Size of the Market: Theory and Evidence from the Legal Services Industry." *J. Law and Econ.* 50:1–44.

Garicano, L., and E. Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Q.J.E.* 121 (4): 1383–1435.

Lentz, R., and D. T. Mortensen. 2008. "An Empirical Model of Growth through Product Innovation." *Econometrica* 76 (6): 1317–73.

Melitz, M. J. 2003. "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Oi, W., and T. Idson. 1999. "Firm Size and Wages." In *Handbook of Labor Economics*, vol. 3, edited by Orley Ashenfelter and David Card, 2165–2214. Amsterdam: Elsevier Sci.

Qian, Y. 1994. "Incentives and Loss of Control in an Optimal Hierarchy." *Rev. Econ. Studies* 61 (3): 527–44.

Rajan, R. G., and J. Wulf. 2006. "The Flattening Firm: Evidence on the Changing Nature of Firm Hierarchies from Panel Data." *Rev. Econ. and Statis.* 88 (4): 759–73.

Rosen, S. 1982. "Authority, Control, and the Distribution of Earnings." *Bell J. Econ.* 13 (2): 311–23.

———. 1986. "The Theory of Equalizing Differences." In *The Handbook of Labor Economics*, vol. 1, edited by Orley Ashenfelter and Richard Layard, 641–92. New York: Elsevier.