The Impact of Market Size on Firm Organization and Productivity*

Grigorios Spanos†

February 11, 2017

Abstract

This paper presents a novel mechanism to explain variation in firm productivity across locations: variation in the degree of stratification into layers, or levels of hierarchy. Initially, I develop a theoretical model with heterogeneous firms, endogenous organizational choice, and endogenous markups that respond to the size of the market. The model yields two implications. First, firms in larger markets organize with a greater number of layers. Second, because of more layers, firms in larger markets are more productive. I then use administrative French data to examine the model's implications on non-tradeable sectors. I find evidence consistent with the model. More precisely, I find that firms in denser markets operate with more layers, and that an additional layer is associated with an increase in firm productivity. In the final part of the paper, I assess the role of organization in explaining the productivity gains from operating in denser markets. I find that between 6.2% and 36.5% of the productivity gains are explained by firms having a greater number of layers.

KEYWORDS: firm organization, heterogeneous firms, market size, density, regional disparities, wages, firm productivity.

JEL Codes: D22, L11, L22, L23, J24, R12.

*I wish to thank Kunal Dasgupta, Gilles Duranton and Peter Morrow for excellent supervision. I am indebted to Pierre-Philippe Combes for his comments and kindly providing me with data and many programs, Laurent Gobillon for providing me with historical data, and to Tanguy van Ypersele for his many insightful theoretical comments and suggestions. This paper also benefited from comments and discussions with Nicolas Berman, Kevin Devereux, Stefan Dodds, Camille Hémet, Emeric Henry, Garth Frazer, Luis Garicano, Nicholas Li, Joan Monras, Frank Neffke, Lorenzo Rotunno, Michel Serafinelli, Marc Sangnier, Bernard Sinclair-Desgagné, Brian Tavares, Daniel Trefler, Thierry Verdier and seminar participants at the Canadian Economics Association, the Urban Economics Association, and the International Trade Workshop at the University of Toronto. Finally, I would like to thank the staff of the French National Statistical Institute, INSEE, in Paris, for their kindness and support. A part of this project was conducted during my stay at the Aix-Marseille School of Economics, and funded by the AMSE Pre-Doctoral Research Grant. All remaining errors are my own.

†Aix-Marseille University, Aix-Marseille School of Economics. E-mail: grigorios.spanos@gmail.com

1 Introduction

This paper provides a new mechanism to explain the productivity of firms across locations. More precisely, this study puts forward that firms in larger markets are more productive because they organize with a greater number of layers. I base my argument on a theoretical model that allows market size to affect the organizational decision of firms and I then examine the implications from the model on French firms that operate in service industries with exclusively local demand. In my empirical analysis I find that firms in denser markets operate with a greater number of layers. For instance, a firm in an urban area with a 100% higher density contains roughly an additional 0.101 layers. I also find that organization has a prominent role in explaining the productivity of firms. For example, controlling for industry-urban effects, an additional layer in a firm is associated with an 12.7% increase in value-added per worker. Both results are implied by the model. In the final part of the paper I examine the role organization has in explaining the differences in the productivity of firms across locations. I use several measures of firm productivity and conclude that part of the productivity gains from denser markets are explained by differences in the way firms organize production. For instance, roughly 26% of the value-added per worker gains from denser urban areas are explained by firms having a greater number of layers, a mechanism that until now has not been investigated in the literature.

In general, studying how firms organize production improves our understanding on firms and how they respond to changes in their environment. Moreover, recent work has demonstrated that how firms organize is important for a number of additional economic reasons (see Garicano and Rossi-Hansberg (2014)). First, organization affects the productivity of firms, and second, how firms organize production also determines the number and type of workers they hire as well as the compensation workers receive from firms. Consequently, understanding the organization of firms also improves our ability to understand the productivity of firms and the composition of workers across markets and locations, as well as how both change in response to the economic environment. Keeping with this reasoning, this study is the first to examine how important are differences in the way firms organize production to understanding the differences in the productivity of firms across locations.

The analysis is made up of two components: a theoretical framework that explores how market size affects the organizational decisions of firms and an empirical analysis examining the model's implications using high quality administrative French data.

The theoretical model combines the knowledge-based management hierarchy framework of Garicano (2000) and Caliendo and Rossi-Hansberg (2012) with the endogenous markups framework of Melitz and Ottaviano (2008). Both firm heterogeneity and firm organization are modeled as in Caliendo and Rossi-Hansberg (2012), where entrepreneurs pay a fixed cost to receive a demand draw and then decide on the optimal number of layers in firms, and the number and knowledge of workers in each layer. Endogenous markups, which respond to the level of competition between firms and the size of the market, are introduced using the linear demand system developed by Ottaviano et al. (2002) and extended by Foster et al. (2008) to allow for variety specific demand. Together both elements create a framework with an endogenous distribution of organizations, of firm productivity, and of worker wages, that vary with the size of the market.

In my framework firms are knowledge-based management hierarchies and the structure of production is based on the theory of Garicano (2000) and Caliendo and Rossi-Hansberg (2012). Agents have one unit of available time and production requires labor to be combined with knowledge. Entrepreneurs pay a fixed cost to receive a demand draw, and then create firms, hire workers, decide on their knowledge and the number of layers in firms. Each organization is made up of a number of production workers and managers with specific levels of knowledge. Following Garicano (2000) production workers are always located in the lowest layer of firms, whereas managers occupy all other layers. Furthermore, each organization also consists of a menu of costs and the decision to add a layer is similar to a tradeoff between fixed and variable costs (Caliendo and Rossi-Hansberg (2012)). An extra layer increases costs because workers are compensated for the knowledge they acquire and the time they devote to the firm. At a certain scale, however, an additional layer allows firms to use workers' knowledge more efficiently, lowering average costs in firms. When a firm adds a layer therefore, it is in a sense increasing its fixed costs and lowering its variable costs (Caliendo and Rossi-Hansberg (2012)). In this framework how firms organize production is ultimately determined by the optimal output produced, which depends on the production technology, the level of demand, and the degree of competition between firms.

The model yields two implications which I examine in the empirical section of the paper. First, the model yields a comparison of the distribution of organizations, that is the percent of firms producing with a given number of layers, across locations. An increase in market size increases the level of competition between firms and lowers markups, which forces firms with the lowest variety specific demand draws to exit the market, and the remaining firms to restructure their organization. This yields the prediction that in larger markets the distribution of organizations first-order stochastically dominates the distribution of organizations in smaller markets. This implies that in larger markets, where the level of competition between firms is greater, the average firm will produce with a greater number of layers.

The second implication from the theoretical framework relates market size to firm productivity. In the model firm productivity is determined by the way production is organized. Firms that produce with a greater number of layers organize labor more efficiently, have lower average costs, and are therefore more productive. Together with the first prediction of the model this implies that in larger markets firms are more productive because they have a greater number of layers.

To examine the two implications of the model I use three types of data: i) matched employer-employee data from the Déclarations Annuelles des Données Sociales (DADS); ii) balance sheet data from the Fichier Complet Unifié de Suse (FICUS); and iii) population census data from the Recensement de la Population (RP). With the DADS I construct the organization of firms using the method developed by Caliendo et al. (2015b). With this method there are four types of organizations in the data: one-layer, two-layer, three-layer, and four-layer firms. With the FICUS I construct measures of firm productivity and I use the demographic information from the RP to measure the size and the characteristics of markets.

The empirical analysis is conducted on non-tradeable monopolistically competitive industries with local demand: Clothes and Shoes Retail, Traditional Restaurants, and Hair and Beauty Salons. These industries satisfy the assumptions of the theory and in my analysis I show that the classification of layers developed by Caliendo et al. (2015b) is meaningful and consistent with the model. Further, the definition of local markets is based on two geographical decompositions of mainland France: i) employment areas, which are travel to work areas; and ii) urban areas, which correspond to cities and their suburban areas. I use density to measure the size of local markets.

To examine the first implication from the model I compare the distribution of organizations across markets of different density using the Mann-Whitney stochastic dominance test. My main finding is the distribution of organizations in high density markets first-order stochastically dominates the distribution of organizations in low density markets. For instance, across all three industries a firm drawn at random from an employment area (urban area) with above-median density is 57.5 (59.3) percent more likely to operate with greater number of layers than a random firm from an urban area with below-median density.

While the evidence is consistent with the model, other factors may explain these results. More precisely, it may be the case that the cost of hiring workers and consumers' demand for local services vary across markets, affecting the level of competition and the organization of firms. It is also likely that firms producing with a greater number of layers attract more workers to a market, or that local shocks simultaneously determine firm organization and density. Such issues violate the independence assumption of the Mann-Whitney test and confound estimates of the relationship between density and the organization of firms. To address this concern, I turn to regression analysis and control for the characteristics of local markets. Following Ciccone and Hall (1996), Combes et al. (2008) and Combes et al. (2010) I also use historical data going back to 1831 to instrument for the density of local areas. Accounting for these possible problems, density continues to affect the organization of firms. With all local area controls a firm in an employment area (urban area) with a 100% higher density contains roughly an additional 0.058 (0.101) additional layers. This implies that increasing the density of an employment area (urban area) the size of Lyon, the third (third) most populated in France, to an urban area the size of Paris, the most (most) populated in France, corresponds to an additional 0.301 (0.134) layers in firms.

Another set of factors that may explain these results is the degree of task specialization across markets and firms. More specifically, it may be the case that firms in denser markets appear to have a greater a number of layers because tasks are specialized in denser markets. It is also likely that bigger firms assign workers to specific tasks and consequently seem to have a greater number of layers. To address these concerns, I also control for the degree of occupational concentration in local markets and for the other characteristics of firms, such as size, number of additional occupations, capital and the legal status of firms. Even when accounting for these factors, firms in denser markets continue to organize with a greater number of layers. Moreover, these results further demonstrate that organization is an important characteristic of firms that cannot be controlled for using other observable variables.

Having shown that firms in denser markets operate with a greater number of layers, I then turn to the second implication from the model and examine how important is organization to understanding the productivity of firms, as well as the differences in the productivity of firms across locations. I measure productivity using several approaches, including value-added per worker as well as total factor productivity (TFP) estimated with the methods proposed by Levinsohn and Petrin (2003) and Wooldridge (2009).

With these measures at hand, I first examine the relationship between the number of layers in firms and productivity. Accounting for local area effects, I find that organization is an important component of firm productivity. For example controlling for industry-employment area (industry-urban area) fixed effects, adding a layer in a firm is associated with a 12.4% (12.7%) increase in value-added per worker. This implies that differences in the way firms organize production accounts for roughly 14.3% (13.1%) of the difference in average value-added per worker between firms located in the first and fourth quartiles of the productivity distribution. I then examine the role organization has in explaining the differences in the productivity of firms across locations and again find that organization plays a prominent role. For instance, across employment areas (urban areas) roughly 22.4% (26%) of the elasticity of value-added per worker with respect to density is accounted for by firms having a greater number of layers. Using different measures of firm productivity, controlling for the characteristics of firms and of local markets, and instrumenting for density and the characteristics of local markets with past values, yields different estimates that lead to same conclusion. From the different specifications, I therefore obtain a range of estimates on the role organization has in explaining the productivity of firms across locations. Overall, I find that between 6.2% and 34.2% (11.6% and 36.5%) of the productivity gains from operating in denser employment areas (urban areas) arise from differences in the way firms organize production.

This paper builds on a growing literature on the organization of firms, reviewed in Antras and Rossi-Hansberg (2009) and Garicano and Rossi-Hansberg (2014). Most of the papers in this literature have been concerned with how information technology and international trade affect the organization of firms and the implications for income inequality and firm productivity.[1] Caliendo and Rossi-Hansberg (2012) are the first to embed the knowledge-based management hierarchy model developed by Garicano (2000) into a monopolistic competition framework with heterogeneous firms. In their model, the elasticity of substitution between product varieties does not vary with the size of a local market, and therefore in the absence of trade the distribution of organizations is the same across locations. As shown below, however, the organization of firms changes with the size of local markets. Empirically, Caliendo et al. (2015b) are the first to develop a method to construct the organization of firms with administrative French data. They center their analysis on firms operating in manufacturing sectors and show that this method is meaningful and consistent with the knowledge-based management hierarchy models. As shown in

[1] Papers examining how changes in information technology affect the organization of firms are the following: Garicano (2000), Garicano and Rossi-Hansberg (2004), Garicano and Rossi-Hansberg (2006), Raghuram and Wulf (2006), and Bloom et al. (2014). Papers analyzing how international trade, or offshoring, relate to the organization of firms are the following: Antras et al. (2006), Antras et al. (2008), Guadalupe and Wulf (2010), Dasgupta (2012), Caliendo and Rossi-Hansberg (2012), Caliendo et al. (2012) and Friedrich (2015).

this study, their method is also suitable to firms operating in service sectors. And, Garicano and Hubbard (2007) embed the framework of Garicano (2000) into a two-sector model to examine how market size affects the degree of field specialization and the organization of U.S. law firms. The mechanism put forth in their model is, however, different. In their model, the decision of firms is determined by the degree of uncertainty over the level of demand which decreases with the size of the market. This leads to more specialized firms and a greater share of workers employed in hierarchies. In contrast, this study focuses on the interactions between firms within an industry. In this paper, organizational decisions are determined by the level of competition between firms which increases with the size of the market.

The main contribution of this paper to the existing body of work is to develop and examine the implications from a theoretical framework that allows for market size to affect the organization of firms. This framework provides a testable prediction on the distribution of organizations. It also provides a new mechanism to explain the productivity of firms and the composition of workers across locations. To my knowledge this is the first paper in the literature to model and empirically examine the distribution of organizations across local markets and then to examine the productivity of firms across geographical areas while accounting for their organization.

More generally, the findings in this study are also relevant to a broad literature examining the productivity of firms, reviewed in Holmes and Schmitz (2010) and Syverson (2011). Within this literature several studies have separately explored the importance of market structure (Syverson (2004), Melitz and Ottaviano (2008) and Combes et al. (2012)) and organizational form (Garicano and Heaton (2010) and Caliendo et al. (2015a)) in explaining the observed dispersion in the productivity of firms. This paper contributes to this literature by showing that organization responds to the structure of a market and by disentangling the role of each component in explaining the productivity of firms.

This study is also related to other research. One is a theoretical literature exploring the nature and sources of the productivity gains benefiting firms and workers in denser markets, summarized in Duranton and Puga (2004) and Puga (2010). Another is an empirical literature estimating the magnitude of these gains, reviewed in Rosenthal and Strange (2004), Melo et al. (2009) and Combes and Gobillon (2015).[2] This paper contributes to both strands of research by providing a new micro-founded mechanism that explains the productivity gains observed from firms and

[2]Several studies estimating the magnitude of the gains from denser markets are the following: Ciccone and Hall (1996) and Glaeser and Mare (2001) for the United States, Combes et al. (2008), Combes et al. (2010) and Combes et al. (2012) for France, and Jofre-Monseny et al. (2011) for Spain.

workers in denser markets, namely differences in the way firms organize production, and by empirically examining the strength of this mechanism. Moreover, a point of departure from the existing studies is the nature of the gains. While the majority of studies emphasize economies of scale that are external to workers and firms, this paper's focus is on economies internal to firms.

The paper is organized as follows. Section 2 presents the theoretical model and discusses its main implications. Section 3 introduces the data and Section 4 examines the central implications of the model. All proofs are relegated to the Appendix.

2 Model

The objective of this paper is to examine how firm organization varies with the size of a market. To uncover this relationship, I embed the knowledge-based management hierarchies framework of Garicano (2000) and Caliendo and Rossi-Hansberg (2012) into the monopolistic competition framework with product differentiation and endogenous markups developed by Melitz and Ottaviano (2008). To model firm demand, I use the linear demand system developed by Ottaviano et al. (2002) and extended by Foster et al. (2008) to allow for heterogeneous product demand. Both elements are important. Each organization contains its own menu of costs, and the decision to add a layer is similar to paying higher fixed costs in exchange for lower variable costs. In turn, an increase in market size increases the level of competition between firms and lowers markups, determining which firms remain in the market and the way firms organize production.

2.1 Demand

The economy contains N homogeneous individuals who supply their unit of labor inelastically. Let q_0^c and q_i^c denote individual consumption of the numeraire good and variety i, and Ω the set of varieties available in the economy. Individual utility has the linear quadratic form and is equal to:

$$U^c = q_0^c + \int_\Omega \alpha_i q_i^c di - \frac{\gamma}{2}\int_\Omega (q_i^c)^2 di - \frac{\eta}{2}\left(\int_\Omega q_i^c di\right)^2, \tag{1}$$

where the parameters α_i, η and γ are positive, and α_i varies across varieties. These parameters determine the level of competition between varieties and the numeraire good. The parameter γ measures individuals' preference for product differentiation, while α_i and η measure the extent differentiated varieties are preferred to the numeraire good.[3]

I assume individual demand for the numeraire good is positive.[4] Let $\Omega^* \subset \Omega$ denote the set of varieties with positive measure M produced in equilibrium. Maximizing utility, substituting and isolating terms yields the linear demand of variety $i \in \Omega^*$:

$$q_i = Nq_i^c = \frac{N}{\gamma}\alpha_i - \frac{N}{\gamma}p_i - \frac{N}{\gamma}\frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}), \tag{2}$$

where $\overline{\alpha} = \frac{1}{M}\int_{\Omega^*} a_i di$ and $\overline{p} = \frac{1}{M}\int_{\Omega^*} p_i di$ are the average α_i and prices of varieties in Ω^*. Since q_i cannot be negative varieties with too great a price will not be consumed. In equation (2) the role of α_i also becomes clear. For a given quantity varieties with a greater α_i command a greater price. In other words, α_i shifts the level of demand without changing the slope of the curve.

In this setting the price elasticity of demand is equal to $\epsilon_i = \left|\frac{\partial q_i}{\partial p_i}\frac{p_i}{q_i}\right| = \left[p_i^{max}/p_i - 1\right]^{-1}$, where $p_i^{max} = \alpha_i - \frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$ is the maximum chargeable price. The term $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$ is common to all varieties and changes with the equilibrium of the model whereas α_i is specific to each variety. For a given price, the elasticity of demand is monotonically decreasing with α_i, and therefore firms producing a variety with a higher α_i encounter a more inelastic demand curve. The elasticity of demand also increases with the term $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$. More precisely, the elasticity of demand increases with M or $\overline{\alpha}$, and decreases with $\overline{p}$. As in Melitz and Ottaviano (2008) I define a tougher competitive environment an equilibrium with a greater price elasticity of demand for all varieties, ϵ_i. and as will be shown further below an increase in the size of the market will induce an increase in the term $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$.

2.2 Production

2.2.1 General Framework

Labor is the only input used in production and is inelastically supplied in a competitive market. Each individual has one unit of time and can work in one of three sectors: the homogeneous good sector, the differentiated good sector, or the schooling sector. Because they are identical, in equilibrium all individuals earn the same wage w, irrespective of their occupation and sector of employment.

[3]Varieties are more substitutable as γ approaches zero and a greater α_i or a lower η increase the demand of variety i relative to the numeraire good.

[4]This assumption requires the following condition to hold $I^c > \int_{i\in\Omega^*} p_i q_i^c di$, where I^c denotes individual income. Namely, individuals do not spend all of their income on the differentiated goods.

In the homogeneous good sector each good is produced with a constant returns to scale technology. In the analysis that follows the price of the homogenous good is normalized to 1 and as a result in equilibrium wages are equal to 1 in the economy.

In the differentiated goods sector individuals start new firms. For simplicity, I refer to an agent that starts a new firm as an entrepreneur.[5] To enter the market an entrepreneur first pays a fixed cost f_E, in units of labor, to develop a product. Once the product is developed she obtains a demand draw α from a known cumulative distribution $G(\alpha)$ which determines the demand schedule of her firm. Given the market environment, if the demand draw is low the entrepreneur may decide to immediately exit and not produce any output. Otherwise she creates a firm to produce her product.

I model production in the differentiated goods sector as Garicano (2000) and Caliendo and Rossi-Hansberg (2012). Firms are knowledge-based hierarchies and organize workers to use their knowledge efficiently. In general workers employed in a firm are divided into two broad categories: production workers and managers. Production workers are located in the lowest layers of firms while managers occupy all other layers. In the highest layer of firms is the entrepreneur who performs the same tasks as agents in that layer.[6]

In the differentiated goods sector production requires labor and knowledge. Agents solve problems to produce output and are compensated for any knowledge they acquire by the firm. Every production worker spends her one unit of time generating a problem z, from a known exponential distribution $F(z)$ with parameter $\lambda > 0$. For a given problem drawn, z, a worker can solve the problem if her knowledge set contains z, at which point A units of output are produced. The type of problem a worker draws, however, is unobservable. For a given realization of z, the only information a worker has is if she can solve the problem or not.[7]

In an one-layer firm, a firm with zero layers of managers, if the worker cannot solve the problem nothing is produced, while in a firm with managers this is not the case. If the worker cannot solve the problem she asks her manager one layer above. The manager spend h units of

[5] Throughout the paper, as I use the same production technology as Caliendo and Rossi-Hansberg (2012) I also adopt their terminology with the exception of the labeling of organizations. Caliendo and Rossi-Hansberg (2012) refer to organizations by the number of management layers they have, whereas I refer to organizations by their total number of layers.

[6] In other words if an entrepreneur creates a firm with zero management layers, then she performs the same tasks as a production worker. Otherwise if she creates a firm with a positive number of managerial layers she performs the tasks of a manager in the top layer.

[7] In the literature, this is referred to as the "labeling assumption". This assumption is crucial for the existence of organizations. See Garicano (2000). Furthermore, the parameter λ determines how common problems are in production. When λ is high problems are less costly to solve because the distribution of problems is more concentrated around zero, and as λ approaches infinity knowledge becomes unimportant for production.

time listening to the worker's problem and solves her problem if her knowledge set includes z. If the manager cannot solve the problem then the worker sends the problem to a manager one layer higher. This process continues until the problem is solved and A units of output are produced or the problem reaches the highest layer of the organization, that is occupied by the entrepreneur. The entrepreneur also spends h units of time listening to the worker's problem and solves the problem if her knowledge set includes z. Otherwise the problem remains unsolved and nothing is produced.

The schooling sector is composed of agents who spend their time providing knowledge to agents in the differentiated good sector. I assume that a unit interval of knowledge requires c units of teachers' time in the schooling sector. As teachers earn a wage 1 for their unit of time, the cost of a unit of knowledge is therefore c. In addition, knowledge in the differentiated goods sector is not cumulative: managers must not learn to solve the commonest problems before they can solve the exceptional ones.[8]

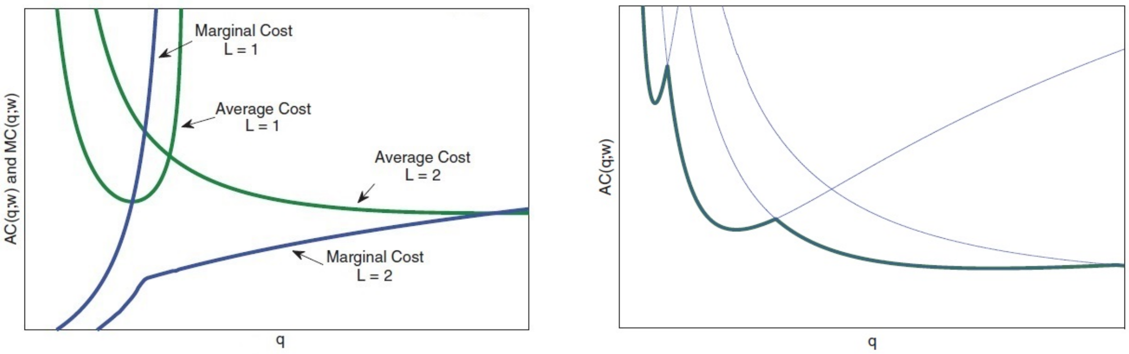
2.3 Entrepreneur's Problem

Consider an entrepreneur that does not exit the differentiated goods market. Given her demand schedule she creates a firm with the objective to maximize profits. In making this decision she also selects the number of layers, as well as the knowledge and number of agents in each layer of the firm. The subsections that follow are concerned with these decisions. In the first subsection I describe the properties of the cost functions which depend on how firms organize production. The cost functions are determined from the cost minimization problem and because it is analyzed in Caliendo and Rossi-Hansberg (2012) I simply summarize results. For details and proofs I refer the reader to their paper. Then in the second part I examine the decision to maximize profits.

2.3.1 Cost Minimization: Properties of Cost Functions

Figure 1a illustrates the marginal and average cost curves as a function of the quantity produced, in one and two-layer firms. In addition Figure 1b illustrates the global average cost curve when the organization of production is endogenous and firms can have either one, two and three-layers. Both figures are reproduced from Caliendo and Rossi-Hansberg (2012) and illustrate the main properties of the cost functions described in this subsection.

[8] That is to solve a problem z, an agent's knowledge set needs to include z, but the agent does not necessarily need to be able solve all problems between 0 and z. In other words, if the length of an agent's knowledge set is t, then the cost of acquiring that knowledge is ct.



(a) Average and Marginal Costs

(b) Global Average Cost

Figure 1: Cost Functions

Fixing the number of organizational layers in a firm the cost functions have the following properties. First, marginal costs are positive and increasing in output which implies that the knowledge of agents in all layers, and the size of each layer is increasing in output. Second, because marginal costs are positive cost functions are strictly increasing. Third, the average cost curves are convex in output, attain their minimum when they intersect the marginal cost curves, and converge to infinity as output approaches zero or infinity.[9]

The intuition for these results is the following. Given the number of layers, an increase in production requires more knowledge in the firm. Some of this additional knowledge is acquired by the entrepreneur, while the remainder is acquired by agents in the other layers. As agents in a layer acquire more knowledge their managers can supervise more of them, which implies that the chosen organizational structure becomes larger and the cost of producing an extra unit of output increases. In addition average costs are not a monotonic function of quantity. When production is small increases in output lead to a less than linear increase in total costs, because the firm does not need to provide too much knowledge to agents in the lower layers of the firm. Average costs are reduced until the minimum efficiency scale (MES) is reached. Beyond the MES, however, the opposite is the case and an increase in output increases the average costs of the firm.

Further, at the level of output where a firm is indifferent between two organizations and

[9]Throughout the paper, I assume, as in Caliendo and Rossi-Hansberg (2012), that the inequality $\frac{c}{\lambda} \leq \frac{h}{1-h}$ holds. This ensures that a firm would rather decrease its number of layers before choosing to hire employees with zero knowledge. See Caliendo and Rossi-Hansberg (2012) for details. Further in an organization with L layers the span of control of managers at layer $l+1$, which is equal to $\frac{n_L^l}{n_L^{l+1}}$ where n_L^l and n_L^{l+1} are the number of agents in layers l and $l+1$, is increasing in output as well. Another property of the cost function is that it is homogeneous of degree one in wages, while conditional factor demands are homogeneous of degree zero in wages. In this model as wages do not vary, I do not use this property. For a proof of all these statements, see Proposition 1 in Caliendo and Rossi-Hansberg (2012).

adds a layer of management, marginal costs decrease discontinuously. This implies that the knowledge of agents in all layers decreases discontinuously, while the number of employees in each layer increases discontinuously. Intuitively because agents are compensated for their time and knowledge adding a layer of management is costly. By adding a layer of management however, firms allow a manager with the ability to solve less frequent problems to better use her knowledge in the firm. To attain a certain level of output a firm that adds a layer of management requires more agents in its existing layers but with less knowledge. Therefore the firm is able to reduce its cost per worker. In other words for a given level of output, adding a new layer of management is as if a firm is paying higher "fixed"costs in the sense that it hires more agents, and, beyond a level of output it can produce at lower average costs because each agent is required to acquire less knowledge (Caliendo and Rossi-Hansberg (2012)).[10]

Let $C(q)$ denote the global minimum total cost of producing q units of the differentiated variety. The global total cost function of a firm is therefore the lower envelope of the total cost functions of producing q units with a given number of layers and is equal to: $C(q) = \min_{L\geq 1}\{C_L(q)\}$. The global total cost function contains the MES of all organizations as well as the regions around them. This is because the MES of each organization achieves a greater quantity and a lower average cost when the number of layers are increased. Therefore for a given organizational form global cost functions adopt the properties of local cost functions, whereas when the number of layers increase global cost functions are either discontinuous or non-differentiable. All of these results are summarized below.

Summary 1 *The global cost functions have the following properties:*

i. Fixing the number of layers, marginal cost, $MC(q)$, is positive and increasing in output. When the number of layers increases, $MC(q)$ decreases discontinuously.

ii. Fixing the number of layers, average cost, $AC(q)$, is convex and attains the minimum when $MC(q) = AC(q)$. $AC(q)$ is globally continuous but not globally convex and converges to infinity when q approaches zero of infinity.

iii. Fixing the number of layers, total cost, $C(q)$, is continuous and convex. $C(q)$ is globally continuous but not globally convex.

[10]Additional implications are that global marginal costs curves are not a monotonic function of quantity. Also managers' span of control decrease discontinuously with a change in the number of layers. For a proof of these statements, and the ones made in the next paragraph see Proposition 2 in Caliendo and Rossi-Hansberg (2012).

2.3.2 Profit Maximization

Caliendo and Rossi-Hansberg (2012) embed this production framework in a monopolistically competitive model with heterogeneous firms and constant markups. In such a setting, the elasticity of demand does not vary with the size of the market, and so in the absence of trade the organization of firms remains the same across markets of different sizes. Here I depart from Caliendo and Rossi-Hansberg (2012) and combine their production framework with the endogenous markups framework of Melitz and Ottaviano (2008).

Consider the profit maximization problem of an entrepreneur who creates a firm. Given her demand, the entrepreneur competes in a monopolistically competitive market and chooses quantity to maximize the profits of the firm. In making this decision, the entrepreneur implicitly decides on the optimal number of layers in the firm. Given her draw α, the entrepreneur's maximization problem is:

$$\pi(\alpha) = \max_{q} \; \{p(q(\alpha))q(\alpha) - C(q(\alpha))\}\,. \tag{3}$$

Because the global cost function is not a convex function, the first order conditions to (3) only yield a local solution. To solve the maximization problem one proceeds as follows. Holding the number of a layers fixed one determines the optimal profits and then compares these local optimal profits to find the global maximum (Caliendo and Rossi-Hansberg (2012)).[11] In this section, I first describe the local solution to the maximization problem and then I return to the global solution.

Fixing number of layers consider the optimization problem of an entrepreneur with demand draw α. From the first order condition the optimal quantity supplied is:

$$q_L(\alpha) = \frac{N}{2\gamma}\left[\alpha - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) - MC_L(q_L(\alpha))\right] \tag{4}$$

where the subscript L indicates the number of layers is held constant at L. The left-hand side is increasing with respect to quantity while the right-hand side is decreasing. Therefore given the number of layers in a firm, for every demand draw there exists a unique level of quantity, $q_L(\alpha)$, such that equation (4) holds with equality. Because marginal costs also depend on the quantity produced a closed form solution to equation (4) is unavailable. Furthermore the equations

[11] For a given number of layers, the local optimal profit function of a firm is strictly concave in q, $\pi(0) = -1$ and $\lim_{q\to\infty} \pi(q) = -\infty$ and therefore the profit maximization problem of the firm is well-defined.

characterizing optimal prices, markups over marginal costs ($p_L(\alpha) - MC_L(q_L(\alpha))$), markups over average costs ($p_L(\alpha) - AC_L(q_L(\alpha))$), revenues and profits, in terms of the model's parameters are the following:

$$p_L(\alpha) = \frac{1}{2}\left[\alpha - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) + MC_L(q_L(\alpha))\right], \tag{5}$$

$$\mu_L^{MC}(\alpha) = \frac{1}{2}\left[\alpha - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) - MC_L(q_L(\alpha))\right], \tag{6}$$

$$\mu_L^{AC}(\alpha) = \frac{1}{2}\left[\alpha - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) + MC_L(q_L(\alpha))\right] - \frac{C_L(q_L(\alpha))}{q_L(\alpha)}, \tag{7}$$

$$r_L(\alpha) = \frac{N}{4\gamma}\left[\left(\alpha - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p})\right)^2 - MC_L(q_L(\alpha))^2\right], \tag{8}$$

$$\pi_L(\alpha) = \frac{N}{4\gamma}\left[\left(\alpha - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p})\right)^2 - MC_L(q(\alpha))^2\right] - C_L(q_L(\alpha)). \tag{9}$$

Prices and markups depend on the quantity supplied. In addition unlike in many models, because marginal costs are generally not equal to average costs, firm's markups over marginal costs and their markups over average costs are also not equal.

The optimal solutions depend on the demand schedule. Holding the number of layers constant, firms with greater demand produce a larger quantity of the differentiated variety and set higher prices. From equation (4) and the fact that for a given number of layers marginal costs increase with quantity, it follows that $q_L(\alpha)$ is an increasing function of α. Prices also increase with marginal costs, and hence increase with α. Naturally, since quantities and prices increase with the level of demand, revenues are also an increasing function of α.

Firms with greater demand draws, also set greater markups. Holding the number of layers constant, an increase in demand has two opposing effects on markups. First for any given level of quantity produced firms charge a greater price, increasing both $\mu_L^{MC}(\alpha)$ and $\mu_L^{AC}(\alpha)$. Second firms' produce a greater quantity of the differentiated variety causing marginal costs to increase and averages costs to either increase or decrease. In all cases the first effect dominates, and both markups over marginal costs and markups over average costs increase with α. These results are summarized in the following proposition:

Proposition 2 *Holding the number of layers constant, $q_L(\alpha)$, $p_L(\alpha)$, $\mu_L^{MC}(\alpha)$, $\mu_L^{AC}(\alpha)$ and $r_L(\alpha)$, are increasing with respect to α. Moreover, $\pi_L(\alpha)$ is continuous, strictly increasing with respect to α and strictly concave in q.*

Proof. see appendix. ∎

Profits are continuous and increasing with respect to demand draws. For a given α, the slope of the profit function is increasing with the number of layers, and so there always exists an entrepreneur that is indifferent between producing with L or with $L+1$ layers.[12] When she increases the number of layers in the firm, marginal costs decrease discontinuously and the optimal quantity supplied by the firm increases discontinuously as well.[13] In contrast, prices decrease discontinuously when the number of layers is increased and therefore prices are not monotone with respect to α. Although prices decrease discontinuously, the increase in quantities outweighs the change in prices, and revenues increase discontinuously. Finally, when the entrepreneur increases the number of layers in the firm, the decrease in the marginal costs outweighs the decrease in prices, and markups increase discontinuously.

For a given α, let $q(\alpha)$, $p(\alpha)$, $\mu^{MC}(\alpha)$, $\mu^{AC}(\alpha)$, $r(\alpha)$ correspond to the global solution to the entrepreneur's maximization problem. Given her demand draw α, the entrepreneur maximizes profits and decides on the optimal number of layers in the firm. As α increases, the demand schedule changes and beyond a threshold, she restructures the firm and increases the number of layers. When the entrepreneur does not change the number of layers, $q(\alpha)$, $p(\alpha)$, $\mu^{MC}(\alpha)$, $\mu^{AC}(\alpha)$, $r(\alpha)$ have the properties of their corresponding local solutions. When the entrepreneur increases the number of layers, marginal costs decrease discontinuously, which affects the global solutions to the entrepreneur's maximization problem. The profit function $\pi(\alpha)$ is the upper envelope of the local profit functions $\pi_L(\alpha)$. It is continuous, strictly increasing in α and convex. These results are summarized in the following proposition:

[12] Formally, applying the envelope theorem to the profit function it follows that $\frac{\partial \pi_L(\alpha)}{\partial \alpha} = q_L(\alpha)$. Since for any α, $q_L(\alpha)$ is increasing with L it follows that the profit function has a steeper slope for a greater number of layers.

[13] Consider such a firm with demand draw $\alpha_{L,L+1}$. The optimal quantity is chosen so that marginal revenue equals marginal cost. Since for a given quantity, an organization with L layers has a greater marginal cost than an organization with $L+1$ layers, $q_L(\alpha_{L,L+1})$ cannot be the optimal quantity supplied by an organization with $L+1$ layers. It follows that $q_{L+1}(\alpha_{L,L+1}) > q_L(\alpha_{L,L+1})$, and at the optimal quantities $MC_L(q_L(\alpha_{L,L+1})) > MC_{L+1}(q_{L+1}(\alpha_{L,L+1}))$. And since revenues increase discontinuously, total costs are greater when producing with $L+1$ layers, i.e. $C_L(q_L(\alpha_{L,L+1})) < C_{L+1}(q_{L+1}(\alpha_{L,L+1}))$.

Proposition 3 *Holding the number of layers constant, $q(\alpha)$, $p(\alpha)$, $\mu^{MC}(\alpha)$, $\mu^{AC}(\alpha)$, and $r(\alpha)$, are increasing with respect to α. When a firm increases the number of layers $q(\alpha)$, $\mu^{MC}(\alpha)$, $\mu^{AC}(\alpha)$, and $r(\alpha)$ increase discontinuously, while $p(\alpha)$ decreases discontinuously. Moreover, $\pi(\alpha)$ is continuous, strictly increasing with respect to α and convex.*

Proof. see appendix. ■

2.4 Equilibrium

Prior to entering the differentiated goods sector, entrepreneurs purchase a demand draw α from a known cumulative distribution $G(\alpha)$ with support $[\alpha_M, \infty]$, at a fixed cost, f_E, in units of labor. If α_M is small there will always be demand draws that yield negative profits from operating a firm, and so the mass of entrepreneurs that exit will be positive. Therefore, given the mass of entrants in the market, M_E, a demand draw α_D exists such that all entrepreneurs with draw $\alpha < \alpha_D$ choose not to create a firm. For the marginal entrepreneur that is indifferent between entering and exiting the market, the profits of her firm are equal to zero. This yields the zero profit condition:

$$\pi(\alpha_D, M) = 0. \tag{10}$$

For the firm earning zero profit, equation (10) also implies that the price of a unit supplied is equal to average costs. The optimal quantity a firm with demand draw α_D produces is the solution to the equation:

$$q_D = \frac{N}{2\gamma}\left[\alpha_D - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) - MC(q_D)\right], \tag{11}$$

which provides an expression for the term $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$:

$$\frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) = \alpha_D - MC(q_D) - \frac{2\gamma}{N}q_D. \tag{12}$$

Substituting (12) back into (9) yields an expression for the operating profits of a firm, conditional on successful entry, as a function of the endogenously determined demand and quantity cutoffs, and the entrepreneur's demand draw:

$$\pi(\alpha, \alpha_D, q_D) = \left[\frac{2\gamma}{N} q_D + MC(q_D) - \alpha_D + \alpha - \frac{\gamma}{N} q(\alpha)\right] q(\alpha) - C(q(\alpha)). \tag{13}$$

Because the equilibrium involves solving for α_D and q_D, profits are also denoted as a function of the demand and quantity cutoffs. For the firm with demand draw α_D, equation (10) is equal to:

$$\begin{aligned} \pi(q_D) &= \frac{\gamma}{N} q_D^2 + MC(q_D) q_D - C(q_D), \\ &= 0. \end{aligned} \tag{14}$$

Equation (14) is independent of α_D. Therefore the solution to equation (14) is independent of the endogenously determined demand cutoff and the mass of firms operating in the differentiated goods sector.

The expected profits from entry, Π^e, is equal to the expected profits before entrepreneurs know their demand schedule minus the fixed entry cost. Unrestricted free entry implies that firms' expected profits, Π^e, are zero which yields the equilibrium free entry condition:

$$\int_{\alpha_D} \pi(\alpha, \alpha_D, q_D) dG(\alpha) = f_E. \tag{15}$$

Finally, the mass of firms operating in equilibrium is the mass of entrants with successful entry. The mass of entrants is therefore equal to: $M_E = M/(1 - G(\alpha_D))$.

The equilibrium is a set of values, q_D, α_D, and M, that solve equations (12), (14) and (15), given the parameters of the model and the distribution of demand draws $G(\alpha)$.[14] Unlike the parameters α and γ, in the differentiated goods sector η only affects the equilibrium of the model through the aggregate term. A condition on η ensures that consumers have a positive demand for the homogeneous good. I prove that:

[14] In the appendix, I also show how firms' optimal quantity, prices, revenues, and markups can be written as a function of α_D and q_D. By rewriting the model using equation (12), I am able to solve the model and obtain results with only limited assumptions on the distribution of demand draws $G(\alpha)$.

Proposition 4 *If $\eta > \underline{\eta}$, then there exists a unique equilibrium.*

Proof. see appendix. ■

2.5 Comparative Statics with respect to Market Size

The comparative statics of primary interest is how changes in the size of the market, N, affect firm outcomes and in particular their organization. From the implicit function theorem it follows:

$$\frac{\partial \alpha_D}{\partial N} = -\frac{\partial \Pi^e/\partial N + \partial \Pi^e/\partial q_D \; \partial q_D/\partial N}{\partial \Pi^e/\partial \alpha_D}. \tag{16}$$

In the appendix I show that the numerator of expression (16) is positive and the denominator is negative. An increase in the size of the market leads to an increase in the demand cutoff α_D. Intuitively, when N increases there are two effects. First there is the direct effect: holding the number of entrants fixed, when the size of the market increases firms increase their sales and profits. Second there is an indirect effect: increased profits for entrants implies that potential profits increase as well, raising the expected value of entry. To maintain the equilibrium condition that expected profits from entry are zero, the term $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$ rises, and under a restriction on the parameter η, M rises in order to lower expected profits.[15] In equilibrium this lowers the demand for firms' sales, increases competition and raises the bar for successful entry. Therefore, bigger markets induce tougher selection.[16] This result is summarized in the proposition below:

Proposition 5 *An increase in N induces an increase in q_D, α_D and $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$. In addition, if $\overline{\eta} > \eta$, then an increase in N induces an increase in M.*

Proof. see appendix. ■

[15] Otherwise it may be the case that $(\overline{\alpha} - \overline{p})$ rises and M decreases with N. Further in the appendix, I also show that there always exists an η in the set $\left[\underline{\eta}, \overline{\eta}\right]$. Finally, whether the number of varieties increases or decreases with the size of the market does not change the results below.

[16] Note that in bigger markets, the term $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$ increases, inducing a decrease in p_i^{max} and an increase in the elasticity of demand, ϵ_i. Further from equation (14), it follows that an increase in N also induces an increase in the quantity produced by the marginal firm, q_D. However this does not imply that firms with demand draw α_D increase their quantity in larger markets. As I have formulated the equilibrium of the model, the equilibrium consists of a mapping between the set of demand draws, α's, and the set of quantities, q's, that depends on N and the other parameters of the model, and amounts to solving for the lower bound of both sets. How this mapping changes along its support determines how firms' production changes with the size of the market. In fact the response to operating in a bigger market is heterogeneous across firms. Appendix C contains simulations illustrating the change in quantities, prices, markups, revenues and profits with N.

Now, consider an entrepreneur with a demand draw $\alpha_{L,L+1}$ that is indifferent between organizing production using L and $L+1$ layers. The distance between $\alpha_{L,L+1}$ and α_D decreases with N. There are two effects that explain this result. First, an increase in N raises the bar for successful entry, increasing α_D. Second, greater competition in bigger markets implies that markups are lower, reducing profits.[17] To increase profits, however, entrepreneurs can re-organize firms. An additional layer increases "fixed"costs, however it also decreases marginal costs and, if the firm's scale is sufficiently large, lowers average costs, enabling some firms to increase markups and profits. Therefore firms producing at a large scale will be re-organized and add additional layers.[18] This result is summarized below:

Proposition 6 *The distance between the marginal entrepreneur with demand draw, α_D, and the entrepreneur that is indifferent between two organizational forms, $\alpha_{L,L+1}$, is decreasing with respect to N. In addition, the change in the distance between α_D and $\alpha_{L,L+1}$, increases with L.*

Proof. see appendix. ■

Further let Λ_N represent the discrete cumulative distribution of organizations in the economy. Given the model parameters, the probability mass of firms producing with at most L layers is equal to:

$$\Lambda_N(L) = \frac{G(\alpha_{L,L+1}) - G(\alpha_D)}{1 - G(\alpha_D)}. \tag{17}$$

Now consider how the distribution changes with N. When the size of the market changes both the numerator and denominator in equation (17) are affected. Alone Proposition 6 does not provide enough information to stochastically order the distribution of organizations across markets of different sizes. If the distribution of demand draws has a non-increasing hazard rate, however,

[17] There is evidence that prices and markups are lower in bigger markets. In recent empirical work, using barcode data and correcting for biases in the measurement of the price index, Handbury and Weinstein (2014) show that prices of food are lower in bigger U.S. cities. In France, using balance sheet data Bellone et al. (2014) estimate firm-level markups and conclude that markups decrease with density.

[18] This result, however, does not imply that the $\alpha_{L,L+1}$'s decrease with the size of the market. When N increases the mapping between demand draws and quantities changes as well. What is important is their relative distance with respect to the demand draw of the marginal entrepreneur. Further the proof to Proposition 6 provides a lower and an upper bound to the change in $\alpha_{L,L+1}$ with respect to N. Both bounds may be positive or negative and depend on the quantity initially supplied and the parameters of the model. In addition, the proof to Proposition 6 shows that the quantity produced by the firm with L layers, $q_L(\alpha_{L,L+1})$ decreases with N, while the quantity produced by the firm with $L+1$ layers, $q_{L+1}(\alpha_{L,L+1})$ increases with N.

the cumulative distribution of organizations in bigger markets first-order stochastically dominates the distribution in smaller markets.[19] This result is summarized in the following proposition:

Proposition 7 *Suppose $N^{'} > N$. If the distribution of demand draws $G(\alpha)$ has a non-increasing hazard rate, then the cumulative distribution of organizations $\Lambda_{N'}$ first-order stochastically dominates Λ_N.*

Proof. see appendix. ■

Proposition 7 contains the main theoretical result of the paper. In larger markets there will be a greater mass of firms producing with a greater number of layers. It is important to note that first-order stochastic dominance is a strong characterization of the distribution of organizations across locations. More precisely, Proposition 7 implies that the distribution of organizations in larger markets will second-order stochastically dominate the distribution of organizations in smaller markets, as well as that firms in larger markets on average produce with a greater number of layers.

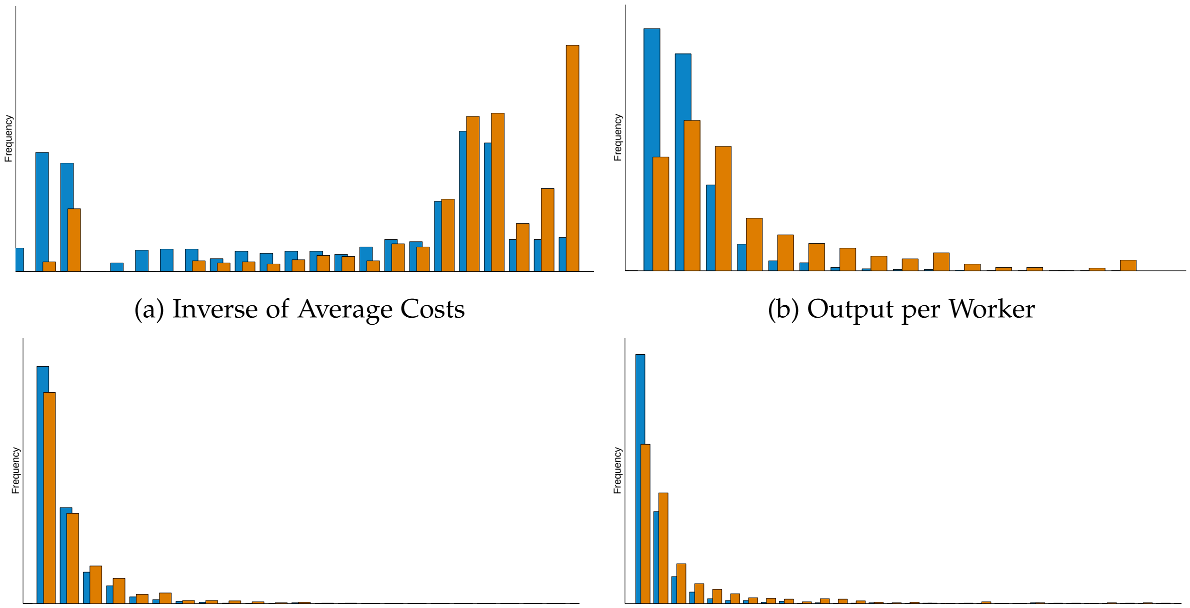
Furthermore, Proposition 7 also has implications for how the distributions of productivity, knowledge and incomes differ across markets. Firms that produce with a greater number of layers are more productive, and the knowledge workers acquire, as well as their income, depend on the organization of the firm they are employed in, as well as their position within it.

2.6 Numerical Simulations

This section uses numerical simulations to illustrate the model's implications on how the distributions of productivity, knowledge and income differ across markets of different size. The figures presented in this section are from two simulations of the model. In both simulations demand is drawn from a Pareto distribution, and the model parameters are identical in both simulations except for the size of the market. Distributions with blue bars are from the smaller market. A complete description of the simulations and additional results are presented in Appendix C.

Figure 2 illustrates how productivity varies across markets. Figures 2a and 2b contain measures of quantity-based productivity, that is productivity measured using firm output, whereas

[19] The Pareto distribution, regularly used in the heterogeneous firm literature, satisfies the non-increasing hazard rate property. Further, this assumption is less binding than in standard heterogeneous firm models. The distribution of demand draws is a deep parameter of the model, and because firms optimally choose their organization, the distributions of productivity and firm size do not necessarily inherit the non-increasing hazard rate property. Further the intuition for non-increasing hazard rate assumptions is following: holding L fixed, it ensures that the mass that is lost from firms changing their organization and producing with $L^{'} \neq L$ is greater than the mass that is gained from firms that change their organization and now produce with L layers.



(a) Inverse of Average Costs

(b) Output per Worker

(c) Revenue per Labor Costs

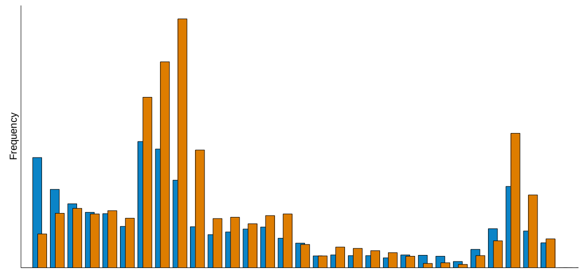
(d) Revenue per Worker

Figure 2: Measures of Productivity

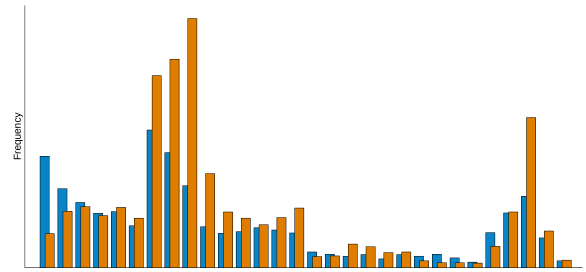
figures 10a and 10b contain measures of revenue-based productivity. All figures show there are different shares of small, medium and high productivity firms across markets. The larger market contains a smaller fraction of low and medium productivity firms and a greater mass of high productivity firms. All measures therefore lead to the same conclusion: in bigger markets firms are on average more productive.

Figure 3 illustrates the distributions of knowledge and income across markets. The larger market has a greater share of agents with intermediate and high levels of knowledge. This is again due to a greater share of firms producing with a greater number of layers. When firms add an additional layer of managers to their organization, the knowledge of workers in the existing layers decreases. Yet at the same time, firms employ more intermediate managers and so there are more agents with intermediate levels of knowledge in the economy. In the aggregate the second effect dominates and the mass of agents with low levels of knowledge is reduced in the bigger market. Panel (b) presents the distributions of income. Results are similar, which is to be expected because workers are compensated for their time and the knowledge they acquire in the firm. To summarize, both figures lead to the conclusion that in bigger markets workers are on average more skilled and earn higher incomes.

It is important to note these findings are robust. They are a direct result of the production framework of the model, and the effect a tougher competitive environment has on firms' or-



(a) Knowledge



(b) Income

Figure 3: Distributions of Knowledge & Income

ganization. Firms producing with a greater number of layers, organize their workforce more efficiently, reduce their average costs and consequently hire more skilled workers and are more productive.

This study therefore establishes a new mechanism to explain the differences in the productivity of firms, worker skills and worker wages across markets and locations. Namely, firms in larger markets are more productive because they organize production with a greater number of layers. Similarly in larger markets workers are more skilled and earn a greater income because they are employed in firms with a greater number of layers. The following sections examine the models' predictions and how important this new mechanism is to understanding the productivity differences of firms across locations.

3 Data and Identification

The empirical analysis of the paper is conducted in the year 2004 on French firms that operate in industries where demand is determined at a local level. The next subsections describe the data and discuss several identification concerns.

3.1 Data Sources

The data in this study are from three French sources. The first is the Déclarations Annuelles des Données Sociales (DADS) that is built from mandatory reports collected by the French National Statistical Institute for Statistics and Economic Studies (INSEE). The DADS is an annual matched employer-employee dataset, it is nearly exhaustive and contains information on all workers who earn a salary in France. For each employee there is information on the total number of hours worked during the year, his annual salary and occupation, as well as the establishment and firm of employment. For each establishment there is information on the geographical area and industry of operation, as well as the industry of the parent firm.

The second source is the Fichier Complet Unifié de Suse (FICUS) which is provided by INSEE. The FICUS dataset is built from mandatory fiscal declarations, it is nearly exhaustive, and contains annual balance sheet information on firms, such as total sales, material assets, and value added, as well as information on their legal structure, such as their legal status and whether they belong to a business group.

And, the third source is the Recensement de la Population (RP) which is a septennial census of the French population and is also provided by INSEE. For the year 1999, the RP is exhaustive and contains demographic information on all individuals and households in France.

3.2 Firm Organization and Productivity

To construct the organization of firms I use the method developed by Caliendo et al. (2015b) and classify employees into organizational layers using the first-digit of occupational codes from the DADS.[20] A unique feature of the French occupational codes is that they preserve a ranking between workers in firms. Occupational code 2 contains workers with the highest level of authority and is composed of owners who receive a salary from the firm and CEOs, occupational code 3 contains senior managers and directors, occupational code 4 contains supervisors and in general employees with a higher level of responsibility than ordinary workers, and occupational codes 5 and 6, which are grouped together to form one category, contain ordinary workers. With these occupational categories a firm can be divided into as many as four distinct layers. Further to measure the total number of layers in firms I follow Caliendo et al. (2015b) and Caliendo et al. (2015a) and only focus on the relative ranking between occupations. In other words, a firm reporting ℓ occupational categories will have ℓ layers. For example, a firm with workers in occupations 2,3 and 4 is a three-layer firm. Not every firm has an employee in every occupational category however, and so there are four types of organizations in the data: one-layer, two-layer, three-layer and four-layer firms.

To measure the productivity of firms I combine balance sheet data from FICUS with measures on the size and wage bill of firms from the DADS. With this information I construct different

[20] In the knowledge-based management theory of firms a layer represents a group of workers with similar wages, who have similar skills and perform tasks at a similar level of authority (Caliendo et al. (2015b)). Further Caliendo et al. (2015b) provide evidence for the manufacturing sector that this classification is meaningful and consistent with the concept of layer in the knowledge-based management theory of firms. A section below provides similar evidence.

measures of revenue-based productivity. One measure of labor productivity, is directly calculated from the data and is simply value-added per worker. The other measures of firm productivity are Total Factor Productivity (TFP) estimated using the approaches proposed by Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015),Wooldridge (2009) and Caliendo et al. (2015a). Additional details are provided in Section 4.2.

3.3 Data Restrictions

I impose several restrictions on the data to properly examine the model's implications. First, the model is silent on whether a firm can have one or many establishments and to be consistent with the theory, I aggregate the data to the level of the firm. Second, often in very small firms although the owner works in the firm, he does not disclose a salary, and so he does not appear in the DADS (Bacheré (2015)). This creates measurement error in my baseline estimates, biasing results, and to deal with this issue I only retain firms with at least 8 employees.[21]

And third, the theory is about single product firms operating in a monopolistically competitive industry and competing in a local market. For this reason I restrict the analysis to firms that only operate in one local market and in one of the following industries: Clothes and Shoes Retail, Traditional Restaurants, and Hair and Beauty Salons. The model's general assumptions are applicable to these industries. The sectors examined in this study are non-tradeable sectors with relatively local demand. The assumption of monopolistic competition also applies reasonably well to these sectors. Even in small markets there are many firms operating in these sectors, each firm provides a differentiated product, and because of their location and quality of service each firm retains some market power.

3.4 Definition of Local Markets

The theory relates market size to organization. One difficulty with examining the theory's implications is to properly identify local markets. In this paper, I separate France into local markets using two geographical decompositions of mainland France: employment areas (*zones d'emploi*)

[21]More precisely, when I examine the second implication of the model, there will be a downward bias in my estimates because firms with unreported workers will have a higher productivity and less layers. This problem cannot be corrected using information from FICUS. In addition, in the French data, value added per worker is higher in firms with a single employee than in other small firms, decreases as the number of employees increases and it tends to stabilize at 4 employees (see Bacheré (2015) for details). In the sectors examined in this study, value-added per worker tends to stabilize at about 5 employees and, value-added per hour also exhibits the same pattern and tends to stabilize at about 8 employees. Furthermore, in the Appendix I take a more strict interpretation of the model, and retain firms all firms with employees in adjacent layers starting with ordinary workers. The results remain the same.

and urban areas (*unité urbaines*). France's decomposition into employment areas is based on local labor markets and each municipality (*commune*) within France belongs to a single employment area. Employment areas are travel to work areas and are composed of geographical spaces in which most of the inhabitants reside and work within the area. Most employment areas correspond to a city and its surrounding area, or to a metropolitan area. In contrast, urban areas are geographical regions with at least 2,000 inhabitants over spaces not separated by more than 200 meters. Urban areas are contiguously built-up spaces, correspond to small cities and their suburban areas, and do not cover all of France (see Briant et al. (2010) and Combes et al. (2008) for further information on geographical decompositions of mainland France).

Both geographical decompositions provide a reasonable definition of local markets. It is sensible to assume that most individuals eat, shop, and visit salons within their area of residence and employment, and so the relevant definition of a local market is employment areas. To the extent that this definition of a market is too broad, and people perform these activities within the city they reside in, defining markets at the urban level is more appropriate. As shown below defining local markets using either geographical decomposition leads to the same conclusions. An alternative concern is that both definitions are too broad and these markets encompass several submarkets. This problem should be of minor concern because in regressions both the dependent and independent variables are averages and are affected in the same way from defining markets too broadly (see Briant et al. (2010)).

A second difficulty is measuring the size of local markets. Within a geographical area the theory abstracts from any spatial dimensions and implicitly assumes that varieties are accessible to all consumers. In the theory market size however, determines the local level of competition. For this reason I measure the size of markets using density in the year 1999, which is defined as the total population residing in an area divided by the area's surface, measured in hectares (1 hectare equals 0.01 kilometers squared). Density is the more suitable measure in this setting, because it accounts for the level of concentration of economic activity and contains more appropriate information on the level of competition in local markets.[22]

[22]Many researchers have related density to the size of markets, such as Ciccone and Hall (1996), Cambpell and Hopenhayn (2005), Sato et al. (2012) and Combes et al. (2012).

3.5 Identification and Instrumental Variables

The theory is static and involves comparisons of firms across markets. To examine the theory's implications the empirical analysis of the paper is conducted in the year 2004 and mainly relies on cross-sectional variation in the size of local markets. I now discuss additional identification concerns.

The theory is about the effect market size has on firm organization and productivity. The empirical analysis of the paper mainly relies on estimating equations of the following form, at some level ℓ, in the year 2004:

$$Y_{\ell,a} = \alpha + \gamma \log density_a + X_{\ell,a}\beta + \epsilon_{\ell,a}, \tag{18}$$

where $Y_{\ell,a}$ is a variable based on the characteristics of firms, $density_a$ is the local density of an area a in the year 1999, and $X_{\ell,a}$ contains additional controls. Depending on the implication being examined, equation (18) is estimated at the market-industry level or at the level of firms.

One concern with the analysis in equation (18) is that density is endogenous. For instance, firms producing with a greater number of layers have more jobs to fill and so may attract more workers to a local market. Alternatively, there may be local shocks that are not accounted for in the estimation, determining both market size and firm outcomes. Following Ciccone and Hall (1996), Combes et al. (2008) and Combes et al. (2010) I instrument for the size of local markets using historical measures of the log of density from as far back as 1831. This strategy is defensible as long as the there is some persistence in the spatial distribution of the French population, and the local determinants of organization today are not related to the size of local markets in the past (Combes et al. (2008), Combes et al. (2010)).[23] To the extent that changes in technology, in the methods of production and changes in the theory of management over the years have changed the structure of firms, using very long lags makes this hypothesis plausible (see Combes et al. (2010) for a more detailed discussion on changes in the French economy since 1831).

Furthermore the theory has several restrictive assumptions which must be addressed when examining its empirical implications. An important assumption of the theory is a homogeneous goods sector that determines the wage of a unit of labor. This assumption implies the sectors included in the analysis have to be small relative to local markets, in the sense that firms do not employ a large fraction of the local labor force. Unlike most industries, the sectors in this study

[23] In other words, given a set of exogenous controls X_a, an instrument Z must satisfy the following two conditions: 1- it must be relevant, i.e. $cov(\log density_a, Z_a|X_a) \neq 0$ and 2- it must be exogenous, i.e. $cov(\epsilon_{\ell,a}, Z_a) = 0$.

satisfy this condition. Across employment areas (urban areas) industries on average employ 0.6 (0.9) percent of the local labor and at the 99*th* percentile the share of workers that are employed in an industry is 4.7 (9.97) percent. This condition also has implications for my identification strategy. Any controls of the characteristics of local markets are unlikely to be determined by the sectors examined, because they amount to a small fraction of the local level of economic activity.

Even though firms should not be able to determine the local wage of a unit of labor, changes in local wages affect the organization of firms. First, local wages determine the cost of hiring workers and consequently determine the cost schedule faced by firms from operating in a market.[24] In some specifications I therefore control for the extent that wages and incomes vary across local markets. To calculate the local cost of hiring workers I use wages in the other sectors of a local market. According to the model worker earnings depend on organization. It is most likely the theory applies equally well to other sectors, and local wages do not reflect the true cost of a unit of labor because they depend on the organization of firms. One way to address this issue is to estimate area fixed effects from a Mincerian regression that controls for skills, occupations, and location.[25] These fixed effects do not contain the returns to skills, nor do they reflect information on workers' position within firms, and so are a suitable measure of the local cost of a unit of labor.

Second, wages also determine incomes which govern consumer expenditures and the level of demand.[26] In some specifications I therefore control for local incomes. From the DADS I calculate the median annual salary from the set of individuals who reside in a local market and are employed in sectors not examined in this study. Because annual salary depends on skills, the number of hours worked in a year, and some individuals are not employed in their area of residence, this variable is not collinear with the cost of a unit of labor.

A second assumption of the theory is that preferences are identical across markets. This

[24]Given its level of demand, the objective of firms is to organize production efficiently. For a given level of output and holding the level of demand fixed, if a unit of labor is costlier in denser markets, some firms will be able to reduce their costs by simultaneously reducing the knowledge of employees in the lower layers of the organization and by increasing the number of managerial layers.

[25]To obtain an estimate of average hourly wages excluding skills I do the following. For the years 1998-2000 I keep all male workers born in October and not working in one of the industries examined in this study, and retain the estimated area fixed effects from the following equation:

$$\ln wage_{it} = \alpha + x_{it}\beta + occ_i + ind_j + g_a + \epsilon_i,$$

where $wage_{it}$ is the hourly wage of worker i in year t, x_{it} contains a quintic polynomial of a worker's age and time fixed effects, ind_j are industry fixed effects, occ_i are occupation fixed effects, and g_a are area fixed effects.

[26]Higher wages imply higher incomes. Holding the cost of labor fixed, a model with a higher income per consumer is isomorphic to a model with a larger market and a unit of income.

assumption is important when trying to identify the relationship between the size of local markets and the organization of firms. In the theory preferences can have a direct effect on firms, and Appendix A.3 shows that changes in preferences can yield similar economic outcomes to changes in the size of markets.[27] Furthermore in urban economics it is a well-known hypothesis that agents sort into locations in part because of their preferences (see for example Holmes and Sieg (2015) and Rosenthal and Ross (2015)). It is also well-known that young individuals, as well as immigrants live in denser markets, and to the extent that their consumption habits are different, changes in density may correlate with changes in consumer preferences. For this reason, in my analysis I control for the demographic composition of local markets. To account for variation in consumer preferences across markets I include the following demographic variables, constructed from the RP data in the year 1999: the share of the local population between the ages 25 and 59, the unemployment rate among local active individuals, and the share of the local population born outside of France.

To isolate the effect of market size on firm organization and productivity, both assumptions imply that additional variables are required in my estimating equations. This raises a concern with my estimation strategy because these additional variables may also be endogenous. Because the sectors examined in this study are small relative to the economic activity of local markets, it is unlikely that they determine local wages and the demographic composition of local markets. Simultaneity bias should therefore not be an issue of concern. Nonetheless, local shocks may determine both groups of variables and the outcome of firms. For this reason, I measure the additional variables in the year 1999. Insofar as these shocks are of short duration and not correlated over time, these variables can be treated as exogenous. However, because the nature of the shocks and their persistence is unknown, I also instrument for the additional variables using lagged values.

To summarize, the empirical analysis of this paper is conducted in the year 2004, and focuses on firms with at least 8 workers, operating in one monopolistically competitive industry and competing in one local market. In my empirical analysis I rely on cross-sectional variation in the

[27] Appendix A.3 shows that given a market of size, N, any new equilibrium that is generated by a proportional increase in the size of the market is isomorphic to an equilibrium with a market of the original size, N, but with different preferences. Furthermore both assumptions imply that additional variables are required in my estimating equations. From the perspective of the model, including controls for preferences, incomes and the cost of a unit of labor, are justified. Nonetheless, it is also very likely that a general formulation of the model with heterogeneous agents and an endogenous location choice, would imply that density is the fundamental parameter determining the characteristics of local markets. Taking a position on this issue, however, is beyond this paper, and for this reason, in the empirical section I consistently present results with and without controls for the characteristics of local markets.

Table 1: General Summary Statistics

	Employment Areas	Urban Areas
Number of Areas	341	1,289
Average Density	2.63	3.02
Median Density	0.743	2.13
St. Dev. of Density	12.47	2.93
Total Number of Firms	27,508	24,197
Clothes and Shoes Retail	2,879	2,819
Traditional Restaurants	20,620	17,605
Hair and Beauty Salons	4,009	3,773
Average Number of Firms per Area	80.66	18.77
Median Number of Firms per Area	33	3
St. Dev. of Number of Firms per Area	221.60	186.46
Corr. B/T Density and Number of Firms	0.863	0.421
Average Number of Establishments Per Firm	1.09	1.09
Median Number of Establishments Per Firm	1	1
St. Dev. of Number of Establishments Per Firm	0.410	0.434
Corr. B/T Density and Local Market Controls		
Cost of a Unit of Labor	0.043	0.055
Median Annual Salary	0.288	0.062
Share of Population Between 25 and 59	0.336	0.161
Share of Population Born Outside France	0.362	0.205
Share of Population Unemployed	0.030	0.197

Notes: Summary statistics across local markets.

size of local markets and include a set of controls to account for any other local factors that can also determine the number of layers and the productivity of firms. To deal with the endogeneity of market size, I use historical data to instrument for the density of local markets. Appendix B.1 contains additional details on the construction of the data and the variables used in this study. The next subsections present descriptive statistics from the sample.

3.6 Descriptive Statistics

Table 1 presents some general statistics of the data. The data cover 341 employment areas and 1,289 urban areas in France. Across all employment areas (urban areas) there are 27,508 (24,197) firms, the average density of an area is 2.63 (3.02) inhabitants per hectare squared, and each area contains on average 80.66 (18.77) firms. The correlation between density and number of firms is 0.863 (0.421) across employment areas (urban areas). The correlations between density and the characteristics of local markets are positive as well. In both samples, the majority of firms operate

Table 2: Summary Statistics of Organizations

	Number of Firms	Average Value Added	Average Number of Workers	Average Number of Hours	Average Hourly Wage
Total Number of Layers					
One-Layers	10,051	149.74	12.53	8,609	6.86
Two-Layers	12,509	232.33	17.28	12,922	7.79
Three-Layers	4,808	486.61	29.54	22,504	9.27
Four-Layers	590	888.29	45.35	36,678	10.66

Notes: Summary statistics of firms.

Table 3: Summary Statistics of Organizations

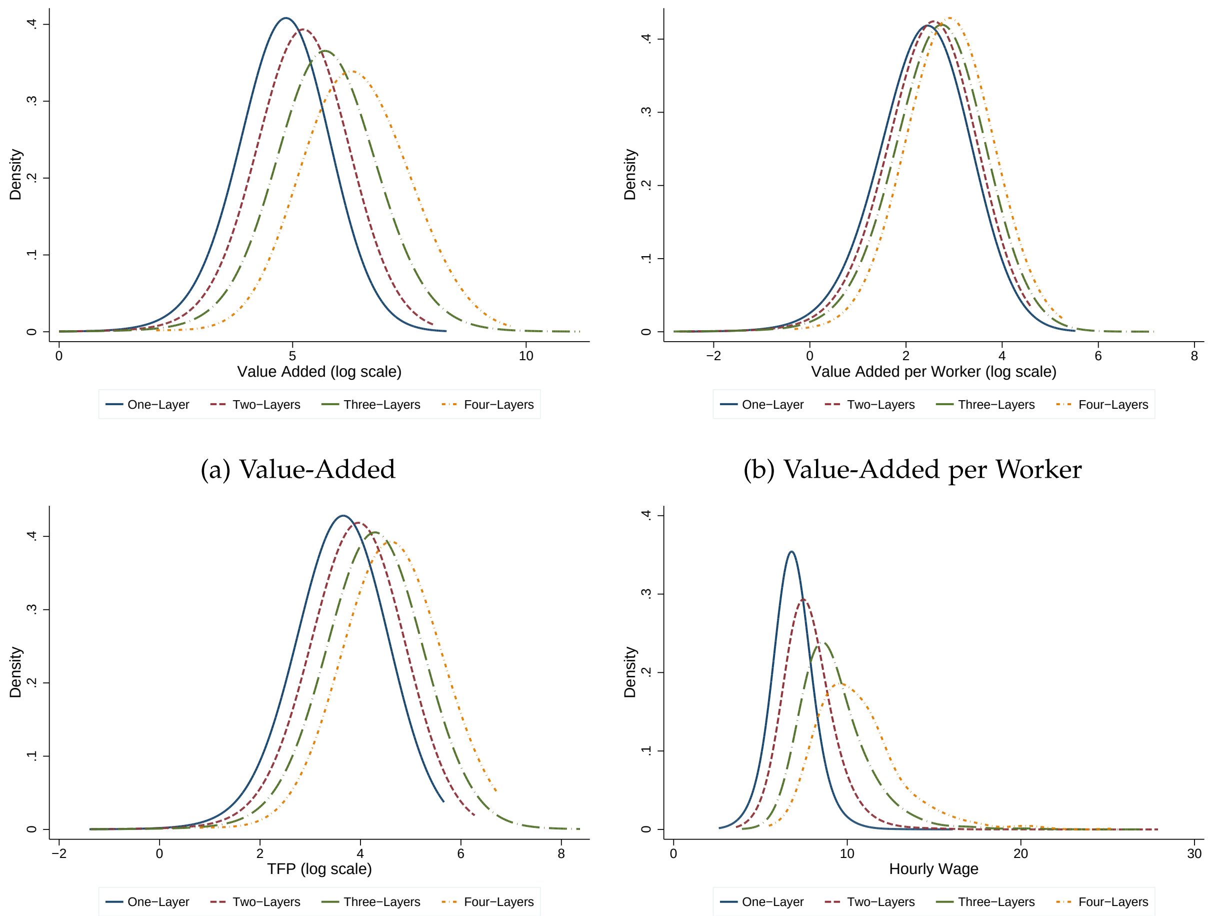
	$N_L^l \leq N_L^{l-1}, \forall l$	$N_L^2 \leq N_L^1$	$N_L^3 \leq N_L^2$	$N_L^4 \leq N_L^3$
Total Number of Layers				
Two-Layers	0.989	0.989	–	–
Three-Layers	0.857	0.962	0.894	–
Four-Layers	0.728	0.922	0.828	0.967
	$w_L^{l-1} \leq w_L^l, \forall l$	$w_L^1 \leq w_L^2$	$w_L^2 \leq w_L^3$	$w_L^3 \leq w_L^4$
Total Number of Layers				
Two-Layers	0.795	0.795	–	–
Three-Layers	0.649	0.931	0.712	–
Four-Layers	0.618	0.971	0.876	0.755

Notes: Summary statistics of firms.

in the Traditional Restaurants sector which accounts for 75 percent of firms across employment areas and 73 percent of firms across urban areas. The samples in the data have a different number of firms because while all geographical areas of mainland France belong to an employment area, urban areas do not contain every region within France. For simplicity, I will henceforth refer to the sample where local markets are defined using employment areas as Sample A, and the sample where markets are defined using urban areas as Sample B.

3.6.1 Descriptive Statistics: Firm Organization

Table 2 groups firms in Sample A by their number of layers, and compares observable characteristics. Statistics are presented for Sample A only, because in Sample A the geographical decomposition of France is exhaustive. Firms producing with a greater number of layers are larger, in terms of value-added, number of workers, and number of hours, and pay higher hourly wages. For example, the average value-added of four-layer firms is 888.29 euros, while the average value-added of one-layer firms is 149.74 euros. Figures 4a, 4b, 4c and 4d further plot the kernel density of value-added, value-added per worker, TFP estimated using the method by Levinsohn and Petrin (2003) and workers, and hourly wages. In all three figures there is a ranking of dis-



(a) Value-Added

(b) Value-Added per Worker

(c) TFP

(d) Hourly Wage

Figure 4: Kernel Density Distributions Across Organizations

tributions with firms operating with a greater number of layers being larger and paying higher wages. This ranking is consistent with this paper's model and more generally the management hierarchy models of Garicano (2000) and Caliendo and Rossi-Hansberg (2012).

Table 4: Summary Statistics Across Geographical Areas

	Number of Firms	Average Number of Layers	Average Value Added	Average Number of Workers	Average Number of Hours	Average Hourly Wage
Sample A: Employment Areas						
All	27,508	1.85	260.66	18.29	13,531	7.77
Below Median	4,428	1.66	186.20	15.54	10,762	7.25
Above Median	23,080	1.88	274.95	18.82	14,062	7.87
Sample B: Urban Areas						
All	24,197	1.88	274.62	18.76	14,198	7.84
Below Median	2,054	1.63	215.30	14.89	9,706	7.40
Above Median	22,143	1.90	280.12	19.12	14,615	7.88

Notes: Summary statistics of firms across local markets.

3.6.2 Descriptive Statistics: Firms Across Locations

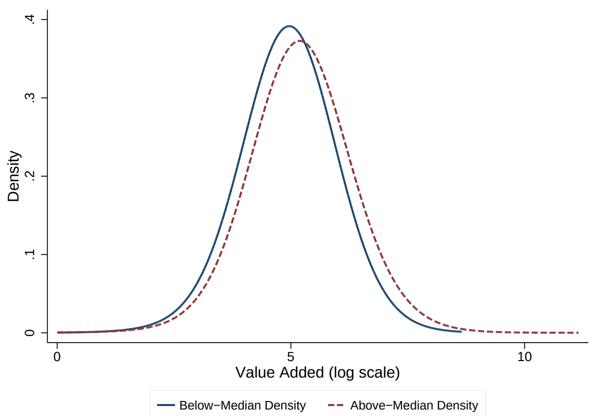
Table 4 groups local markets by their density and reports summary statistics across the different halves of the density distribution. The evidence shows that firms in denser markets have more organizational layers, are larger in terms of value-added, number of workers and number of hours worked, and pay higher wages. For example in Sample A (Sample B), the average number of layers in firms located in markets with below-median density is 1.66 (1.63) while the average number of layers in firms operating in markets with above-median density is 1.88 (1.90). This evidence is consistent with model.

Figures 5a, 5b, 5c, and 5d plot the kernel density of value-added, two measures for firm productivity, value-added per worker and TFP estimated using the method by Levinsohn and Petrin (2003) and workers, as well hourly wages, separately for firms operating in employment areas with below and above median density. There is a clear ranking of distributions with firms operating in denser markets being more productive and workers employed in denser markets earning higher wages. These findings are consistent with the empirical literature examining firm productivity and worker earnings across locations (see for example Combes and Gobillon (2015) and the references therein).

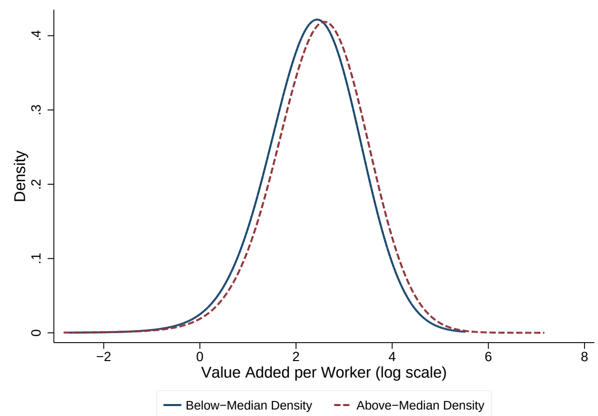
4 Empirical Analysis

4.1 Firm Organizations Across Locations

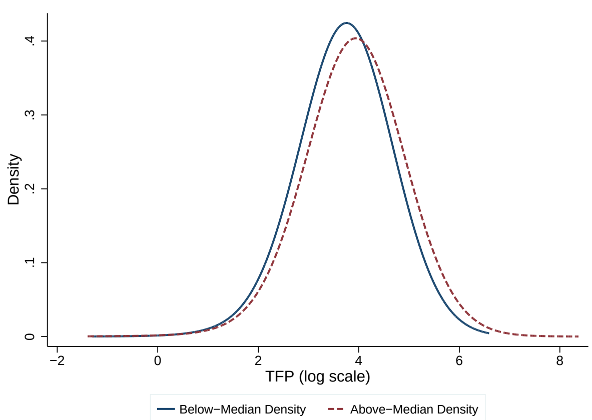
This section examines the first main prediction of the model, in denser markets firms organize with a greater number of layers. The analysis proceeds in stages. First, I group local markets by



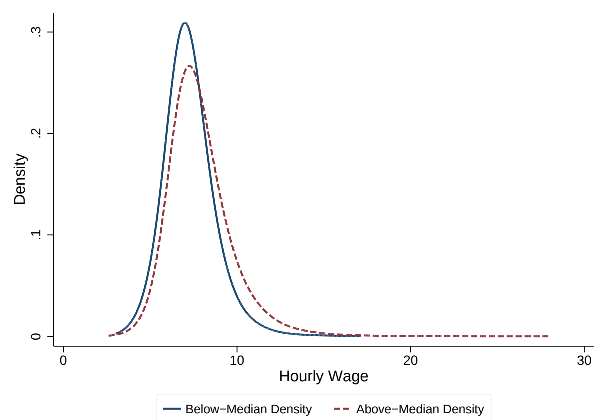
(a) Value-Added



(b) Value-Added per Worker



(c) TFP



(d) Hourly Wage

Figure 5: Kernel Density Distributions Across Markets

Table 5: Distribution of Organizations Across Geographical Areas

	Number of Firms	One Layers	Two Layers	Three Layers	Four Layers
Sample A: Employment Areas					
All	27,508	0.365	0.438	0.174	0.021
Below Median	4,428	0.464	0.412	0.114	0.009
Above Median	23,080	0.346	0.443	0.186	0.023
Sample B: Urban Areas					
All	24,197	0.347	0.443	0.184	0.023
Below Median	2,054	0.474	0.422	0.097	0.005
Above Median	22,143	0.335	0.445	0.192	0.025

Notes: Distribution of organizations across local markets.

Table 6: Mann-Whitney Distribution Tests

	Sample A: Employment Areas		Sample B: Urban Areas	
	Null Hypothesis: Distributions are Equal	Probability Above Median > Below Median	Null Hypothesis: Distributions are Equal	Probability Above Median > Below Median
ALL Industries	0.000	0.575	0.000	0.593
Clothes and Shoes Retail	0.001	0.547	0.264	0.527
Traditional Restaurants	0.000	0.578	0.000	0.598
Hair and Beauty Salons	0.000	0.585	0.000	0.595

Notes: Results of Mann-Whitney stochastic dominance test. The null hypothesis is that both distributions are equal. Columns 1 and 3 report the p-values of the test. Columns 2 and 4 reports the probability that a random draw of an organization from an areas with below-median density is greater than a random draw of an organization from an areas with above-median density. Industries are grouped together based on their 3-digit Nace Rev 1.1 classification code.

their density and test non-parametrically whether the cumulative distribution of organizations in denser markets first-order stochastically dominates the distribution of organizations in less dense markets. Second, I use regression analysis to examine the distribution of organizations, controlling for the characteristics of local markets and considering possible endogeneity problems between market size and organization. And third, I use regression analysis to examine the relationship between organization and the size of local markets at the firm level, controlling for the characteristics of firms, the characteristics of local markets and addressing endogeneity problems.

4.1.1 Ordinal Stochastic Dominance Tests

Table 5 compares the distribution of organizations across the different halves of the density distribution. In denser markets a greater share of firms produce with a greater number of layers. For example in Sample A, the top half of the distribution has a 1.4 percent greater share of four-layer firms, an 7.2 percent greater share of three-layer firms and a 3.1 percent greater share of two-layer firms relative to the bottom half of the density distribution. Results from Sample B are similar.

Table 6 further compares distributions and examines whether the cumulative distribution of organization in denser markets first-order stochastically dominates the distribution in less dense markets, using the Mann-Whitney test.[28] The underlying hypothesis of the test is both distributions are the same, while the alternative is one distribution has systematically larger values than the other. Columns 1 and 2 in Table 6 compare the distribution of organizations between employment areas with above-median and below-median density, while Columns 3 and 4 compare the distribution of organizations across urban areas. Columns 1 and 3 report the p-values of the tests. In nearly every case the null hypothesis is rejected at conventional levels of significance.[29] Columns 2 and 4 contain the probability that a firm chosen at random from an area with above-median density has a greater number of layers than a random firm from an area with below-median density. The results indicate that the distribution of organizations in denser markets first-order stochastically dominates the distribution in less dense markets. For example when grouping all industries together, the Mann-Whitney test indicates that a random firm from an employment area with above-median density is 57.5 percent more likely to have a greater number of layers than a random firm from an employment area with below-median density. To the extent that firms with a greater number of layers are also more productive, these results imply tangible differences in the productivity of firms across locations.

[28]The Mann-Whitney test relies on three assumptions: 1- both samples are random, 2- across and within samples observations are independent of one another, and 3- the response variable is ordinal.

[29]Moreover, across urban areas in the Clothes and Shoes Retail sector the p-value of the test statistics is greater than 0.10 and the Mann-Whitney test fails to reject the null hypothesis. In fact, because there are fewer firms in this sector and the sample of firms in both groups is very unequal, the power of the Mann-Whitney test is lower in the Clothes and Shoes Retail sector (see Zimmerman (1987)). Nevertheless, the probabilities reported in Table 6 for the Clothes and Shoes Retail sector are comparable to the other sectors, suggesting that a random firm from a denser market is more likely to have a greater number of layers.

Table 7: Regressions Results Across Employment Areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep: Percent One-Layer Firms										
log density	-0.050	-0.050	-0.029	-0.035	-0.021	-0.052	-0.045	-0.024	-0.015	-0.022
	$(0.006)^a$	$(0.006)^a$	$(0.006)^a$	$(0.008)^a$	$(0.008)^b$	$(0.004)^a$	$(0.005)^a$	$(0.007)^a$	$(0.008)^c$	$(0.009)^b$
Dep: Percent Two-Layer Firms										
log density	0.024	0.025	0.016	0.014	0.016	0.026	0.020	0.014	0.006	0.018
	$(0.006)^a$	$(0.006)^a$	$(0.007)^b$	$(0.008)^c$	$(0.009)^c$	$(0.005)^a$	$(0.005)^a$	$(0.007)^b$	(0.008)	$(0.009)^b$
Dep: Percent Three-Layer Firms										
log density	0.022	0.022	0.012	0.020	0.006	0.020	0.020	0.006	0.006	0.004
	$(0.005)^a$	$(0.005)^a$	$(0.005)^b$	$(0.006)^a$	(0.007)	$(0.003)^a$	$(0.004)^a$	(0.005)	(0.005)	(0.008)
Dep: Percent Four-Layer Firms										
log density	0.003	0.002	0.000	0.000	-0.001	0.005	0.004	0.002	0.001	-0.000
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	$(0.001)^a$	$(0.001)^a$	(0.001)	(0.001)	(0.002)
Sample	A	A	A	A	A	A	A	A	A	A
Method	OLS	OLS	OLS	OLS	OLS	OLS	2SLS	OLS	2SLS	2SLS
First-Stage Statistics										
Partial R-squared	-	-	-	-	-	-	0.747	-	0.734	0.206
SW F-Statistic	-	-	-	-	-	-	198.02	-	110.49	9.86
KP Wald F-Statistic	-	-	-	-	-	-	198.02	-	110.49	5.20
Over-Identification Test (p-value)										
Percent One-Layer Firms	-	-	-	-	-	-	0.612	-	0.134	0.143
Percent Two-Layer Firms	-	-	-	-	-	-	0.845	-	0.279	0.320
Percent Three-Layer Firms	-	-	-	-	-	-	0.648	-	0.346	0.698
Percent Four-Layer Firms	-	-	-	-	-	-	0.104	-	0.075	0.269
R-squared										
Percent One-Layer Firms	0.311	0.312	0.323	0.316	0.325	0.473	-	0.506	-	-
Percent Two-Layer Firms	0.061	0.065	0.063	0.065	0.069	0.121	-	0.138	-	-
Percent Three-Layer Firms	0.152	0.153	0.157	0.155	0.167	0.409	-	0.430	-	-
Percent Four-Layer Firms	0.039	0.040	0.041	0.042	0.044	0.271	-	0.294	-	-
Wage Controls*	No	Yes	No	No	Yes	No	No	Yes	Yes	Yes
Income Controls	No	No	Yes	No	Yes	No	No	Yes	Yes	Yes
Demographic Controls	No	No	No	Yes	Yes	No	No	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	949	949	949	949	949	408	408	408	408	408

*Notes: a,b,c: significant at the 1%, 5% and 10% level. *: indicates variables are treated as exogenous in regressions. Clustered standard errors at the employment area level are reported in parentheses. This table reports regression results for equation (19). Each column displays the estimate from a separate regression. Density measures the local density of an employment area. Industry fixed effects are at the three-digit Nace Rev 1.1. level. Wage controls contain the cost of a unit of labor. Income controls contain the median the annual salary of individuals residing in an employment area. Demographic controls contain the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. Column (7) only instrument for density using the log of density measured in 1831 and 1881. Column (9) only instrument for density using the log of density measured in 1831, 1851, 1881 and 1901. Column (10) instruments for density and local characteristics using the following variables: the log of density measured in 1831, 1851, 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, and the average person per household in 1949.*

Table 8: Robustness Checks Across Employment Areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep: Percent One-Layer Firms										
log density	−0.037	−0.015	−0.037	−0.023	−0.031	−0.016	−0.032	−0.019	−0.047	−0.019
	(0.005)^a	(0.009)^c	(0.005)^a	(0.009)^b	(0.005)^a	(0.009)^c	(0.005)^a	(0.008)^b	(0.005)^a	(0.009)^b
Dep: Percent Two-Layer Firms										
log density	0.019	0.017	0.014	0.019	0.015	0.016	0.013	0.017	0.014	0.009
	(0.005)^a	(0.009)^b	(0.005)^a	(0.008)^b	(0.006)^b	(0.009)^c	(0.006)^b	(0.009)^c	(0.006)^b	(0.009)
Dep: Percent Three-Layer Firms										
log density	0.013	−0.001	0.018	0.004	0.012	0.000	0.014	0.002	0.026	0.011
	(0.004)^a	(0.007)	(0.004)^a	(0.008)	(0.004)^a	(0.007)	(0.004)^a	(0.007)	(0.003)^a	(0.006)^c
Dep: Percent Four-Layer Firms										
log density	0.004	−0.000	0.004	−0.000	0.003	−0.000	0.004	−0.000	0.005	−0.000
	(0.001)^a	(0.002)	(0.001)^b	(0.002)	(0.001)^a	(0.002)	(0.001)^a	(0.002)	(0.001)^a	(0.002)
Sample	A	A	A	A	A	A	A	A	A	A
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
First-Stage Statistics										
Partial R-squared	0.738	0.204	0.730	0.180	0.728	0.201	0.731	0.191	0.755	0.225
SW F-Statistic	195.91	9.81	187.18	9.81	165.52	9.80	177.60	9.15	218.89	11.19
KP Wald F-Statistic	195.91	5.07	187.18	5.13	165.52	5.16	177.60	4.70	218.89	5.47
Over-Identification Test (p-value)										
Percent One-Layer Firms	0.856	0.081	0.735	0.209	0.960	0.129	0.863	0.036	0.798	0.083
Percent Two-Layer Firms	0.788	0.299	0.740	0.281	0.725	0.279	0.591	0.215	0.733	0.307
Percent Three-Layer Firms	0.648	0.460	0.694	0.711	0.947	0.687	0.999	0.482	0.679	0.048
Percent Four-Layer Firms	0.113	0.248	0.127	0.274	0.122	0.273	0.133	0.267	0.169	0.340
Wage Controls*	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Income Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	408	408	408	408	408	408	408	408	375	375

*Notes: a,b,c: significant at the 1%, 5% and 10% level. *: indicates variables are treated as exogenous in regressions. Clustered standard errors at the employment area level are reported in parentheses. This table reports regression results for equation (19). Each column displays the estimate from a separate regression. Density measures the local density of an employment area. Industry fixed effects are at the three-digit Nace Rev 1.1. level. Wage controls contain the cost of a unit of labor. Income controls contain the median the annual salary of individuals residing in an employment area. Demographic controls contain the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. Columns (1), (3), (5), (7) and (9) only instrument for density using the log of density measured in 1831 and 1881. Columns (2), (4), (6), (8) and (10) instrument for density and local characteristics using the following variables: the log of density measured in 1831, 1851, 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, and the average person per household in 1949. Columns (1) and (2) control for the average number of additional occupations in firms which is equal to:* $\ln(\#of occupations - \#of layers + 1)$. *Columns (3) and (4) control for the degree of task specialization using the HHI index. Columns (5) and (6) control for the average log number of workers in firms. Columns (7) and (8) control for the average log number of hours in firms. Columns (9) and (10) retain only firms with adjacent layers and with production workers.*

4.1.2 Market Level Regressions Results: Distribution of Organizations

The Mann-Whitney test is reliable if, apart from density, all markets have the same characteristics. Markets are different along many dimensions, however, and these dimensions may be correlated with density and firm organization. Another concern, common in the urban economics literature, is that local shocks simultaneously determine firm organization and density. Both issues imply that the independence assumption of the Mann-Whitney test, that observations are independent both across and within samples, may be violated. This section therefore examines the model's prediction using regression analysis and addresses these additional concerns.

The simplest way to compare the distribution of organizations is to estimate the relationship between density and the share of firms producing with a given number of layers across industries and locations. More precisely, for industry j in area a, I estimate the following equation:

$$p^{L}_{j,a} = \alpha + \gamma \log density_a + X_{j,a}\beta + \epsilon_{j,a}, \tag{19}$$

where $p^{L}_{j,a}$ is the percent of firms in industry j operating in area a and producing with L layers, $density_a$ is the local density of an area a in the year 1999, and $X_{j,a}$ contains industry and area controls discussed above. According to the model, we should expect the estimates of γ to be negative for low values of L and positive for higher values.

Table 7 reports regression results with standard errors clustered at the local market level. The unit of observation is an industry in a local market and every entry in the table reports a result from a separate regression. Further to have a meaningful comparison of the distribution of organizations we require a large number of observations in each local market. Because the samples are small relative to the number of geographical areas, this condition is not always satisfied. In particular, when local markets are based on urban areas there are few firms operating in markets with below median density. The analysis, therefore, is only conducted on Sample A, where local markets are based on employment areas.

Column 1 only controls for industry. The estimates have the expected sign and are significant at conventional levels when the dependent variable is the share of one-layer, two-layer and three-layer firms. The distribution of organizations in employment areas with a 100% higher density contains roughly 3.4% fewer one-layer firms, 1.7% more two-layer firms and 1.5% more three-layer firms. Increasing the density of an employment area the size of Lyon, the third most populated in France, to an employment area the size of Paris, the most populated in France, corresponds to a 13.4 percent increase in the average number of layers in firms.

Columns 2-5 contain local area controls. Column 2 controls for the cost of a unit of labor, Column 3 controls for income, and Column 4 controls for the demographic composition of employment areas. Relative to Column 1 the estimated coefficients on density in Column 2 are identical, implying that differences in the local cost of hiring workers do not play a prominent role in explaining the distribution of firms across locations. In contrast, in Columns 3 and 4 the magnitude of the estimated coefficients on density are lower than in Column 1. This is not surprising because both income and preferences can affect the level of demand, and consequently the organization of firms. Nonetheless, density continues to affect the distribution of organizations. Column 5 considers all controls together, and suggests that the distribution of organizations in an employment area with a 100% higher density contains roughly 1.5% fewer one-layer firms and 1.1% more two-layer firms.

The analysis in Columns 1-5 contain all industry-employment area pairs in the data. One issue is that for some pairs there are too few firms operating in a market. As a result, the dependent variables contain measurement error which reduces the precision of the point estimates. To address this issue and to make comparisons between OLS and IV results, Columns 6-10 retain industry-employment area pairs with at least 10 firms and with historical values.

Columns 6 and 7 return to the specification in Column 1. Columns 6 restricts the sample to industry-employment area pairs with at least 10 firms and with historical values. The estimates have the expected sign and are significant at conventional levels. The magnitude of the coefficients on density are almost identical to Column 1, implying that differences in the samples do not affect results. Column 7 instruments for density using long-lagged values measured in 1831 and 1881 with two-stage least squares (2SLS). When the dependent variable is the share of one-layer and two-layer firms the magnitude of density is lower than in Column 6, implying that unobservable local shocks may be affecting both the size of markets and the organization of firms. Density continues to explain the distribution of organizations. The 2SLS results indicate that an employment area with a 100% higher density contains roughly 3.1% fewer one-layer firms, 1.3% more two-layer firms, 1.3% more three-layer firms and 0.2% more four-layer firms. Increasing the density of an employment area the size of Lyon to an employment area the size of Paris corresponds to a 12.5 percent increase in the average number of layers in firms. It is also important to note that in Column 7 the instruments are relevant. The partial R-squared from the first-stage is high, the instruments are not weak according to the tests developed by

Sanderson and Windmeijer (2016) and Kleibergen and Paap (2006), and in all cases they also pass the Sargan-Hansen over-identification test.[30]

Columns 8, 9 and 10 return to the specification in Column 5. Column 8 restricts the sample to industry-employment area pairs with at least 10 firms and with historical values, and the estimates are again nearly identical to Column 5. An employment area with a 100% higher density contains roughly 1.6% fewer one-layer firms and 1.4% more two-layer firms. Column 9 only instruments for density using historical values. The estimates have the expected sign, however, density loses significance when the dependent variable is the share of two-layer firms. Column 10 further instruments for density and the local area controls except for the cost of a unit of labor, because the hypothesis that the cost of a unit of labor is exogenous cannot be rejected at conventional levels. Throughout the paper this will always be the case. Relative to Column 8 the coefficients on density are statistical indistinguishable in Column 10, and the results indicate that an employment area with a 100% higher density contains 1.5% fewer one-layer firms and 1.2% more two-layer firms. Furthermore although they pass the Sargan-Hansen over-identification test, the instruments in Column 10 fall below the rule of thumb advocated by Staiger and Stock (1997). As a robustness check, the model in Column 10 has also been estimated with continuously updating GMM (CUE) which simulation evidence suggests is less sensitive to weak instruments than 2SLS (see Hahn et al. (2004) and Baum et al. (2007)). The results, not reported in the table, are similar to Column 10. An employment area with a 100% higher density contains 1.5% fewer one-layer firms and 1.4% more two-layer firms.

Overall the evidence is consistent with the model. Denser markets have a smaller share of one-layer firms and a greater share of two-layer, three-layer and four-layer firms. These findings suggest that the distribution of organizations in denser markets first-order stochastically dominates the distribution in less dense markets.

Table 8 reports a set of robustness results. For each set of results, the odd columns in Table 8 return to the specification in Column 7 from Table 7, which controls for industry and instruments for density, and the even columns return to the specification in Column 10 which controls for the characteristics of local markets and instruments for density, local incomes, and the demographic composition of local markets. Furthermore, throughout Table 8 the additional controls are treated

[30]Throughout the paper, to assess whether the instruments are weak I do not report the test statistic developed by Cragg and Donald (1993), and rely on the rule of thumb of Staiger and Stock (1997) that the F-statistic should be greater than 10 instead of the results from Stock and Yogo (2005). This is because in my regressions the standard errors are clustered at the market area level. See for additional discussions of these issues.

as exogenous variables, because the hypothesis that they are exogenous cannot be rejected at conventional levels.

Columns 1-4 control for the degree of specialization in markets. Tasks are more specialized in denser markets (see Stigler (1951), Garicano and Hubbard (2007) and Duranton and Jayet (2011)) and to the extent that this is reflected in the French occupational codes, my analysis may be confounding a greater number of layers with more task specialization. Moreover, it is important to note that one implication of the theory is that the division of labor is greater in denser markets. The theory can explain differences in the degree of specialization in vertical tasks across markets, that is tasks within a firm that have different levels of authority, however the model abstracts from any type of horizontal specialization, or tasks within a firm with the same level of authority. Even though a previous section showed that the classification of firms into layers is meaningful and consistent with the theory, Columns 1-4 control for occupations. Columns 1 and 2 control for the average number of additional occupations in firms across employment areas whereas Columns 3 and 4 control for the degree of occupational concentration in local markets measured with the Herfindahl-Hirschman Index (HHI) at the 2-digit occupational level.[31] Columns 1 and 3 only instrument for density, while Columns 2 and 4 instrument for density, local incomes, and the demographic composition of local markets. In comparison to Columns 7 and 10 from Table 7 the magnitude of the estimated coefficients on density is lower, however, density continues to affect the distribution of firms. The conclusions therefore remain the same and are consistent with the model. Column 4 indicates that an employment area with a 100% increase in density contains roughly 1.5% fewer one-layer firms and 1.3% more two-layer firms.

Columns 5-8 control for the size of firms. A second concern is that because firms in denser markets are larger, they hire a greater number of workers and assign them to specific tasks, which increases the number of possible occupations in firms (see Becker and Murphy (1992) and Garicano and Hubbard (2007)). Although Columns 1-4 control for occupations, Columns 5-8 return to Columns 7 and 10 in Table 7 and also control for the average size of firms. Columns 5-6 measure the size of firms using workers, while Columns 7-8 use hours. It should be noted, however, results with firm level controls should be interpreted with caution. In the model both

[31]To measure the number of additional occupations in firms, I use the 2-digit level of the occupational codes, and account for the number of layers in firms. In other words, for a firm with z occupations and L number of layers, I measure the number of additional occupations as: $\ln(z - L + 1)$. The implicit assumption is that a firm with one-layer should only have 1 occupation, a two-layer firm 2 occupations, and so forth. To measure the degree of occupational concentration in cities I construct a Herfindahl-Hirschman Index (HHI) at the 2-digit occupational level. The correlation between the index and the density is -0.099, implying that there is less occupational concentration in denser markets. This is what one would expect if there is more task specialization in denser markets.

firm organization and the size of firms is driven by an underlying demand parameter and it is uncertain where the source of variation that separately identifies both variables in the data is coming from. Despite this concern, in all cases density continues to play a prominent role in explaining the distribution of organizations. Column 8 indicates that an employment area with a 100% increase in density contains 1.3% fewer one-layer firms and 1.1% more two-layer firms. Additionally, the findings in Columns 5 to 8 suggests that organization is an important dimension to understanding firms, that cannot be accounted for by simply controlling for the size of firms.

Columns 9-10 use a different classification of the number of layers in firms. A final concern is that the definition of organizations is too broad and not consistent with the model. More precisely, in the model firms always have ordinary workers and increase their number of layers by hiring a worker in the layer above. Even though the occupational codes preserve a hierarchical ranking between workers in firms, to address this concern, Columns 9 and 10 restrict the data to firms with adjacent layers and with ordinary workers. In both cases, density continues to affect the distribution of organizations. The findings in Column 10 indicate that an employment area with a 100% increase in density contains 1.3% fewer one-layer firms and 0.7% more three-layer firms.

The remaining robustness checks are reported in the Appendix in Table B1. Here I simply summarize the main findings. First, Table B1 retains only single establishment firms. Second, Table B1 also restricts the sample to independent firms, that is firms that do not belong to a business group. And third Table B1 also controls for the amount of capital in firms, along with the size of firms. In all cases, the conclusions do not change. Density continues to affect the distribution of organizations and the findings remain consistent with the model.

4.1.3 Firm Level Regressions Results: Firm Organization

The previous subsection examined the theory's implication on the distribution of organizations. One issue however, is that when local markets are based on urban areas there are few firms operating in markets with below median density. This subsection therefore examines the relationship between density and organization at the level of the firm. It is important to note the results here are related to a weaker implication of the model, namely that firms in denser markets on average produce with a greater number of layers.

To examine the relationship between density and organization at the level of firms, for firm i operating in industry j and in area a, I estimate the following equation:

Table 9: Firm Level Regression Results

Dependent Variable: Number of Layers	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample A: Employment Areas										
log density	0.084	0.063	0.041	0.028	0.086	0.084	0.030	0.026	0.030	0.023
	$(0.003)^a$	$(0.006)^a$	$(0.002)^a$	$(0.007)^a$	$(0.002)^a$	$(0.003)^a$	$(0.007)^a$	$(0.012)^b$	$(0.013)^b$	$(0.008)^a$
Sample	A	A	A	A	A	A	A	A	A	A
Method	OLS	OLS	OLS	OLS	OLS	2SLS	OLS	2SLS	2SLS	2SLS
First-Stage Statistics	-	-	-	-	-	-	-	-	-	-
Partial R-squared	-	-	-	-	-	0.916	-	0.348	0.307	0.305
SW F-Statistic	-	-	-	-	-	1,175	-	12.23	11.82	12.18
KP Wald F-Statistic	-	-	-	-	-	1,175	-	8.576	8.599	8.554
Over-Identification Test (p-value)	-	-	-	-	-	0.467	-	0.396	0.083	0.146
R-squared	0.075	0.076	0.078	0.079	0.077	-	0.081	-	-	-
Sample Size	27,508	27,508	27,508	27,508	25,791	25,791	25,791	25,791	25,791	25,791
Sample B: Urban Areas										
log density	0.146	0.114	0.089	0.080	0.161	0.155	0.080	0.078	0.085	0.042
	$(0.024)^a$	$(0.014)^a$	$(0.013)^a$	$(0.011)^a$	$(0.016)^a$	$(0.022)^a$	$(0.010)^a$	$(0.012)^a$	$(0.013)^a$	$(0.012)^a$
Sample	B	B	B	B	B	B	B	B	B	B
Method	OLS	OLS	OLS	OLS	OLS	2SLS	OLS	2SLS	2SLS	2SLS
First-Stage Statistics	-	-	-	-	-	-	-	-	-	-
Partial R-squared	-	-	-	-	-	0.715	-	0.360	0.370	0.360
SW F-Statistic	-	-	-	-	-	22.84	-	12.78	12.58	12.78
KP Wald F-Statistic	-	-	-	-	-	22.84	-	7.919	7.967	7.920
Over-Identification Test (p-value)	-	-	-	-	-	0.282	-	0.109	0.100	0.268
R-squared	0.070	0.073	0.077	0.077	0.066	-	0.072	-	-	-
Sample Size	24,192	24,192	24,192	24,192	21,605	21,605	21,605	21,605	21,605	21,605
Wage Controls*	No	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes
Income Controls	No	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes
Demographic Controls	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: a,b,c: significant at the 1%, 5% and 10% level. *: indicates variables are treated as exogenous in regressions. Clustered standard errors at the local market level are reported in parentheses. This table reports regression results for equation (20). Each column displays the estimate from a separate regression. Density measures the local density of an employment area. Industry fixed effects are at the four-digit Nace Rev 1.1. level. Wage controls contain the cost of a unit of labor. Income controls contain the median the annual salary of individuals residing in an employment area. Demographic controls contain the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. In Sample A, Column (6) only instruments for density using the log of density measured in 1851 and 1881. In Sample B, Column (6) only instruments for density using the log of density measured in 1881 and 1901. In Sample A, Column (8) instruments for density and local characteristics using the following variables: the log of density measured in 1851 and 1881, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a telephone in their residence, and the share of the population in 1968 with a bathtub or shower in their residence. In Sample B, Column (8) instruments for density and local characteristics using the following variables: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an urban area in 1993, the share of the local population between the ages 25 and 54 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a telephone in their residence, the share of the population in 1968 with a bathtub or shower in their residence, the share of the population in 1968 with a toilet in their residence, and the share of the population in 1968 with heating in their residence. Column (9) add as a control the HHI index of occupational specialization at the city level. Column (10) add as a control the size of firms, in logs, measured using workers.*

$$ORG_{i,a} = \alpha + \gamma \log density_a + X_{j,a}\beta + Z_i\delta + \epsilon_{j,a}, \tag{20}$$

where $ORG_{i,a}$ is the number of layers in a firm, Z_i contains firm level controls that are not accounted for in the theory, and the remaining variables were defined in the previous section.

Table 9 reports regression results with standard errors are clustered at the local market level. The unit of observation is a firm in a local market and every entry in the table reports a result from a separate regression. The first panel in Table 9 reports results from Sample A, where local markets are based on employment areas, while the second panel reports regression results from Sample B, where local markets correspond to urban areas.

Column 1 only controls for industry. The estimates of density have the expected sign and are significant at conventional levels. A firm in an employment area (urban area) with a 100% higher density contains nearly 0.058 (0.101) additional layers. This implies that increasing the density of an employment area (urban area) the size of Lyon, the third (third) most populated in France, to an employment area (urban area) the size of Paris, the most (most) populated in France, corresponds to an additional 0.301 (0.134) layers in firms.

Columns 2-4 sequentially control for the characteristics of local markets. Column 2 controls for both local incomes and the cost of a unit of labor, while Column 3 controls for the demographic composition of local markets. In both Columns 2 and 3 the estimated coefficients on density are lower than in Column 1, because the characteristics of local markets affect the organization of firms. Column 4 considers all local area controls together. Density continues to affect the organization of firms and the results suggest that a firm in an employment area (urban area) with a 100% higher density contains roughly 0.019 (0.055) additional layers.

Columns 5 to 8 deal with instrumenting for density and the characteristics of local markets. Columns 5 and 6 return to the specification in Column 1. To allow for comparisons between OLS and IV results, Column 5 restricts the sample to firms operating in local markets with historical values. The coefficients on density are positive and significant, and relative to Column 1 the magnitude of density is slightly larger. Column 6 uses past values as instruments and estimates the model with 2SLS. In both samples, the coefficients on density are significant and statistically indistinguishable from Column 5. A firm in an employment area (urban area) with a 100% higher density contains roughly 0.058 (0.107) additional layers. The instruments in Column 6 are also valid, in the sense that they are not weak and they pass the Sargan-Hansen over-identification test.

Columns 7 and 8 come back to Column 4. Column 7 restricts the sample to local markets with historical values and in comparison to Column 4 the estimated coefficients are nearly the same. Column 8 further instruments for density, local incomes, and the demographic composition of local markets using lagged values. Column 8 does not instrument for the cost of a unit of labor because the hypothesis it is exogenous cannot be rejected at conventional levels. Across employment areas and urban areas the coefficients on density are statistically identical to Column 7. The findings indicate that firms in denser markets operate with a greater number of layers, consistent with the theory above. A firm in an employment area (urban area) with a 100% higher density will on average contain roughly 0.018 (0.054) additional layers, implying that increasing the density of an employment area (urban area) the size of Lyon to an employment area (urban area) the size of Paris corresponds to an additional 0.093 (0.071) layers in firms. The instruments in Column 6 pass the Sargan-Hansen over-identification test, however they fall below the rule of thumb advocated by Staiger and Stock (1997). As a robustness check, the model has also been estimated with continuously updating GMM (CUE). The results, not reported in the table, are similar.

For reasons discussed in the previous section, Columns 9-10 report a set of robustness results. Column 9 controls for the degree of occupational concentration in local markets using the HHI index, and Column 10 controls for the size of firms measured using workers. In every case the estimates are positive and significant at conventional levels. The remaining robustness check are reported in Tables B2 and B3 in the Appendix. Table B2 also controls for the number of additional occupations in firms, the size of firms using hours, and the level of capital in firms. Table B3 considers a different classification of firms, retains single establishment firms, and firms that are not part of a business group, and also controls for the legal status of firms. In all cases the conclusions remain the same. Firms in denser markets continue to produce with a greater number of layers.

To summarize, the evidence is consistent with the theory. Even when controlling for the characteristics of local markets and of firms, there is strong evidence that firms in denser market operate with a greater number of layers. Firm organization therefore varies with the size of the market and is an additional dimension to understanding differences between firms. The next section examines how important is organization to the productivity of firms.

Table 10: Value Added per Worker Regression Results

	VA per Worker (1)	VA per Worker (2)	VA per Worker (3)	VA per Worker (4)	VA per Worker (5)	VA per Worker (6)	VA per Worker (7)	VA per Worker (8)	VA per Worker (9)	VA per Worker (10)	VA per Worker (11)
Sample A: Employment Areas											
organization	0.147	0.122	0.124	0.088	0.169	0.120	0.172	0.123	0.119	0.125	0.124
	(0.021)^a	(0.012)^a	(0.013)^a	(0.008)^a	(0.017)^a	(0.013)^a	(0.016)^a	(0.017)^a	(0.014)^a	(0.014)^a	(0.014)^a
Sample	A	A	A	A	A	A	A	A	A	A	A
Method	OLS	OLS	OLS	OLS	2SLS	OLS	OLS	OLS	OLS	OLS	OLS
First-Stage Statistics	-	-	-	-	-	-	-	-	-	-	-
Partial R-squared	-	-	-	-	0.386	-	-	-	-	-	-
SW F-Statistic	-	-	-	-	1,995	-	-	-	-	-	-
KP Wald F-Statistic	-	-	-	-	1,995	-	-	-	-	-	-
Over-Identification Test (p-value)	-	-	-	-	0.075	-	-	-	-	-	-
R-squared	0.096	0.134	0.164	0.152	-	0.126	0.140	0.247	0.158	0.154	0.162
Sample Size	27,508	27,508	27,508	10,107	10,107	27,508	27,508	27,508	24,133	25,662	26,281
Sample B: Urban Areas											
organization	0.142	0.131	0.127	0.082	0.163	0.123	0.175	0.131	0.122	0.128	0.129
	(0.019)^a	(0.017)^a	(0.017)^a	(0.010)^a	(0.021)^a	(0.016)^a	(0.020)^a	(0.021)^a	(0.016)^a	(0.018)^a	(0.018)^a
Sample	B	B	B	B	B	B	B	B	B	B	B
Method	OLS	OLS	OLS	OLS	2SLS	OLS	OLS	OLS	OLS	OLS	OLS
First-Stage Statistics	-	-	-	-	-	-	-	-	-	-	-
Partial R-squared	-	-	-	-	0.386	-	-	-	-	-	-
SW F-Statistic	-	-	-	-	2,147	-	-	-	-	-	-
KP Wald F-Statistic	-	-	-	-	2,147	-	-	-	-	-	-
Over-Identification Test (p-value)	-	-	-	-	0.173	-	-	-	-	-	-
R-squared	0.088	0.156	0.193	0.128	-	0.106	0.122	0.223	0.192	0.190	0.193
Sample Size	24,197	24,197	24,197	8,825	8,825	24,197	24,197	24,197	21,128	22,615	23,000
No. of Occ.	No	No	No	No	No	Yes	No	No	No	No	No
Firm Size	No	No	No	No	No	No	Yes	Yes	No	No	No
Capital	No	No	No	No	No	No	No	Yes	No	No	No
Legal Status	No	No	No	No	No	No	No	Yes	No	No	No
Industry FE	Yes	Yes	No	No	No	No	No	No	No	No	No
Area FE	No	Yes	No	No	No	No	No	No	No	No	No
Area-Industry FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: a,b,c: significant at the 1%, 5% and 10% level. Clustered standard errors at the local market level are reported in parentheses. Regression results for equation (23). Each column displays the estimate from a separate regression. Industries fixed effects are at the 4-digit Nace Rev 1.1. Columns (5) instruments for the number of layers in firms using the measure from the years 1998 and 2002. In Columns (7) and (8) firm size is measured using workers. Column (9) restricts the sample to firms with adjacent layers starting from layer 1. Column (10) restricts the sample to single establishment firms and Column (11) restricts the sample to independent firms.

4.2 Productivity and Firm Organizations Across Locations

This section deals with the second main prediction of the model. Namely, in denser markets firms are more productive because they organize with a greater number of layers. The analysis consists of three parts. First, I discuss how I measure firm productivity. Second, I show that organization is important for understanding the productivity of firms. Third, I examine the role of organization in explaining the productivity of firms across locations relative to other mechanisms from the literature.

4.2.1 Measuring Firm Productivity

Computing firm productivity is essential to examine the second implication from the model. To construct measures of firm productivity I combine balance sheet information from FICUS with measures on the size and wage bill of firms from the DADS. It is important to note that because the balance sheet data do not contain information on prices and units produced, my analysis is limited to measures of revenue-based productivity. My analysis mainly presents results with number of workers as a measure of the size of firms, because I suspect errors in the number of hours reported in service firms.

To assess the importance of organization in the productivity of firms, in my analysis I use several approaches to measure firm productivity. First, because the firms in this study are in service industries, I begin my analysis with a measure of labor productivity, defined as value-added per worker.[32] This measure is used in the empirical literature examining the productivity of service firms and is closely related to Figure 10b from the model simulations.

Second, I adopt a more structural approach and measure productivity as Total Factor Productivity (TFP).[33] In this case, I use the standard methods in the literature to obtain estimates of firm productivity. More precisely, for firm i in year t, I assume that value-added, va_{it}, is equal to:

[32] Indeed there has been a long-standing debate about the appropriate measure of firm productivity in service firms. See Haskel and Sadun (2009) for a discussion on retail firms. One issue with value-added per worker, however, is that it ignores the other inputs used in production. See Syverson (2011) for a discussion on the different measures of productivity and the references therein.

[33] In terms of the model, this would imply that A is heterogeneous across firms. A change in A would have two effects on firms. First, for a given level of quantity, because they require less labor to produce a unit of output, firms with a greater A will have a smaller number of layers. Second, for a given demand curve, firms with a greater A will produce a greater quantity, which will force them to produce with a greater number of layers. Which effect dominates depends on the assumptions in the model. Furthermore, in terms of data, we have seen that firms in denser markets produce with a greater number of layers, and these firms tend to be more productive. As results, this implies that the second effect dominates in the sectors examined in this study.

$$\ln va_{it} = \alpha_k \ln k_{it} + \alpha_l \ln l_{it} + \psi_t + \epsilon_{it}, \tag{21}$$

where k_{it} is the capital of firm i at time t, l_{it} denotes the amount of labor used, measured using either workers or hours, and ψ_t are year fixed effects. The residual, ϵ_{it}, is the sum of the productivity of firm i at time t, ω_{it}, and an independently and identically distributed error term, η_{it}. To address concerns with endogenous factor inputs equation (21) is estimated using the methods proposed by Levinsohn and Petrin (2003) and Wooldridge (2009).[34] Throughout my analysis, I assume that productivity follows a first-order Markov process of the form $\omega_{it} = E[\omega_{it}|\omega_{it-1}] + \xi_{it}$, where ξ_{it} is an innovation term that is independently and identically distributed and uncorrelated with past values of labor and capital. Capital is assumed to be predetermined at time t, and labor is a free variable determined at time t and correlated with the innovation term, ξ_{it}.

And third, I estimate TFP using a method adapted from Caliendo et al. (2015a). The theory in this paper takes the view that organization is a choice made by firms and that organization determines the efficiency of labor. In other words, the number of units of output produced from a unit of labor depends on the number of layers in firms. Organization, however, does not determine A, the number of units of output produced when a problem is solved, which most closely corresponds to measures of firm TFP. Under the assumption that entrepreneurs can devote less than their full unit of labor to a firm, for firm i at time t, Caliendo et al. (2015a) show that value-added is equal to:

$$\ln va_{it} = \alpha_k \ln k_{it} + \alpha_C C(O_{it}, w) + \alpha_O ORG_{it} + \psi_t + \epsilon_{it}, \tag{22}$$

where O_{it} is the number of problems solved by workers in firm i at time t, $C(O_{it}, w)$ is the firm's wage bill at time t, and remaining terms are defined above. Equation (22) is also estimated using the methods proposed by Levinsohn and Petrin (2003) and Wooldridge (2009) to deal with endogenous factor inputs. To identify the parameters α_k, α_C and α_O using the structural methods I assume that organization and the wage bill are free variables, determined at time t, and correlated with the innovation term ξ_{it}. Additionally, I assume that the innovation term is uncorrelated with past values of the wage bill and the number of layers in firms. The benefit of using this last approach is that it explicitly takes into consideration the number of layers in firms and controls

[34]Further, throughout the analysis similar results are obtained when using the methods put forth by Olley and Pakes (1996) and Ackerberg et al. (2015). These results are available upon request.

for the quality of labor in the estimation of firm TFP.[35] For these reasons, it is my preferred measure of TFP. I will only return to this measure in the last part of this section, when examining the role of organization in explaining the productivity of firms across locations. All measures of TFP are estimated over the period 2002-2006. Appendix B provides additional details.

4.2.2 Organization and Productivity

To examine the relationship between productivity and the organization of firms I estimate the following equation:

$$\ln \phi_{i,a} = \alpha + \vartheta ORG_{i,a} + \theta_{j,a} + Z_i\delta + \epsilon_{i,j,a}, \tag{23}$$

where $\ln \phi_{i,a}$ is the productivity of firm i operating in area a measured either as value-added per worker or TFP estimated from equation (21), $ORG_{i,a}$ is the number of layers firm i, $\theta_{j,a}$ are industry and location interaction effects, and Z_i are firm controls. According to the theory we expect estimates of ϑ to be positive.

Table 10 reports results with productivity measured as value-added per worker and standard errors clustered at the local market level. The unit of observation is a firm in a local market and every entry in the table reports a result from a separate regression. The top panel of Table 10 report results from Sample A, where local markets correspond to employment areas, while the bottom panel reports regression results from Sample B, where local markets are based on urban areas.

Column 1 only controls for the industry of firms. In each case the relationship between the number of layers in firms and productivity is positive and significant at conventional levels. In Sample A (Sample B) an extra layer in a firm is associated with a 14.7% (14.2%) increase in valued-added per worker.

Column 2 contains industry and location fixed effects. Relative to Column 1, the magnitude of the estimated coefficients of organization are lower. This is not surprising because density affects both the organization and the productivity of firms. The relationship between the number of layers in firms and productivity, however, remains positive and significant. Adding a layer in a

[35] Although equation (22) is very similar to Caliendo et al. (2015a) there is a difference. In their paper Caliendo et al. (2015a) take the position that organization not only determines the efficiency of labor but also firm TFP, which is maps into the parameter A of the model. More precisely, Caliendo et al. (2015a) assume that the productivity of firms, ω_{it} follows the following process: $\omega_{it} = \rho\omega_{it-1} + \beta ORG_{it} + \xi_{it}$. The number of layers in a firm do not only have a contemporaneous effect on TFP, but also past organizational structures can also affect current TFP. See Caliendo et al. (2015a) for details.

firm is associated with a 12.2% (13.1%) increase in valued-added per worker in Sample A (Sample B).

Column 3 saturates the model and contains industry and location interaction terms, which account for many of the possible characteristics of local markets. Even with a full set of industry and location controls organization continues to be an important determinant of firm productivity. In each case, the magnitude of the coefficients is almost identical to Column 2 and in Sample A (Sample B) adding a layer in a firm is associated with a 12.4% (12.7%) increase in valued-added per worker.

Moreover, it may be the case that firm productivity and organization are simultaneously determined, or that contemporaneous firm level shocks affect both variables. To address this concern, Columns 4-5 deal with instrumenting for organization using the number of layers in firms from the years 1998 and 2002. To compare OLS and IV results, Column 4 returns to the specification in Column 3 and restricts the sample to firms with lagged values. In Sample A (Sample B) an extra layer in a firm is associated with a 8.8% (8.2%) increase in valued-added per worker. Column 5 instruments for the organization of firms using the number of layers in firms from the years 1998 and 2002. The results lead to the same conclusion. Organization continues to have a role in explaining the productivity of firms. For example, accounting for all possible characteristics of local markets an additional layer in a firm is associated with a 16.9% (16.3%) increase in valued-added per worker. The instruments pass the Sargan-Hansen over-identification test at the 5% percent level of significance and the magnitude of the estimated coefficients is nearly twice the magnitude reported in Column 4. However, if the productivity of firms follows a Markov process then lagged values of organization are likely to still be endogenous. For this reason and because of the limited number of available variables that can serve as instruments for the organization of firms, in the remaining analysis I simply report OLS regression results with the caution that organization may be endogenous.

Columns 6-11 return to the specification in Column 3 and report a series of robustness results. Columns 6-8 control for the characteristics of firms. Column 6 controls for the number of additional occupations in firms, Column 7 controls for the size of firms, and along with firm size, Column 8 also controls for the level of capital and the legal status of firms. In every case, the relationship between organization and productivity remains positive and significant. For example, Column 8 reports that in Sample A (Sample B) adding a layer in a firm is associated with a 12.3% (13.1%) increase in value-added per worker. Furthermore, Columns 9-11 restrict the sam-

Table 11: Additional Regression Results

	LP Workers (1)	LP Hours (2)	WD Workers (3)	WD Hours (4)
Sample A: Employment Areas				
organization	0.291	0.177	0.304	0.168
	(0.016)[a]	(0.014)[a]	(0.016)[a]	(0.014)[a]
Sample	A	A	A	A
Method	OLS	OLS	OLS	OLS
Area-Industry FE	Yes	Yes	Yes	Yes
R-squared	0.199	0.821	0.185	0.806
Sample Size	23,840	23,840	23,840	23,840
Sample B: Urban Areas				
organization	0.295	0.183	0.307	0.174
	(0.024)[a]	(0.022)[a]	(0.024)[a]	(0.022)[a]
Sample	B	B	B	B
Method	OLS	OLS	OLS	OLS
Area-Industry FE	Yes	Yes	Yes	Yes
R-squared	0.186	0.835	0.171	0.822
Sample Size	21,002	21,002	21,002	21,002

Notes: a,b,c: significant at the 1%, 5% and 10% level. Clustered standard errors at the local area level are reported in parentheses. OLS regression results for equation (23). Each column displays the estimate from a separate regression with industry-location fixed effects. Industries are at the 4-digit Nace Rev 1.1. Regressions are different based on the method use to estimate productivity. Columns (1) and (2) use the method by Levinsohn and Petrin (2003). Columns (3) and (4) use the method by Wooldridge (2009).

ple of firms. Column 9 restricts the sample to firms with adjacent layers and ordinary workers, Column 10 restricts the sample to single establishment firms, and Column 11 restricts the sample to independent firms. In every case, the estimated coefficients on organization are positive and significant, and their magnitudes are comparable to the estimates in Column 3. The conclusions therefore remain the same and organization continues to be positively associated with the productivity of firms.

Up to here the findings indicate that firms with a greater number of layers are more productive. Table 11 further examines whether the findings are robust across different measures of firm productivity. Columns 1 and 2 measure productivity using the method from Levinsohn and Petrin (2003) which accounts for endogenous factor inputs. In all cases, the estimated coefficients are positive and significant at conventional levels. In Sample A (Sample B) an extra layer in a firm is associated with a 29.1% (29.5%) increase in TFP estimated using workers and a 17.7% (18.3%) increase in TFP estimated using hours. In Columns 3 and 4 the dependent variable is TFP estimated using the method proposed by Wooldridge (2009) which controls for endogenous factor inputs and allows to identify the coefficient on labor in the estimation of equation (21). The relationship between the number of layers in firms and productivity remains positive and significant. The magnitudes on the estimated coefficients on organization are similar to those reported in Columns 1 and 2, where TFP is estimated using the method by Levinsohn and Petrin (2003). An additional layer in a firm is associated with a 30.4% increase in TFP estimated using workers and a 16.8% increase in TFP estimated using hours in Sample A. The corresponding estimates in Sample B are similar. An extra layer is associated with a 30.7% increase in TFP estimated using workers and a 17.4% increase in TFP estimated using hours.

Using different measures of firm productivity therefore does not qualitatively change conclusions. However, relative to the results in Table 10 the estimated magnitudes on the coefficient of organization are larger in Table 11. This is not surprising because although they measure the same economic concept, productivity, these measures are conceptually distinct. For instance, value-added per worker ignores the other inputs used in production while TFP is a residual.[36] Nonetheless, despite the different magnitudes, both sets of measures indicate that firms that organize with a greater number of layers are more productive.

Beyond being statistically significant, these estimates are also quantitatively and economically significant. To emphasize the important role organization has in the productivity of firms, consider firms located in the first and fourth quartiles of the value-added per worker distribution. In Sample A (Sample B) average value-added per worker in the first quartile is 5.72 (5.95), while in the fourth quartile it is equal to 26.19 (26.92). In addition, firms in Sample A (Sample B) on average organize with 1.66 (1.70) layers in the first quartile and 1.78 (2.11) in the fourth. Differences in the way firms organize production therefore accounts for between 12.7% and 13.7% of the difference in the average value-added per worker between firms located in the first and fourth quartiles of the productivity distribution in Sample A, and between 12.5% and 13.6% in Sample B. The corresponding values for TFP in Columns (1), (2), (3) and (4) of Table 11 are roughly 16.4%, 4.9%, 16.6% and 6.2% in Sample A, and 16.5%, 4.3%, 16.4% and 5.1% in Sample B.[37]

Overall, the findings are largely consistent with the model. Firms in denser markets are more productive, in part, because they organize with a greater number of layers. The findings in Tables 10 and 11 are not only statistically significant but economically meaningful as well and further suggest that understanding how firms organize production is important for understanding the differences in the productivity of firms. The following section examines this question.

[36]The sample size in both tables is also different. In unreported results, I restrict the sample of firms with a measure of TFP when estimating the relationship between value-added per worker and the number of layers in firms. The magnitude of the estimated coefficients are nearly identical to those reported in Table 10.

[37]It is important to note that these results most likely provide a lower bound, because large firms that operate in many markets are excluded from the analysis.

Table 12: Second-Stage Value-Added per Worker Regression Results

Dependent Variable from First-Stage	VA per Worker (1)	VA per Worker (2)	VA per Worker (3)	VA per Worker (4)	VA per Worker (5)	VA per Worker (6)	VA per Worker (7)	VA per Worker (8)	VA per Worker (9)	VA per Worker (10)
Model 1: Without Org										
log density	0.049	0.039	0.048	0.048	0.037	0.046	0.044	0.045	0.046	0.048
	$(0.004)^a$	$(0.009)^a$	$(0.004)^a$	$(0.004)^a$	$(0.009)^a$	$(0.009)^a$	$(0.010)^a$	$(0.009)^a$	$(0.009)^a$	$(0.009)^a$
Model 2: With Org										
log density	0.038	0.036	0.037	0.038	0.034	0.043	0.040	0.042	0.043	0.045
	$(0.003)^a$	$(0.009)^a$	$(0.004)^a$	$(0.003)^a$	$(0.009)^a$	$(0.009)^a$	$(0.011)^a$	$(0.009)^a$	$(0.009)^a$	$(0.009)^a$
% Change	22.4	7.6	22.9	20.8	8.10	6.5	9.1	6.6	6.9	6.2
Sample	A	A	A	A	A	A	A	A	A	A
Method	OLS	OLS	OLS	2SLS	OLS	2SLS	2SLS	2SLS	2SLS	2SLS
KP Wald F-Statistic	-	-	-	2,700	-	10.04	10.16	10.04	10.04	9.97
Over-Id Test (p-value): Model 1	-	-	-	0.680	-	0.143	0.110	0.153	0.135	0.051
Over-Id Test (p-value): Model 2	-	-	-	0.736	-	0.214	0.181	0.215	0.202	0.081
R-squared: Model 1	0.611	0.641	0.614	-	0.644	-	-	-	-	-
R-squared: Model 2	0.580	0.610	0.582	-	0.613	-	-	-	-	-
Sample Size	27,508	27,508	26,531	26,531	26,531	26,531	26,531	26,531	26,531	25,850
Model 1: Without Org										
log density	0.073	0.060	0.078	0.087	0.046	0.059	0.068	0.056	0.063	0.067
	$(0.012)^a$	$(0.012)^a$	$(0.013)^a$	$(0.013)^a$	$(0.016)^a$	$(0.020)^a$	$(0.022)^a$	$(0.020)^a$	$(0.020)^a$	$(0.019)^a$
Model 2: With Org										
log density	0.054	0.050	0.057	0.068	0.035	0.048	0.056	0.047	0.055	0.060
	$(0.010)^a$	$(0.011)^a$	$(0.011)^a$	$(0.012)^a$	$(0.016)^b$	$(0.020)^b$	$(0.021)^b$	$(0.020)^b$	$(0.019)^a$	$(0.019)^a$
% Change	26.0	16.6	26.9	21.8	23.9	16.6	17.6	16.0	12.6	11.6
Sample	B	B	B	B	B	B	B	B	B	B
Method	OLS	OLS	OLS	2SLS	OLS	2SLS	2SLS	2SLS	2SLS	2SLS
KP Wald F-Statistic	-	-	-	22.84	-	7.33	7.55	7.33	7.33	7.23
Over-Id Test (p-value): Model 1	-	-	-	0.490	-	0.161	0.203	0.166	0.170	0.090
Over-Id Test (p-value): Model 2	-	-	-	0.383	-	0.184	0.234	0.185	0.221	0.117
R-squared: Model 1	0.425	0.463	0.493	-	0.539	-	-	-	-	-
R-squared: Model 2	0.394	0.429	0.460	-	0.506	-	-	-	-	-
Sample Size	24,197	24,197	21,606	21,606	21,606	21,606	21,606	21,606	21,606	21,043
Second-Stage: All Controls	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Second-Stage: Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: a,b,c: significant at the 1%, 5% and 10% level. Clustered standard errors at the local market level are reported in parentheses. Each entry displays a results from equation (24). Model 1 does not control for the number of layers in firms in the first-stage while Model 2 does. Industry fixed effects are at the four-digit Nace Rev 1.1. level. Controls in the second-stage are the following: the cost of a unit of labor, the median the annual salary of individuals residing in an area, the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. Column (4) instruments for density using density measured in 1881 and 1901. Columns (6)-(10) instrument for density and the characteristics of local markets except for the cost of a unit of labor. In Sample A, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a telephone in their residence, and the share of the population in 1968 with a toilet in their residence. In Sample B, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 54 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a heating in their residence, and the share of the population in 1968 with a bathtub or shower in their residence. Column (7) add as a control the HHI index of occupational specialization at the city level in the second-stage. Column (8) controls for the number of additional occupations in firms in the first-stage. Column (9) controls for the size of firms (measured using hours) in the first-stage. Column (10) controls for the size of firms, the level of capital and the legal status of firms in the first-stage.

4.2.3 Firm Productivity, Organization, and Density

The empirical literature investigating the productivity gains from operating in denser markets, generally examines the elasticity of measures of local productivity with respect to density. The main conclusion emerging from these studies is there are positive productivity gains from operating in denser markets, with estimates in the range of 0.02 to 0.10 (Combes and Gobillon (2015)). These gains are generally attributed to external economies of scale. In contrast, this paper's focus is on economies of scale internal to firms. This section therefore compares the role of organization in explaining the productivity of firms across locations to the role of density.

To assess the important role organization has in explaining the productivity of firms across locations, I adopt a two-stage approach. In the first stage, I estimate equation (23) with industry-area interaction effects. These results were reported in the previous section. In the second-stage I estimate the following equation:

$$\theta_{j,a} = \alpha + \gamma \log density + X_{j,a}\beta + \epsilon_{j,a}, \tag{24}$$

where the dependent variable, $\theta_{j,a}$, is the estimated industry-area interaction effects from the first-stage, $density_a$ is the local density of an area a in the year 1999 and $X_{j,a}$ contains industry and area controls. I conduct this exercise with and without controlling for the number of layers in firms in the first-stage, and compare results. According to the theory, we should expect part of the productivity gains from operating in denser markets to be due to firms operating with a greater number of layers. This implies that we expect to observe a decrease in the elasticity of local productivity with respect to density, γ, when controlling for the organization of firms in the first-stage.

Table 12 reports results from the second-stage with value-added per worker as the dependent variable in the first-stage, and standard errors clustered at the local area level. The top panel presents results from Sample A, where local markets are based on employment areas, while the bottom panel reports results from Sample B, where local markets correspond to urban areas. In both panels each column reports two coefficients, estimated from Models 1 and 2. Model 1 does not control for the organization of firms in the first-stage, while Model 2 reports the elasticity of local productivity with respect to density when the organization of firms is included in the first-stage.

Column 1 only controls for density in the second stage. The estimated relationships are positive and significant at conventional levels. Model 1, which does not control for the organization of firms in the first-stage, reports that across employment areas (urban areas) a 1% increase in density is associated with a 4.9% (7.3%) increase in value-added per worker. These estimates are consistent with the literature which reports an elasticity of local productivity with respect to density in the range of 0.02 and 0.10. Model 2 includes the number of layers in firms in the first-stage. The magnitude on density is now lower, suggesting that part of the productivity gains from operating in denser markets are explained by firms having a greater number of layers. Across employment areas (urban areas) a 1% increase in density is now associated with a 3.8% (5.4%) increase in value-added per worker. This corresponds to roughly a 22% (26%) decrease in the measured elasticity of value-added per worker with respect to density.

Column 2 controls for the characteristics of local markets in the second stage. Density remains positive and significant at conventional levels. In most cases, the magnitude of the estimated coefficients is now lower, suggesting that part of the productivity gains from denser markets are related to the characteristics of local markets. Controlling for the number of layers in firms decreases the elasticity of local productivity with respect to density. Column 2 reports that across employment areas (urban areas) approximately 7.6% (16.6%) of the measured elasticity of value-added per worker with respect to density is related to the organization of firms.

Columns 3-6 deal with instrumenting for density and the characteristics of local markets. Columns 3 an 4 return to the specification in Column 1. Column 3 restricts the sample to local areas with historical values, to allow for comparisons between OLS and IV results. Across both employment areas and urban areas the point estimates are nearly identical to Column 1. Column 4 uses past values as instruments and estimates the model with 2SLS. In each case, the coefficients on density are significant and the magnitudes are statistically indistinguishable from Column 1. Model 1, which does not control for the organization of firms in the first-stage, reports that across employment areas (urban areas) a 1% increase in density is associated with a 4.8% (8.7%) increase in value-added per worker. Model 2, which includes the number of layers in firms in the first-stage reports that across employment areas (urban areas) a 1% increase in density is now associated with a 3.8% (6.8%) increase in value-added per worker. This corresponds to approximately a 21% (22%) decrease in the measured elasticity of value-added per worker with respect to density across employment areas (urban areas).

Columns 5 and 6 come back to Column 2. Column 5 restricts the sample to local markets with historical values. In comparison to Column 2 the estimated coefficients are nearly the same across employment areas, however they are lower across urban areas. Column 6 further instruments for density, local incomes, and the demographic composition of local markets using lagged values. Across employment areas and urban areas the coefficients on density are statistically identical to Column 2. The findings indicate that across employment areas (urban areas) approximately 6.5% (16.6%) of the measured elasticity of value-added per worker with respect to density is related to the organization of firms. In unreported results, the model in Column 6 has also been estimated with continuously updating GMM (CUE), because the instruments fall below the rule of thumb advocated by Staiger and Stock (1997). The findings are similar.

Columns 7-10 assess the robustness of the results. Column 7 returns to the specification in Column 6 and also controls for the degree of occupational concentration in local markets using the HHI index. When including organization in the first-stage, the elasticity of value-added per worker with respect to density decreases by approximately 9.1% (17.6%) across employment areas (urban areas). Column 8 instead controls for the number of occupations in firms in the first-stage. The findings are similar. Across employment areas (urban areas) the density elasticity decreases by nearly 6.6% (16.0%). Column 9 controls for the size of firms in the first-stage. The elasticity of value-added per worker with respect to density decreases by approximately 6.9% (12.6%) across employment areas (urban areas). Column 10 further controls for the size of firms, the level of capital and the legal status of firms in the first-stage. Across employment areas the elasticity of value-added per worker with respect to density decreases by roughly 6.2% and across urban areas the density elasticity decreases by 11.6%.

Several robustness results are also reported in Tables B4 and B5 in the Appendix. Table B4 restricts the sample to firms with adjacent layers and ordinary workers, to single establishment firms, and to firms that are not part of a business group. In every case, controlling for the number of layers in firms decreases the density elasticity. Table B5 considers two alternative specifications, which are estimated in one-stage. One specification includes indicator variables for the number of layers in firms, and the second organization-industry fixed effects. In all cases, the estimated magnitudes on density are almost identical to Table 12, and thus the findings are robust to the different ways one can control for the organization of firms.

Table 13: Second-Stage TFP Regression Results

Dependent Variable from First-Stage	LP Workers (1)	LP Workers (2)	LP Hours (3)	LP Hours (4)	WD Workers (5)	WD Workers (6)	WD Hours (7)	WD Hours (8)
Model 1: Without Org								
log density	0.070	0.050	0.048	0.038	0.071	0.050	0.046	0.036
	(0.006)^{a}	(0.013)^{a}	(0.006)^{a}	(0.010)^{a}	(0.006)^{a}	(0.013)^{a}	(0.006)^{a}	(0.010)^{a}
Model 2: With Org								
log density	0.046	0.044	0.034	0.034	0.047	0.043	0.033	0.033
	(0.005)^{a}	(0.012)^{a}	(0.005)^{a}	(0.010)^{a}	(0.005)^{a}	(0.012)^{a}	(0.005)^{a}	(0.009)^{a}
% Decrease	34.2	12.0	29.1	10.5	33.8	14.0	28.2	8.3
Sample	A	A	A	A	A	A	A	A
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
KP Wald F-Statistic	2,653	7.16	2,653	7.16	2,653	7.16	2,653	7.16
Over-Id Test (p-value): Model 1	0.181	0.428	0.409	0.268	0.186	0.438	0.460	0.259
Over-Id Test (p-value): Model 2	0.139	0.659	0.388	0.369	0.142	0.675	0.445	0.346
Sample Size	22,997	22,997	22,997	22,997	22,997	22,997	22,997	22,997
Model 1: Without Org								
log density	0.124	0.107	0.092	0.080	0.126	0.108	0.088	0.077
	(0.012)^{a}	(0.018)^{a}	(0.011)^{a}	(0.014)^{a}	(0.012)^{a}	(0.018)^{a}	(0.011)^{a}	(0.013)^{a}
Model 2: With Org								
log density	0.080	0.083	0.064	0.066	0.080	0.084	0.062	0.063
	(0.009)^{a}	(0.017)^{a}	(0.009)^{a}	(0.013)^{a}	(0.009)^{a}	(0.017)^{a}	(0.009)^{a}	(0.013)^{a}
% Decrease	35.4	22.4	30.4	17.5	36.5	22.2	29.5	18.1
Sample	B	B	B	B	B	B	B	B
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
KP Wald F-Statistic	22.07	7.09	22.07	7.09	22.07	7.09	22.07	7.09
Over-Id Test (p-value): Model 1	0.405	0.066	0.171	0.017	0.359	0.062	0.159	0.018
Over-Id Test (p-value): Model 2	0.635	0.085	0.216	0.024	0.572	0.078	0.194	0.025
Sample Size	18,769	18,769	18,769	18,769	18,769	18,769	18,769	18,769
Second-Stage: All Controls	No	Yes	No	Yes	No	Yes	No	Yes
Second-Stage: Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes:a,b,c: significant at the 1%, 5% and 10% level. Clustered standard errors at the local market level are reported in parentheses. Each entry displays a results from equation (24) with TFP estimated using equation (21). Model 1 does not control for the number of layers in firms in the first-stage while Model 2 does. Industry fixed effects are at the four-digit Nace Rev 1.1. level. Controls in the second-stage are the following: the cost of a unit of labor, the median the annual salary of individuals residing in an area, the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. Columns (1), (3), (5) and (7) instrument for density using density measured in 1881 and 1901. Columns (2), (4), (6) and (8) instrument for density and the characteristics of local markets except for the cost of a unit of labor. In Sample A, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 having access to heating in their residence, and the share of the population in 1968 with a sanitary installation in their residence. In Sample B, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 54 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a heating in their residence, and the share of the population in 1968 with a bathtub or shower in their residence.

Table 13 assesses the robustness of results using other measures of firm productivity. For each set of results, the odd columns control for industry and instrument for density, and the even columns control for the characteristics of local markets and instrument for density, local incomes, and the demographic composition of local markets. In Columns 1-4, the dependent variable in the first-stage is TFP estimated using the approach proposed by Levinsohn and Petrin (2003). Columns 1-2 estimate TFP using workers to proxy for the size of firms, while Columns 3-4 use with hours. In every case, the coefficients on density are positive and significant at conventional levels, and including organization in the first-stage decreases the elasticity of local TFP with respect to density. Across employment areas the decrease ranges from 10.5% to 34.2%, and across urban areas it ranges from 17.5% to 35.4%. In Columns 5-8, the dependent variable in the first-stage is TFP estimated with the method from Wooldridge (2009). Columns 5-6 estimate TFP using workers to proxy for the size of firms and Columns 7-8 use with hours. The findings are similar. The estimated coefficients are positive and significant, and controlling for organization in the first-stage decreases the density elasticity. Across employment areas the decrease ranges from 8.3% to 33.8%, and across urban areas it ranges from 18.1% to 36.5%.

To summarize, part of the productivity gains from operating in denser markets are explained by differences in the way firms organize production. When accounting for the number of layer in firms, the elasticity of local value-added per worker with respect to density decreases by 6.2% to 22.9% across employment areas, and by 11.6% to 26.9% across urban areas. Using different measures of firm productivity does not qualitatively change conclusions. Controlling for the number of layers in firms decreases the elasticity of local TFP with respect to employment area density (urban area density) by as little as 8.3% (18.1%) to as much as 34.2% (36.5%). The estimated effect of organization, however, is greater than when firm productivity is measured as value-added per worker. This is to be expected because as shown in the previous section, when productivity is measured as TFP, organization plays a more prominent role in the productivity of firms.

The estimates reported in this section are also quantitatively and economically significant. To emphasize what these numbers imply, consider increasing the density of an employment area (urban area) the size of Lyon to the size of Paris. Differences in the way firms organize production accounts for between 6.8% and 24.0% of the percent increase in value-added per worker in Sample A, and between 10.7% and 26.7% in Sample B. The corresponding values with TFP are similar: when we increase the density of an area the size Lyon to the density the size of Paris, differences in the way firms organize production accounts for between 8.8% and 37.1% of the percent increase in TFP in Sample A, and between 18% and 37.9% in Sample B.[38]. The final section compares estimates with the method adapted from Caliendo et al. (2015a).

4.2.4 Comparing TFP Estimates

Another way to assess the relevance of organization is to compare the relationship between local productivity and density using the different measures of TFP estimated above, with the method described in equation (22) and adapted from Caliendo et al. (2015a). The benefit of using the latter approach to estimate TFP is that it directly accounts for the number of layers in firms.

Table 14 presents results and reports 2SLS estimates of the density elasticity using the method adapted from Caliendo et al. (2015a), my preferred estimate of firm TFP. The table has the same structure as previous tables: the top panel presents results from Sample A, while the bottom panel reports results from Sample B. Furthermore to remain consistent with the previous analysis, Table 14 estimates the relationship between local productivity and density using the two-step approach, without controlling for the number of layers in firms in the first-stage. The estimates obtained are compared to the corresponding density elasticities from Model 1 in Table 13. Standard errors are clustered at the local market level.

In Columns 1 and 2 the dependent variable in the first-stage is TFP estimated with the specification adapted from Caliendo et al. (2015a) and using the method proposed by Levinsohn and Petrin (2003). Column 1 only controls for density in the second-stage. The point estimates are positive and significant at conventional levels. Across employment areas (urban areas) a 1% increase in density is associated with a 2.8% (5.3%) increase in TFP. Relative to the magnitudes from Model 1, in Columns 1 and 3 of Table 13, across employment areas the estimated relationship is approximately 60.0% lower when TFP is estimated using workers, and 41.6% lower when TFP is estimated using hours. Across urban areas the decrease in magnitudes is similar. The elasticity of productivity with respect to density decreases by nearly 57.2% and 42.3%. Column 2 controls for density and the characteristics of local markets. Across employment areas (urban areas) a 1% increase in density is now associated with a 3.0% (6.3%) increase in TFP. In comparison to the estimates reported in Columns 2 and 4 of Table 13, this corresponds to roughly a 40.0% (41.1%)

[38] In Sample A (Sample B) the relative density of Paris to Lyon is 36.09 (2.50). Alternatively, moving from the average density of markets in the first quartile of the density distribution to the fourth yields similar magnitudes. In Sample A (Sample B) their relative densities are equal to 26.33 (7.10). Between 6.7% and 23.9% of the percent increase in value-added per worker in Sample A, and between 11.1% and 27.4% in Sample B, is accounted for by the organization of firms. Between 8.8% and 36.9% of the percent increase in TFP in Sample A, and between 18.7% and 38.3% in Sample B, is accounted for by the organization of firms.

and 21.0% (21.2%) decrease in the density elasticity across employment areas (urban areas). These results again suggest that organization is important to understanding the productivity of firms across locations.

Columns 3 and 4 estimate TFP with the specification adapted from Caliendo et al. (2015a) and using the method proposed by Wooldridge (2009) to account for endogenous inputs. Column 3 only controls for density and reports that across employment areas (urban areas) a 1% increase in density is associated with a 2.1% (4.1%) increase in TFP. Across employment areas, this leads to a 70.4% and 54.3% decrease in the density elasticity from Columns 5 and 7 of Table 13. The conclusions are similar across urban areas. Accounting for the organization of firms decreases the density elasticity by roughly 67.4% and 53.4%. Column 4 controls for density and the characteristics of local markets. Across employment areas (urban areas) a 1% increase in density is associated with a 2.6% (5.4%) increase in TFP. In comparison to the estimates from Columns 6 and 8 of Table 13 the magnitude on density is approximately 48.0% (50.0%) and 27.7% (29.8%) lower across employment areas (urban areas).

To sum up, across employment areas (urban areas) taking into account the organization of firms in the estimation of TFP, decreases the elasticity of local productivity with respect to density by as little as 21.0% (21.2%) and as much as 70.4% (67.4%). These magnitudes are greater than reported in the previous section, because they involve a comparison between different specifications of the production function. Nonetheless, in broad terms, they imply that when we increase the density of an employment (urban area) the size of Lyon to the size of Paris, between 22.2% and 73.0% (23.7% and 62.6%) of percent increase in local TFP is accounted for by changes in the number of layers in firms.

Overall, the findings are largely consistent with the model. Firms in denser markets are more productive, in part, because they organize with a greater number of layers. The findings are not only statistically significant but economically meaningful as well, and suggest that the mechanism provided in this paper, differences in the way firms organize production, is one reason that firms in denser markets are more productive.

5 Conclusion

This paper provides a novel mechanism to explain the differences in the productivity of firms across locations. Namely, firms in denser markets are more productive because they organize

Table 14: Productivity Regression Results

Dependent Variable from First-Stage	CMORH LP (1)	CMORH LP (2)	CMORH WD (3)	CMORH WD (4)
Sample A: Employment Areas				
log density	0.028	0.030	0.021	0.026
	(0.004)[a]	(0.008)[a]	(0.004)[a]	(0.007)[a]
% Decrease Using Workers	60.0	40.0	70.4	48.0
% Decrease Using Hours	41.6	21.0	54.3	27.7
Sample	A	A	A	A
Method	2SLS	2SLS	2SLS	2SLS
First-Stage Statistics				
KP Wald F-Statistic	2,653	7.16	2,653	7.16
Over-Id Test (p-value)	0.364	0.025	0.442	0.012
Sample Size	22,997	22,997	22,997	22,997
Sample B: Urban Areas				
log density	0.053	0.063	0.041	0.054
	(0.005)[a]	(0.012)[a]	(0.004)[a]	(0.012)[a]
% Decrease Using Workers	57.2	41.1	67.4	50.0
% Decrease Using Hours	42.3	21.2	53.4	29.8
Sample	B	B	B	B
Method	2SLS	2SLS	2SLS	2SLS
First-Stage Statistics				
KP Wald F-Statistic	22.07	7.09	22.07	7.09
Over-Id Test (p-value)	0.110	0.012	0.092	0.012
Sample Size	18,769	18,769	18,769	18,769
Second-Stage: All Controls	No	Yes	No	Yes
Second-Stage: Industry FE	Yes	Yes	Yes	Yes

Notes: a,b,c: significant at the 1%, 5% and 10% level. Clustered standard errors at the local market level are reported in parentheses. Each entry displays a results from equation (24) with TFP estimated using equation (22). % Decrease using workers calculates the difference in the elasticity relative to estimates of TFP using workers and not controlling for the organization of firms. % Decrease using workers calculates the difference in the elasticity relative to estimates of TFP using hours and not controlling for the organization of firms. Industry fixed effects are at the four-digit Nace Rev 1.1. level. Controls in the second-stage are the following: the cost of a unit of labor, the median the annual salary of individuals residing in an area, the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. Columns (1) and (3) instrument for density using density measured in 1881 and 1901. Columns (2) and (4) instrument for density and the characteristics of local markets except for the cost of a unit of labor. In Sample A, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 having access to heating in their residence, and the share of the population in 1968 with a sanitary installation in their residence. In Sample B, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 54 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a heating in their residence, and the share of the population in 1968 with a bathtub or shower in their residence.

with a greater number of layers. I present both theoretical and empirical evidence to justify this claim. The theory of the paper relies on a model that combines the knowledge-based management hierarchies framework of Garicano (2000) and Caliendo and Rossi-Hansberg (2012) with the monopolistic competition framework developed by Ottaviano et al. (2002) and Melitz and Ottaviano (2008). The essence of the argument is that increasing the number of layers in firms involves a tradeoff between fixed and variable costs. Changes in the size of the market determines markups and the relative tradeoff between fixed and variable costs. The model yields two implications on firms that depend on the size of the market. First, firms in bigger markets organize with a greater number of layers, and second, firms in bigger markets are more productive because they operate with a greater number of layers.

Using administrative French data I then examine the model's implications on monopolistically competitive sectors with exclusively local demand: Clothes and Shoes Retailers, Hair and Beauty Salons and Traditional Restaurants. The empirical strategy relies on exploiting cross-sectional variation in the size of local markets. To address endogeneity issues I instrument for the size of markets using historical values, and control for other factors that can affect the organization of firms. The empirical evidence is consistent with the predictions of the model. I first find that firms in denser markets operate with a greater number of layers. Second, I find that organization is an important component of firm productivity. And third, I assess the role of organization in explaining the productivity of firms across locations. Using a variety of measures of firm productivity, I conclude that part of these gains are explained by differences in the organization of firms. For instance, between 6.2% and 36.5% of the value-added per gains from operating in denser urban areas are explained by firms having a greater number of layers. This mechanism and its relative importance to understanding the economic activity of firms across locations, has until now not been investigated in the literature.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83, 6: (2015) 2411–2451.

Antras, Paul, Luis Garicano, and Esteban Rossi-Hansberg. "Offshoring in a Knowledge Economy." *Quarterly Journal of Economics* 121, 1: (2006) 31–77.

———. *Organizing Offshoring: Middle Managers and Communication Costs*. Cambridge, US: Harvard University Press, 2008, e. helpman, d. marin, t. vierdier eds. edition.

Antras, Paul, and Esteban Rossi-Hansberg. "Organizations and Trade." *Annual Review of Economics* 43–64.

Bacheré, Hervé. "Les Très Petites Entreprises, 2 Millions d' Unités Très Diverses." *Insee Focus* 24.

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. "Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing." *Stata Journal* 7, 4.

Becker, Gary S., and Kevin M. Murphy. "The Division of Labor, Coordination Costs, and Knowledge." *The Quarterly Journal of Economics* 107, 4: (1992) 1137–1160.

Bellone, Flora, Patrick Musso, Lionel Nesta, and Frederic Warzynski. "International Trade and Firm-Level Markups when Location and Quality Matter." *Journal of Economic Geography* forthcoming.

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen. "The Distinct Effects of Information Technology and Communication Technology on Firm Organization." *Management Science* 60, 12: (2014) 2859–2885.

Briant, Anthony, Pierre-Philippe Combes, and Miren Lafoucade. "Dots to Boxes: Do the Size and Shape of Spatial Units Jeopardize Economic Geography Estimations?" *Journal of Urban Economics* 67, 3: (2010) 287–302.

Caliendo, Lorenzo, Luca David Mion, Giordano Opromolla, and Esteban Rossi-Hansberg. "Productivity and Organization in Portuguese Firms." Working Paper 21811, National Bureau of Economic Research, 2015a.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. "Anatomy of French Production Hierarchies." *NBER* Working Paper No. 18259.

———. "Anatomy of French Production Hierarchies." *Journal of Political Economy* 123, 4: (2015b) 809–852.

Caliendo, Lorenzo, and Esteban Rossi-Hansberg. "The Impact of Trade on Organization and Productivity." *Quarterly Journal of Economic* 127, 3: (2012) 1393–1467.

Cambpell, Jeffrey R., and Hugo A. Hopenhayn. "Market Size Matters." *The Journal of Industrial Economics* 53, 1: (2005) 1–25.

Ciccone, Antonio, and Robert E. Hall. "Productivity and the Density of Economic Activity." *American Economic Review* 86, 1: (1996) 54–70.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. "Spatial Wage Disparities: Sorting Matters." *Journal of Urban Economics* 63, 2: (2008) 723–742.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sebastien Roux. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection." *Econometrica* 80, 6: (2012) 2543–2594.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sebastien Roux. "Estimating Agglomeration Economies with History, Geology, and Worker Effects." In *Agglomeration Economics*, edited by Edward L. Glaeser, University of Chicago Press, 2010, chapter 1, 15âĂŞ66.

Combes, Pierre-Philippe, and Laurent Gobillon. "The Empirics of Agglomeration Economies." In *Handbook of Regional and Urban Economics*, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, North-Holland, 2015, volume 5, chapter 5, 247–348.

Cragg, John G., and Stephen Donald. "Testing Identifiability and Specification in Instrumental Variable Models." *Econometric Theory* 9, 2: (1993) 222–240.

Dasgupta, Kunal. "Learning and Knowledge Diffusion in a Global Economy." *Journal of International Economics* 87, 2: (2012) 323–336.

Duranton, Gilles, and Hubert Jayet. "Is the Division of Labour Limited by the Extent of the Market? Evidence from French Cities." *Journal of Urban Economics* 69, 1: (2011) 56–71.

Duranton, Gilles, and Diego Puga. "Micro-Foundations of Urban Agglomeration Economies." In *Handbook of Regional and Urban Economics*, edited by J. Vernon Henderson, and Jacques-Francois Thisse, North-Holland, 2004, volume 4, chapter 10, 2063–2117.

Foster, Lucia, John Haltiwanger, and Chad Syverson. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98, 1: (2008) 394–425.

Friedrich, Benjamin. "Trade Shocks, Firm Hierarchies and Wage Inequality.", 2015. Working Paper.

Garicano, Luis. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economics* 108, 5: (2000) 874–904.

Garicano, Luis, and Paul Heaton. "Information Technology, Organization, and Productivity in the Public Sector: Evidence from Police Departments." *Journal of Labor Economics* 28, 1: (2010) 167–201.

Garicano, Luis, and Thomas N. Hubbard. "Managerial Leverage is Limited by the Extent of the Market: Hierarchies, Specialization, and the Utilization of Lawyer's Human Capital." *Journal of Law and Economics* 50, 1: (2007) 1–43.

Garicano, Luis, and Esteban Rossi-Hansberg. "Inequality and the Organization of Knowledge." *American Economic Review Papers and Proceedings* 50, 1: (2004) 197–202.

———. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics* 121, 4: (2006) 1383–1435.

———. "Knowledge-based Hierarchies: Using Organizations to Understand the Economy." *Working Paper* .

Glaeser, Edward L., and David C. Mare. "Cities and Skills." *Journal of Labor Economics* 19, 2: (2001) 316–342.

Guadalupe, Maria, and Julie M. Wulf. "The Flattening Firm and Product Market Competition: The Effect of Trade Liberalization on Corporate Hierarchies." *American Economic Journal: Applied Economics* 2, 4: (2010) 105–127.

Hahn, Jinyong, Jerry Hausman, and Guido Kuersteiner. "Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations." *Econometrics Journal* 7, 1: (2004) 272–306.

Handbury, Jessie, and David E. Weinstein. "Goods Prices and Availability in Cities." *Review of Economic Studies* forthcoming.

Haskel, Jonathan, and Raffaella Sadun. "Regulation and UK Retailing Productivity: Evidence from Micro Data." *IZA* DP No. 4028.

Holmes, Thomas J., and James A. Jr. Schmitz. "Competition and Productivity: A Review of the Evidence." *Annual Review of Economics* 2: (2010) 105–127.

Holmes, Thomas J., and Holger Sieg. "Structural Estimation in Urban Economics." In *Handbook of Regional and Urban Economics*, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, North-Holland, 2015, volume 5, chapter 2, 69–114.

Jofre-Monseny, Jordi, Raquel Marin-Lopez, and Elisabet Viladecans-Marsal. "The Mechanisms of Agglomeration: Evidence from the Effect of Inter-Industry Relations on the Location of New Firms." *Journal of Urban Economics* 70, 2-3: (2011) 61–74.

Kleibergen, Frank, and Richard Paap. "Generalized Reduced Rank Tests Using the Singular Value Decomposition." *Journal of Econometrics* 133, 1: (2006) 97–126.

Levinsohn, James, and Kamil Petrin. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies* 70, 2: (2003) 317–342.

Melitz, Marc J., and Giancarlo I.P. Ottaviano. "Market Size, Trade and Productivity." *Review of Economic Studies* 75, 1: (2008) 295–316.

Melo, Patricia C., Daniel J. Graham, and Rober B. Noland. "A Meta-Analysis of Estimates of Urban Agglomeration Economies." *Regional Science and Urban Economics* 39, 3: (2009) 332–342.

Olley, G. Steven, and Ariel Pakes. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64, 6: (1996) 1263–1297.

Ottaviano, Gianmarco P.I., Takatoshi Tabuchi, and Jacques-Francois Thisse. "Agglomeration and Trade Revisited." *International Economic Review* 43, 2: (2002) 409–435.

Puga, Diego. "The Magnitude and Causes of Agglomeration Economies." *Journal of Regional Science* 50, 1.

Raghuram, Rajan G., and Julie Wulf. "The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies." *The Review of Economics and Statistics* 88, 4: (2006) 759–773.

Rosen, Sherwin. “Specialization and Human Capital.” *Journal of Labor Economics* 1, 1: (1983) 43–49.

Rosenthal, Stuart S., and Stephen L. Ross. “Change and Persistence in the Economic Status of Neighborhoods and Cities.” In *Handbook of Regional and Urban Economics*, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, North-Holland, 2015, volume 5, chapter 16, 1047–1120.

Rosenthal, Stuart S., and William C. Strange. “Evidence on the Nature and Sources of Agglomeration Economics.” In *Handbook of Regional and Urban Economics*, edited by J. Vernon Henderson, and Jacques-Francois Thisse, North-Holland, 2004, volume 4, chapter 49, 2119–2171.

Sanderson, Eleanor, and Frank Windmeijer. “A weak instrument -test in linear {IV} models with multiple endogenous variables.” *Journal of Econometrics* 190, 2: (2016) 212–221.

Sato, Yasuhiro, Takatoshi Tabuchi, and Yamamoto Kazuhiro. “Market Size and Entrepreneurship.” *Journal of Economic Geography* 12, 6: (2012) 1139–1166.

Staiger, Douglas, and James H. Stock. “Instrumental Variables Regression with Weak Instruments.” *Econometrica* 65, 3: (1997) 557–586.

Stigler, George J. “The Division of Labor is Limited by the Extent of the Market.” *Journal of Political Economy* 59, 3: (1951) 185–193.

Stock, James H., and Motohiro Yogo. “Testing for Weak Instruments in Linear IV Regression.” In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by Donald W. K. Andrews, and James H. Stock, Cambridge University Press, 2005, chapter 5, 80–108.

Syverson, Chad. “Market Structure and Productivity: A Concrete Example.” *Journal of Political Economy* 112, 6: (2004) 1181–1222.

———. “What Determines Productivity?” *Journal of Economic Literature* 49, 2: (2011) 326–365.

Wooldridge, Jeffrey M. “On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables.” *Economics Letters* 104, 3: (2009) 112–114.

Zimmerman, Donald W. “Comparative Power of Student t Test and Mann-Whitney U Test for Unequal Sample Sizes and Variances.” *The Journal of Experimental Education* 55, 3: (1987) 171–174.

A Appendix

A.1 Cost Minimization Problem

Consider a one layer organization, created by a self-employed entrepreneur producing output q. Because she is the only worker in the firm, she spends her time generating production problems. To reduce her costs and increase her productivity she will always choose to acquire knowledge that allows her to solve the commonest problems before learning how to solve rarer problems. Her knowledge set is therefore from 0 to z_1^1, she is able to solve $F(z) = 1 - \exp^{-\lambda z_1^1}$ fraction of problems, and the expected output of the organization is: $A\left[1 - \exp^{-\lambda z_1^1}\right]$. Therefore z_1^1 is equal to: $\frac{1}{\lambda}\ln\left(\frac{A-q}{A}\right)$. In a firm agents are compensated for their one unit of time and the knowledge they acquire, and so the total cost to the firm is: $w(cz_1^1 + 1)$.

Now consider a firm with L layers, 1 layer of production workers and $L-1$ layers of managers producing output q. Garicano (2000) characterizes the cost efficient way for a firm to organize its production. The optimal organization will have the following properties. First, the frequency of problems that agents can solve is decreasing with their position in the firm. And second, organizations will never contain layers with overlapping intervals of knowledge. In other words production workers will learn to solve problems from 0 to z_L^1, and managers in layer l will learn to solve problems from z_L^{l-1} to z_L^l. The intuition for these results is the following. The objective of organizations is to better match problems with agents who know the solution. By allowing agents in the lower layers of an organization to solve common problems, a firm ensures that managers with the knowledge to solve infrequent problems avoid common problems, and are able to leverage their knowledge over more problems.[39] By not allowing the knowledge of the different layers to overlap, firms reduce redundancies in their organization and their overall costs.

In an organization with L layers each agent has one unit of time, and as production workers are the only agents to draw problems the number of problems to be solved in the organization is equal to the number of agents in layer 1, n_L^1. The number of managers in each layer of an organization is determined by three factors: the time managers devote to listening to a problem, h, the number of problems generated in the organization, and the fraction of problems agents in the layers below cannot solve. In particular the size, n_L^l, of managers in layer l is given by the following equation: $hn_L^1\exp^{-\lambda Z_L^{l-1}} = n_L^l$, where $Z_L^{l-1} = \sum_{l=1}^{l-1} z_L^l$ is the cumulative knowledge at layer l.[40] In addition the total output of an organization with L layers is determined by four components: number of production problems created in the firm, the fraction of problems that are solved in the organization, the costs of communication between managers and production workers, and the units of output that are created when a problem is solved. In other words, the total expected output of an organization with L layers is equal to: $An_L^1\left[1 - \exp^{-\lambda Z_L^L}\right]$.

In the firm, agents are compensated for their one unit of time and the knowledge they acquire. For a given number of layers L and a given quantity q, the entrepreneur decides the number of employees to hire in each layer, n_L^l, and the knowledge of the employees at a given layer, z_L^l, with the objective to minimize costs. The cost minimization problem of an organization with $L \geq 1$

[39] The efficient organizational structure is related to Rosen (1983): "the incentives for specialization ... arise from increasing returns to the utilization of human capital." The cost of acquiring knowledge is independent of how the knowledge is used in the firm. A firm can therefore increase its efficiency by organizing production in such a way that maximizes the utilization rate of each agents' knowledge.

[40] From the equations that characterize the size of each layer, one can derive an expression for the number of workers in the layer below that a manager supervises, the span of control of managers in layer l. See Caliendo and Rossi-Hansberg (2012) for these expressions.

layers is, therefore, the following:

$$C_L(q) = \min_{\{n_L^l, z_L^l\} \geq 0} \sum_{l=1}^{L} n_L^l [c z_L^l + 1] \tag{25}$$

subject to

$$A \left[1 - \exp^{-\lambda Z_L^L}\right] n_L^1 \geq q, \tag{26}$$

$$n_L^l = n_L^1 h \exp^{-\lambda Z_L^{l-1}} for L \geq l \geq 2, \tag{27}$$

$$n_L^L = 1, \tag{28}$$

where $Z_L^l = \sum_{l=1}^{l} z_L^l$ is the cumulative knowledge at layer l. As production workers are the only agents in the organization to draw problems, there are in total n_L^1 problems in the firm. The first constraint indicates that the total output produced by the firm has to be at least q units of output. The second constraint determines the size of each layer l while the last constraint ensures that the entrepreneur supplies all of her time to the firm.[41]

Furthermore for a given number of layers, L, the marginal cost function of a firm is equal to:

$$MC_L(q) = \frac{ch}{\lambda A} \exp^{\lambda z_L^L}. \tag{29}$$

The cost function of the firm with two layers is equal to:

$$C_2(q) = \frac{c}{\lambda} \left(\frac{h}{A} \exp^{\lambda z_2^2} q + \left(1 - \frac{\exp^{\lambda z_2^1}}{h}\right) + \lambda z_2^2 + \frac{\lambda}{c} \right), \tag{30}$$

where z_2^2 and z_2^1 are the solutions to the cost minimization problem, and cost function of the firm with $L \geq 3$ layers is equal to:

$$C_L(q) = \frac{c}{\lambda} \left(\frac{h}{A} \exp^{\lambda z_L^L} q + \left(1 - \exp^{\lambda z_L^{L-1}}\right) + \lambda z_L^L + \frac{\lambda}{c} \right), \tag{31}$$

where z_L^L and z_L^{L-1} are the solutions to the cost minimization problem. Furthermore the average cost function of a firm with L layers is: $AV_L(q) = \frac{C_L(q)}{q}$.

[41] As explained in Caliendo and Rossi-Hansberg (2012) the last constraint is not just a normalization. It ensures that organizations cannot be replicated at the minimum efficient scale.

A.2 Proofs

A.2.1 Proof of Proposition 2

We first derive the results for quantity $q(\alpha)$. We have:

$$\frac{\partial q(\alpha)}{\partial \alpha} = \frac{N}{2\gamma}\left[1 - \frac{\partial MC(q(\alpha))}{\partial q(\alpha)}\frac{\partial q(\alpha)}{\partial \alpha}\right].$$

Rearranging yields:

$$\frac{\partial q(\alpha)}{\partial \alpha} = \frac{\frac{N}{2\gamma}}{1 + \frac{\partial MC(q(\alpha))}{\partial q(\alpha)}\frac{N}{2\gamma}}.$$

Since within layers $\frac{\partial MC(q(\alpha))}{\partial q(\alpha)} > 0$, quantity is increasing with α.

Now moving onto prices. Since

$$\frac{\partial p(\alpha)}{\partial \alpha} = \frac{1}{2}\left[1 + \frac{\partial MC(q(\alpha))}{\partial q(\alpha)}\frac{\partial q(\alpha)}{\partial \alpha}\right],$$

and within layers, marginal costs are increasing with quantity, and quantity is increasing with α, $p(\alpha)$ is increasing with respect to α.

Now moving onto markups over marginal costs. Since

$$\frac{\partial \mu^{MC}(\alpha)}{\partial \alpha} = \frac{1}{2}\left[1 - \frac{\partial MC(q(\alpha))}{\partial q(\alpha)}\frac{\partial q(\alpha)}{\partial \alpha}\right],$$

and by substituting the expression for $\frac{\partial q(\alpha)}{\partial \alpha}$, within layers $\mu(\alpha)$ is increasing with respect to α.

Now moving onto markups over average costs. By definition markups over marginal costs, and markups over average costs are equal to:

$$\mu^{MC}(\alpha) = p(\alpha) - MC(q(\alpha)),$$

$$\mu^{AC}(\alpha) = p(\alpha) - AC(q(\alpha)).$$

It therefore follows that:

$$MC(q(\alpha)) - AC(q(\alpha)) = \mu^{AC}(\alpha) - \mu^{MC}(\alpha).$$

Marginal costs are increasing with quantity, while average cost curves are convex and attain their minimum when they intersect their associated marginal cost curve. It therefore follows that:

$$\frac{\partial\left[MC(q(\alpha)) - AC(q(\alpha))\right]}{\partial \alpha} = \frac{\partial\left[MC(q(\alpha)) - AC(q(\alpha))\right]}{\partial q(\alpha)}\frac{\partial q(\alpha)}{\partial \alpha} > 0,$$

which in turn implies that:

$$\frac{\partial \mu^{AC}(\alpha)}{\partial \alpha} > \frac{\partial \mu^{MC}(\alpha)}{\partial \alpha}.$$

Since $\frac{\partial \mu^{MC}(\alpha)}{\partial \alpha} > 0$ it follows that $\frac{\partial \mu^{AC}(\alpha)}{\partial \alpha} > 0$.

Now moving onto revenues. Since

$$\frac{\partial r(\alpha)}{\partial \alpha} = q(\alpha)\frac{\partial p(\alpha)}{\partial \alpha} + p(\alpha)\frac{\partial q(\alpha)}{\partial \alpha},$$

it follows that within layers, revenues are increasing with respect to α.

Now moving onto profits. From the maximization problem we know that,

$$\pi(\alpha) = \left[\alpha - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) - \frac{\gamma}{N}q^*(\alpha)\right] q^*(\alpha) - C(q^*(\alpha)),$$

where $*$ denotes the optimal quantities chosen. By the envelope theorem,

$$\frac{\partial \pi(\alpha)}{\partial \alpha} = q^*(\alpha).$$

Thus profits are increasing with respect to α. Since the optimal quantity produced, $q^*(\alpha)$ is increasing with the number of layers L, it follows that the slope of the profit function $\pi(\alpha)$ is increasing with L. And since,

$$\frac{\partial^2 \pi(\alpha)}{\partial^2 \alpha} = \frac{\partial q^*(\alpha)}{\partial \alpha}.$$

$\frac{\partial q^*(\alpha)}{\partial \alpha}$ is positive, profits are convex.

I now show that holding the number of layer fixed, profits are concave with respect to q. We know that

$$\pi(\alpha) = \quad p(\alpha)q(\alpha) - C(q(\alpha)).$$

Substituting in for $p(\alpha) = \alpha - \frac{2\gamma}{N}q - \frac{\eta M}{\eta M + \gamma}(\overline{\alpha} - \overline{p})$, and taking the second derivative with respect to q yields:

$$\frac{\partial^2 \pi(\alpha)}{\partial^2 q} = -\frac{2\gamma}{N} - \frac{\partial MC(q(\alpha))}{\partial q},$$

which is negative. Thus profits are concave in q.

A.2.2 Proof of Proposition 3

I first derive the results with respect to quantity $q(\alpha)$. Since

$$\frac{\partial q(\alpha)}{\partial MC} = -\frac{N}{2\gamma},$$

when the firm increases the number of layers, marginal costs decrease discontinuously and quantity increases discontinuously.

Now onto prices. Since

$$\frac{\partial p(\alpha)}{\partial MC} = \frac{1}{2},$$

when the firm increases the number of layers, marginal costs decrease discontinuously, and thus prices decrease discontinuously as well.

Now moving on to markups over marginal costs. Since

$$\frac{\partial \mu^{MC}(\alpha)}{\partial MC} = -\frac{1}{2},$$

when the firm increases the number of layers, marginal costs decrease discontinuously, and thus markups over marginal costs increase discontinuously as well.

Now moving onto markups over average costs. Markups over average costs are also equal to: $\mu^{AC}(\alpha) = \pi(\alpha)/q(\alpha)$. The numerator of the derivative of this expression with respect to marginal costs is equal to:

$$q(\alpha)\frac{\partial \pi(\alpha)}{\partial MC} + \pi(\alpha)\frac{\partial q(\alpha)}{\partial MC}.$$

The first term is equal to: $q(\alpha)\frac{\partial \pi(\alpha)}{\partial MC} = q(\alpha)\left[-\frac{N}{2\gamma}MC(q(\alpha)) - q(\alpha)\right]$, where I have used the expression for $\frac{\partial r(\alpha)}{\partial MC}$ derived below. The second term is equal to $\pi(\alpha)\frac{\partial q(\alpha)}{\partial MC} = -\pi(\alpha)\frac{N}{2\gamma}$. Using the expression for profits, $\pi(\alpha) = \frac{\gamma}{N}q(\alpha)^2 + q(\alpha)MC(q(\alpha)) - C(q(\alpha))$, and eliminating common terms implies that the numerator is equal to:

$$-\frac{q(\alpha)^2}{2} - \frac{N}{2\gamma}C(q(\alpha)),$$

which is negative. The denominator of the derivative of $\mu^{AC}(\alpha) = \pi(\alpha)/q(\alpha)$ with respect to marginal costs is equal to $q(\alpha)^2$ which is always positive. It therefore follows that when the firm increases the number of layers, marginal costs decrease discontinuously, and thus markups over average costs increase discontinuously as well.

Now moving onto revenues. Since

$$\frac{\partial r(\alpha)}{\partial MC} = -\frac{N}{2\gamma}MC(q(\alpha)),$$

as the firm increases the number of layers, marginal costs decrease discontinuously, revenues increase discontinuously as well.

Now moving onto profits. To show that profits are strictly increasing with respect to α and convex the arguments are similar to the proof of Proposition 2. Here I simply show that profits are continuous. Consider an entrepreneur that is indifferent between producing with layers L and $L+1$. Then it follows:

$$\pi_L(\alpha) = \pi_{L+1}(\alpha).$$

Since, within layers profits are continuous and when an entrepreneur is indifferent between layers L and $L+1$ profits are equal, thus profits are globally continuous with respect to α.

A.2.3 Proof of Proposition 4

The equilibrium of the model is determined from the zero-profit condition and the free-entry condition:

$$\pi(\alpha_D, M) = 0, \tag{32}$$

$$\int_{\alpha_D} \pi(\alpha, M) dG(\alpha) = f_E, \tag{33}$$

where M denotes the mass of firms operating in equilibrium and α_D is the demand draw of the entrepreneur that is indifferent between entering and exiting the market.

First I transform the equilibrium to be a function of q_D and α_D. From the first order condition of the firm's maximization problem, for a given α, quantity is determined by the equation:

$$q(\alpha) = \frac{N}{2\gamma}\left[\alpha - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) - MC(q(\alpha))\right].$$

For a given α there exists a unique quantity $q(\alpha)$, that is a solution to the expression of above. Rewriting this equation yields an expression of the term $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$ as a function the demand draw, α, and the optimal quantity produced, $q(\alpha)$:

$$\frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) = \alpha - MC(q(\alpha)) - \frac{2\gamma}{N} q(\alpha).$$

Substituting this expression in the profit of the firm yields:

$$\pi(\alpha, q(\alpha)) = \frac{\gamma}{N} q(\alpha)^2 + MC(q(\alpha))q(\alpha) - C(q(\alpha)).$$

Doing the same for the marginal firm yields the expression for profit:

$$\pi(q_D) = \frac{\gamma}{N} q_D^2 + MC(q_D)q_D - C(q_D).$$

For a firm with demand draw α, I rewrite quantities, prices, markups and revenues as a function of q_D, p_D, μ_D and r_D, and the parameters of the model. This yields the following expressions:

$$\begin{aligned}
q(\alpha) &= q_D + \frac{N}{2\gamma}\left[\alpha - \alpha_D + MC(q_D) - MC(q(\alpha))\right], \\
p(\alpha) &= p_D + \frac{1}{2}\left[\alpha - \alpha_D - MC(q_D) + MC(q(\alpha))\right], \\
\mu(\alpha) &= \mu_D + \frac{1}{2}\left[\alpha - \alpha_D + MC(q_D) - MC(q(\alpha))\right], \\
r(\alpha) &= r_D + q_D \frac{1}{2}\left[\alpha - \alpha_D - MC(q_D) + MC(q(\alpha))\right] + p_D \frac{N}{2\gamma}\left[\alpha - \alpha_D + MC(q_D) - MC(q(\alpha))\right] \\
&\quad + \frac{N}{4\gamma}\left[(\alpha - \alpha_D)^2 - (MC(q_D) - MC(q(\alpha)))^2\right].
\end{aligned}$$

The equilibrium is now determined by the solution to the following three equations:

$$\frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) = \alpha_D - MC(q_D) - \frac{2\gamma}{N} q_D. \tag{34}$$

$$ZCP = \pi(\alpha_D, q_D) = 0, \tag{35}$$

$$FE = \int_{\alpha_D} \pi(\alpha, \alpha_D, q_D) dG(\alpha) - f_E = 0, \tag{36}$$

Here M, q_D and α_D are variables that are to be determined. Note that the solution to equation (35) depends only on q_D and the parameters of the model. Given the solution to equation (35), equation (36) is only a function of α_D and the parameters of the model. Finally, once q_D and α_D are both determined, N is determined equation (34). Therefore to prove that a solution exists, I need to show that there exists a q_D and α_D such that equations (35) and (36) are satisfied, and that a unique solution exists to (34).

First, I show that a solution to equation (35) exists. First, consider the slope of the profit function:

$$\frac{\partial \pi(q_D)}{\partial q_D} = \frac{2\gamma}{N} q_D + q_D \frac{\partial MC(q_D)}{\partial q_D} > 0.$$

When q_D is sufficiently large, $MC(q_D) > AC(q_D)$ and it follows that $\pi(q_D) = \frac{2\gamma}{N} q_D^2 + q_D MC(q_D) - C(q_D) > 0$. Since $\lim_{q_D \to 0} C(q_D) = w$ and $\lim_{q_D \to 0} MC(q_D) = 0$ it follows that $\lim_{q_D \to 0} \pi(q_D) < 0$. Therefore, there exists a unique q_D exists such that $\pi(q_D) = 0$.

Second, consider the equation (36). Using Leibniz's integral rule, the slope of the free-entry condition is:

$$\frac{\partial FE}{\partial \alpha_D} = -\pi(\alpha, \alpha_D, q_D) dG(\alpha_D) + \int_{\alpha_D} \frac{\partial \pi(\alpha, \alpha_D, q_D)}{\partial \alpha_D} dG(\alpha).$$

The first term by definition is equal to zero. I now show that the second term is positive. Using the expression for profits, and after eliminating common terms, it follows that:

$$\frac{\partial \pi(\alpha, \alpha_D, q_D)}{\partial \alpha_D} = \left[\frac{2\gamma}{N} q(\alpha) + q(\alpha) \frac{\partial MC(q(\alpha))}{\partial q(\alpha)} \right] \frac{\partial q(\alpha)}{\partial \alpha_D},$$

where from the expression of quantity it follows that:

$$\frac{\partial q(\alpha)}{\partial \alpha_D} = -\frac{\frac{N}{2\gamma}}{1 + \frac{N}{2\gamma} \frac{\partial MC(q(\alpha))}{\partial q(\alpha}} < 0.$$

Therefore FE is downward sloping. Further, when $\alpha_D = \alpha_M$, $FE > 0$, and in the limit, when α_D approaches infinity $\lim_{\alpha_D \to \infty} FE < 0$. Hence, there exists a unique α_D such that $FE = 0$.

For a given number of layers L, there exists a solution to $\pi_L(\alpha_D^L, q_D^L) = 0$, and thus there are a discrete set of potential solutions. I now show that from this set, there is only one combination of α_D^L, q_D^L that satisfies the equilibrium.

Suppose not. Consider two possible solutions α_D^L, q_D^L and $\alpha_D^{L+1}, q_D^{L+1}$, associated with organizations with L and $L+1$ layers respectively. Without loss of generality, assume that $\alpha_D^L < \alpha_D^{L+1}$. In this case it follows that $\pi_L(\alpha_D^L, q_D^L) = \pi_{L+1}(\alpha_D^{L+1}, q_D^{L+1}) = 0$. By Proposition 2 it follows that all firms with demand draws in the interval $\left[\alpha_D^L, \alpha_D^{L+1}\right]$, will earn positive profits producing with an organization with L layers. By Proposition 2 it also follows that for the entrepreneur with demand draw α_D^{L+1}, $\pi_L(\alpha_D^{L+1}) > \pi_{L+1}(\alpha_D^{L+1}) = 0$, and so he will earn positive profits producing with an organization with L layers. Therefore $\alpha_D^{L+1}, q_D^{L+1}$ is not an equilibrium solution.

I now show that a unique M exists that satisfies equation (34). Since prices can be written as a function of α_D and q_D, it follows that $\overline{p}$ is solely a function of α_D and q_D. Hence because the right-hand side of equation (34) is constant, while the left-hand side is increasing with respect to M, a unique solution exists and M is equal to:

$$M = \frac{\gamma}{\eta} \frac{\alpha_D - MC(q_D) - \frac{2\gamma}{N} q_D}{(\overline{\alpha} - \overline{p}) - \alpha_D + MC(q_D) + \frac{2\gamma}{N} q_D}$$

In equilibrium labor markets also clear. Labor is used for several purposes, as workers in the homogeneous sectors, as workers and managers in the differentiated good sector, as teachers, and to design new products. Let H be the mass of workers in the homogeneous good sector. As the total mass of agents in the economy is given by N, the labor market clearing condition is given by:

$$H + \frac{M}{1 - G(\alpha_D)} \left[f_E + \int_{\alpha_D} C(q(\alpha)) dG(\alpha) \right] = N. \tag{37}$$

I now show that if $\eta > \underline{\eta}$ both the homogeneous and differentiated goods will be produced in equilibrium. For simplicity, I define $K = \alpha_D - MC(q_D) - \frac{2\gamma}{N} q_D$ and $B = \int_{\alpha_D} \frac{\gamma}{N} q(\alpha) \frac{g(\alpha)}{1 - G(\alpha_D)}$. First consider the term $\overline{\alpha} - \overline{p}$. This can be rewritten as:

$$\begin{aligned} \overline{\alpha} - \overline{p} &= \int_{\alpha_D} \left[\alpha - \left[\alpha - K - \frac{\gamma}{N} q(\alpha) \right] \right] \frac{g(\alpha)}{1 - G(\alpha_D)} \\ &= K + B. \end{aligned}$$

It then follows that:

$$\begin{aligned} \frac{\eta M}{\gamma + \eta M} (\overline{\alpha} - \overline{p}) &= \frac{\eta M}{\gamma + \eta M} [K + B] \\ &= K. \end{aligned}$$

And by isolating terms it follows that:

$$M = \frac{\gamma K}{\eta B}. \tag{38}$$

Next consider the equilibrium condition:

$$N - \frac{M}{1 - G(\alpha_D)} \left[[1 - G(\alpha_D)] f_E + \int_{\alpha_D} C(q(\alpha)) dG(\alpha) \right] > 0.$$

which can simply be rewritten as $N > M\overline{r}$. By substituting in the expression for M from above, it follows that:

$$\eta > \frac{\gamma K \overline{r}}{NB}. \tag{39}$$

Thus if $\eta > \underline{\eta} = \frac{\gamma K \bar{r}}{NB}$ both the homogeneous and differentiated goods will be produced in equilibrium.

A.2.4 Proof of Proposition 5

Consider an increase in N. The proof proceeds in steps. I first show that q_D and α_D increase with N. I then show that the term $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$ increases with N. And finally, I show under what conditions M increases with N.

From the zero-profit equation, it follows that:

$$0 = -\frac{\gamma}{N^2}q_D^2 + \frac{2\gamma}{N}q_D\frac{\partial q_D}{\partial N} + q_D\frac{\partial MC(q_D)}{\partial q_D}\frac{\partial q_D}{\partial N},$$

which after rearranging terms yields the result:

$$\frac{\partial q_D}{\partial N} = \frac{\frac{\gamma}{N^2}q_D}{\frac{2\gamma}{N} + \frac{\partial MC(q_D)}{\partial q_D}} > 0, \tag{40}$$

Since the denominator and numerator are both positive.

Now, consider the equation characterizing the expected profits of entry V^e:

$$\int_{\alpha_D} \pi(\alpha, \alpha_D, q_D)dG(\alpha) = f_E.$$

From this equation, it follows that:

$$\frac{\partial \alpha_D}{\partial N} = -\frac{\partial V^e/\partial N + \partial V^e/\partial q_D\, \partial q_D/\partial N}{\partial V^e/\partial \alpha_D}. \tag{41}$$

In the proof of Proposition 4, I showed that the denominator in equation (41) is negative. I now show that the numerator is positive. The profit of a firm with demand draw α is:

$$\pi(\alpha, \alpha_D, q_D) = \left[\frac{2\gamma}{N}q_D + MC(q_D) - \alpha_D + \alpha - \frac{\gamma}{N}q(\alpha)\right]q(\alpha) - C(q(\alpha)). \tag{42}$$

Using the envelope theorem and taking the derivative with respect to N and q_D yields:

$$\begin{aligned}
\frac{\partial \pi(\alpha, \alpha_D, q_D)}{\partial N} + \frac{\partial \pi(\alpha, \alpha_D, q_D)}{\partial q_D}\frac{\partial q_D}{\partial N} &= \left[\frac{2\gamma}{N}q_D + MC(q_D) - \alpha_D + \alpha - \frac{\gamma}{N}q(\alpha)\right]q(\alpha) - C(q(\alpha)) \\
&= \left[-\frac{2\gamma}{N^2}q_D + \frac{2\gamma}{N}\frac{\partial q_D}{\partial N} + \frac{\partial MC(q_D)}{\partial N}\frac{\partial q_D}{\partial N} + \frac{\gamma}{N^2}q(\alpha)\right]q(\alpha) \\
&= \left[\frac{\gamma}{N^2}(q(\alpha) - q_D)\right]q(\alpha),
\end{aligned} \tag{43}$$

where here α_D is held fixed and I substituted for $\frac{\partial q_D}{\partial N}$ using equation (40). Therefore since

$$\frac{\partial V^e}{\partial N} + \frac{\partial V^e}{\partial q_D}\frac{\partial q_D}{\partial N} = \int_{\alpha_D}\left[\frac{\partial \pi(\alpha, \alpha_D, q_D)}{\partial N} + \frac{\partial \pi(\alpha, \alpha_D, q_D)}{\partial q_D}\frac{\partial q_D}{\partial N}\right]dG(\alpha)$$

the numerator in equation (41) is positive and α_D is increasing with respect to N.

Rearranging the first order condition of the firm's maximization problem gives:

$$\alpha_D - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) = \frac{2\gamma}{N} q_D + MC(q_D),$$

and taking the derivative of this expression with respect to N yields:

$$\frac{\partial \alpha_D}{\partial N} - \frac{\partial \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p})}{\partial N} = -\frac{\gamma}{N^2} q_D.$$

Hence it follows that:

$$\frac{\partial \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p})}{\partial N} = \frac{\partial \alpha_D}{\partial N} + \frac{\gamma}{N^2} q_D > 0.$$

I now show that if $\overline{\eta} > \eta$ the mass of firms in the differentiated goods sector, M, will increase with N. In the proof of Proposition 4, I showed that M can be rewritten as:

$$M = \frac{\gamma K}{\eta B}.$$

where $K = \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p})$ and $B = \int_{\alpha_D} \frac{\gamma}{N} q(\alpha) \frac{g(\alpha)}{1 - G(\alpha_D)}$. Taking the derivative of this expression with respect to N yields:

$$\frac{\partial M}{\partial N} = \frac{\gamma}{\eta B^2}\left[B \frac{\partial K}{\partial N} - K \frac{\partial B}{\partial N}\right].$$

Hence $\frac{\partial M}{\partial N}$ is positive if and only if $B\frac{\partial K}{\partial N} - K\frac{\partial B}{\partial N}$ is positive. I do not know the sign of $\frac{\partial B}{\partial N}$. If it is negative then it automatically follows that $\frac{\partial M}{\partial N}$ is positive, and I do not have to make an assumption on η. However assume that $\frac{\partial B}{\partial N}$ is positive. Then using the expression for M from above yields:

$$\frac{\frac{\partial K}{\partial N}}{\frac{\partial B}{\partial N}} > \frac{K}{B} = \frac{M\eta}{\gamma},$$

and by isolating η, one obtains following inequality:

$$\frac{\gamma \frac{\partial K}{\partial N}}{M \frac{\partial B}{\partial N}} > \eta. \tag{44}$$

Hence if $\overline{\eta} = \frac{\gamma \frac{\partial K}{\partial N}}{M \frac{\partial B}{\partial N}} > \eta$, the mass of firms in the differentiated goods sector, M, will increase with N.

I now show that there always exists an η such that $\overline{\eta} > \eta > \underline{\eta}$. Substituting the expressions for both terms yields:

$$\frac{\gamma K \overline{r}}{NB} < \frac{\gamma \frac{\partial K}{\partial N}}{M \frac{\partial B}{\partial N}}. \tag{45}$$

After rearranging terms it follows that:

$$\frac{M\bar{r}}{N} < \frac{B\frac{\partial K}{\partial N}}{K\frac{\partial B}{\partial N}}.$$

By assumption, for η to be in the set $\left[\underline{\eta}, \overline{\eta}\right]$ the following two conditions must simultaneously hold:

$$M\bar{r} < N \qquad \& \qquad K\frac{\partial B}{\partial N} < B\frac{\partial K}{\partial N}.$$

where by assumption $\frac{\partial B}{\partial N}$ is positive. Hence from these two conditions it follows that:

$$\frac{M\bar{r}}{N} < 1 < \frac{B\frac{\partial K}{\partial N}}{K\frac{\partial B}{\partial N}}.$$

Therefore there always exists an $\eta \in \left[\underline{\eta}, \overline{\eta}\right]$.

A.2.5 Proof of Proposition 6

Consider an entrepreneur with demand draw $\alpha_{L,L+1}$ that is indifferent between two organizational forms L and $L+1$. Then it follows that her profits are the same and:

$$\pi_L(\alpha_{L,L+1}, N) = \pi_{L+1}(\alpha_{L,L+1}, N).$$

In this section, I first show how q_L and q_{L+1} change with respect to N. Then I examine how $\alpha_{L,L+1}$ changes with N. Finally, I examine how $\alpha_{L,L+1}$ changes relative to $\alpha_{L+1,L+2}$ with respect to N.

Substituting the expression for profits, implies that:

$$\frac{\gamma}{N}q_L^2 + MC_L(q_L)q_L - C_L(q_L) = \frac{\gamma}{N}q_{L+1}^2 + MC_{L+1}(q_{L+1})q_{L+1} - C_{L+1}(q_{L+1}),$$

where q_L and q_{L+1} are the quantities produced by the entrepreneur with demand draw $\alpha_{L,L+1}$ when she is producing with L or $L+1$ layers. Taking the derivative of this expression with respect to N and eliminating any common terms yields:

$$-\frac{\gamma}{N^2}q_L^2 + \frac{2\gamma}{N}q_L\frac{\partial q_L}{\partial N} + q_L\frac{\partial MC_L}{\partial q_L}\frac{\partial q_L}{\partial N} = -\frac{\gamma}{N^2}q_{L+1}^2 + \frac{2\gamma}{N}q_{L+1}\frac{\partial q_{L+1}}{\partial N} + q_{L+1}\frac{\partial MC_{L+1}}{\partial q_{L+1}}\frac{\partial q_{L+1}}{\partial N}. \tag{46}$$

Since q_L is the optimal quantity supplied by the entrepreneur, it satisfies the first order condition of the firm's maximization problem:

$$q_L = \frac{N}{2\gamma}\left[\alpha_{L,L+1} - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) - MC_L(q_L)\right].$$

Similarly because q_{L+1} is the optimal quantity supplied by the entrepreneur, it satisfies the equation:

$$q_{L+1} = \frac{N}{2\gamma}\left[\alpha_{L,L+1} - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) - MC_{L+1}(q_{L+1})\right].$$

From these two expressions it follows that:

$$MC_L(q_L) + \frac{2\gamma}{N} q_L = MC_{L+1}(q_{L+1}) + \frac{2\gamma}{N} q_{L+1}. \tag{47}$$

Taking the derivative of equation (47) with respect to N yields:

$$\frac{\partial MC_L(q_L)}{\partial q_L}\frac{\partial q_L}{\partial N} - \frac{2\gamma}{N^2} q_L + \frac{2\gamma}{N}\frac{\partial q_L}{\partial N} = \frac{\partial MC_{L+1}(q_{L+1})}{\partial q_{L+1}}\frac{\partial q_{L+1}}{\partial N} - \frac{2\gamma}{N^2} q_{L+1} + \frac{2\gamma}{N}\frac{\partial q_{L+1}}{\partial N}. \tag{48}$$

Multiplying equation (48) by q_{L+1}, substituting this expression into equation (46), and rearranging yields:

$$\frac{\partial q_L}{\partial N}\left[\frac{2\gamma}{N}\left(q_L - q_{L+1}\right) + \left(q_L - q_{L+1}\right)\frac{\partial MC_L}{\partial q_L}\right] = \frac{\gamma}{N^2}\left(q_{L+1} - q_L\right)^2. \tag{49}$$

Since q_L is less than q_{L+1} the term on the right hand-side is positive, while the expression in brackets on the left-hand side is negative. Hence from equation (49) we have:

$$\frac{\partial q_L}{\partial N} < 0. \tag{50}$$

Performing the same steps as above, but multiplying equation (48) by q_L yields:

$$\frac{\partial q_{L+1}}{\partial N}\left[\frac{2\gamma}{N}\left(q_{L+1} - q_L\right) + \left(q_{L+1} - q_L\right)\frac{\partial MC_{L+1}}{\partial q_{L+1}}\right] = \frac{\gamma}{N^2}\left(q_{L+1} - q_L\right)^2. \tag{51}$$

which implies:

$$\frac{\partial q_{L+1}}{\partial N} > 0. \tag{52}$$

Hence for the two quantities q_L and q_{L+1} such that an entrepreneur is indifferent between two organizational forms, q_L is decreasing with N while q_{L+1} is increasing with respect to N. Therefore when controlling for market size, larger firms will have more layers.

I now examine how $\alpha_{L,L+1}$ changes with respect to N. The first order condition of the firm's maximization problem can be rewritten as:

$$q_L = \frac{N}{2\gamma}\left[\alpha_{L,L+1} - \alpha_D + \alpha_D - \frac{\eta M}{\gamma + \eta M}(\overline{\alpha} - \overline{p}) - MC_L(q_L)\right].$$

Taking the derivative of this expression with respect to N and isolating common terms yields:

$$\frac{\partial q_L}{\partial N}\left[1 + \frac{N}{2\gamma}\frac{\partial MC_L(q_L)}{\partial q_L}\right] = \frac{q_L}{N} + \frac{N}{2\gamma}\frac{\partial\left(\alpha_{L,L+1} - \alpha_D\right)}{\partial N} + \frac{N}{2\gamma}\frac{\partial\left(\alpha_D - \frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})\right)}{\partial N}. \tag{53}$$

The expression in brackets is positive and $\frac{\partial q_L}{\partial N}$ is negative, so the left-hand side in equation (53) is negative. Also since:

$$\frac{q_L}{N} + \frac{N}{2\gamma}\frac{\partial\left(\alpha_D - \frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})\right)}{\partial N} = \frac{q_L}{N} - \frac{q_D}{2N} > 0,$$

the distance between $\alpha_{L,L+1}$ and α_D decreases with N. Namely,

$$\frac{\partial\left(\alpha_{L,L+1} - \alpha_D\right)}{\partial N} < 0.$$

I now proceed to analyze how $\alpha_{L,L+1}$ changes with N. Returning to the first order condition of the firm's maximization problem, taking the derivative of this expression with respect to N, and isolating common terms yields:

$$\frac{\partial q_L}{\partial N}\left[1+\frac{N}{2\gamma}\frac{\partial MC_L(q_L)}{\partial q_L}\right]=\frac{q_L}{N}+\frac{N}{2\gamma}\frac{\partial \alpha_{L,L+1}}{\partial N}-\frac{N}{2\gamma}\frac{\partial\left(\frac{\eta M}{\gamma+\eta M}(\overline{\alpha}-\overline{p})\right)}{\partial N}.$$

Since the term on the left-hand side is negative, we have:

$$\frac{\partial\left(\frac{\eta M}{\gamma+\eta M}(\overline{\alpha}-\overline{p})\right)}{\partial N}-\frac{2\gamma}{N^2}q_L>\frac{\partial \alpha_{L,L+1}}{\partial N}.$$

which provides an upper bound to $\frac{\partial \alpha_{L,L+1}}{\partial N}$. Performing the same exercise with respect to q_{L+1} implies:

$$\frac{\partial \alpha_{L,L+1}}{\partial N}>\frac{\partial\left(\frac{\eta M}{\gamma+\eta M}(\overline{\alpha}-\overline{p})\right)}{\partial N}-\frac{2\gamma}{N^2}q_{L+1}.$$

which provides a lower bound to $\frac{\partial \alpha_{L,L+1}}{\partial N}$.

Consider two entrepreneurs who are indifferent producing with two types of organizations. Denote the demand draw of the entrepreneur who is indifferent between L, $L+1$ layers as $\alpha_{L,L+1}$ and the demand draw of the entrepreneur who is indifferent between $L+1$, $L+2$ layers as $\alpha_{L+1,L+2}$. I now examine how $\alpha_{L,L+1}$ and $\alpha_{L+1,L+2}$ change relative to each other with respect to N. From the first order condition of the firm's maximization problem it follows that:

$$q_L(\alpha_{L,L+1})=\frac{N}{2\gamma}\left[\alpha_{L,L+1}-\frac{\eta M}{\gamma+\eta M}(\overline{\alpha}-\overline{p})-MC_L(q_L(\alpha_{L,L+1}))\right],$$

where $q_L(\alpha_{L,L+1})$ is the quantity supplied by the firm with demand draw $\alpha_{L,L+1}$ using an organization with L layers. This implies that:

$$\frac{\eta M}{\gamma+\eta M}(\overline{\alpha}-\overline{p})=\alpha_{L,L+1}-MC_L(q_L(\alpha_{L,L+1}))-\frac{2\gamma}{N}q_L(\alpha_{L,L+1}).$$

Returning to the first order condition of the firm's maximization problem with demand draw $\alpha_{L+1,L+2}$ and substituting in the expression for $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha}-\overline{p})$ from above, it follows that:

$$q_{L+1}(\alpha_{L+1,L+2})=q_{L+1}(\alpha_{L,L+1})+\frac{N}{2\gamma}\left[\alpha_{L+1,L+2}-\alpha_{L,L+1}+MC_{L+1}(q_{L+1}(\alpha_{L,L+1}))-MC_{L+1}(q_{L+1}(\alpha_{L+1,L+2}))\right],$$

where $q_{L+1}(\alpha_{L+1,L+2})$ is the quantity supplied by the firm with demand draw $\alpha_{L+1,L+2}$ using an organization with $L+2$ layers. Taking the derivative of this expression with respect to N and isolating common terms yields:

$$\frac{\partial q_{L+1}(\alpha_{L+1,L+2})}{\partial N}\left[\frac{2\gamma}{N}+\frac{\partial MC_{L+1}(q_{L+1}(\alpha_{L+1,L+2}))}{\partial q_{L+1}(\alpha_{L+1,L+2})}\right]-\frac{\partial q_{L+1}(\alpha_{L,L+1})}{\partial N}\left[\frac{2\gamma}{N}+\frac{\partial MC_{L+1}(q_{L+1}(\alpha_{L,L+1}))}{\partial q_{L+1}(\alpha_{L,L+1})}\right]=$$

$$\frac{2\gamma}{N^2}(q_{L+1}(\alpha_{L+1,L+2})-q_{L+1}(\alpha_{L,L+1}))+\frac{\partial \alpha_{L+1,L+2}}{\partial N}-\frac{\partial \alpha_{L,L+1}}{\partial N}.$$

Since $\frac{\partial q_{L+1}(\alpha_{L+1,L+2})}{\partial N} < 0$ and $\frac{\partial q_{L+1}(\alpha_{L,L+1})}{\partial N} > 0$, the term on the left-hand side is negative, and therefore:

$$-\frac{2\gamma}{N^2}(q_{L+1}(\alpha_{L+1,L+2}) - q_{L+1}(\alpha_{L,L+1})) > \frac{\partial \alpha_{L+1,L+2}}{\partial N} - \frac{\partial \alpha_{L,L+1}}{\partial N}.$$

This provides an upper bound to $\frac{\partial \alpha_{L+1,L+2}}{\partial N} - \frac{\partial \alpha_{L,L+1}}{\partial N}$. Since $q_{L+1}(\alpha_{L+1,L+2}) - q_{L+1}(\alpha_{L,L+1}) > 0$, it follows that:

$$\frac{\partial \alpha_{L+1,L+2}}{\partial N} - \frac{\partial \alpha_{L,L+1}}{\partial N} < 0. \tag{54}$$

Therefore the distance between the demand draws $\alpha_{L,L+1}$ and $\alpha_{L+1,L+2}$ decreases with N. Adding and subtracting $\frac{\partial \alpha_D}{\partial N}$ on the left-hand side, isolating terms and taking into account that $\frac{\partial \alpha_{L+1,L+2}}{\partial N} - \frac{\partial \alpha_D}{\partial N}$ and $\frac{\partial \alpha_{L,L+1}}{\partial N} - \frac{\partial \alpha_D}{\partial N}$ are both negative, implies that the absolute change of $\alpha_{L+1,L+2}$ relative to α_D is greater than the change of $\alpha_{L,L+1}$ relative to α_D.

I can provide a lower bound for this expression as well. Using the same argument as above, but replacing $q_{L+1}(\alpha_{L+1,L+2})$ with the quantities produced by the entrepreneur with demand draw $\alpha_{L+1,L+2}$ using an organization of $L+2$ layers, $q_{L+2}(\alpha_{L+1,L+2})$, and replacing $q_{L+1}(\alpha_{L,L+1})$ with the quantities produced by the entrepreneur with demand draw $\alpha_{L,L+1}$ using an organization of L layers, $q_L(\alpha_{L,L+1})$ yields:

$$\frac{\partial \alpha_{L+1,L+2}}{\partial N} - \frac{\partial \alpha_{L,L+1}}{\partial N} > -\frac{2\gamma}{N^2}(q_{L+2}(\alpha_{L+1,L+2}) - q_L(\alpha_{L,L+1})).$$

A.2.6 Proof of Proposition 7

The probability mass of firms producing with at most L layers is:

$$\Lambda_N(L) = \frac{[G(\alpha_{L,L+1}) - G(\alpha_D)]}{1 - G(\alpha_D)}, \tag{55}$$

with $\Lambda_N(L_{max}) = 1$ for some L_{max} which is the maximum number of layers in a firm. Taking the derivative of this expression with respect to N yields:

$$\frac{\partial \Lambda_N(L)}{\partial N} = \frac{[1 - G(\alpha_D)]\left[g(\alpha_{L,L+1})\frac{\partial \alpha_{L,L+1}}{\partial N} - g(\alpha_D)\frac{\partial \alpha_D}{\partial N}\right] + [G(\alpha_{L,L+1}) - G(\alpha_D)]\, g(\alpha_D)\frac{\partial \alpha_D}{\partial N}}{[1 - G(\alpha_D)]^2}.$$

Eliminating common terms, adding and subtracting $[1 - G(\alpha_D)]\, g(\alpha_{L,L+1})\frac{\partial \alpha_D}{\partial N}$, yields:

$$\begin{aligned}\frac{\partial \Lambda_N(L)}{\partial N} &= \frac{[1 - G(\alpha_D)]\, g(\alpha_{L,L+1})\frac{\partial \alpha_{L,L+1}}{\partial N} - [1 - G(\alpha_{L,L+1})]\, g(\alpha_D)\frac{\partial \alpha_D}{\partial N}}{[1 - G(\alpha_D)]^2} \\ &= \frac{[1 - G(\alpha_D)]\, g(\alpha_{L,L+1})\left[\frac{\partial \alpha_{L,L+1}}{\partial N} - \frac{\partial \alpha_D}{\partial N}\right] + [[1 - G(\alpha_D)]\, g(\alpha_{L,L+1}) - [1 - G(\alpha_{L,L+1})]\, g(\alpha_D)]\frac{\partial \alpha_D}{\partial N}}{[1 - G(\alpha_D)]^2}.\end{aligned} \tag{56}$$

The denominator is always positive, so the sign of $\frac{\partial \Lambda_N(L)}{\partial N}$ depends on the numerator. Previous sections have shown that $\frac{\partial \alpha_{L,L+1}}{\partial N} - \frac{\partial \alpha_D}{\partial N}$ is negative, and that $\frac{\partial \alpha_D}{\partial N}$ is positive. Hence the numerator in equation (56) will be negative if the following condition holds:

$$[1 - G(\alpha_D)]\, g(\alpha_{L,L+1}) \leq [1 - G(\alpha_{L,L+1})]\, g(\alpha_D),$$

which can be rewritten as

$$\frac{g(\alpha_{L,L+1})}{[1 - G(\alpha_{L,L+1})]} \leq \frac{g(\alpha_D)}{[1 - G(\alpha_D)]}. \tag{57}$$

Equation (57) is the hazard rate of the distribution of demand draws, $G(\alpha)$. Thus as long as $G(\alpha)$ has a non-increasing hazard rate, it follows that the probability mass of firms producing with at most L layers, $\Lambda_N(L)$ will be decreasing with respect to N. Therefore, if $N^{'} > N$, it follows that the distribution of layers in economy $N^{'}$, $\Lambda_{N'}$, will first order stochastically dominate the distribution of layers in economy N, Λ_N.

A.3 Isomorphisms

In this section I prove a simple theorem that relates to models that use Ottaviano-Tabuchi-Thisse preferences (Ottaviano et al. (2002)). I demonstrate that there is a correspondence between many market equilibria in this class of models, which makes it difficult to empirically test the model's prediction. In particular, the theorem implies that for any proportional change in the size of the market, there exists an equilibrium with the exact same outcomes, but that is derived from different parameters of the utility functions. This results stresses the importance of including demographic controls in regressions so as to proxy for consumers' preferences. In this section, I restrict my analysis to the utility function described in the paper with heterogeneous α, however analogous results can be derived from a model with constant α, as in Melitz and Ottaviano (2008).

Theorem 8 *Consider the utility function (1). Let α, η and γ be parameters that govern agents' utility, and let N denote the size of the market. Then for any constant $k > 0$ the following parameters yield exactly same equilibrium outcomes: α, ηk and γk and Nk.*

Proof. The proof relies on showing that the demand curve is the same in both frameworks. Consider the model with the parameters α, η, γ and N. Then a firm's demand curve is equal to:

$$p = \alpha + \frac{\eta M}{\eta M + \gamma}(\overline{\alpha} - \overline{p}) - \frac{\gamma}{N}q \tag{58}$$

where M is the mass of varieties in the economy. Replacing the parameters η, γ and N by the parameters ηk, γk and Nk , yields the exact same demand curve. ■

Theorem 8 implies that any equilibrium derived from a model with Ottaviano-Tabuchi-Thisse utility and parameters $\{\alpha, \eta, \gamma, N\}$ is isomorphic to the set of equilibria from models with parameters $\{\alpha, \eta k, \gamma k, Nk\}$. This theorem has implications when researchers conduct comparative statics and take their model to the data. In particular consider the implications of an increase N in a closed economy. Theorem 8 implies the following:

Corollary 9 *Consider the utility function (1). Let α, η and γ be parameters that govern agents' utility, and let N denote the size of the market. Consider the case of an increase in the size of the market of the sort $N^* = Nk$ with $k > 0$. Then the new equilibrium is identical to the following equilibrium in a market of size N and with parameters α, $\frac{\eta}{k}$ and $\frac{\gamma}{k}$.*

Corollary 9 implies that for any equilibrium generated from an increase in a market's size is isomorphic to an equilibrium where the market's size does not change, but where consumers have different preferences. Without controlling for agents' preferences, any estimates on how market size, N, affects economic outcomes across different regions, will be misidentified.

B Data Appendix

B.1 Data Construction

B.1.1 DADS - Exhaustive Cross-Sections

I begin with the matched employer-employee data. The years that are used in this study are from 1997-2007.The DADS is nearly exhaustive, and for a given year uniquely identify each worker, as well as firms and plants within firms. For every worker the data contains limited information on their job, such as number of hours, number of days and annual salary, as well as some demographic information, such as their age, gender and municipality of residence, and munipality of employment. For each firm, the data contain information on its primary industry of operation.

I first remove from the data observations with missing identifiers, and observations that have missing information for one of the following variables: number of days, number of hours, net and brut salaries, occupation and municipality. Second, using the industrial codes (NACE Rev 1.1.) I then restrict the data to firms that operate in one of the following industries: Clothes and Shoes Retailers (NACE codes 524C and 524E), Traditional Restaurants (NACE code 553A) and Hair and Beauty Salons (NACE codes 930D and 930E). And third I match each municipality to an employment area and when possible to an urban urea, using the 1999 geographical definitions from INSEE, creating Samples A and B. Within each sample for every firm, I then calculate the number of areas that it operates in, and drop firms that operate in more than one location.

For each firm, I construct its organization using the 1st-digit of the cs-occupational codes. With this method workers can be classified into as many as four layers. In my main results I adopt a strict interpretation of the model, and only retain firms with adjacent layers starting from layer 1. As a robustness check I use a more lenient interpretation, which is counts the total number of layers in firms and retains firms with layer 1. In contrast to the strict interpretation of the model, the lenient interpretation allows for gaps between the layers in firms.

From each Sample, I retain from the DADS information on the number of layers in firms, the total number of workers, the total number of hours, and the total wage bill of firms. Also, I retain information on the location of firms, and their industry of operation.

B.1.2 FICUS

FICUS contains balance sheet information as well as some information on the structure of firms firms, and is nearly exhaustive. The years that are used in this study are from 1997-2007. For a given year the data uniquely identifies firms. From this data I remove duplicate observations and drop firms with missing or negative values for the following variables: value added, sales, total number of workers and salaries paid.

From FICUS I retain the following information: value-added, sales, capital, whether the firm belongs to a business group, and the legal status of the firm. I then match this information with the Sample A and B from DADS using the firm identifiers from both datasets, and remove any unmatched observations.

B.1.3 RP

The RP data is a septennial census of the French population. In this study I construct variables from they year 1999, I use the year 1968 to construct instruments for the corresponding variables. For the year 1999, the RP is exhaustive and contains demographic information on all individuals and households in France. For each individual, there is information on the household he belongs to, his characteristics such as age, gender, education, employment status, and nationality, and the location of the household at the municipal level. I again match each municipality to an employment area and when possible to an urban urea, using the 1999 geographical definitions from INSEE, creating Samples A and B. Within each sample, I then construct measures of the fraction of individuals who are between the ages of 25 to 64, the share of unemployed workers and the share of immigrants in markets. This information is merge to Samples A and B from the DADS.

B.1.4 DADS - Panel

To create controls for the cost of a unit of labor, and the income of individuals in firms I use the panel dimension of the DADS. From 1976-1993 this dataset contains information on all workers born in October in even numbered years, and from 1993-2007 contains information on all workers born in October. The main difference between the panel dimension of the DADS and the exhaustive dimension, is that in the panel individuals can be tracked over time. The variables in both datasets are generally the same.

To estimate the cost of a unit of labor, for the years 1998-2000, I first match each municipality of work to an employment area or and urban area. I then retain male workers born in October and not working in one of the industries examined in this study, and estimated the following equation:

$$\ln wage_{it} = \alpha + x_{it}\beta + occ_i + ind_j + g_a + \epsilon_i,$$

where $wage_{it}$ is the hourly wage of worker i in year t, x_{it} contains a quintic polynomial of a worker's age and time fixed effects, ind_j are industry fixed effects (Nace Rev 1.1 at the 3-digit level), occ_i are cs-occupation fixed effects (1st-digit level), and g_a are area fixed effects. I then retain the fixed effects g_a. I also do this for lagged years to created instruments for the cost of a unit of labor. In particular, I use the years $1976-1979$ and the years $1984-1986$.

I then proceed to obtain a measure of the income of individuals in a geographical area. For the year 1999, I match each commune of residence to an employment area, and whenever possible to an urban area. I then drop individuals that are employed in one of the sectors examined in this study, and I calculate the average annual salary across areas of residence. Because information on residence is not known prior to 1993, I do this for the year 1993.

B.1.5 Occupational Description

To get a sense of the hierarchies of firms, table 15 presents a description of the occupations that are associated with each occupational category in firms. It is important to note, however, that although the classification of occupational codes is detailed, it does not permit one to observe the tasks workers perform, nor to observe the chain of command within the firm.

Table 15: Description of Occupations

	Job Description	Examples
Layer Four	This occupational category includes the actual shopkeepers and owners when they are employees of their own trade, and corporate officers of a business.	Owners, CEO, CFO.
Layer Three	This occupational category includes employees occupying an executive or senior managerial position within a business.	Store Director, Head Chef, Department Head, General Manager.
Layer Two	This occupational category includes employees occupying supervisory position within a business.	Bar Manager, Second Chef, Warehouse Manager, Sales Manager.
Layer One	This occupational category contains employees performing manual or administrative work who are either skilled and unskilled. Workers with degrees in the same field as their profession are considered skilled occupations. The rest are unskilled occupations.	Servers, Cooks, Dishwashers Barmen, Kitchen Helpers, Bus Boys, Receptionists, Cashiers, Estheticians, Hairdressers, Merchandisers, Warehouse Workers, Clerk, Sales Personnel.

Notes: List of occupations in the different layers in firms.

B.2 Additional Tables

Table B1: Additional Robustness Checks Across Employment Areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep: Percent One-Layer Firms										
log density	−0.048	−0.024	−0.043	−0.016	−0.041	−0.022	−0.031	−0.017	−0.031	−0.019
	(0.004)^{a}	(0.009)^{a}	(0.004)^{a}	(0.009)^{c}	(0.005)^{a}	(0.008)^{b}	(0.005)^{a}	(0.009)^{c}	(0.005)^{a}	(0.008)^{b}
Dep: Percent Two-Layer Firms										
log density	0.023	0.020	0.020	0.016	0.017	0.018	0.015	0.017	0.013	−0.017
	(0.006)^{a}	(0.009)^{b}	(0.005)^{a}	(0.009)^{c}	(0.006)^{a}	(0.009)^{b}	(0.006)^{b}	(0.009)^{c}	(0.006)^{b}	(0.009)^{c}
Dep: Percent Three-Layer Firms										
log density	0.020	0.002	0.016	0.001	0.018	0.004	0.012	0.000	0.014	0.002
	(0.004)^{a}	(0.008)	(0.004)^{a}	(0.008)	(0.004)^{a}	(0.007)	(0.004)^{a}	(0.007)	(0.004)^{a}	(0.007)
Dep: Percent Four-Layer Firms										
log density	0.004	0.001	0.005	−0.000	0.004	−0.000	0.003	−0.000	0.004	−0.000
	(0.001)^{a}	(0.002)	(0.001)^{a}	(0.002)	(0.001)^{a}	(0.002)	(0.001)^{b}	(0.002)	(0.001)^{a}	(0.002)
Sample	A	A	A	A	A	A	A	A	A	A
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
First-Stage Statistics										
Partial R-squared	0.748	0.208	0.746	0.207	0.743	0.196	0.728	0.193	0.755	0.188
SW F-Statistic	182.31	10.09	196.67	9.83	184.76	9.51	163.91	9.52	178.91	9.13
KP Wald F-Statistic	182.31	5.35	196.67	5.19	184.76	5.02	163.91	5.05	178.91	4.69
Over-Identification Test (p-value)										
Percent One-Layer Firms	0.896	0.533	0.585	0.209	0.808	0.129	0.986	0.096	0.863	0.032
Percent Two-Layer Firms	0.707	0.425	0.760	0.281	0.749	0.279	0.694	0.251	0.592	0.209
Percent Three-Layer Firms	0.996	0.798	0.415	0.711	0.764	0.687	0.968	0.659	0.999	0.457
Percent Four-Layer Firms	0.338	0.429	0.322	0.274	0.141	0.273	0.129	0.275	0.132	0.270
Wage Controls*	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Income Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	375	375	402	402	408	408	408	408	408	408

*Notes: a,b,c: significant at the 1%, 5% and 10% level. *: indicates variables are treated as exogenous in regressions. Clustered standard errors at the employment area level are reported in parentheses. This table reports regression results for equation (19). Each column displays the estimate from a separate regression. Density measures the local density of an employment area. Industry fixed effects are at the three-digit Nace Rev 1.1. level. Wage controls contain the cost of a unit of labor. Income controls contain the median the annual salary of individuals residing in an employment area. Demographic controls contain the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. Columns (1), (3), (5), (7) and (9) only instrument for density using the log of density measured in 1831 and 1881. Columns (2), (4), (6), (8) and (10) instrument for density and local characteristics using the following variables: the log of density measured in 1831, 1851, 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, and the average person per household in 1949. Columns (1) and (2) retain only single establishment firms. Columns (3) and (4) retain only independent firms. Columns (5) and (6) control for the average log of capital. Columns (7) and (8) control for average log of capital and the average log number of workers in firms. Columns (9) and (10) control for average log of capital and the average log number of hours in firms.*

Table B2: Robustness Checks - Firm Level Regression Results

Dependent Variable: Number of Layers	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample A: Employment Areas												
log density	0.073	0.020	0.085	0.030	0.061	0.023	0.054	0.028	0.073	0.030	0.025	0.029
	(0.003)^a	(0.011)^c	(0.004)^a	(0.013)^b	(0.002)^a	(0.008)^a	(0.002)^a	(0.009)^a	(0.002)^a	(0.010)^a	(0.008)^a	(0.009)^a
Sample	A	A	A	A	A	A	A	A	A	A	A	A
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
First-Stage Statistics	-	-	-	-	-	-	-	-	-	-	-	-
Partial R-squared	0.916	0.348	0.918	0.305	0.915	0.348	0.914	0.348	0.915	0.348	0.348	0.348
SW F-Statistic	1,164	12.11	1,175	11.82	1,157	12.18	1,163	12.16	1,160	12.09	12.13	12.12
KP Wald F-Statistic	1,164	8.576	1,175	8.599	1,157	8.554	1,163	8.532	1,160	8.497	8.488	8.475
Over-Identification Test (p-value)	0.602	0.292	0.239	0.093	0.463	0.146	0.330	0.489	0.536	0.183	0.083	0.420
Sample Size	27,508	27,508	27,508	27,508	25,791	25,791	25,791	25,791	25,129	25,129	25,129	25,129
Sample B: Urban Areas												
log density	0.132	0.061	0.160	0.085	0.114	0.042	0.102	0.040	0.134	0.068	0.043	0.040
	(0.018)^a	(0.011)^a	(0.025)^a	(0.013)^a	(0.022)^a	(0.012)^a	(0.019)^a	(0.011)^a	(0.022)^a	(0.011)^a	(0.011)^a	(0.011)^a
Sample	B	B	B	B	B	B	B	B	B	B	B	B
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
First-Stage Statistics	-	-	-	-	-	-	-	-	-	-	-	-
Partial R-squared	0.714	0.360	0.696	0.370	0.713	0.360	0.711	0.360	0.713	0.360	0.360	0.360
SW F-Statistic	22.61	12.79	17.32	12.58	22.17	12.78	19.93	12.78	22.14	12.69	12.68	12.68
KP Wald F-Statistic	22.61	7.921	17.32	7.967	22.17	7.920	19.93	7.916	22.14	7.824	7.820	7.823
Over-Identification Test (p-value)	0.242	0.130	0.231	0.100	0.598	0.266	0.260	0.159	0.568	0.125	0.216	0.165
Sample Size	21,605	21,605	21,605	21,605	21,605	21,605	21,605	21,605	21,043	21,043	21,043	21,043
Wage Controls*	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes
Income Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: a,b,c: significant at the 1%, 5% and 10% level. *: indicates variables are treated as exogenous in regressions. Clustered standard errors at the local market level are reported in parentheses. This table reports regression results for equation (20). Each column displays the estimate from a separate regression. Density measures the local density of an employment area. Industry fixed effects are at the four-digit Nace Rev 1.1. level. Wage controls contain the cost of a unit of labor. Income controls contain the median the annual salary of individuals residing in an employment area. Demographic controls contain the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. In Sample A, Columns (1), (3), (5), (7), (9) only instrument for density using the log of density measured in 1851 and 1881. In Sample B, Columns (1), (3), (5), (7), (9) only instrument for density using the log of density measured in 1881 and 1901. In Sample A, Columns (2), (4), (6), (8), (10), (11) and (12) instrument for density and local characteristics using the following variables: the log of density measured in 1851 and 1881, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a telephone in their residence, and the share of the population in 1968 with a bathtub or shower in their residence. In Sample B, Columns (2), (4), (6), (8), (10), (11) and (12) instrument for density and local characteristics using the following variables: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an urban area in 1993, the share of the local population between the ages 25 and 54 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a telephone in their residence, the share of the population in 1968 with a bathtub or shower in their residence, the share of the population in 1968 with a toilet in their residence, and the share of the population in 1968 with heating in their residence. Columns (1) and (2) add as a control the number of additional occupations in firms, measured in logs. Columns (3) and (4) add as a control the HHI index of occupational specialization at the city level. Columns (5) and (6) add as a control for the size of firms, in logs, measured using workers. Columns (7) and (8) add as a control for the size of firms, in logs, measured using hours. Columns (9) and (10) add as a control for the level of capital in firms, in logs. Column (11) controls for capital and the number of workers in firms, both in logs. Column (12) controls for capital and the number of hours in firms, both in logs.*

Table B3: Additional Robustness Checks - Firm Level Regression Results

Dependent Variable: Number of Layers	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample A: Employment Areas								
log density	0.090	0.030	0.083	0.028	0.079	0.021	0.063	0.034
	(0.004)[a]	(0.012)[b]	(0.003)[a]	(0.012)[b]	(0.004)[a]	(0.011)[c]	(0.002)[a]	(0.010)[a]
Sample	A	A	A	A	A	A	A	A
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
First-Stage Statistics	-	-	-	-	-	-	-	-
Partial R-squared	0.916	0.349	0.917	0.348	0.915	0.347	0.914	0.347
SW F-Statistic	1,190	11.95	1,163	12.58	1,157	12.31	1,140	12.12
KP Wald F-Statistic	1,190	8.460	1,163	9.255	1,157	8.813	1,140	8.550
Over-Identification Test (p-value)	0.480	0.156	0.608	0.477	0.463	0.271	0.488	0.134
Sample Size	22,606	22,606	24,038	24,038	24,611	24,611	25,791	25,791
Sample B: Urban Areas								
log density	0.164	0.083	0.153	0.082	0.141	0.070	0.119	0.067
	(0.025)[a]	(0.013)[a]	(0.020)[a]	(0.013)[a]	(0.021)[a]	(0.013)[a]	(0.017)[a]	(0.013)[a]
Sample	B	B	B	B	B	B	B	B
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
First-Stage Statistics	-	-	-	-	-	-	-	-
Partial R-squared	0.715	0.362	0.717	0.360	0.713	0.358	0.711	0.359
SW F-Statistic	23.23	13.42	23.43	12.77	22.82	12.87	22.50	12.85
KP Wald F-Statistic	23.23	8.293	23.43	7.929	22.82	7.999	22.50	7.959
Over-Identification Test (p-value)	0.312	0.041	0.220	0.100	0.276	0.129	0.388	0.196
Sample Size	18,778	18,778	20,099	20,099	20,454	20,454	21,605	21,605
Wage Controls*	No	Yes	No	Yes	No	Yes	No	Yes
Income Controls	No	Yes	No	Yes	No	Yes	No	Yes
Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes

*Notes: a,b,c: significant at the 1%, 5% and 10% level. *: indicates variables are treated as exogenous in regressions. Clustered standard errors at the local market level are reported in parentheses. This table reports regression results for equation (20). Each column displays the estimate from a separate regression. Density measures the local density of an employment area. Industry fixed effects are at the four-digit Nace Rev 1.1. level. Wage controls contain the cost of a unit of labor. Income controls contain the median the annual salary of individuals residing in an employment area. Demographic controls contain the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. In Sample A, Columns (1), (3), (5) and (7) only instrument for density using the log of density measured in 1851 and 1881. In Sample B, Columns (1), (3), (5) and (7) only instrument for density using the log of density measured in 1881 and 1901. In Sample A, Columns (2), (4), (6) and (8) instrument for density and local characteristics using the following variables: the log of density measured in 1851 and 1881, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a telephone in their residence, and the share of the population in 1968 with a bathtub or shower in their residence. In Sample B, Columns (2), (4), (6) and (8) instrument for density and local characteristics using the following variables: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an urban area in 1993, the share of the local population between the ages 25 and 54 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a telephone in their residence, the share of the population in 1968 with a bathtub or shower in their residence, the share of the population in 1968 with a toilet in their residence, and the share of the population in 1968 with heating in their residence. Columns (1) and (2) restrict the sample to firms with adjacent layers and with ordinary workers. Columns (3) and (4) restrict the sample to single establishment firms. Columns (5) and (6) restrict the sample to firms that do not belong to a business group. Columns (7) and (8) add as a control the legal status of firms.*

Table B4: Robustness: Second-Stage Value-Added Regression Results

Dependent Variable from First-Stage	VA per Worker (1)	VA per Worker (2)	VA per Worker (3)	VA per Worker (4)	VA per Worker (5)	VA per Worker (6)
Model 1: Without Org						
log density	0.049	0.053	0.049	0.046	0.049	0.044
	$(0.004)^a$	$(0.009)^a$	$(0.003)^a$	$(0.009)^a$	$(0.004)^a$	$(0.009)^a$
Model 2: With Org						
log density	0.038	0.049	0.039	0.043	0.039	0.042
	$(0.003)^a$	$(0.009)^a$	$(0.003)^a$	$(0.009)^a$	$(0.004)^a$	$(0.009)^a$
% Change	22.4	7.5	20.4	6.5	20.4	4.5
Sample	A	A	A	A	A	A
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
KP Wald F-Statistic	2,795	9.93	2,700	8.27	2,720	8.11
Over-Id Test (p-value): Model 1	0.727	0.127	0.976	0.141	0.711	0.149
Over-Id Test (p-value): Model 2	0.791	0.219	0.968	0.217	0.775	0.229
Sample Size	23,277	23,277	24,743	24,743	25,341	25,341
Model 1: Without Org						
log density	0.089	0.062	0.093	0.064	0.087	0.051
	$(0.013)^a$	$(0.021)^a$	$(0.013)^a$	$(0.019)^a$	$(0.015)^a$	$(0.021)^b$
Model 2: With Org						
log density	0.068	0.050	0.073	0.052	0.068	0.041
	$(0.012)^a$	$(0.021)^b$	$(0.011)^a$	$(0.019)^a$	$(0.013)^a$	$(0.020)^b$
% Change	23.5	19.3	21.5	18.7	21.8	19.6
Sample	B	B	B	B	B	B
Method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
KP Wald F-Statistic	23.23	7.68	23.43	7.27	22.82	7.47
Over-Id Test (p-value): Model 1	0.419	0.144	0.470	0.143	0.368	0.149
Over-Id Test (p-value): Model 2	0.334	0.173	0.360	0.152	0.267	0.163
Sample Size	18,778	18,778	20,099	20,099	20,454	20,454
Second-Stage: All Controls	No	Yes	No	Yes	No	Yes
Second-Stage: Industry FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: a,b,c: significant at the 1%, 5% and 10% level. Clustered standard errors at the local market level are reported in parentheses. Each entry displays a results from equation (24). Model 1 does not control for the number of layers in firms in the first-stage while Model 2 does. Industry fixed effects are at the four-digit Nace Rev 1.1. level. Controls in the second-stage are the following: the cost of a unit of labor, the median the annual salary of individuals residing in an area, the share of the local population between the ages 25 and 59, the share of the local population born outside of France, and the share of active workers who are unemployed. Columns (1), (3), and (5) instrument for density using density measured in 1881 and 1901. Columns (2), (4) and (6) instrument for density and the characteristics of local markets except for the cost of a unit of labor. In Sample A, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a telephone in their residence, and the share of the population in 1968 with a toilet in their residence. In Sample B, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 54 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a heating in their residence, and the share of the population in 1968 with a bathtub or shower in their residence. Columns (1) and (2) restrict the sample to firms with adjacent layers and with ordinary workers, Columns (3) and (4) restrict the sample to single-establishment firms, and Columns (5) and (6) restrict the sample of independent firms.

Table B5: Alternative Specification: Second-Stage Value-Added Regression Results

Dependent Variable from First-Stage	VA per Worker (1)	VA per Worker (2)	VA per Worker (3)	VA per Worker (4)	VA per Worker (5)	VA per Worker (6)
Model 3: Within Org						
log density	0.038	0.035	0.037	0.038	0.034	0.042
	$(0.003)^a$	$(0.009)^a$	$(0.003)^a$	$(0.003)^a$	$(0.009)^a$	$(0.009)^a$
Model 4: Within Org-Industry						
log density	0.038	0.036	0.037	0.038	0.034	0.044
	$(0.003)^a$	$(0.009)^a$	$(0.003)^a$	$(0.003)^a$	$(0.009)^a$	$(0.009)^a$
Sample	A	A	A	A	A	A
Method	OLS	OLS	OLS	2SLS	OLS	2SLS
KP Wald F-Statistic	-	-	-	2,454	-	10.05
Over-Id Test (p-value): Model 3	-	-	-	0.761	-	0.209
Over-Id Test (p-value): Model 4	-	-	-	0.815	-	0.197
R-squared: Model 3	0.108	0.112	0.102	-	0.112	-
R-squared: Model 4	0.110	0.114	0.110	-	0.114	-
Sample Size	27,508	27,508	26,531	26,531	26,531	26,531
Model 3: Within Org						
log density	0.054	0.050	0.057	0.068	0.036	0.050
	$(0.009)^a$	$(0.011)^a$	$(0.009)^a$	$(0.011)^a$	$(0.016)^b$	$(0.020)^b$
Model 4: Within Org-Industry						
log density	0.055	0.050	0.058	0.069	0.036	0.051
	$(0.009)^a$	$(0.011)^a$	$(0.010)^a$	$(0.012)^a$	$(0.016)^b$	$(0.020)^b$
Sample	B	B	B	B	B	B
Method	OLS	OLS	OLS	2SLS	OLS	2SLS
KP Wald F-Statistic	-	-	-	21.93	-	5.62
Over-Id Test (p-value): Model 3	-	-	-	0.392	-	0.204
Over-Id Test (p-value): Model 4	-	-	-	0.351	-	0.212
R-squared: Model 3	0.095	0.101	0.091	-	0.097	-
R-squared: Model 4	0.097	0.103	0.094	-	0.100	-
Sample Size	24,197	24,197	21,606	21,606	21,606	21,606
All Controls	No	Yes	No	No	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: a,b,c: significant at the 1%, 5% and 10% level. Clustered standard errors at the local area level are reported in parentheses. Model 3 regresses value-added per worker (in logs) on organization fixed effects and density. Model 4 regresses value-added per worker (in logs) on organization-industry fixed effects and density. Industries are defined at the 4-digit Nace Rev 1.1. level. Column (1) only controls for density and organization. Column (2) also controls for the local characteristics of markets. Column (3) restricts the sample to observations with historical values. Column (4) instruments for density only. Column (5) restricts the sample to observations with past-values. Column (6) instruments for density and the characterstics of local markets. Column (3), instruments for density using density measured in 1881 and 1901. Column (6) instruments for density and the characteristics of local markets except for the cost of a unit of labor. In Sample A, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 39 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a telephone in their residence, and the share of the population in 1968 with a toilet in their residence. In Sample B, the instruments are the following: the log of density measured in 1881 and 1901, the median annual salary of individuals residing in an employment area in 1993, the share of the local population between the ages 25 and 54 in 1968, the share of the local population born outside of France in 1968, and the share of active workers who are unemployed in 1968, the share of the population in 1968 residing in buildings built before 1949, the average person per household in 1949, the share of the population in 1968 having access to a heating in their residence, and the share of the population in 1968 with a bathtub or shower in their residence.

Table B6: Parameter Values

	N	A	h	λ	c	γ	η	k	α_m	f_E
Model 1	500	10	0.42	28	14	5	3	3.95	1	1.75
Model 2	1000	10	0.42	28	14	5	3	3.95	1	1.75

Notes: Parameters used in simulations of Models 1 and 2.

Table B7: Equilibrium Values

	α_D	q_D	$\frac{\eta M}{\gamma+\eta M}(\overline{\alpha}-\overline{p})$	M	$\overline{\alpha}$	$\overline{p}$	$\alpha_{0,1}$	$\alpha_{1,2}$	$\alpha_{2,3}$
Model 1	2.368	6.893	2.069	3.248	3.145	0.013	2.557	3.759	9.432
Model 2	3.294	7.361	3.031	3.375	4.453	0.005	3.379	3.971	6.804

Notes: Equilibrium Values from simulations of Models 1 and 2.

C Model Simulations

C.1 Simulations - Aggregate Economy

This section presents numerical simulations model. There are two economies, one with $N = 500$ and the other with twice its size at $N = 1000$. Demand is drawn from a Pareto distribution with coefficient $k = 3.95$ and with support $[1, \infty]$, so $G(\alpha) = 1 - \alpha^{-3.95}$.[42] The complete set of parameters chosen for each model are listed in Table B6. Because the objective of this study is to examine how market size affects the organization of firms, the remaining parameters chosen in both models are identical.

Table B7 presents equilibrium values. An increase increases the demand draw of the marginal firm, α_D, and the aggregate term, $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha}-\overline{p})$. These results are consistent with Proposition 5 from the previous section. Table B7 also reports the demands draws at which entrepreneurs are indifferent between two organizational forms, $\alpha_{L,L+1}$. Larger markets affect the distribution of organizations through two channels. First, bigger markets induce tougher selection and increase the demand cutoff α_D. And second, because markups over marginal and average costs are lower in bigger markets, firms' are induced to re-organized production in favor of more layers. Because the quantities produced by firms change as well, it is not always the case that the cutoffs $\alpha_{L,L+1}$ decrease in bigger markets. However, the distance between α_D and $\alpha_{L,L+1}$ does decrease with N. This result was proven in Proposition 6.

Figure 6 presents the cumulative distribution of organizations. There is a ranking of distributions. The Pareto distribution always satisfies the non-decreasing hazard rate property, and so it follows from Proposition 7 that the distribution of organizations in Model 2 first-order stochasti-

[42] The parameter k is chosen to relatively large for computational tractability.

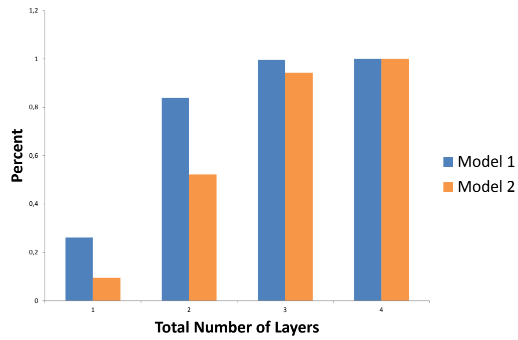


Figure 6: Cumulative Distribution of Organizations

cally dominates the distribution of organizations in Model 1. Applying the Mann-Whitney test to the simulated data confirms the result. It rejects the hypothesis that both distributions are equal at the one percent level (the p-value of the test statistic is 0.000) and reports that the probability that a random firm from Model 2 has a greater number of layers than a random from model 1 is 0.694.

C.2 Simulations - Firm Level Results

Panel (a) illustrates that the response in quantities produced is heterogeneous across firms. An increase in the size of the market has two opposing effects on firms' demand schedules. A direct effect: an increase in N rotates firms' demand curves outwards away from the quantity-axis, increasing demand. And an indirect effect: an increase in market size increases the aggregate term $\frac{\eta M}{\gamma+\eta M}(\overline{\alpha} - \overline{p})$ leading to a downward shift in firms' demand curves, lowering demand. For firms with sufficiently high demand draws the direct effect dominates and they increase production. For firms with low demand draws the indirect effect dominates and they produce less.

Panels (b) and (c) illustrate that firms' markups decrease in bigger markets. For any quantity chosen, both firms' markups over marginal costs and their markups over average costs decrease with N. This effect holds even for firms that change their organization. In addition the impact of lower markups implies that profit per unit decreases, and therefore to attain any given level of profit firms need produce more output. Therefore the mapping from quantities to profits decreases as well.

Panel (d) illustrate that the prices charged by firms' decrease in bigger markets. There are two factors that determine how the prices change in larger markets. First, firms in bigger markets produce different quantities and some firms change their organization, both of which affect firms' marginal costs.[43] Second in larger markets demand is more elastic, markups over marginal costs decrease as demonstrated in panel (b) and prices decrease. Overall the change in markups dominates any changes in marginal costs and prices decrease with N.

Panel (f) illustrates the profits earned by firms may decrease or increase. Two factors determine how profits change in larger markets. First as demonstrated in panel (c), firms in larger markets charge lower markups over average costs which decreases their profits. Second as illustrated in panel (a), the quantity produced by firms changes. Holding markups constant, an increase in the quantity produced increases profits. Firms that reduce their quantity earn lower profits, because both effects work in the direction to lower profits. In contrast firms that increase their quantity the effects work in the opposite direction, and for firms that sufficiently increase their quantity, the second effect dominates and their profits increase.

C.3 Simulations - Productivity Results

I now turn to productivity and investigate how the distribution of productivity is different across locations. To provide a complete description of the model and to relate the model's implications to the empirical literature, I present several measures of firm productivity. The first two measures are based on quantity, and so represent measures of quantity-based productivity, while the last two measures are based on revenues and represent measures of revenue-based productivity.

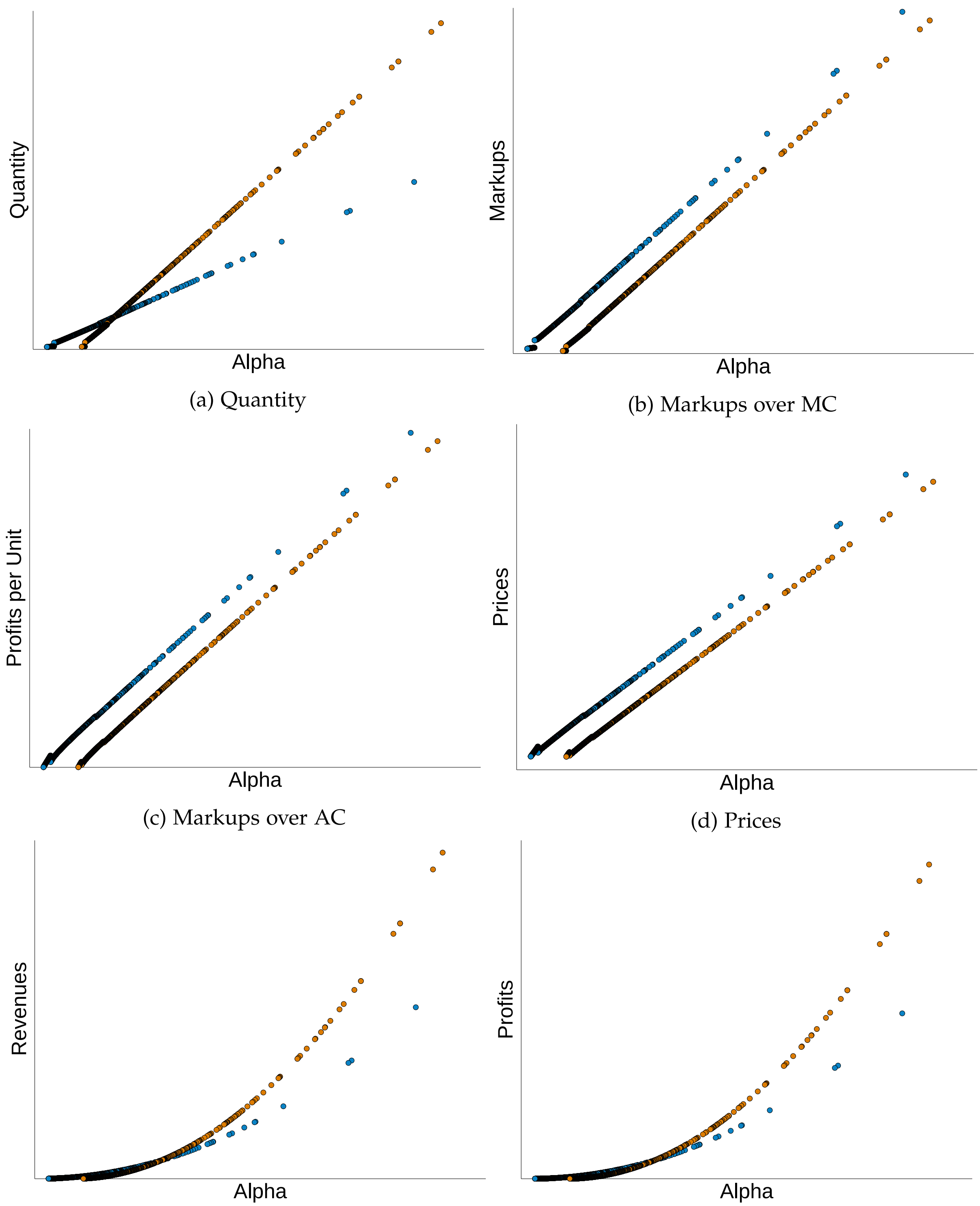
C.3.1 Quantity-Based Productivity Measures

The first measure of quantity-based productivity is obtained from the costs functions of firms. It is the inverse of average costs and is equal to:

$$\phi_1(q) = \frac{q}{C(q)} = \frac{1}{AV(q)}. \tag{59}$$

A previous section characterized the cost function of firms. The important points from that section are the following. First, unlike most models with heterogeneous firms, in this model firms' marginal costs are not equal to their average costs. The only exceptions are at the MES points. Second, because both marginal costs and average costs depend on the quantity produced

[43] Firms that do not change their organization and produce more output increase their marginal costs, while firms that produce less output and do not change their organization decrease their marginal costs. For firms that change their organization the opposite takes places. Since prices are a markup over marginal costs, holding markups constant an increase in marginal costs induces firms to charge higher prices, while a decrease in marginal costs induces firms to charge lower prices.



(a) Quantity

(b) Markups over MC

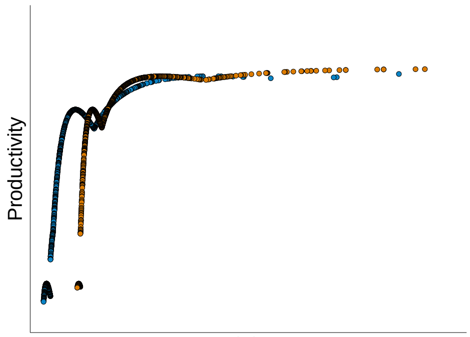
(c) Markups over AC

(d) Prices

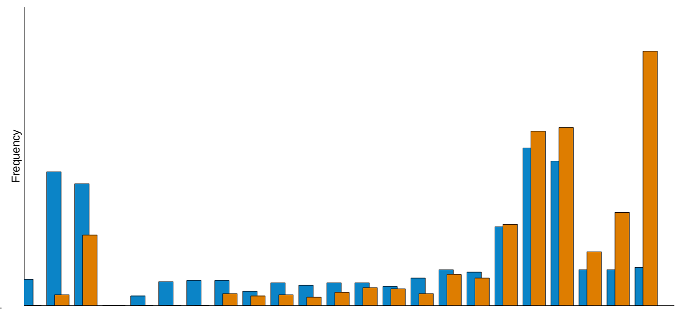
(e) Revenues

(f) Profits

Figure 7: The Impact of Market Size on Individual Producers. Blue (N = 500) Orange (N = 1000).



(a) Productivity



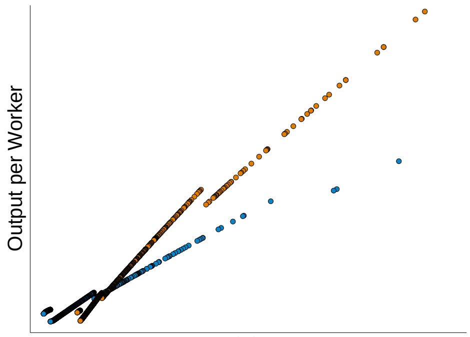
(b) Productivity Distributions

Figure 8: Productivity

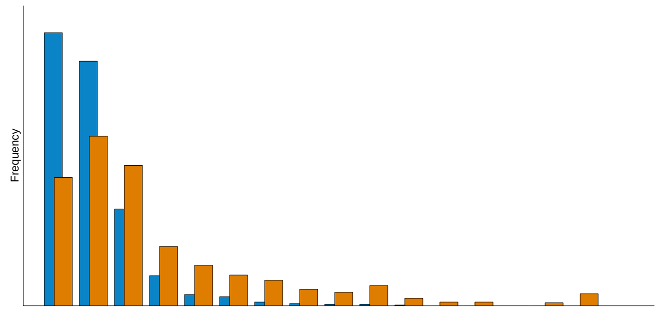
and on the organization of a firm, they are endogenous, and so productivity is endogenous as well. Third average cost are neither constant nor a monotonic function of quantity. This implies that firm-level productivity will also be neither constant nor a monotonic function of quantity. And fourth, the minimum average cost is decreasing with the number of layers, and the level of output that attains the minimum average cost is increasing with the number of layers. In terms of productivity this implies that firms with a greater number of layers can attain a greater productivity. In addition because the quantity produced by firms and their organization depend on the size of the market, the productivity of firms will also depend on the size of the market. Taken together all these assertions indicate that the distribution of productivity will be different across locations.

Figure 8a illustrates the heterogeneous responses in productivity as a result of operating in a bigger market. The productivity of firms with relatively low values of α decreases in larger markets, whereas firms with high demand draws have their productivity increase. Note that this is different from the findings of Caliendo and Rossi-Hansberg (2012), where in a closed-economy, an increase in market size only raises wages and the number of firms, but does not affect the quantities produced by firms, and thus their organizational structure and productivity remain the same.

Figure 8b presents the distribution of productivity from both economies. Although the Pareto distribution is invariant to truncation, the figure shows different shares of small, medium and high productivity firms across markets. The fraction of low productivity and medium productivity firms decreases in the larger market, while the mass of high productivity firms increases. Because markups are lower in bigger markets, firms are induced to reorganize. At the same



(a) Output per Worker



(b) Output per Worker Distributions

Figure 9: Output per Worker

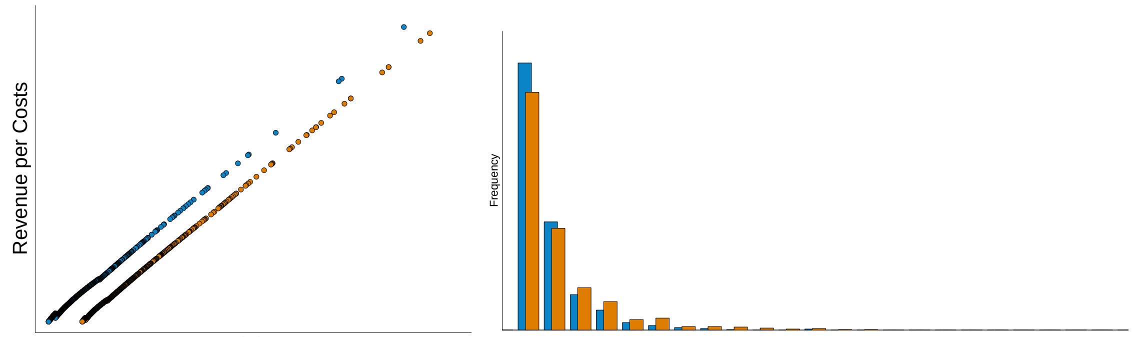
time, because they are induced to produce more output, firms in the middle of the distribution increase their productivity in larger markets, thereby increasing the mass at the tail. In other words, a larger market makes firms in the middle of the distribution more productive because it incentivizes them to reorganize and produce with more layers.

The distribution of productivity in model 2 has a mean of 5.21 and a variance of 0.076. In comparison, in model 1 the mean of the distribution of productivity is 4.94 and its variance is 0.141. Note that these values are qualitatively consistent with empirical studies examining the distribution of productivity across locations. A conclusion emerging from these studies is that the distribution of productivity in denser markets has a higher mean and a lower variance (for example, Syverson (2004)).

The second measure of quantity-based productivity is output per worker and is equal to: $\phi_2(\alpha) = \frac{q(\alpha)}{\sum_{j=1}^{L} n_L^j(\alpha)}$. Figures 9a and 9b present quantity-labor productivity at the level of firms and the distribution across both markets. As Figure 9a illustrates quantity-labor productivity increases for the majority of firms, which are the firms that increased their output in the larger market. Figure 9b presents the distribution of quantity-labor productivity in both economies. In the bigger market the share of low productivity firms decreases while the mass of medium and high productivity firms increases. The distribution of labor productivity in model 2 has a mean of 19.01 and a variance of 279.54 while, in model 1 the distribution of labor productivity has a mean of 9.94 and a variance of 41.66.

C.3.2 Revenue-Based Productivity Measures

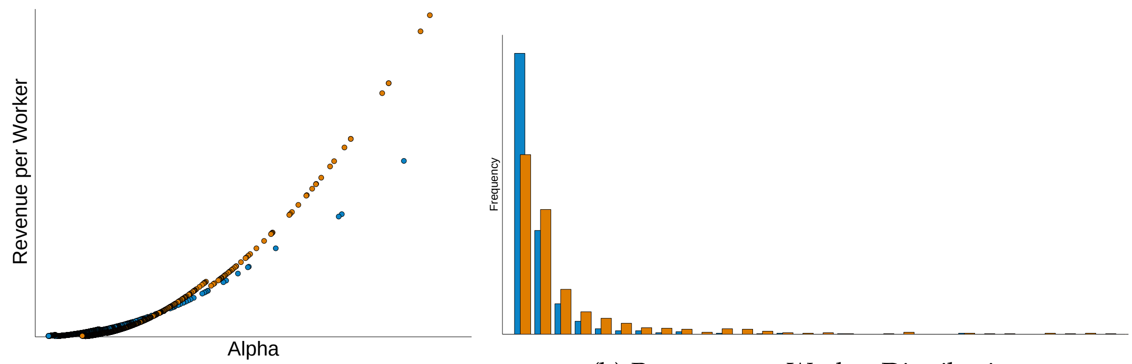
I now turn to measures of revenue-based productivity and examine how they vary across markets. The first measure relates revenues generated from a dollar spent on labor and is simply equal to:



(a) Revenue per Labor Costs

(b) Revenue per Labor Costs Distributions

Figure 10: Revenue per Labor Costs



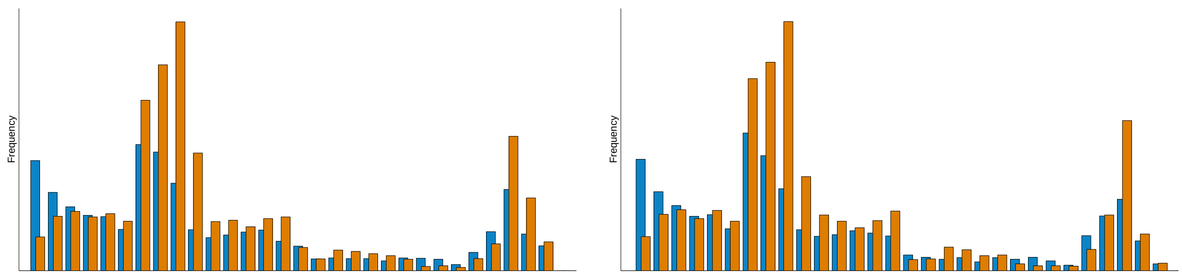
(a) Revenue per Worker

(b) Revenue per Worker Distributions

Figure 11: Revenue per Worker

$\psi_1(\alpha) = \frac{r(\alpha)}{C(\alpha)}$. Figure 10a illustrates that productivity at the level of firms and Figure 10b presents the distribution of productivity across both markets. The results are similar as above. In the bigger market has a greater the mass of medium and high productivity firms. Further in model 1 the distribution of productivity has a mean of 3.32 and a variance of 2.77. In contrast, in model 2 the distribution of productivity has mean of 4.22 and a variance of 4.28, both greater than in model 1. Again these results are qualitatively consistent with empirical studies examining the productivity of firms across locations.

The second measure of productivity based on revenues is revenue-labor productivity and is equal to: $\psi_2(\alpha) = \frac{r(\alpha)}{\sum_{j=1}^{L} n_L^j(\alpha)}$. Conclusions remain the same. In the bigger market the share of low productivity firms decreases while the mass of medium and high productivity firms increases. The distribution of labor productivity in model 2 has a mean of 29.79 and a variance of 128.74 while in model 1 the distribution of labor productivity has a mean of 9.54 and a variance of 20.83.



(a) Knowledge

(b) Income

Figure 12: Distributions of Knowledge & Income

C.3.3 Income and Knowledge

I now turn to income and examine how the distribution of income varies across locations. In the model, because wages are normalized to 1, income is equal to $cz_L^l + 1$ and so the distribution of income is similar to the distribution of knowledge in the economy. Panel (a) in Figure 12 illustrates the distribution of knowledge in the economy. The fraction of agents with intermediate levels of knowledge is bigger in the larger market, while the fraction of agents with low levels of knowledge is smaller. The effect is due to the increased number of firms producing with more layers in the bigger market. Because a bigger market incentivizes firms to add layers of management, the knowledge of existing workers decreases. At the same time, because firms employ more intermediate managers, there are more agents with intermediate levels of knowledge in the economy. The second effect dominates and the mass of agents with low levels of knowledge is reduced in the bigger market.

Panel (b) presents the distribution of income. The distribution of income closely resembles the distribution of knowledge. To draw more meaningful comparisons, in Table B8 I report the mean and variance of the distribution of income from both models. Because these results are again based on numerical simulations, I only draw qualitative conclusions. As reported in Table B8 the distribution of income has a higher mean but a lower variance in the bigger market relative to the smaller market. A higher mean is consistent with empirical studies that examine how wages differ across locations. A conclusion emerging from these studies is that workers earn higher wages in denser markets. The numerical simulations suggest that the model is to be able to qualitatively account for this fact.

The simulations, however, are unable to account for the fact that wage inequality is greater in denser areas, documented in the empirical literature. As there is a very large weight given to

Table B8: Qualitative Comparisons on the Distributions of Income

	Model 1	Model 2
Mean	1.4095	1.4421
Variance	0.3480	0.3264

Notes: Comparisons of distribution of income from model 1 and 2.

firms with low demand draws, perhaps this may due to the value of the Pareto shape parameter k chosen in the simulations.