The Return to Knowledge Hierarchies*

Luis Garicano
University of Chicago

Thomas N. Hubbard
Northwestern University

July 6, 2006

Abstract

Hierarchies allow individuals to leverage their knowledge through others' time. This mechanism increases productivity and amplifies the impact of skill heterogeneity on earnings inequality. To quantify this effect, we analyze the earnings and organization of U.S. lawyers and use the equilibrium model of knowledge hierarchies in Garicano and Rossi-Hansberg (2006) to assess how much lawyers' productivity and the distribution of earnings across lawyers reflects lawyers' ability to organize problem-solving hierarchically. We analyze earnings, organizational, and assignment patterns and show that they are generally consistent with the main predictions of the model. We then use these data to estimate the model. Our estimates let us infer how much production would be lost and how much the earnings distribution would change if lawyers were not able to "vertically specialize." We find that the impact of hierarchy on productivity and earnings distributions is substantial but not dramatic, reflecting the fact that the problems lawyers face are diverse and that the solutions tend to be customized.

*The research in this paper was conducted while the authors were Census Bureau research associates at the Chicago Research Data Center. Research results and conclusions are those of the authors and do not necessarily indicate concurrence by the Bureau of the Census. This paper has been screened to ensure that no confidential data are revealed.

I. INTRODUCTION

When knowledge is embodied in individuals, they must spend time applying it to each specific problem they face and possibly also communicating specific solutions to others. This can make it difficult for individuals to exploit increasing returns associated with knowledge, relative to a situation where knowledge can be encoded in blueprints, as in Romer (1986, 1990). For example, radiologists who are experts at interpreting x-rays generally cannot sell their knowledge in a market like a blueprint; instead, they usually must apply their knowledge to each patient's specific x-ray. A way around this problem is vertical, or hierarchical, specialization where some non-expert radiologists (e.g,. residents) diagnose routine cases and request help from experts in cases they find difficult. Recent work in organizational economics, starting with Garicano (2000), has analyzed how such knowledge hierarchies allow experts to exploit increasing returns from their knowledge by leveraging it through others' time.

What are the returns to "knowledge hierarchies?" In this paper we study this question empirically in a context where production depends strongly on solving problems: legal services. We analyze the earnings and organization of U.S. lawyers, and use the equilibrium model of knowledge hierarchies in Garicano and Rossi-Hansberg (2006) to estimate the returns to specialization that hierarchical production provides lawyers, and the impact this has on earnings inequality among these individuals. We conclude that hierarchical production has a substantial, but not dramatic, effect on lawyers' productivity and the distribution of lawyers' earnings. We believe that our results reflect the nature of production in this context: hierarchical production is valuable, but the return to hierarchy is limited when solutions tend to be highly customized and communication costs are high.

We proceed in two stages. We first propose an equilibrium model of problem-solving hierarchies, and show that the main empirical implications of such a model are consistent with our data, which come from the U.S. Economic Census and contain information on partners' earnings, associates' earnings, and associate-partner ratios at thousands of law offices throughout the United States. We then develop a structural estimation framework, estimate the model's parameters, and use these parameters to infer how much production would be lost if partners were not able to "vertically specialize" by delegating work to associates, and to construct earnings distributions across lawyers, comparing those we observe to those that would obtain if lawyers could not organize hierarchically.

Throughout, our analysis exploits the insight that the organization of production and earnings patterns within industries are jointly determined by the same underlying mechanism: the equilibrium assignment of individuals to firms and hierarchical positions. This equilibrium assignment, in turn, reflects the characteristics of the underlying production

function.[1] (Lucas (1978), Rosen (1982)) This insight contains an important empirical implication: earnings patterns contain a wealth of information that allows researchers to better understand the nature of production in an industry, and in turn, the industry's equilibrium organization. Our empirical analysis applies these ideas at several points, most prominently when we draw inferences about the nature of production from lawyers' earnings and organization, and when we develop a strategy for structural estimation.

In section II we propose a model of hierarchy, which is based on the equilibrium model of knowledge hierarchies with heterogeneous agents in Garicano and Rossi-Hansberg (2006). In this model, production involves the application of individuals' time and knowledge to problem-solving. Individuals have heterogeneous cognitive ability; some individuals can learn to solve problems at lower cost than others. Individuals choose how much knowledge to acquire and whether to work on their own or in hierarchical teams. When individuals work in teams, some individuals may communicate their knowledge to others – thus organizing production hierarchically allows more talented individuals to leverage their knowledge by applying it to others' time. More knowledgeable managers must be matched with more knowledgeable subordinates, as this allows agents to better leverage their knowledge by avoiding dealing with the routine problems others could also solve.

Garicano and Rossi-Hansberg (2006) show that equilibrium assignment in this type of model is characterized by three properties: scale effects associated with managerial skill, positive assortative matching, and strong stratification by skill. That is, better managers work with more and better workers, managers work with workers with dissimilar skill levels, and the least skilled manager is more skilled than the most skilled worker. These assignment patterns generate distinct earnings and organizational patterns. Scale effects and positive assortative matching together imply that managerial earnings, worker earnings, and the worker/manager ratio should be positively correlated; strong stratification implies that managers earn more than workers, even when comparing managers at firms with low worker/manager ratios to workers at firms with high worker/manager ratios. We discuss how these implications differ from those generated by other classes of production functions (such as O-ring functions where skills are strictly and symmetrically complementary (as in Kremer (1993)), or those where skills are strict substitutes (as discussed in Grossman and Maggi (2000)), and non-hierarchical production functions where individuals' skills are complements but affect production asymmetrically (Kremer and Maskin (2004)).

In section IV, we examine lawyers' earnings and organizational patterns, using data from the 1992 Census of Services (see Section III for a description of the data). These data

[1] Rosen notes that "the firm cannot be analyzed in isolation from other production units in the economy. Rather, each person must be placed in his proper niche, and the marriage of personnel to positions and to firms must be addressed directly." (322)

contain law-office-level information about revenues, the number of partners, the number of associates, and associate earnings. We use these data to infer how much partners and associates earn at each office. We examine earnings patterns and draw inferences about the equilibrium assignment of lawyers to each other, to organizational positions, and to firms. Our evidence indicates that, consistent with positive sorting, higher-earning partners work with higher-earning associates. Perhaps more surprisingly, higher-earning associates work in offices with greater associate-partner ratios. These patterns are true both within and across local geographic markets. We also find evidence consistent with strong skill stratification: controlling for their field of specialization, partners in offices with the lowest partner-associate ratios earn more than associates in offices with the highest partner-associate ratios. That is, the least-leveraged partners earn more than associates do, even those associates who work at offices with high partner-associate ratios. These empirical patterns are consistent with the equilibrium assignments generated by our model.

We then consider the implications of our model with respect to the equilibrium assignment of individuals to markets of different sizes. If increasing returns lead highly-skilled managers to work in the largest markets, the equilibrium assignment patterns depicted above imply a distinctive cross-market pattern in which the probability that an individual works in a large market rises, falls, then rises with their skill. This is because as an individual's skill increases, their comparative advantage changes from being a worker supporting a highly-skilled, highly-leveraged manager (who works in a large market) to being a low-leverage manager who works in a small market. We show evidence that suggest such a pattern: the relationship between individual lawyers' earnings and the size of the local market in which they work is non-monotonic. When looking at local markets in a given size range, earnings distributions among lawyers appear to be bimodal, with the location of both modes in the earnings distributions increasing as one moves from smaller to larger local markets. These spatial patterns are easily rationalized by production functions such as those we propose that involve limited quality-quantity substitution in individuals' human capital, and in which cross-matching obtains in equilibrium.

Thus overall, we find a production function like the one we propose fits reasonably the main aspects of the data.

In section V, we move from testing to estimation, and propose an econometric framework in which we can estimate this type of production function from equilibrium earnings and organization data. This framework exploits close connections between equilibrium assignment models and the hedonics literature. Our econometric framework exploits two crucial features of the model. First, leverage is a sufficient statistic for worker skill Second, the productivity of a hierarchical team, per unit of productive time, is determined only by the manager's skill. These two features allow us to avoid some of the main difficulties

involved in hedonic models. As a result, we can obtain consistent estimates of the crucial parameter in our model: the time cost of team production. We show that we can identify this parameter from the ratio between team average product and the marginal cost of leverage. This, in turn allows us to recover the team production function and generate counterfactuals that indicate what lawyers would produce and earn absent hierarchical production.

We find that hierarchical production increases lawyers' productivity substantially: it increases output by at least one-third, relative to non-hierarchical production in which there is no vertical specialization within offices. We also find that hierarchies expand substantially earnings inequality, increasing the ratio between the 95th percentile and median earnings among lawyers from 3.1 to 4.5, mostly by increasing the earnings of the very highest percentile lawyers in business and litigation-related segments, and leaving relatively unaffected the earnings of the less leveraged lawyers. Though these effects are substantial, we believe them to be far smaller than in other sectors of the economy. We discuss the source of these differences and what they may mean for production in the service sector in the paper's conclusion.

We see the contribution of the paper as methodological as well as substantive. Methodologically, we wish to reintroduce the idea that earnings patterns say a lot about the nature of human capital intensive production and about the underlying reasons for industries' equilibrium organization. This idea has been underexploited, in part because of the lack of data sets that contain not only information about individuals' earnings, but also on their position within their firms' organization and their firms' characteristics.[2] To exploit these patterns requires combining equilibrium analysis with organizational models. Evidence on who works with whom and in what capacity can be enormously informative, but inferences from such evidence must be based on equilibrium models since such models allow assignments to be based on individuals' comparative rather than absolute advantage.

Before jumping to our analysis, a few caveats are in order. Our approach, which emphasizes and exploits labor market equilibria, does not come for free. We largely abstract from most of the incentive issues that dominate the organizational economics literature, as well as many of the details of internal labor markets. We also must place restrictions on agent heterogeneity so that our equilibrium does not involve sorting on multiple dimensions. The returns to this approach are considerable, however, as it provides for a tractible equilibrium model from which we can estimate the impact of organization (or, equivalently, the impact of vertical specialization) on lawyers' output and the distribution of lawyers earnings. In short, this approach allows us to develop a

[2]It might also reflect an intellectual separation between the fields of labor economics and industrial organization that Rosen (1982) was trying to bridge.

first estimate of the return to hierarchy.

The paper is structured as follows. Section II discusses the general existing theoretical results on equilibrium assignment under different assumptions about production, and about scale of operations effects. In section III we describe our data. Section IV analyzes earnings patterns in legal services in light of these models. Section V discusses and presents the estimates of our structural model and analyzes how much hierarchical production affects lawyers' output and the earnings distribution among lawyers. Section VI concludes.

II. HIERARCHIES, ASSIGNMENT, AND HETEROGENEITY

II.1. A problem-solving hierarchy

We develop a simplified version of the model of hierarchy in Garicano and Rossi-Hansberg (2006), with only two layers of hierarchy and exogenous knowledge. All agents are endowed with a skill level z and with one unit of time. The population is described by a given distribution of skill, $G(z)$, with density function $g(z)$. Production involves the application of these agents' time and knowledge to solving clients' problems. Skill is unidimensional and vertical; z can thus be thought of as an index that reflects the share of client problems that an agent can solve. Thus, more-skilled agents can solve a greater share of these problems than less-skilled agents, and the problems that a less-skilled agent can solve are a subset of those that a more-skilled agent can solve. Throughout this paper, we normalize the skill units z to dollars, so that an agent with skill z can solve problems with dollar value z. We will then think of z as the earnings an agent can attain when working on his own.

Agents can either work on their own or form hierarchical teams. Hierarchical teams are comprised of a manager with skill z_m and n workers with skill z_w.[3] We assume that managers can apply their knowledge toward problems that workers cannot solve by themselves, but that managers must spend time communicating with the workers when they do so. Less knowledgeable workers require more help per worker, and thus the span of the manager, n, is limited by the knowledge of the workers through the manager's time constraint. This time constraint imposes that the workers' skill, z_w, and the worker/manager ratio, n, are linked by a function of the form $n(z_w)$, with $n'(z_w) > 0$.[4]

[3] Hierachical teams will be optimal when matching problems and knowledge is difficult (see Garicano 2000).

[4] In Garicano's original model, $F(z)$ is the probability that an agent can solve a problem and $(1-F(z))$ is the probability that he asks for help, and each time a worker asks for help costs the manager a share h of this time endowment. Since a manager has 1 unit of time, the number of workers who may work under this manager is given by $n(z_w) = 1/(h(1 - F(z_w)))$. For our empirical purpuses, the specific relation between n and z_w is irrelevant, and we simply will write $n(z_w)$.

The production function of a hierarchical team is given by:

$$y = z_m f(n(z_w))$$

This function has two terms: z_m, the manager's skill, and $f(n(z_w))$, the amount of time team members spend in direct production. The function $f(n)$ is a mapping from the team members' time endowment $(n+1)$ to actual productive time available. $f(n)$ accounts for the possibility that these two quantities should differ if hierarchical production requires agents to spend time communicating or coordinating. Throughout the rest of the paper, we will assume that $f(n) = (n+1)^\theta$, so that n units of worker time and 1 unit of manager time results in $(n+1)^\theta$ units of time spent in production.[5] We assume $\theta < 1$, so that hierarchical production is costly in terms of the time agents spend in production. The specification implies that if an agent works on his own $(n = 0)$, then $f(0) = 1$; individuals working on their own do not incur communication or coordination costs.

The top part of Figure 1 depicts production under nonhierarchical production, in which agents work on their own. The left side of this panel depicts the time and knowledge of $(n+1)$ agents. The lines depict these agents' time endowments, the shaded regions depict these agents' knowledge. n of these agents have knowledge z_w, 1 has knowledge z_m. Assume that each of these agents confront a set of problems that vary in their difficulty, and that each of these sets requires one unit of agent time to handle. These $(n+1)$ sets of problems are depicted on the right. Under nonhierarchical production, each of these agents simply handles the problems they themselves confront. Output of each of the n lower-skilled agents would be z_w and output of the higher-skilled agent would be z_m. Total output would be $z_m + nz_w$.

The bottom part of this Figure depicts hierarchical production. Total output is $z_m(n+1)^\theta$, the product of the manager's skill and the time the $(n+1)$ agents are able to spend in production. Output per unit of productive time is improved, relative to autarchic production, because problems are allocated to workers and managers according to their comparative advantage; workers handle the easiest problems the group confronts, while managers handle the hardest ones. This improvement is the benefit of hierarchical production; the drawback is that hierarchical production involves a loss in time spent in production.

We note here that this production function has several key elements. First, individuals of different skills are not perfect substitutes to one another; difficult problems can only be solved by highly-skilled agents. Second, managers' and workers' skills are complementary.

[5] We choose this flexible functional form for convenience in the econometric estimation. In Garicano (2000) and Garicano and Rossi-Hansberg (2006), the manager specializes in problem solving and thus the cost of team production is his unit of time. That is $f(n) = n$, rather than $f(n) = (n+1)^\theta$ as here.

Third, output is asymmetrically sensitive to managerial and worker skill; the asymmetric sensitivity follows naturally from the fact that teams are formed by workers and a manager, whose skill increases the productivity of all of the workers to which this skill is applied. Last, and related, managers can exploit scale effects associated with their human capital.

II.2. Equilibrium

Obtaining an equilibrium in this economy implies solving a continuous assignment problem with two twists relative to standard assignment problems. First, who is assigned to whom is not a given, but an equilibrium outcome. In standard assignment models this identity is assumed. In contrast, here we are "marrying" a mass of workers with a mass of managers– *where those roles and masses are not given by assumption.* Second, agents can decide not to be matched and produce on their own.[6]

To solve the assignment problem, note first that optimality requires positive sorting, that is, workers with more knowledge must be assigned to managers with more knowledge. The reason is that there is a complementarity between knowledge of workers and managers through the manager's time constraint. A more knowledgeable manager will spread his higher knowledge over a larger number of workers, and that requires workers to be more knowledgeable so that they do not ask an excessive number of questions.[7]

To characteize the equilibrium in this economy, we need to describe three objects: first, the allocation of agents to positions – workers, managers, and "unleveraged" indiviudals who are neither managers nor workers; second, the team composition – i.e., the matching between workers and managers and the number of workers per manager; and third, the earnings function. All these objects form an equilibrium, where earnings are such that agents do not want to switch either teams or positions.

The equilibrium is characterized by a pair of thresholds (z^*, z^{**}), such that all agents with knowledge $z < z^*$ become production workers, all agents with knowledge above z^{**} become managers, and those in between are "unleveraged."

Then suppose a mass n of workers with knowledge z_w and a mass 1 of managers with knowledge z_m are matched together in a team. For this to be an equilibrium it must be the case that the assignment maximizes managerial rents, that is, that the manager would not be better off matching with either less knowledgeable or more knowledgeable workers. Manager's rents are given by

$$R(z_m) = \max_{z_w} z_m f(n(z_w)) - w(z_w)\, n(z_w) \tag{1}$$

[6] We sketch only the equilibrium construction, see Garicano and Rossi-Hansberg (2006) and Garicano and Rossi-Hansberg (2004) for the details.

[7] Formally, $\partial^2 y / \partial z_w \partial z_m = n'(z_w) > 0$.

It follows that a necessary condition for the assignment to be an equilibrium is that workers wages' equal the marginal value of workers' knowledge, that is,

$$z_m f'(n(z_w)) = w(z_w) + \frac{w'(z_w)}{n'(z_w)} n(z_w) \tag{2}$$

In words, the marginal value of an increase in leverage is the skill of the manager times the increase in effective time. The marginal cost is the extra wage cost w, plus the increase in wages driven by the need for more skilled workers required by larger teams, w'.

A second equilibrium condition is the market clearing one. Given wages and earnings, the supply and demand of production workers equalize, namely,

$$\int_0^{z_w} g(z)\, dz = \int_{m(0)}^{m(z_w)} n\left(m^{-1}(z)\right) g(z)\, dz \text{ for all } z_w \leq z^*, \tag{3}$$

where the matching function $m(z_w)$ denotes the knowledge of the manager assigned to workers with knowledge z_w. Since (3) holds for all $z_w \leq z_1$, we can differentiate with respect to z_w, to obtain

$$m'(z_w) = \frac{1}{n(z_w)} \frac{g(z_w)}{g(m(z_w))}, \tag{4}$$

which, together with $m(0) = z^{**}$ and $m(z^*) = 1$, determines the equilibrium assignment function $m(z)$.

Finally, the occupational choice of agents must be optimal. Given equilibrium assignment and wage functions we can determine the rents of a manager with skill z_m, $R^*(z_m)$. A worker can always choose to become self-employed and get z. Thus, equilibrium earnings are given by $U(z) = \max\{z, R^*(z), w(z)\}$. This implies that the marginal production worker (the most knowledgeable one) must be indifferent between being a production worker or being self-employed, $w(z^*) = z^*$, and the marginal manager (the least knowledgeable one) must be indifferent between being a manger and being self-employed, $R^*(z^{**}) = z^{**}$. These conditions allow us to solve for the earnings and assignment functions. Figure 2 presents a graphical depiction of the resulting earnings function.

We will use equation 2 and the equilibrium relationship between associates' earnings and n to estimate the production function But before doing this, we proceed to extract some empirical implications from this model that we can take to the data.

II.3. Implications for equilibrium assignment and earnings patterns

Equilibrium assignment under this "hierarchical production function" has three important characteristics, as discussed in Garicano and Rossi-Hansberg (2006).[8] First, it involves positive sorting, which follows directly from the complementarity between managerial and worker skill. Intuitively, a more highly-skilled manager has a comparative advantage in working with more highly-skilled workers, since such workers allow the managers to apply their human capital to a greater amount of worker time. Second, since $n' > 0$, positive sorting implies that there exist scale of operations effects: more highly-skilled managers manage larger teams. Third, the equilibrium involves strong stratification: that is, in equilibrium there must exist some skill level such that all agents of skill below a given level are workers, and all of those above that level are working on their own or are managers. This is less straightforward in this case, but an analogous proof to the one in Garicano-Rossi-Hansberg (2006) holds. Informally, the gist of the argument is as follows. Suppose that in a equilibrium a worker a with skill z_w^a were more skilled than a manager in a different team, b, with skill z_m^b. This would mean that a problem faced by b that he cannot solve remains unsolved, while it does pay to solve a problem faced by a which he cannot solve. But the fact that $z_m^b < z_w^a$ means that the problems a cannot solve are harder, and thus it cannot be optimal to ask those questions, and this is assignment is not optimal.[9]

Proposition 1 *(Garicano and Rossi-Hansberg 2006) Equilibrium assignment with hierarchical production functions has the following properties.*

1. *Positive sorting. More highly skilled managers work with more highly skilled workers.*
2. *Scale of operations effects. More highly skilled managers manage larger teams of workers.*
3. *Stratification. The least skilled manager is more skilled than the most skilled worker.*

These characterisitics are summarized in Figure 2, which characterizes the resulting agents' equilibrium earnings as a function of their skill. Individuals below z^* are associates, while those above z^{**} are leveraged partners. The slope of this earnings function increases discontinuously at z^{**}; this reflects the impact of leverage on partners' earnings, which in turn is determined by $f(.)$.

[8] To be precise, all of these results are obtained in GRH (2006) with a production functions of the form zn,rather than $zf(n)$,and with a specific form for $n(z_w)$ given by the (hierarchical) nature of problem solving. As we argue below, analogous results hold in this case.

[9] Note that the argument for stratification requires that agents may form three layers hierarchies, which we have not considered formally above for simplicity. A formal proof would be considerably more involved, and is presented in Garicano and Rossi-Hansberg (2006).

II.4. Alternative production functions and equilibrium assignment patterns

The previous subsections brought us from a production function to implications about equilibrium assignment. Such a path is a common theme of the literature on equilibrium assignment: the nature of the interaction in production between individuals' skills determines the equilibrium assignment of individuals in an economy to each other (and thus to productive units) and to positions. This, in turn, shapes the equilibrium organization of production and the distribution of earnings. Thus different forms of production functions result in different implications for earnings and assignments. We discuss some alternative production functions and assignment implications in what follows. Although we keep the discussion quite informal and focus on presenting the ideas behind these results,[10] the reader should keep in mind an economy with a continuum of agent types and a type space that is a compact subset of the real line, where there is a continuous probability distribution over types. This discussion points out how the assignment patterns generated by our model summarized in Proposition 1 are distinct from those generated by production functions contemplated elsewhere in the literature. The following section will examine earnings and assignment patterns in the context of law firms in light of this discussion.

Non-hierarchical production Production functions differ in how agents' skills interact – in particular, whether production is supermodular or submodular in agents' abilities – and whether production is symmetrically or asymmetrically sensitive to individuals' abilities.

First, consider production functions with symmetric complementarities (Becker, 1981, 1993), such as $y = z_1 z_2$. These production functions capture situations where for example all tasks have to be accomplished for success, and skill determines the probability of success on a given task Kremer (1993). Individuals' willingness to pay to be paired with an individual with a given talent level is increasing with their own talent, and thus these production functions produce self-matching or segregation in equilibrium – those in each team have equal ability. This self-matching stands in contrast with the cross-matching that obtains in our model.[11]

Second, consider production functions that are submodular in individuals' abilities, so

[10] See Sattinger (1993) for a good review of the literature on this topic and Legros and Newman (2002) for the formal exposition of a set of general conditions characterizing positive and negative assortative matching in equilibrium.

[11] Note that equilibrium assignment between managers and workers in our model never involves self-matching - a worker never has as much skill as his manager. To see this, note that an agent with skill z who works on his own earns z. A team of $n+1$ such agents working together in a hierarchy with one acting as manager and n acting as workers earns $zf(n(z))$, which is less than they would earn if each worked on its own, $(n+1)z$. When workers are identical, the team produces less than all the workers would produce on their own, and thus it is not formed. Equilibrium assignment therefore must involve some cross-matching.

that individuals' abilities are substitutes. This would be the case, for example, if only the best idea or most skillful execution matters; other ideas or efforts turn out to be redundant. These production functions imply negative assortative matching – the more able the manager the less able the workers (See Proposition 3 in Legros and Newman (2002) for a precise statement of this result).[12] This is unlike our model, in which there is positive sorting in equilibrium.

Third, consider production functions that are supermodular in individuals' abilities, but in which production is asymmetrically sensitive to individuals' abilities, such as the function proposed by Kremer and Maskin (1997):

$$y = z_1^{\theta} z_2^{1-\theta}, \tag{5}$$

with $\theta > 1/2$, so that production is more sensitive to the ability of the individual assigned task 1 than task 2. Kremer and Maskin show that equilibrium assignment may involve either self-matching or cross-matching, depending on the support of the distribution of skills, and in the cross-matching outcome stratification obtains: all agents above a given ability threshold work in task 1, and all agents below it work in task 2. The combination of complementarity and asymmetry in Kremer and Maskin's production function is also present in our model, so it follows that it shares some of its most important implications: positive sorting and stratification. However, Kremer and Maskin's analysis differs from ours because their production function involves two agents. It therefore cannot generate implications with respect to the match between individuals and worker/manager ratios. In our model, asymmetric sensitivity arises precisely because of the way production can be organized: one agent's talent, the manager's, can affect the productivity of all of those with whom he or she works. This leads to an important implication that is not part of Kremer and Maskin's analysis: part 2 of Proposition 1, which concerns scale of operations effects.

Hierarchical Production Scale effects associated with managers' human capital is not a novel concept. A long-standing literature, starting with Simon (1957), and including papers by Mayer (1960), Lucas (1978), Calvo and Weillisz (1979), Rosen (1982) and Waldman (1984), has proposed that the reason that the distribution of income is more skewed than the underlying distribution of skills lies in how resources are allocated to

[12] Suppose, for example, that production requires two individuals, and output takes place if and only if at least one individual knows the solution to a particular problem, and suppose that the probability that individual i knows the solution is z_i.Then output is given by the submodular production function $y = 1 - (1 - z_1)(1 - z_2)$. A production function like this was first suggested by Sah and Stiglitz (1986) in the context of project screening within a "polyarchy:" a project is approved if at least one division head likes it.

individuals.[13] In these models, higher-ability managers raise the productivity of the resources they are assigned more than lower-ability managers. As a result, in equilibrium, more able managers are allocated more resources, and this leads the marginal value of their ability to increase faster than if they were working on their own.

Production functions in this literature have the generic structure:

$$y = z_m f(n) \tag{6}$$

where z_m is managerial human capital and n is the manager's span of control, which, depending on the model, may be the number of workers (Lucas, 1978), efficiency units of labor, i.e., total units of skill managed (Rosen, 1982), or physical capital. In these models, managerial human capital z_m shifts up the marginal product of the workers or capital they are assigned, but managers' span of control is generally limited implicitly or explicitly by managers' time.

Equilibrium assignment patterns in these models share aspects of our model. In particular, they involve scale of operations effects: more skilled managers are assigned more resources to manage in equilibrium. As a result, the distribution of earnings is more skewed than the distribution of skills. However, this class of models has generally assumed perfect substitutability among the resources managed by the manager, so that only the quantity of resources, and not the quality of which they are composed, matters.[14] Absent an element of imperfect substitutability between workers of different skill, these models do not allow for a full analysis of either the equilibrium assignment of individuals to each other or of earnings distributions; if skilled and unskilled workers are perfect substitutes, in equilibrium managers should be indifferent between working with a few relatively skilled workers or many unskilled workers. Assignment patterns between individual managers and workers would then be indeterminate. Our model allows for a more complete analysis of assignment and earnings patterns because it combines imperfect substitutability of the form in (5), with scale effects of the form in (6), and this completeness facilitates our structural estimation below.

This type of production functions are most applicable in human capital intensive industries, where the most important inputs are individuals' skills and where organizational structures are designed to exploit these skills. It is thus natural to think that optimiz-

[13] See Gabaix and Landier (2006) for a modern application of this type of theory to trends in CEO pay.

[14] For example, in models where productive resources are human capital, either only the total number of workers matters (as in Lucas (1978)) or workers of different skill are perfect substitutes (as in Rosen (1982)). In Waldman's (1984) more general model, no restrictions on the interaction between managers and workers skills are imposed, but that allows only to characterize the correlation between ability levels and hierarchical position and the fact that the wage distribution is more skewed to the right than the ability distribution. The specific model he analyzes does not allow for complementarities between worker and manager skill and as a result has equilibria with workers more skilled than managers.

ing the utilization of human capital is an important concern in the production of legal services; we turn to an analysis of earnings patterns in this industry.

III. DATA

The data are from the 1992 Census of Services. Along with standard questions about revenues, employment, and other economic variables, the Census asks a large sample of law offices questions about the number of individuals in various occupational classes that work at the office and payroll by occupational class. For example, it asks offices to report the number of partners or proprietors, the number of associate lawyers, and the number of non-lawyers that work at the office. It also asks payroll by occupational class: for example, the total amount associate lawyers working at the office are paid. These questions elicit the key variables in our analysis. Other questions ask offices to report the number of lawyers that specialize in each of 13 fields of the law (e.g., corporate law, tax law, domestic law) and the number of lawyers who work in multiple fields. These variables allow us to control for the field composition of lawyers at various points in our analysis.

These data have several aspects that lend themselves to an analysis of equilibrium assignment. They cover an entire, well-defined human-capital-intensive industry in which organizational positions have a consistent ordering across firms, and allow us to construct estimates of individuals' earnings at the organizational position*office level at a large number of firms. This allows us to explore how individuals' earnings are related to others with whom they work, their organizational position, and characteristics of the firm and market in which they work. Data that allows one to connect individuals' earnings with firm characteristics across firms is not common, and it is even less common to be able to connect earnings with individuals' organizational position. These data have shortcomings, however: whether they contain information about organizational positions depends on firms' legal form of organization, they do not directly report partners' earnings, and at best they provide information on earnings at the organizational position*office level rather than the individual level. We next discuss these shortcomings and how we address them.

Responses and Firms' Legal Form of Organization Responses to some of the Census' questions have different meanings, depending on the office's legal form of organization. The reason for this is that all lawyers are legally considered associate lawyers at offices that are legally organized as "professional service organizations" (PSOs) such as limited liability corporations. This is true even though lawyers at these offices distinguish among themselves in the same way they do at offices legally organized as partnerships:

some are partners and others are associates. The variables the Census collects thus differ between PSOs and partnerships.

Table 1 summarizes these differences. The data report the number of lawyers (and non-lawyers) regardless, but distinguish between partners and associates only at partnerships. The data report payroll of all lawyers at PSOs (since all lawyers are legally associates), but only the payroll of associate lawyers at partnerships. The data do not directly report the earnings of partners at partnerships, since these individuals are legally owners rather than employees; their earnings are not considered payroll. The data contain revenues, as reported from tax forms, for all offices, but not non-payroll expenses ("overhead").

The data on partnerships are advantageous because they are disaggregated within establishments; they distinguish between partners and associates. This disaggregation is important for our analysis, both because it allows us to examine the implications and estimate parameters of the hierarchical production model described above and more generally because it brings the analysis closer to the individual level. However, the data on partnerships do not directly report partners' earnings. To use these observations, we must therefore generate estimates of partners' earnings based on the data we have at hand. We next describe how we do so.

Estimating Partners' Earnings Partnerships commonly pay out to partners their earnings net of expenses during the year. Thus, earnings per partner at office i, R_i, can be depicted by the identity:

$$R_i = (TR_i - w_i n_i p_i - x_i l_i - oh_i)/p_i$$

where TR_i is total revenues at office i, w_i is average associate earnings at office i, n_i is associates per partner, p_i is the number of partners, x_i is non-lawyer earnings per lawyer, $l_i = p_i(1 + n_i)$ is the number of lawyers, and oh_i is overhead. This can be rewritten as:

$$R_i + oh_i/p_i = (TR_i - w_i n_i p_i - x_i l_i)/p_i$$

The data on partnerships contain the variables on the right side of this expression. Thus, we observe the sum of partners' earnings and operating expenses. We do not observe R_i and oh_i separately for partnerships; our task is to distinguish between these.

The above identity also implies:

$$oh_i = TR_i - (R_i p_i + w_i n_i p_i) - x_i l_i$$

The observations of PSOs contain each of the three terms on the right hand side, and thus allow us to impute overhead for each of these offices.

Our approach for estimating partners' earnings is to use the data from the PSOs to develop estimates of overhead for each of the partnerships in the data. By the identity above, estimates of operating expenses imply estimates of partners' earnings.

The Census' Operating Expenses Survey provides evidence on the nature of law offices' "overhead" expenses.[15] A significant share of these expenses are closely connected to payroll; these include "employers' cost of fringe benefits": the firm's contribution to Social Security, health insurance, retirement plans, and so on. These expenses amount to about 15% of payroll in the aggregate; additional evidence from Altman Weil's 1994 Survey of Law Firm Economics indicates that this 15% figure is consistent across firms. Other expenses are more closely associated with running the office: for example, leasing and rental payments (on average, about 23% of overhead), and office supplies and phone and communication expenses (combined, about 10%). In general, some office expenses are closely related to the location and employment size of the office (e.g., rent), others are more closely related to how much business takes place (e.g., communication). This evidence shapes our specification of law offices' overhead expenses below.

Overhead at PSOs We use the data from PSOs extensively to examine what affects overhead, in light of our previous knowledge of the structure of law firms' costs. In particular, we are mindful of the following:

- "Non-payroll fringe benefits" are consistently about 15% of payroll.
- Operating expenses increase with the office's scale; some elements with the number of people in the office and some with the amount of business.
- Some operating expenses such as rent should be higher in larger markets.
- Offices' cost structure might differ depending on whether they serve businesses or individuals (e.g., the former might involve more travel or business development expenses). It also might differ depending on whether they are involved in "transactional" or litigation work.

We account for the first of these by simply assuming that fringe benefits are 15% of payroll for all offices, which allows our data to be used to explain variation in $oh_i^* = TR_i - 1.15 * [(R_i p_i + w_i n_i p_i) - x_i l_i]$. We specify oh_i^* as a function of market size, revenues, and the number of individuals working at the office ("employment"), interacting market size and employment to allow for the fact that additional office space may be more costly in larger markets. Furthermore, we define three classes of offices and allow all of our coefficients to vary across these classes. We define "litigation" offices as those with at least

[15] Bureau of the Census (1996).

one lawyer specializing in a litigation-intensive field (negligence, insurance), and classify the remainder as "business, non-litigation" and "individual, non-litigation" depending on whether the office's primary source of revenues is from businesses or individual clients.

We found that a relatively parsimonious specification, which does not include full interactions, fits the data nearly as well as one that included the full interactions. We report the coefficient estimates from this specification in Table 2.[16] We allow the intercept term to vary with indicator variables that correspond to the employment size of the county in which the office is located, include interactions between employment and these market size measures. The coefficient estimates imply that the fixed overhead cost of a very small law office is on the order of $40,000, and that this does not vary significantly with market size. The interactions suggest that the overhead associated with each additional individual is about $6,500 but this tends to be much greater in very large markets. We allow the coefficient on revenues to differ depending whether the office has a litigation specialist and if not, whether it serves businesses or individuals.[17] The estimates indicate that the relationship is strongest for "business, non-litigation offices" and weakest for "individual, non-litigation offices." The estimates indicate that overhead increases by about $0.12-$0.17 with each $1.00 increase in revenues. The differences across office classes are statistically significant.

The R-squared for this regression, 0.78, is high. We found that more detailed specifications, including those that include county fixed effects instead of the market size dummies and that interact "office class" with the employment variables, increase the R-squared by very small amounts and generate almost exactly the same distributions in lawyers earnings as those reported later in this paper.[18]

Comparing Estimated and Actual Earnings Distributions A first step is to compare distributions from actual data and the estimates using only the PSOs: if we apply the procedure to the PSOs, do we obtain a distribution close to what we started from?

The left side of Table 3 reports the results from this exercise. We compute the distribution of lawyers' earnings across offices, weighting each office by the number of lawyers, among PSOs. The median is $96,000; the 10^{th} and 90^{th} percentiles are $48,000 and

[16] We included [*employment*-2] rather than *employment* in these regressions. Our sample only contains observations of offices with positive employment, thus the smallest office in our sample has two individuals: a lawyer plus a non-lawyer. This normalization allows us to interpret the intercepts in terms of the fixed cost of operating a very small office.

[17] We do not include the "office class" dummies separately. The fixed cost associated with a very small office should not vary with the segment in which it operates (e.g., rent should not differ), and this imposes this as a restriction.

[18] This likely reflects that (a) the cost of office space varies little across most counties, and (b) the relationship between operating expenses and employment – which largely reflects costs associated with office space, furniture, computer equipment, etc. – indeed should not vary depending on the details of what a law office does.

$179,000, respectively. The second columns report these percentiles when using the predicted values generated by the overhead regression. The two distributions are extremely similar at all of the quantiles. Our estimates match the mean by construction, but the fact that they match the quantiles well implies that our specification is able to capture much of the within- and across-market variation in overhead expenses among offices in this sample.

The right side compares estimates of the (imputed) earnings distribution among lawyers in partnerships and proprietorships with those generated from other Census data that contain individual earnings data: the Census' Public Use Microdata Sample (PUMS).

The PUMS data contain individuals' response from the 1990 Census of Population. We use the 5% State Sample. Among other things, the Census asks individuals their occupation, the industry in which they work, their usual hours of work, the number of weeks they worked in the previous year, and their business and salary income. We extract observations of full-time lawyers working out of law offices.[19] We convert all dollar amounts to 1991 dollars to make them comparable to those reported in the Census of Services data.

A drawback to the PUMS data is that the earnings data are top-coded. Individuals' business income is top-coded if it exceeds $90,000; their salary income is top-coded if it exceeds $140,000. Thus, earnings distributions derived from PUMS reflect actual responses only below $90,000, which is approximately $99,000 in 1991 dollars. About two-thirds of lawyers in the PUMS have earnings less than this level.

The first column on the right side of the table reports quantiles of lawyers' earnings distribution generated from the PUMS data. We report these for the 10^{th}-60^{th} percentiles because the earnings data are top-coded above these levels. The median lawyer in our PUMS subsample earned $71,442.

The second column reports estimates derived from the partnerships and proprietorships in our Census data, using estimates of overhead generated from the specification in Table 3. The distribution generated by this method tracks that generated by the PUMS data fairly closely, though the estimates are consistently $5,000-$10,000 higher than the PUMS quantiles in the middle of the distribution. This comparison suggests that our estimates of partner pay might be somewhat high. Our main analysis will revolve around how much lower partner pay would be, absent hierarchical production: the difference between our estimates of partner pay and a counterfactual. The counterfactuals we present are just transformations of estimated partner pay, so if our estimate of partner pay is somewhat high, so will be the counterfactual. Our estimate of the difference between the two will

[19] We extract observations of lawyers who worked out of law offices (rather than as judges or as in-house counsel), and eliminate those reporting that they were not in the labor force, whose usual hours were less than 40 hours per week, and who worked fewer than 46 weeks during the previous year. We also eliminate individuals younger than 25 or older than 70 years.

be affected far less.

Aggregation of Individuals' Earnings Our data do not allow us to distinguish among associates or among partners who work at the same office. This aspect of our data limits our analysis of equilibrium assignment patterns: we cannot examine the matching among partners and among associates. In other work, we find evidence suggestive of positive assortative matching across firms within these organizational positions. In Garicano and Hubbard (forthcoming), we use data from the "blue page" listings of law offices throughout Texas from the Martindale-Hubbell directory of lawyers, and show that partners work disproportionately with other partners who obtained their degree at a similarly-ranked law school, and with other partners with similar experience levels. Similar patterns hold for associates.

Our discussions of earnings patterns and what they imply about the nature of human-capital-intensive production will downplay assortative matching among partners and among associates, simply because we cannot investigate it empirically here. We suspect that there is positive assortative matching within organizational positions, and that it might take the form of self-matching, but further research with individual-level earnings data is necessary to determine whether this is the case. Such research would lend further insights on equilibrium assignment and the nature of human-intensive-production in this context.

Our analysis will also tend to understate earnings heterogeneity across lawyers, because at best we can examine earnings at the organizational position*office level rather than at the individual level. Although most of the earnings heterogeneity among partners (or among associates) is across offices rather than within offices, there is sometimes quite a bit of within-office heterogeneity, especially at very large offices where there are different tiers of partners.[20] This would lead us to understate the very highest quantiles when we construct earnings distributions. However, Census disclosure restrictions constrain what we can report, and we do not report any quantiles above the 95th percentile for this reason. We therefore do not think this issue has a large impact on the results we present and discuss below.

IV. EARNINGS PATTERNS AND HIERARCHICAL PRODUCTION

We next test the hypotheses derived from the model (Proposition 1); this tees up the structural estimation to follow, which takes the model's assumptions as maintained. Some of the hypotheses are intuitive, and can be generated by a broad class of production functions – (e.g., positive correlation between partner and associate earnings). Others are less obvious, such as the correlation between associate earnings and the associate/partner

[20]It should be noted, however, that multiple tiers of partners is currently only common among very large firms, and in 1992, the time of our data, it was much less common than it is now.

ratio and strong occupational stratification in earnings. We proceed in what follows from the weaker to the more powerful tests.

Associates and Partners' Earnings Are Positively Correlated. Our first evidence comes from simple regressions of average associate earnings within an office on average partner earnings within an office, using offices with at least one associate. Results are in Table 4. Panel A reports the coefficient on ln(partner earnings) in five regressions. In the first, there are no controls. The coefficient is positive and significant. The point estimate of 0.343 indicates that, on average, associate earnings are 34% higher where average partner earnings are 100% higher. The second column includes a vector of field controls; this vector includes the share of lawyers in the office that specialize in each of 13 fields (e.g., corporate law, probate law). The third and fourth control for geographic market differences. In the third, we include a vector of five dummies that correspond to the employment size of the county in which the office is located;[21] in the fourth, we instead include county fixed effects. The fifth column controls for the office's scale in terms of partners by including $partners$, $partners^2$, and $partners^3$. The coefficient on *ln(partner earnings)* decreases when including the field and market controls, indicating that part of the raw correlation captures cross-field and cross-market differences in average earnings. The result in the fourth column, which includes county fixed effects, indicates that associates' and partners' earnings are positively correlated within as well as between markets. The coefficient decreases only slightly when we control for the office's scale in terms of partners, indicating that the correlation between partner and associate earnings does not reflect that both partners and associates tend to earn more in offices with more partners. Throughout, the coefficient on *ln(partner earnings)* remains positive and significant; the coefficient in the last column indicates that on average, associate pay is 18% higher at offices where partner pay is twice as high.

We have run analogous specifications using various subsamples of the data. We ran them separately for "business" and "individual" offices depending on whether 50% of revenues come from clients who are businesses or individuals. We ran "within market" specifications, running them separately for offices in several very large counties (Manhattan, Los Angeles County, etc.). Finally, we conducted "within field" specifications by using only offices where all lawyers work in a particular field (e.g., patent law). We find that the results we report above are robust: partner and associate earnings are positively correlated in these specifications as well.

These results show strong evidence that associate earnings are higher at offices where partner earnings are higher. While they need not necessarily reflect that associates'

[21] These correspond to the following employment size categories: 20,000-100,000, 100,000-200,000, 200,000-400,000, 400,000-1,000,000, and greater than 1,000,000.

and partners' ability is positively correlated – a positive correlation in earnings could be driven by office-level demand shocks (everyone receives a bonus in good years) – they are consistent with production functions that generate positive assortative matching in equilibrium. Such functions include those in which individuals' skills are complements, but not those in which they are substitutes.

Associates' and Partners' Earnings Are Positively Correlated with Associate-Partner Ratios. We next investigate whether associates' and partners' earnings are higher at offices where the associate/partner ratio is higher. Panel B in Table 4 reports results from specifications where we regress ln(partner earnings) on ln(associates/partner). In each of the specifications, the coefficient on ln(associates/partner) is positive and significant. In Panel C, the dependent variable is instead ln(associate earnings). Once again, the point estimates are positive and significant.[22] Using the results from the last column and comparing offices where one has an associate/partner ratio that is twice as high as the other, average partner pay is 32% higher and average associate pay is 11% higher at the office with the higher associate/partner ratio. The elasticity between partner earnings and the associate/partner ratio is about three times that between associate earnings and the associate/partner ratio.

Our results provide evidence consistent with a key implication of hierarchical production functions: that comparing earnings among individuals who are at the same hierarchical rank, those who work in groups with more lower-level individuals per upper-level individual earn more. This result also shows that the correlation between associate and partner earnings reported in the previous subsection do not just reflect transitory earnings shocks, unless these shocks also lead associate/partner ratios to change.

Stratification. An important aspect of equilibrium assignment under our model is that it should lead to stratification. In this context, this implies that all associates should be less able than any partner. The evidence above suggests that more able associates work in offices with higher associate/partner ratios, as do more able partners. Thus, in this context stratification requires in particular that partners in offices with low associate/partner ratios have higher ability than associates in offices with high associate/partner ratios.

We examine this by investigating the ordering of lawyers' earnings. We classify lawyers according to whether they are partners or associates, and the associate/partner ratio of their office. Regarding the latter, we create four categories: less than 0.5, between 0.5 and 1.0, between 1.0 and 2.0, and greater than 2.0. This divides lawyers into eight categories. We refer to the associate categories as A1-A4, and the partner categories as

[22] We have run analogous specifications using the subsamples we discuss in the previous subsection and find very similar results.

P1-P4. We then examine the ordering of lawyers' earnings across these categories. An ordering corresponding to occupational stratfication would be: A1, A2, A3, A4, P1, P2, P3, P4. An ordering corresponding to self-matching would be: A1, P1, A2, P2, A3, P3, A4, P4.

Our specifications take the form of ordered logits, where:

$$\begin{aligned} P_1 =& 1 - \Lambda(\beta w_i - \alpha_1) \\ P_j =& \Lambda(\beta w_i - \alpha_{j-1}) - \Lambda(\beta w_i - \alpha_j),\ j = 2, ..., N-1 \\ P_N =& \Lambda(\beta w_i - \alpha_N) \end{aligned}$$

P_j is the probability that lawyer i is in position j in the specified ordering. For the occupational stratification ordering, position 1 is A1, position 2 is A2, and so on. These probabilities are a function of lawyer i's earnings w_i, and thresholds α_j. We estimate this model using different orderings, and compare orderings' explanatory power using Vuong's (1989) non-nested hypothesis test.

Our earnings data are at the level of individuals who work at the same office and organizational position; at offices with both partners and associates, there are two observations. We weight each observation by the number of lawyers the observation represents. We let the thresholds α_j vary across fields and counties, allowing them to be linear functions of the share of lawyers in the office who are in each of the 13 fields in our data, and a vector of county fixed effects. This allows relationships between earnings and organizational position to vary across fields and across markets. We impose the constraint $\alpha_j > \alpha_{j-1}$ so the model is well-defined.

Table 5 reports Vuong test statistics when comparing the occupational stratification specification with other specifications. Under the null hypothesis that specifications fit the data equally well, the Vuong test statistic is distributed N(0,1). Like the previous two subsections, in Panel A we use only data from offices with at least one associate. From the log-likelihood values, the occupational stratification specification fits the data better than that in the second row, in which associates at offices with high associate/partner ratios "outrank" partners at offices with low associate/partner ratios. The Vuong test statistic of 8.83 is easily greater than the critical value of 1.96 for a size 0.05 test, indicating that one can reject the null that the specifications fit the data equally well in favor of the alternative that the occupational stratification specification fits better. This test reflects that, controlling for market size and lawyers' fields, associates at offices where the associate/partner ratio is high earn less than partners at offices where this ratio is low. Associates not only tend to earn less than partners in their office, but also than partners more generally.

The final row of the Panel reports test statistics when comparing the occupational stratification specification with a specification that uses the "self-matching" ordering The results indicate that the former fits the data significantly better than the latter, which is not a surprise given the results reported in the previous row.

While some of these patterns may reflect aspects of the labor market not captured by our model,[23] the broad consistency of these earnings patterns with Proposition 1 provides some assurance as we move away from testing the model's hypotheses to estimating its parameters structurally.

Panel B expands the analysis by including partners at offices with no associates – "unleveraged partners" – in the analysis for the first time. Occupational stratification implies that such individuals should rank above all associates but beneath all other partners. We examine this by comparing this to other orderings. The results using all counties indicate that the stratification specification does not fit best; specifications in which unleveraged partners are out ranked by some classes of associates fit significantly better. The rest of the table explores this further by splitting the sample according to whether the employment size of the county is less than 400,000.[24] The results indicate that the occupational stratification ordering fits best for lawyers in the small and medium sized counties (though not statstically significantly better than some of the other orderings), but fits poorly for lawyers in the largest ones. What this reflects is that the distribution of earnings among unleveraged partners in these large counties has a long lower tail. One possible explanation for this is that the small share of lawyers in these counties that work on their own are disproportionately working part-time, but absent data on hours this is only a conjecture. In any case, there is no evidence that the stratification result extends to unleveraged partners in the nation's largest counties.

Earnings Distributions and Local Market Size. Before moving to estimation, we report an additional piece of evidence relevant to assignment patterns: the distribution of lawyers' earnings across differently-sized local markets. The assignment of individuals to markets may reflect the equilibrium assignment of individuals to each other. One possible pattern in the assignment of individuals to markets is a simple one implied by Rosen (1981): in situations where there is limited substitution between the quality and quantity of human capital, "superstar effects" could lead skill and the size of the market in which they work to be positively associated throughout their respective domains.[25] However, if

[23] For example, dynamic aspects may lead associates' equilibrium wages to be lower than they would be in a static equilibrium. If so, this would bias the results toward finding stratification and hence lower the power of this test.

[24] As noted in the table, only about 40 counties were above this threshold as of 1992. Counties near this level include Hillsborough County, FL (Tampa) and Orange County, FL (Orlando).

[25] Rosen writes that an important implication of his analysis is that "it is monetarily advantageous to work in a larger overall market; and it is increasingly advantageous the more talented one is...the best

the equilibrium assignment of individuals to each other involves cross-matching, like in our model, one would not expect such a pattern. Individuals who tend to work in the largest markets would include not only those with the greatest skill, but also individuals in the middle of the skill distribution whose comparative advantage is working under experts. Under cross-matching, skill and market size would not be positively associated throughout their respective domains, even in the presence of "superstar effects." At some point in the skill distribution, as an individual's skill increases, their comparative advantage would change from being a worker supporting a highly-skilled, highly-leveraged manager (who works in a large market) to being a low-leverage manager who works in a small market. When the equilibrium assignment of individuals to each other involves cross-matching, this could lead the relationship between individuals' skill and the size of the market in which they work to be non-monotonic.

Figure 3 depicts how the earnings distribution across lawyers varies with market size. We construct the Figure in the following way. We first compute earnings deciles across our entire sample, and classify lawyers according to the decile in which they fall. We then classify lawyers according to the employment size of the county in which they work. We then construct histograms that characterize the distribution of lawyers across earning deciles, within each of the six market size categories. We show these distributions across earnings deciles rather than earnings because it provides a useful benchmark: if the earnings distribution is the same across markets, then the histograms would depict a uniform distribution within each market size category. Departures from uniform indicate earnings ranges in which lawyers are over- and under-represented within these market size categories. Actual earnings distributions are highly positively skewed; the fact that earnings ranges are much wider in the upper than lower deciles is a manifestation of this skewness.

This Figure shows an interesting pattern. Although higher-earning lawyers tend to work in larger markets, earnings and market size do not appear to be positively associated throughout their domains. Within market size categories, the earnings distributions tend to be bimodal, with each of the modes increasing as market size increases; there is a lower hump that moves from the 2nd to the 8th decile, and an upper hump that moves from the 7th to the 10th decile as one moves from the upper to the lower panels. In Appendix A we discuss and report a regression version of this Figure that controls for lawyers' fields. The evidence is similar, if not stronger.

In closing this section, we note that the cross-matching implied by hierarchical production functions implies other interesting non-monotonicities. For example, it implies that as individuals' talent increases, neither the worker/manager ratio nor the talent level (as proxied by the skill of the most able member) of the group in which they work increase

doctors, lawyers, and professional athletes should be found in the largest cities." (1981:855)

monotonically, even though such production functions imply that worker skill, managerial skill, and the number of workers per manager are strictly complementary. This is because individuals' assignment to positions changes at some point as their skill increases – they change from being a worker to being a manager – and when this happens, their equilibrium assignment to each other changes dramatically. Individuals go from working with people at the top of the skill distribution to working with people at the bottom of the skill distribution.[26] We think these non-monotonicities could lend interesting insights into individuals' career progressions in some contexts, and one could test for them using longitudinal data that follows individuals over time. We also note that these sorts of non-monotonicities follow straightforwardly from the principle of comparative advantage, but are only evident when one combines analysis of the production function with an equilibrium model.

V. STRUCTURAL ESTIMATION

The previous section provided evidence of earnings patterns that are largely consistent with the equilibrium assignment patterns implied by our model. With this in hand, we now lean more heavily on the model, using it as a maintained assumption to estimate the parameters of this function structurally. We will then use these parameter estimates to infer the return to hierarchy: how much does lawyers' ability to organize production hierarchically, as in Figure 1, increase lawyers' productivity and how does it affect the distribution of earnings across lawyers? How much of earnings heterogeneity reflects this versus skill differences?

We start the section with a discussion of the econometric issues involving hedonic estimation of the model developed in Section II. We then specify our empirical model.

V.1. Preliminaries: Estimation Issues and Hedonics

In the context of lawyers, the equilibrium assignment depicted in Section 2 is the result of lawyers choosing with whom to work, given their skill z. The equilibrium earnings of a lawyer who becomes a partner equal:

$$R = \max_{z_w} z_m(n(z_w)+1)^{\theta} - w\left(z_w\right) n\left(z_w\right) \tag{7}$$

where z_m is the partner's skill and z_w is this partner's associates' skill.

Setting aside the endogenous matching between agents and positions, the problem these lawyers face, and the competitive equilibrium that arises, is analogous to that analyzed

[26] From these individuals' perspective, this might involve the end of their apprenticeship; our analysis depicts the conditions in which apprenticeships are an equilibrium outcome.

in the hedonics literature starting with Rosen (1974). Heterogeneous managers choose among workers of different skills in a competitive labor market. The section of the earnings function plotted in Figure 2 that lies below z^*, which corresponds to $w(z)$, is interpretable as a hedonic wage surface that relates the wage associates receive to their ability. As in Rosen (1974), this surface reflects a competitive equilibrium in which heterogeneous suppliers (associates) match with heterogeneous buyers (partners). We use the properties of this equilibrium to estimate θ. Once θ is known, we can then compute what earnings distributions would look like and how much production would be lost, absent hierarchical production.

In Figure 3, we recast this section of Figure 2 in terms of the hedonics literature. Define $\varphi(z; z_m, \theta, R) = z_m(n(z)+1)^\theta - R$ as a partner's bid curve; this is a partner with skill z_m's willingness to pay for an associate with skill z, at a given level of partner earnings R. This bid curve is analogous to a consumer's expenditure function. In the competitive equilibrium depicted in Figure 2, the surface $w(z)$ is the locus of tangency points between heterogeneous partners' bid curves and heterogeneous associates' offer curves, which are analogously defined. In equilibrium, $\varphi'(z; z_m, \theta, R) = w'(z)$; partners' marginal willingness to pay for skill is equal to the marginal price of skill. We depict two bid curves for lower and higher-skilled managers in Figure 3. In equilibrium, both of these curves are tangent to the wage-skill surface.

The usual approach in the hedonic literature is to first estimate the function $w(z)$ and then use the first order conditions $\varphi'(z; z_m, \theta, R) = w'(z)$ to estimate the parameters of the production function. In our context, this would mean finding a proxy for skill (e.g. education or experience) and then proceed with the estimation as if it was z. We cannot do this in our context because we observe none of the standard proxies of human capital such as education or experience. But even if we did observe such variables, this method would suffer from a well-known problem. Although we would not observe skill perfectly, market participants would. This would imply that agents sort on variables we would not directly observe. In terms of the econometrics, this would lead to biased estimates of the parameters of interest. Measurement error in the key independent variable in our first-stage regression, skill, would lead us to incorrectly estimate the marginal price of skill, and this in turn would contaminate our estimate of θ.

However, a key property of the labor market equilibrium in our model not only makes it feasible to estimate such a model using our data, but also substantially limits the problem of unobserved skill. This key property is that $n(z)$ is invertible: n, the number of associates per partner, increases monotonically with associate skill; thus, n is a sufficient statistic for associate skill. In terms of the hedonics, this allows us to reformulate the problem in terms of the supply and demand for leverage rather than the supply and demand for skill.

We depict this in Figure 4. The vertical axis, $P(n) = w(n)n$, is the price of leverage faced by a manager. A manager's bid curve $\varphi(n; z_m, \theta, R)$ is now stated in terms of his willingness to pay for leverage rather than skill. Writing the bid curves in terms of leverage rather than skill affords two advantages. First, n is a variable we observe directly in the data – it is the number of associates per partner. Second, as long as there continues to be unidimensional sorting on skill, there is now no problem associated with sorting on unobservables – we observe a sufficient statistic, n, which summarizes all relevant aspects of skill, including both that which is captured in usual proxies and that which is not.[27] The invertibility property means that the quantities n and $w\prime(n)$ summarize a lot of information in equilibrium. n is not only the number of associates per partner, but is also an error-free index of associates' skill. Likewise $w\prime(n)$ is not only the marginal price of leverage, but is also a monotonic transformation of the marginal price of skill.[28]

Formally, the transformation allows us to write the first order condition in Section II, equation (2), for the worker's wages,

$$z_m f'(n(z_w)) = w(z_w) + \frac{w'(z_w)}{n'(z_w)} n(z_w), \tag{8}$$

in terms of n as:

$$z_m f'(n) = w(n) + w'(n)\, n \tag{9}$$

Recasting the problem in this way is particularly beneficial because our data contain more information than the usual data used in a hedonic estimation exercise, and this substantially mitigates the usual identification problems that arise in the hedonics literature. (See the discussion in Epple (1987).) Figure 5 depicts the marginal price schedule and the marginal benefit curves (the first derivative of the bid curves) that correspond to what we depict in Figure 4. The standard identification problem in hedonic studies is that one observes a locus of equilibrium marginal prices and quantities, but this alone is not sufficient to trace out the slopes of the marginal benefit curves – and thus estimate the parameters of the production function. One generally needs supply-side shifters of $P\prime(n)$ to trace these out, but it is generally very hard to find shifters of $P\prime(n)$ that, in equilibrium, are independent of the demand schedules $\varphi'(n; z_m, \theta, R)$. Our data, however,

[27] Our exploitation of the invertibility of $w(n)$ is similar to Olley and Pakes' (1995) use of the invertibility of the investment function in productivity estimation. In both cases, the idea is that if theory implies that an agent's decision variable increases monotonically with an unobserved variable, an arbitrary increasing function of the decision variable substitutes for the unobserved variable.

[28] Note that this is true under a wide variety of additional assumptions that might change the wage schedule or the match between associates and partners, but not the invertibility of $w(n)$. For example, if more skilled associates are willing to work for less for more skilled partners, this would change the shape of the equilibrium wage-leverage surface but n would continue to be a sufficient statistic for associate skill, since this would merely reinforce positive sorting.

contain information on R^*, partners' equilibrium earnings. This additional information allows us to determine whether these marginal benefit curves are steep or flat, because it allows us to compare the marginal to the average benefits of leverage; if the marginal and average benefits are equal, this implies that the marginal benefit curve is flat. We can thus estimate the parameters of interest without the supply-side shifters usually needed in hedonic estimation.

We next specify the empirical model. The fact that n is a sufficient statistic for associate skill is valuable for the reasons we describe above, but presents a small challenge in specification, since it eliminates an important reason why the data will not fit the model exactly. The theoretical model in Section II implies that both partners' earnings and associates' earnings are deterministic functions of n, but we must account for the fact that this will not be true in the data. The stochastic elements of our model aim to relax the deterministic relationships between earnings and skill in the theoretical model, while maintaining the deterministic relationships between skill and n that give rise to invertibility. This requires assumptions on these stochastic elements that preserve *unidimensional sorting*. These assumptions restrict the individuals in our model to be unidimensional, fully characterized by a single parameter z that indexes their skill.

V.2. Empirical Model

The timing of the model follows. First, partners choose the number of associates they work with to maximize expected earnings, subject to the wage-leverage surface they face. Partners' uncertainty at this point in time is over the demand they will receive for their services. Second, uncertainty, and partners' earnings, are realized.[29]

The Wage-Leverage Surface..—

We assume throughout that, within a particular labor market, $w_i(n) = w(n) + \xi_i$: the wage-leverage (or, equivalently, wage-skill) surface faced by partners at office i is the market wage-leverage surface plus some random variable ξ_i, which we think of as a compensating differential that accounts for differences in working conditions, and any other factor that would shift associates' offer curves for working at office i. This random element is assumed to be attached to the office and, conditional on n, affect the wage level but not the marginal price of leverage. One implication of this is that ξ_i shifts all associates' offer curves up or down equally.[30]

We specify the market wage-leverage surface as a polynomial in n, controls for the field

[29] This timing assumption implies that human capital here, much like physical capital in much of the productivity literature, is fixed in the short run – firms cannot adjust this in response to demand shocks. We view this as reasonable in this context, in which there is a distinct hiring season for associates.

[30] This implies that the slope of the wage-leverage surface for a given n is the same for all partners; we discuss the implications of this below in the next subsection.

composition of lawyers in office i, and a full set of county fixed effects. In practice, we found little additional explanatory power when adding terms in n beyond quadratic, or allowing $w(n)$ to differ across fields. Suppressing the controls, the first stage regression assumes:

$$w(n) = \beta_0 + \beta_1 n + \beta_2 n^2 \tag{10}$$

Thus the marginal wage is:

$$w_i'(n) = \beta_1 + 2\beta_2 n$$

We therefore estimate the wage-leverage surface by regressing average associate earnings at office i on a quadratic in the associate/partner ratio at the office and a set of the above controls. The coefficients on n allow us to construct an estimate of the marginal price of leverage, $w_i'(n)$, for partners at each office.

Partners' Bid Curves.—

Let R_i index a partner's earnings in office i, w_i represent an associate's pay in office i, and n_i equal associates per partner, or leverage, at office i. We assume that partners in the same office are similarly-skilled. A partner's earnings in office i are then given by:

$$\begin{aligned} R_i(z_{mi}, n) &= \widehat{z}_{mi}(n_i + 1)^\theta - w_i(n_i)n_i - c_i(n_i) \\ &= \varepsilon_i z_{mi}(n_i + 1)^\theta - w_i(n_i)n_i - c_i(n_i) \end{aligned} \tag{11}$$

where $\widehat{z}_{mi} = \varepsilon_i z_{mi}$, z_{mi} is the partner's skill, and $(n_i + 1)^\theta$ is the effective team time.

This function has two differences with the objective function in Section II; first we account for the fact that production involves factors other than lawyers with the term $c_i(n_i)$. As discussed in Section III, these factors include nonlawyers and overhead. Second, we introduce ε_i, an office level i.i.d term with $E(\varepsilon_i) = 1$, to account for the fact that an office's revenues is not a deterministic function of z_{mi} and n_i. We envision this as an office-level demand shock that reflects that clients' demands for legal services are stochastic, and this leads to uncertainty about the number of hours lawyers bill during the year, but there may be other interpretations of this term. In any case, these shocks are realized only after organizational decisions are made, and thus affects partners' earnings but not the organizational equilibrium.

Recall that θ is the elasticity of effective time with respect to the size of the team. If $\theta = 1$, there are constant returns, and adding associates does not diminish the effective time per lawyer. If $\theta < 1$, adding associates does so. θ is a measure of hierarchies' coordination costs, and is the parameter of interest in estimation; if this is known, then we can use partners' earnings equation (11) to back out $R_i(z_{mi}, 0)$ and $\widehat{z}_{mi}$, what the earnings

and revenues, respectively, of a partner at office i would be, if unleveraged. This, in turn, allows us to infer how much hierarchical production affects productivity – the gains to vertical specialization – and the distribution of earnings across lawyers.

Partner i's problem is:

$$\max E(R_i) = z_{mi}(n_i + 1)^{\theta} - w_i(n_i)n_i - c_i(n_i) \tag{12}$$

The first-order condition to this problem is:

$$z_{mi}\theta(n_i + 1)^{\theta-1} = w_i'(n_i)n_i + w_i + c_i'(n_i) \tag{13}$$

Solving for z_{mi} and substituting the expression into (11) yields:

$$\frac{[R_i + w_i n_i + c_i(n_i)]}{(n_i + 1)}\theta\frac{1}{\varepsilon_i} = w_i'(n_i)n_i + w_i + c_i'(n_i)$$

The left side of this equation is the marginal benefits of leverage; the right is the marginal costs. These are equal at the point where partner i's bid curve is tangent to the wage-leverage surface he or she faces. To see where identification of θ comes from, it is useful to write this equilibrium relationship as:

$$\theta = \frac{MC_i}{AR_i}\varepsilon_i \tag{14}$$

where AR_i and MC_i are the average revenues per lawyer and the marginal cost of leverage at office i. Taking logs and rearranging, we obtain:

$$\ln AR_i - \ln MC_i = -\ln\theta + \ln\varepsilon_i \tag{15}$$

$$\ln AR_i - \ln[w_i'(n_i)n_i + w_i + c_i'(n_i)] = -\ln\theta + \ln\varepsilon_i \tag{16}$$

We use this equation to estimate θ. The first term of our dependent variable is just revenues per lawyer, which we observe in our data. The second term contains two variables we do not observe directly $w_i'(n_i)$ and $c_i'(n_i)$. The first of these terms is the marginal price of leverage; as noted above, we use our coefficient estimates of the wage-leverage surface regression described above to construct an estimate of $w_i'(n_i)$ for every office.

We obtain an estimate of $c_i'(n_i)$ from data on nonlawyer payroll and our estimate of the overhead equation in Section 3. Letting p_i equal the number of partners at office i, we

specify:

$$\begin{aligned} c_i(n_i) &= (x_i l_i + oh_i)/p_i \\ &= x_i(1 + n_i) + oh_i/p_i \end{aligned}$$

The first term is nonlawyer payroll per partner. We assume that x_i, nonlawyer pay per lawyer, is constant. As above, oh_i represents overhead expenses. Therefore,

$$c_i'(n_i) = x_i + oh_i\prime/p_i$$

$c_i'(n_i)$ has two parts. One is that hiring an associate requires hiring support staff as well; we assume that it requires hiring a proportionate amount of support staff, which implies an increase in nonlawyer pay of x_i. The other part is the increase in overhead. Following Section 3, the increase in overhead includes increases in fringe benefits – 15% of the additional lawyer and nonlawyer payroll associated with hiring an associate. It also includes the increase in space, computer equipment, etc. that goes along with increasing the employment size of the office. We use the coefficients on employment in the overhead regression to estimate this for every office, remembering that the employment increase that comes with hiring an additional associate includes a proportionate amount of additional support staff as well.

Estimating θ.—

Following equation (16), we derive a preliminary estimate of θ by simply regressing the difference between the log of revenues per lawyer and the log of our estimate of the marginal cost of leverage, described above, on the field shares of lawyers in each office.[31] Including the field shares on the right side allows the coordination costs of hierarchy to vary across different fields of the law. The estimate is preliminary because it is biased, even asymptotically: applying Jensen's inequality, $E(\varepsilon_i) = 1$ implies $E(\ln \varepsilon_i) \neq 0$. If ε_i is distributed log-normally with parameters μ and σ^2, the assumption $E(\varepsilon_i) = 1$ implies $\ln \varepsilon_i$ is distributed $N(-\sigma^2/2, \sigma^2)$, and thus an OLS estimate of $-\ln \theta$ is biased by $-\sigma^2/2$. This bias is very small, however. Following the discussion in Goldberger (1968) and van Garderen (2001), we have estimated this equation using maximum likelihood under the assumption of log-normality to obtain consistent estimates of θ. The estimates of θ are almost identical to those we report; they are lower by about 0.02 relative to a mean value of about 0.70. The bias in the estimates that we report here is very small relative to

[31] As we discuss below, one has to adjust revenues to take into account that overhead increases with revenues. Our dependent variable uses revenues net of the associated overhead rather than gross revenues.

the estimates themselves, and we have found that accounting for it implies little change in the results from our counterfactual exercises. Future versions of this paper will report results derived from consistent estimates of θ.[32]

Dynamics and the Returns to Hierarchy.—

An important assumption in the model in Section 2 is that agents maximize their current-period earnings, given their skill. This assumption underlies our empirical specification as well. This specification of agents' objective rules out dynamic aspects of the labor market, including that individuals value working with higher-skilled agents because it provides them future benefits, for example in the form of better training. Such dynamic aspects are very realistic in our context, but are difficult to incorporate directly in our equilibrium model. We can, however, analyze how our estimates would change if our assumption that agents maximize current period earnings were replaced by one in which agents working as associates value working with higher-skilled partners.

If we have misspecified agents' tastes in this way, this will tend to bias *downward* our estimates of the return to hierarchy, and our estimates can therefore be thought of as a lower bound. In terms of the hedonics, the "current period earnings" objective function ensures that there exists a single market wage-leverage surface (represented by equation (10) off of which all agents optimize, and consequently that the marginal price schedule $w\prime(n)$ is the same for all managers, irrespective of their skill. If instead associates are willing to work for less for higher-skilled partners, then our estimate of $w\prime(n)$ will understate the marginal price of leverage faced by each individual manager; a manager of a given skill will find it more expensive to increase leverage because it will cost him more at the margin to outbid a slightly-higher-skilled manager for a slightly-higher-skilled associate. Our estimates of the marginal cost of leverage will then be too low. If so, this will lead our estimate of θ, which is identified by the ratio of marginal cost and average revenues, also to be biased downward. This, in turn, will lead us to underestimate the returns to leverage, since a too-low θ implies that we overstate the extent to which the potential returns to leverage are eaten up by coordination costs.

We discuss the quantitative impact this has on our estimates below. To preview, our investigations lead us to believe that leads to only a small bias on our estimates of the returns to hierarchy.

V.3 Estimation

We begin by estimating the $w(n)$ surface, using offices with at least one associate. Results are reported in Table 6. Our estimates imply that $w'(n)$ is positive, and increases

[32] We do not report them here because they have not been released by the Bureau of the Census.

slightly with n although the latter is not statistically significant. The wage-leverage surface is convex, but only slightly so. Ignoring the very small second-order term, the estimates imply that increasing the associate/partner ratio by one is associated with a $3,254 increase in average associate pay.

In Table 7, we report the mean and various quantiles of marginal cost implied by this estimate. We also report analogous figures for the various components of marginal cost of leverage. On average, the marginal cost of hiring an additional associate is $145,000, though there is wide dispersion across offices. Our estimates imply that, on average, the pay the additional associate receives is only 43% of the marginal cost of adding an associate. A significant share of the marginal cost is made up of the additional associate's support staff (on average, 27%), the cost of the associate and staff's fringe benefits (10%), and the cost of the additional overhead (17%). Although the wage-leverage surface is convex, the fact that it is not far from linear and leverage levels tend to be low implies that the marginal price of leverage, $w'(n)n$, makes up a very small part of the marginal cost of leverage throughout our sample.

We also report various quantiles of average revenues per lawyer across offices in our sample. Comparing this to our estimated marginal cost exercise foreshadows our estimate of θ below, which is identified by the ratio of estimated marginal cost and average revenues. As one would expect, the distribution of average revenues per lawyer is highly skewed across offices. All of the quantiles are 55-65% higher than our estimates of marginal cost, suggesting that $f(n, \theta)$ is not constant returns: θ will be less than one.[33]

The right side of Table 6 reports our estimates of the equation (15). We allow $-\ln\theta$ to be a linear function of the field shares of the lawyers in the office: if the ratio between (estimated) marginal costs and average revenues varies systematically with lawyers' field, θ will differ across offices. The omitted field in this specification is "general practice," lawyers who work in more than one of the Census-defined fields. The estimate on the constant implies a value of θ of 0.75: for an office consisting only of general practitioners (the omitted category), moving from $n = 0$ to $n = 1$ increases the time to which the partner's knowledge is apply by ($2^{0.75}$-1), or 68%. In other words, hiring your first associate is like adding two-thirds of an extra body to your group in terms of how it affects the group's time in production. Relative to a situation where two lawyers work on their own, hierarchical production decreases the time these lawyers spend in production by 16%. Our estimate of θ is lowest – about 0.52 – for an office with all negligence-plaintiff lawyers, and highest – about 0.81 – for an office with only specialists in probate law, suggesting that the coordination costs associated with hierarchies tend to be high for

[33] θ is not simply the ratio between marginal costs and average revenues reported here: recall that we estimate that each dollar of revenues is associated with about 15 cents in overhead (more for business-non-litigation offices, less for individual non-litigation offices). A rough estimate of θ is therefore the product of this ratio and $(1 - 0.15)$.

the former and low for the latter.

Our estimates reflect that average revenues per lawyer are generally high relative to the estimated marginal cost of leverage, which indicates that there must be diminishing returns on the revenue side associated with hiring additional associates. Our theoretical model attributes these diminishing returns to coordination costs which reduce the time lawyers spend in directly-productive activities; however, they may reflect other unmodeled factors as well, such as decreasing returns from the time partners spend in business-generating activities. What is important for our analysis is quantifying θ rather than distinguishing between the various possible sources of diminishing returns, as this set of parameters serves as an input to the counterfactuals that we discuss next.

V.4 The Returns to Vertical Specialization

We first use our estimates to quantify the returns to hierarchical production. Our counterfactual is this. Suppose the match between clients and offices stayed the same. However, suppose partners and associates were not allowed to divide work according to their talents with the associates handling the easier tasks; suppose instead that division of labor within the office was arbitrary. Production would be lost because the match between lawyers and problems would be less efficient: associates would be given tasks they cannot handle but their partner could. What would be the value of the lost production?

Consider this calculation for a hierarchy with one partner and n associates, referring again to Figure 1. This office's revenues, which are observed in the data, are $TR_i = \widehat{z}_{mi}(1+n_i)^{\theta}$. Absent the division labor, the office's revenues would equal $\widehat{z}_{mi} + n_i\widehat{z}_{wi}$, where $\widehat{z}_{wi} = \varepsilon_i z_{wi}$. In expectation this quantity is less than $\widehat{z}_{mi} + n_i w_i$, because $w_i > z_{wi}$: from Figure 2, in expectation, associates earn more as associates than they would if they worked on their own. A lower bound for the increase in the value of production afforded by vertical specialization at office i, averaged across the lawyers in the office, is therefore $\widehat{z}_{mi}((1+n_i)^{\theta} - 1) - n_i w_i)/(n_i + 1)$. We calculate this quantity for every office in our sample, exploiting the fact that $\widehat{z}_{mi} = TR_i/(1+n_i)^{\theta}$ and using our estimate of θ from the production function estimation. We also calculate this quantity under the assumption that $\theta = 1$, which corresponds to constant returns to leverage. We therefore compare actual revenues per lawyer against two benchmarks. One is revenues per lawyer if vertical specialization were prohibited within offices: this provides evidence on the *achieved* returns from vertical specialization The other is revenues per lawyer if vertical specialization were allowed and there were no coordination costs. This provides evidence on the *potential* returns from vertical specialization (but which coordination costs limit).

Table 8 reports the results of this analysis. We include offices with and without associates in the analysis, though of course the returns to hierarchy are zero for offices

without associates. Average revenues per lawyer in our sample equals $227,000. We estimate that they would be $171,000 if the division of labor were arbitrary. From Table 8, vertical specialization associated with hierarchies increases productivity in the U.S. legal services industry by at least 33%. This ranges considerably across offices. We calculated the distribution of the percentage increase across offices (weighted by the number of lawyers). The 90th percentile is 63%; the median is 29%. The final column in Table 8 reports analogous estimates for the $\theta = 1$ case – no coordination costs associated with hierarchical production. These estimates imply that revenues per lawyer, holding constant the matching between lawyers and between clients and firms, would increase to about $278,000, implying that coordination costs prevent lawyers from achieving about 1/2 of the potential gains from vertical specialization.

Our estimates thus imply that organizing production hierarchically increases productivity in legal services substantially – by at least 33%. The overall returns to hierarchy appear to be substantial in this human-capital-intensive industry.

We have examined the robustness of this result to the possibility that the labor market equilibrium might be affected by dynamics not present in this model. As discussed above, dynamic elements that lead associates to be willing to work for less under more skilled partners lead us to understate the marginal price of leverage, and thus the marginal cost of leverage, faced by any particular partner. We report in Table 7 that the marginal price of leverage is very low for nearly all partners, only $4,000 on average. We explored the robustness of our estimates by assuming that the marginal price of leverage is two, four, and ten times as much as our estimates imply. Our estimates of the returns to leverage do increase – as discussed above, our previous results are a lower bound – but not by much. Assuming that the marginal price of leverage is ten times what we estimate – $40,000 rather than $4,000 on average, our estimates imply that hierarchical organization increases productivity by 40% rather than 33%.

The reason such large differences in marginal price of leverage have small effects on our estimates is straightforward. Increasing the marginal price of leverage, even by a large amount, implies a much smaller percentage change in the marginal cost of leverage and a moderate increase in our estimate of θ. Even after the change, the estimate implies significant decreasing returns to leverage for most offices. Furthermore, recall that n is small at most offices. A moderate increase in estimated the returns to leverage in an industry where most entites are low-leverage to begin with implies a very small change in the estimates of the returns to leverage that are in fact achieved.[34]

[34]It has a simillarly small effect on our estimate of how hierarchy affects earnings distributions, a topic we discuss in the next subsection.

V.5 Hierarchy and Earnings Distributions

We next use our estimates of θ and equation (11) to derive estimates of $R_i(z_{mi}, 0) = \widehat{z}_{mi} - c_i(0)$ at offices with associates: this is what partners at these offices would earn, absent hierarchical production. This differs from $\widehat{z}_{mi}$ because it accounts for the costs of operating a zero-associate office. We estimate $\widehat{z}_{mi}$ the same way we do in the previous subsection. From section V.2, $c_i(0) = x_i + oh_i/p_i$, the sum of nonlawyer pay per lawyer and overhead per partner. We estimate oh_i for each office using the coefficients in the overhead equation. We compute quantiles of the distribution of these variables across the leveraged partners in our sample, and compare them to quantiles associated with our observations of partner pay.

Figure 7 reports twenty quantiles of partner pay and $R_i(z_{mi}, 0)$, using only partners in offices with at least one associate; the difference between the two curves reflects the effect of leverage on the earnings of individuals who are, in fact, leveraged. Median earnings among lawyers in this group are \$168,000. Our estimates imply that, absent hierarchical production, the median instead would be \$135,000, about 20% lower. $R_i(z_{mi}, 0)$ is about 20% lower than partner pay thoughout most of the Figure, one exception being that it is 26% lower at the 95th percentile. Considering only leveraged partners, lawyers' ability to leverage their knowledge through working with associates does increase earnings inequality in absolute terms – the difference between the 95th percentile and 50th percentile earnings increases from \$227,000 to \$327,000. But because it largely shifts the earnings distribution among these lawyers proportionately, the ratio between the 95th percentile and 50th percentile earnings increases by only a small amount, from 2.7 to 2.9. This relatively small proportionate shift reflects our finding of fairly strong diminishing returns to leverage: even very highly skilled lawyers find it difficult to leverage their knowledge via associates because it costs them in terms of time in production. As a consequence, hierarchical production does not amplify skill-based earnings inequality nearly as much as it would, if leverage did not involve a sacrifice of time in production.

Figure 8 extends the analysis to all lawyers, not just leveraged partners, as we include unleveraged partners and associates in the construction of our earnings distributions. This Figure depicts the distribution of lawyer pay and "estimated pay absent hierarchies." "Estimated pay absent hierarchies" equals $R_i(z_{mi}, 0)$ for leveraged partners, as before. It equals actual pay for unleveraged partners – we observe what these individuals did earn when unleveraged. For associates, we also assume that "estimated pay absent hierarchies" equals their actual pay. This is a biased estimate for the reason described above: these individuals earn more as associates than they would absent hierarchies. Thus, since associates tend to be below the median earnings, quantiles of "estimated pay absent hierarchies" below the median will tend to be upward-biased. This will have little effect

on our analysis, however, because we are most interested in upper tail of this distribution and how it compares to that of the partner pay distribution.

The Figure indicates that, when looking across all lawyers, hierarchical production tends to make earnings distributions more skewed. The difference between this and the previous Figure reflects the simple fact that over half of lawyers are unleveraged – they are either unleveraged partners or associates – and the vast majority of these lawyers are below the 70th percentile in both of these distributions. Our estimates indicate that hierarchical production increases median earnings from $72,000 to $77,000, by 7%, but increases 95th percentile earnings by 55%. The ratio between the 95th percentile and median earnings increases from 3.1 to 4.5, indicating that hierarchical production does indeed amplify skill-based earnings inequality when one looks at all lawyers, not just leveraged partners.

Finally, Figure 9 depicts these distributions separately for lawyers in the the three classes of offices we defined earlier: "business, non-litigation," "individual, non-litigation," and "litigation" offices. The Figure indicate that hierarchical production has a similiar effect on the earnings distribution among lawyers in "business non-litigation" and "litigation" offices, increasing the ratio between the 95th percentile and median earnings from about 3.0 to about 4.2. The estimates suggest that skill-based earnings inequality is greater among "business-non-litigation" lawyers,[35] but that hierarchical production amplifies this inequality similarly. Lawyers in "individual, non-litigation" offices look much different. Absent hierarchies there is much more inequality, although lawyers in these offices tend to earn much less than lawyers in the other classes of offices. In part due to a long lower tail, the ratio between the 95th percentile and the median is 5.0. Hierarchical production increases this Figure only marginally, to 5.6. The returns to hierarchy are low in this segment of the industry, and this is reflected in low levels of leverage, even among the relatively small share of lawyers in this segment who are leveraged partners, and in the fact that average revenues per lawyer among offices with associates in this segment tend to be low. The latter implies a low return to hierarchy, even when the marginal cost of leverage is low, because it implies that the partner's skill cannot be high.

Combined, these results provide evidence on how organizational structure – here, hierarchical production – can have a substantial effect on productivity through gains to vertical specialization and on the earnings distribution, especially in segments where individuals are confronted with problems with varying degrees of complexity. We estimate that, holding the match between problems and offices constant, lawyers' ability to organize work hierarchically increases revenue-weighted output by at least one-third. Outside

[35] There is an important caveat to this statement: we are not reporting earnings above the 95th percentile, to avoid disclosure problems associated with Census microdata. In any given year, the highest-earning lawyers in the U.S. tend to be specialists in litigation who receive a share of the proceeds from a large case.

of the sector that deals within individuals' "non-litigation-related" demands, it also significantly amplifies earnings inequality. However, regarding the latter, it is important to place this amplification result in context. Lawyers *can* exploit increasing returns associated with their knowledge, but our estimates imply that it is very difficult for them to do so – there are sharply diminishing returns to leverage. This difficulty may be typical of human-capital-intensive production where leverage implies applying one person's knowledge to others' time. When the scaling up is over physical capital – as in Lucas (1978) – the organizational problems associated with leverage may be much less severe, and one might expect to see much more earnings inequality.

VI. CONCLUSION

Earnings and assignments contain important information about the nature of production and the value of organization that has been empirically ignored by organizational economists until now. Using this information requires embedding organizations in an equilibrium model. We have taken a first step towards exploiting this information by embedding an organizational model in a labor market equilibrium with heterogeneous individuals. This step has costs, as it leads us to abstract from many details of internal labor markets that are the focus of much of the organizational economics literature, in particular, how organizations respond to the problem of providing individuals incentives. But it also generates enormous benefits, in allowing us to exploit previously underexploited information to quantify an effect that organization has on productivity and earnings distributions.

Specifically, we study how much hierarchical production increases lawyers' productivity and amplifies skill-based earnings inequality. We have done this in two stages. First, we have explored the empirical implications of an equilibrium model of a hierarchy in a knowledge economy developed by Garicano and Rossi-Hansberg (2006). The model embeds a production function characterized by limited substitution between the quality and quantity of individuals' human capital, complementarity between individuals' skills, and increasing returns to knowledge in an economy with heterogeneous agents. We show that the model captures reasonably well some important empirical regularities concerning lawyers' earnings and organization, such as positive assortative matching, scale effects associated with managerial skill, and stratification by skill. Second, we then take this model as a maintained assumption and estimate its parameters in order to construct counterfactual productivity and earnings distributions – what lawyers would produce and earn if it were not possible for highly-skilled lawyers to leverage their talent by working with associates. We conclude that hierarchies expand substantially the productivity of lawyers: they increase aggregate output by at least one-third, relative to non-hierarchical

production in which there is no vertical specialization within offices. We also find that hierarchies expand substantially earnings inequality, increasing the ratio between the 95th percentile and median earnings among lawyers from 3.1 to 4.5, mostly by increasing substantially earnings of the very highest percentile lawyers in business and litigation-related segments, and leaving relatively unaffected the earnings of the less leveraged lawyers.

Reflecting on our results, we conjecture that while hierarchies contribute substantially to productivity and earnings inequality in our context, their effect on productivity and especially earnings might be far smaller than in other contexts. In industries where production is more physical-capital intensive, top-level managers sometimes earn multiples in the hundreds of times of what their subordinates earn, and they control enormous organizations (see Gabaix and Landier, 2006). We speculate that the complexity and customization of problem-solving in law firms limits the ability of agents to leverage their human capital: production requires some agent to spend time on each problem, even if they have solved a similar problem in the past. If this speculation is correct, then reductions in communication costs that expand the ability of managers to leverage their talent are likely to have a smaller impact on productivity and earnings inequality in professional services than in other sectors where problems are more routinzed and where, as a result, knowledge can be exploited in a less time-intensive manner. Future work is necessary in order to uncover systematic differences in the return to knowledge across sectors and to link such differences to the characteristics of the knoweldge involved. Time and knowledge are both scarce inputs, and exploiting increasing returns associated with knowledge depends critically on how much time must be expended in doing so.

REFERENCES

Becker, Gary. *Treatise on the Family.* Cambridge: Harvard University Press, 1981.

Bureau of the Census, *1992 Census of Service Industries, Subject Series: Capital Expenditures, Depreciable Assets, and Operating Expenses.* Government Printing Office: Washington, 1996.

Calvo, Guillermo A. and Stanislaw Weillisz. "Supervision, Loss of Control, and the Optimum Size of the Firm." *Journal of Political Economy*, Vol. 86, No. 5 (October 1978): 943-952.

Calvo, Guillermo A. and Stanislaw Weillisz "Hierarchy, Ability, and Income Distribution." *Journal of Political Economy*, Vol. 87, No. 5 (October 1979): 991-1010.

Garicano, Luis. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*, Vol. 108, No. 5 (October 2000): 874-904.

Garicano, Luis and Thomas N. Hubbard. "Specialization, Firms, and Markets: The Division of Labor Within and Between Law Firms." NBER Working Paper 9719, May 2003.

Garicano, Luis and Thomas N. Hubbard. "Hierarchical Sorting and Learning Costs: Theory and Evidence from the Law," *Journal of Economic Behavior and Organization*, forthcoming.

Garicano, Luis and Esteban Rossi-Hansberg. "Organization and Inequality in a Knowledge Economy." Forthcoming, *Quarterly Journal of Economics*.

Goldberger, Arthur S. "The Interpretation and Estimation of Cobb-Douglas Functions." *Econometrica*, Vol. 36, No. 3-4 (July 1968): 464-472.

Grossman, Gene M. and Giovanni Maggi. "Diversity and Trade." *American Economic Review*, Vol. 90, No. 5 (December 2000): 1255-1275.

Kremer, Michael. "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics*, Vol. 108, No. 3 (August 1993): 551-575.

Kremer, Michael and Eric Maskin. "Wage Inequality and Segregation by Skill." NBER Working Paper No. 5718, August 1996.

Legros, Patrick and Andrew Newman. "Assortative Matching in a Non-Transferable World." CEPR Discussion Paper No. 3469, July 2002.

Lucas, Robert E., Jr. "On the Size Distribution of Business Firms." *Bell Journal of Economics*, Vol. 9, No. 2 (Autumn 1978): 508-523.

Meyer, Thomas A. "The Distribution of Ability and Earnings." *The Review of Economics and Statistics, Vol.42, N. 2,* (May 1960): 189-195.

Olley, G. Steven and Ariel Pakes. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica,* Vol.64, No. 6, (November 1996): 1263-1297.

Qian, Yingyi. "Incentives and Loss of Control In an Optimal Hierarchy." *Review of Economic Studies*, Vol. 61, No. 3 (July 1994): 527-544.

Radner, Roy and Timothy Van Zandt. "Information Processing in Firms and Returns to Scale." *Annales d'Economie et de Statistique*, Vol. 25-26 (January-June 1992): 265-298.

Romer, Paul M. "Increasing Returns and Long-Run Growth." *Journal of Political Economy*, Vol. 94, No. 5 (October 1986): 1002-1037.

Romer, Paul M. "Endogenous Technological Change." *Journal of Political Economy*, Vol. 98, No. 5 (October 1990): S71-S102.

Rosen, Sherwin. "The Economics of Superstars." *American Economic Review*, Vol. 71, No. 5 (December 1981): 845-858.

Rosen, Sherwin. "Authority and the Distribution of Earnings." *Bell Journal of Economics*, Vol. 13, No. 2 (Autumn 1982): 311-323.

Sah, Raaj and Joseph Stiglitz, "The Architecture of Economic Systems: Hierarchies and Polyarchies," *American Economic Review*, Vol. 76, No. 4 (September 1986): 716-727.

Sattinger, Michael. "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature*, Vol. 31, No. 2 (June 1993):831-880.

Simon, Herbert A. "The Compensation of Executives." *Sociometry*, Vol. 20, No. 1 (March 1957), 32-35.

van Garderen, Kees Jan. "Optimal Prediction in Loglinear Models." *Econometrica*, Vol. 104, No. 1 (August 2001): 119-140.

Vuong, Quang H. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica*, Vol. 57, No. 2 (March 1989): 307-333.

Waldman, Michael. "Worker Allocation, Hierarchies, and the Wage Distribution." *Review of Economic Studies*, Vol. 51, No. 1 (January 1984): 95-109.

APPENDIX A. REGRESSION VERSION OF FIGURE 6

Table A1 depicts a regression version of this Figure and tests whether the relationships depicted in the Figure are statistically significant. The Table reports results from six regressions. These regressions take the form:

$$y_i = \alpha + \delta_2 D_{i2} + ... + \delta_{10} D_{i10} + \gamma Z_i + \varepsilon_i$$

In the first column, y_i is a dummy variable that equals one if lawyer i works in a county with fewer than 20,000 employees and zero otherwise, D_{ij} is a dummy variable that equals one if lawyer i's earnings are at least decile j, and Z_i is a vector including the share of lawyers in the office who are in each of the 13 fields in our data.[36] The other columns contain analogous specifications using the dummy variables that equal one if lawyer i works in each of the five other market size categories we construct. The sum of the coefficients in the rows equals zero by construction, since the estimates in any one of the rows are implied by the other five. Like in our analysis of stratification, our observations are at the office*organizational position level, and all specifications weight observations using the product of the number of lawyers the observation represents and the Census sampling weight associated with the office. The variables of interest in these specifications are the δ_i's, which indicate whether the share of lawyers in decile i is greater or less than that in decile $i-1$.

These regressions indicate that the patterns depicted in Figure 3 are statistically significant for the most part, and are robust to controlling for systematic differences in lawyers' earnings across fields. For example, the coefficients in the first column of Table A1 indicate a significant decrease, then a significant increase as one moves down the table. Similar statistically significant changes in sign appear in the other columns as well.

We think these patterns are interesting, though they are admittedly not dispositive. It would be far better to conduct this analysis with individual-level earnings data. Some of the clustering of earnings may be due to the fact that our office-level data forces us to ignore heterogeneity in earnings among associates and among partners who work in the same office. This would be a particular problem in situations where much of the earnings heterogeneity across associates and across partners within local markets is within rather than across firms. We have investigated this by conducting a similar analysis using lawyers data from the PUMS database described above. The problem with using the PUMS data for this exercise is that it is top-coded above $99,000 1991 dollars, and therefore allows us to construct earnings distributions only for roughly the bottom two-thirds of the distribution. We analyzed these data, and found a similar pattern to that

[36] Including the latter controls for cross-field differences in lawyers' earnings, but the patterns in the coefficients change little when excluding this vector.

in our data: the within-market-size earning distributions exhibit a mode that increases with market size, similar to the lower mode in Figure 3 though less pronounced. The fact that it is less pronounced might reflect the difference between using individual- and office*organizational position data. This evidence leads us to believe that the patterns we depict are not just an artifact of aggregation, though aggregation might exaggerate these patterns. Top-coding prevents us from investigating whether, like in our data, there is an increasing upper mode when using the PUMS data

Table 1
Data Reported for Legal Services Establishments in the 1992 Census of Services
by Legal Form of Organization

	PSOs	Partnerships and Proprietorships
Lawyers	Yes	Yes
Partners	No	Yes
Associates	No	Yes
Lawyers' Earnings	Yes	No
Partners' Earnings	No	No
Associates' Earnings	No	Yes
Non-Lawyers	Yes	Yes
Non-Lawyers' Earnings	Yes	Yes
Revenues	Yes	Yes
Operating Expenses (other than payroll)	No	No

Table 2
Overhead, Employment, and Revenues

Offices That Are Legally Organized As Professional Service Organizations (N=10438)

Dependent Variable: (Revenues - Payroll)

C	**39.881**	Employment	**6.474**		
	(13.370)		**(1.515)**		
Market Size Dummies		Market Size*Employment Interactions		Office Class*Revenues Interactions	
20K-100K	-3.438	20K-100K*Employment	2.111	Business, Non-Litigation*Revenues	**0.168**
	(16.275)		(1.599)		**(0.004)**
100K-200K	10.441	100K-200K*Employment	1.201	Individual, Non-Litigation*Revenues	**0.124**
	(16.745)		(1.555)		**(0.006)**
200K-400K	10.175	200K-400K*Employment	2.729	Litigation*Revenues	**0.143**
	(14.903)		(1.515)		**(0.004)**
400K-1M	6.085	400K-1M*Employment	1.623		
	(14.628)		(1.518)		
More than 1M	-3.860	More than 1M*Employment	**7.776**		
	(15.083)		**(1.539)**		

R-Squared: 0.78

Market size dummies are defined in terms of total employment in the county in which the offices is located.
Employment is the total number of individuals (lawyers and non-lawyers) working in the office, minus 2.
Offices is defined as "litigation" if at least one lawyer specializes in a litigation intensive field (negligence or insurance).
Non-litigation offices are defined as "business" or "individual" according to whether their revenues come primarily from businesses or individuals.

Bold indicates rejection of the hypothesis b=0 using a one-tailed t-test of size 0.05.

Table 3
Comparison of Earnings Distributions Using Actual Data and Estimates

Variable	Actual Earnings	Estimated Earnings	Actual Earnings	Estimated Earnings
Percentile				
10th	48	45	28	13
20th	60	63	38	36
30th	74	76	49	53
40th	86	89	59	66
50th	96	98	71	77
60th	107	110	86	95
70th	122	124		121
80th	142	145		168
90th	179	183		256
Sample	Census, Offices Organized As PSOs	Census, Offices Organized As PSOs	PUMS 5% State Sample Privately Practicing Lawyers	Census, Offices Organized As Partnerships or Proprietorships

Source: 1992 Census of Services, authors' calculations

All earnings are reported in thousands of 1991 dollars.
For PSOs, the unit of observation is the office; earnings data are the average pay per lawyer in the offices. The distribution is then taken across offices, weighting by the number of lawyers in the office. For partnerships and proprietorships, we calculate averages across partners and across associates separately within the same office. We then construct the earnings distribution using both of these observations, weighting by the number of lawyers they represent.

Table 4
Comovement of Associate Pay, Partner Pay, and Associates/Partner

Partnerships and Proprietorships with at Least One Associate

Panel A: Dependent Variable: ln(associate pay) (N=5130)

ln(partner pay)	**0.343 (0.007)**	**0.267 (0.007)**	**0.208 (0.007)**	**0.193 (0.007)**	**0.176 (0.007)**
R-squared	0.30	0.40	0.51	0.67	0.69

Panel B: Dependent Variable: ln(partner pay) (N=5130)

ln(associates/partner)	**0.453 (0.015)**	**0.370 (0.014)**	**0.317 (0.015)**	**0.323 (0.017)**	**0.322 (0.017)**
R-squared	0.15	0.31	0.33	0.50	0.52

Panel C: Dependent Variable: ln(associate pay) (N=5319)

ln(associates/partner)	**0.255 (0.011)**	**0.202 (0.010)**	**0.114 (0.009)**	**0.069 (0.010)**	**0.107 (0.010)**
R-squared	0.09	0.31	0.44	0.60	0.64
Controls	None	Field Shares	Field Shares, Market Size Dummies	Field Shares, County Dummies	Field Shares, County Dummies, Partners, Partners**2, Partners**3

Bold indicates rejection of the hypothesis b=0 using a one-tailed t-test of size 0.05.

Table 5
Vuong Tests of Occupational Stratification

Panel A: Lawyers in Partnerships and Proprietorships With at Least One Associate

Ordering								-LogL	Vuong Test Statistic
A1	A2	A3	A4	P1	P2	P3	P4	17171	
A1	A2	A3	P1	A4	P2	P3	P4	17491	8.83
A1	P1	A2	P2	A3	P3	A4	P4	19807	12.07

Panel B: Lawyers in Partnerships and Proprietorships

Ordering									All Counties		Counties w/<400K Employment		Counties w/>400K Employment	
A1	A2	A3	A4	P0	P1	P2	P3	P4	23897		14943		8445	
A1	A2	A3	P0	A4	P1	P2	P3	P4	23817	-2.18	14965	1.21	8375	-3.06
A1	A2	P0	A3	A4	P1	P2	P3	P4	23545	-3.25	14992	1.34	8210	-3.15
A1	P0	A2	A3	A4	P1	P2	P3	P4	23447	-2.93	15013	1.19	8150	-2.93
P0	A1	A2	A3	A4	P1	P2	P3	P4	23520	-2.28	15131	2.55	8120	-3.03

This table reports Vuong test statistics when comparing the occupational stratification specification in the first row to alternative specifications. The null is that the specifications fit the data equally well. Under the null, this statistic is distributed N(0,1). See Vuong (1989) for details.

The specifications are ordered logits, where lawyers are classified according to their occupational position and the associate/partner ratio in their office. The categories A1-A4 correspond to associates in offices where this ratio is less than 0.5, at least 0.5 but less than 1.0, at least 1.0 but less than 2.0, and greater than 2.0, respectively. The categories P1-P4 correspond to partners classified analogously. P0 in Panel B is partners in offices without associates.

The ordered logits predict lawyers' classification as a function of their earnings. All specifications allow threshold "alphas" to vary with field shares and county employment size dummies (see text for how these are defined).

The unit of observation is at the occupation*office level (partners or associates at a given office). In Panel A, N=10,950, which reflects that there are two observations for each of the 5475 partnerships and proprietorships with at least one associate in our data. In Panel B, N=14,906, which reflects that, in addition, there are 3,956 offices without associates.

Approximately 40 counties had >400K employment as of 1992; counties with approximately 400K employees include Hillsborough County, FL (Tampa) and Orange County, FL (Orlando).

Table 6
The Wage-Leverage Surface, Production Function Estimates

Partnerships and Proprietorships With At Least One Associate (N=5319)

	Wage-Leverage Surface Estimates	Production Function Estimates
Constant		**0.289**
		(0.009)
Associates/Partner	**3.254**	
	(1.522)	
(Associates/Partner)**2	0.084	
	(0.384)	
Share(Banking Law Specialist)	**13.646**	0.007
	(3.225)	(0.033)
Share(Corporate Law Specialist)	**41.389**	0.016
	(2.962)	(0.024)
Share(Insurance Law Specialist)	**8.911**	**-0.049**
	(1.830)	**(0.018)**
Share(Negligence-Defense Specialist)	**9.507**	**0.091**
	(1.902)	**(0.018)**
Share(Patent Law Specialist)	**26.802**	**0.177**
	(2.816)	**(0.027)**
Share(Government Law Specialist)	**19.478**	**0.127**
	(3.675)	**(0.039)**
Share(Environmental Law Specialist)	**29.261**	0.078
	(5.402)	(0.060)
Share(Real Estate Law Specialist)	**13.745**	0.008
	(2.571)	(0.031)
Share(Tax Law Specialist)	**29.741**	**0.209**
	(5.642)	**(0.055)**
Share(Criminal Law Specialist)	-1.704	0.057
	(2.974)	(0.040)
Share(Domestic Law Specialist)	-3.798	**0.139**
	(3.742)	**(0.047)**
Share(Negligence-Plaintiff Specialist)	**13.701**	**0.369**
	(1.835)	**(0.023)**
Share(Probate Law Specialist)	4.887	-0.080
	(4.432)	(0.051)
Share(Other Specialist)	**13.300**	**0.061**
	(1.464)	**(0.014)**
R-Squared	0.48	0.06

The dependent variable in the wage-leverage surface regression is average associate pay in the office.

The dependent variable in the production function is ln(revenues/lawyer*(1-K))-ln(MC), where K is the coefficient on revenues in the overhead regression for the office (0.168, 0.124, and 0.148 for business non-litigation, individual non-litigation, and litigation offices, respectively), and MC is the estimated marginal cost of leverage for the office. The coefficients reported here correspond to -ln(theta) in the text. The 0.289 coefficient estimate for the constant implies an estimate of theta of 0.749 for an office of general practitioners (the omitted category).

Bold indicates rejection of the hypothesis b=0 using a one-tailed t-test of size 0.05.

Table 7
Revenues Per Lawyer and the Estimated Marginal Cost of Leverage

Partnerships and Proprietorships With At Least One Associate (N=5319)

			Components of Estimated Marginal Cost				
Percentile	Revenues/ Lawyer	Estimated MC	Associate Pay	Nonlawyer Pay	Benefits	Overhead	Marginal Price of Leverage
	TR/(n+1)	*MC*	*w*	*x*	*0.15*(w+x)*	*oh*'/p*	*w'n*
10th	124	80	30	18	8	16	1
25th	173	107	44	26	11	18	2
50th	227	138	61	37	15	21	3
75th	297	175	77	49	19	27	5
90th	372	213	96	61	23	36	7
Mean	247	145	62	39	15	24	4
Relative to Estimated MC	1.70	1.00	0.43	0.27	0.10	0.17	0.03

Table 8
The Returns to Vertical Specialization

Partnerships and Proprietorships (N=9283)

	Revenues/Lawyer		
	Absent Vertical Specialization (upper bound)	Actual	Constant Returns to Leverage
	[g(zm)+nw]/(n+1)	*TR/(n+1)*	*g(zm)*
Percentile			
10th	77	83	89
25th	114	139	150
50th	163	209	248
75th	209	288	357
90th	258	374	485
Mean	171	227	278
Mean, Relative to Absent Returns to Specialization	1.00	1.33	1.63

All figures are reported in thousands of 1991 dollars.
The "absent vertical specialization" figures are an upper bound because associate wages overstate the value of their production, absent hierarchy. Comparing these to actual revenues per lawyer thus is a lower bound on the returns to vertical specialization.

Table A1
Regressions of Market Size Dummies on Lawyers' Earnings Decile

	Dependent Variable *Dummy Variable That Equals One If Office Is Located In a County Where Number of Employees Is:*					
	0-20,000	*20K-100K*	*100K-200K*	*200K-400K*	*400K-1M*	*Over 1M*
Earnings Above 1st Decile	0.006 (0.009)	**0.079** **(0.012)**	**0.054** **(0.011)**	**0.102** **(0.015)**	**-0.201** **(0.016)**	**-0.041** **(0.014)**
Earnings Above 2nd Decile	**-0.027** **(0.009)**	**-0.028** **(0.012)**	-0.006 (0.011)	-0.002 (0.015)	**0.089** **(0.016)**	-0.026 (0.014)
Earnings Above 3rd Decile	**-0.029** **(0.009)**	**-0.039** **(0.012)**	-0.019 (0.011)	**0.076** **(0.015)**	0.003 (0.016)	0.010 (0.014)
Earnings Above 4th Decile	**-0.029** **(0.009)**	**-0.046** **(0.012)**	**-0.041** **(0.011)**	**-0.098** **(0.015)**	**0.190** **(0.016)**	0.024 (0.013)
Earnings Above 5th Decile	**0.022** **(0.009)**	**0.045** **(0.012)**	-0.003 (0.011)	**-0.049** **(0.015)**	**-0.122** **(0.016)**	**0.106** **(0.014)**
Earnings Above 6th Decile	-0.007 (0.009)	**0.040** **(0.012)**	0.001 (0.011)	**0.037** **(0.015)**	**-0.110** **(0.016)**	**0.038** **(0.013)**
Earnings Above 7th Decile	0.002 (0.009)	0.009 (0.012)	**0.045** **(0.011)**	0.005 (0.015)	-0.027 (0.016)	**-0.035** **(0.014)**
Earnings Above 8th Decile	-0.016 (0.009)	**-0.061** **(0.012)**	-0.014 (0.011)	0.007 (0.015)	**0.169** **(0.016)**	**-0.085** **(0.013)**
Earnings Above 9th Decile	0.008 (0.009)	-0.013 (0.012)	**-0.034** **(0.011)**	**-0.095** **(0.015)**	-0.001 (0.016)	**0.136** **(0.013)**

This table reports results from six regressions; these are the regression versions of Figure 1. In the first column, the dependent variable is a dummy that equals one if the lawyers the observation represents work in a county with fewer than 20,000 employees. The independent variables are a series of dummies that indicate where in the overall distribution these lawyer lie. These dummies are defined so the coefficients represent changes relative to the previous category. The -0.029 point estimate in the third row, first column indicates that, moving from the 3rd earnings decile to the 4th lowers the probability a lawyer works in a very small market by 2.9 percentage points.

The row sums of the point estimates are zero by construction; the point estimates any one of the rows are implied by the other five.

All regressions include a vector of controls that include the shares of lawyers in the office who are in each of the 13 Census-defined fields. The estimates presented here differ little when excluding these controls.

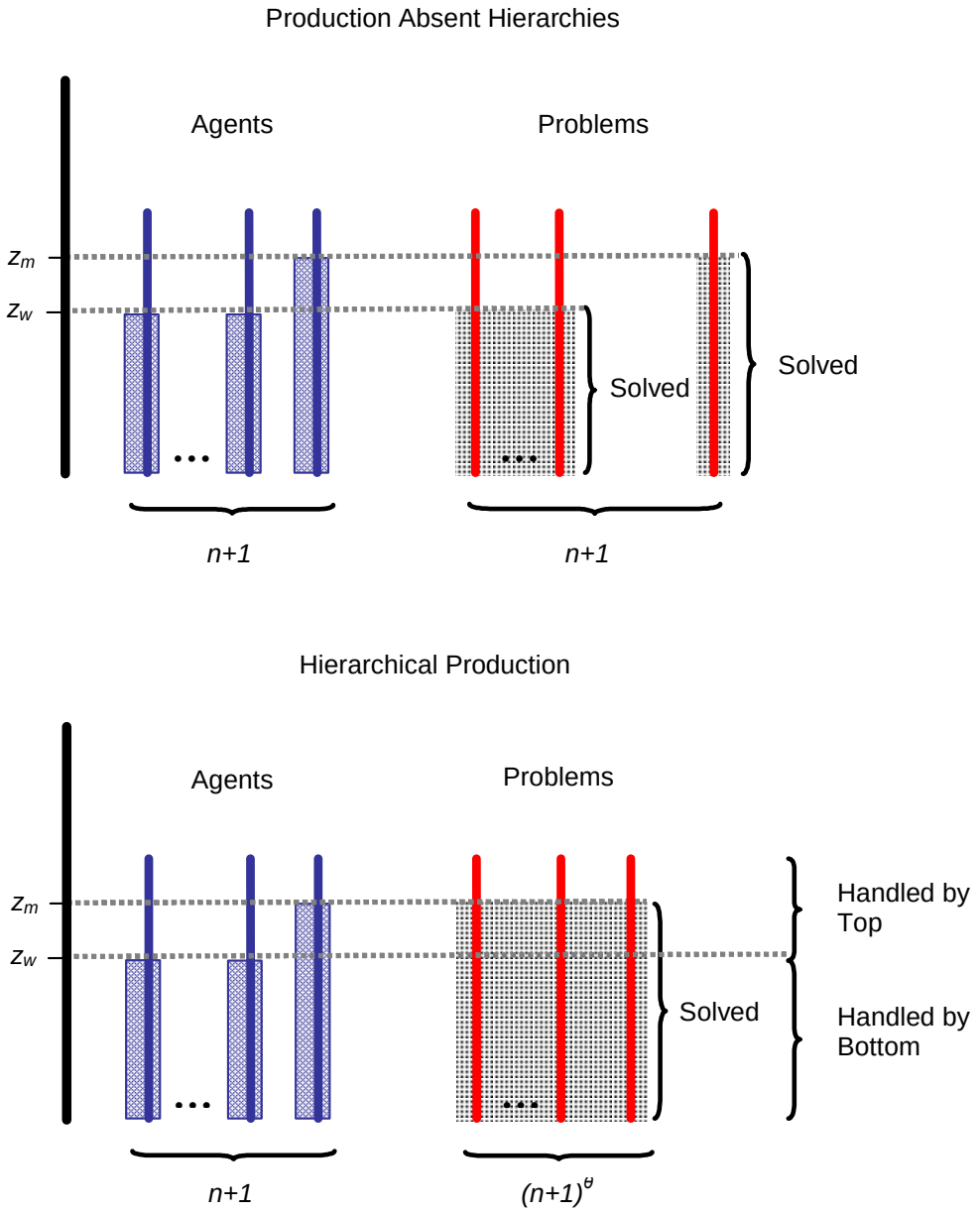


Figure 1. Non-Hierarchical and Hierarchical Production. The panel depicts production absent hierarchies; sets of problems are allocated to lawyers arbitrarily and each lawyer applies their time and knowledge toward whatever set they confront. Output is $z_m + nz_w$.The bottom panel depicts output under hierarchical production. The $n+1$ lawyers have $(n+1)^\theta$ units of effective time to solve problems. Lawyers divide work so that the n associates handle the easiest parts and the partner handles the hardest parts of the problems the group confronts. Output is $z_m(n+1)^\theta$.

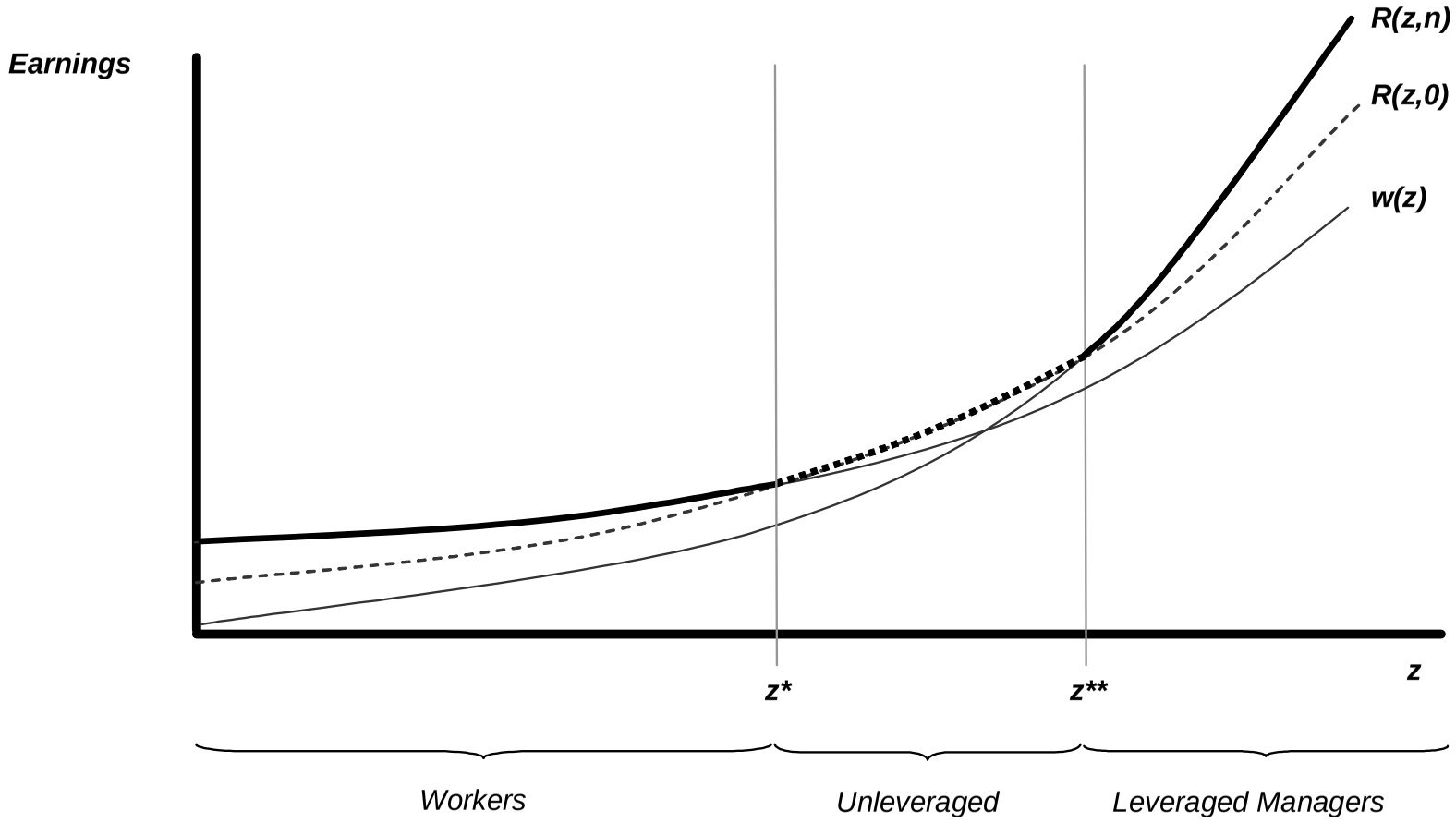
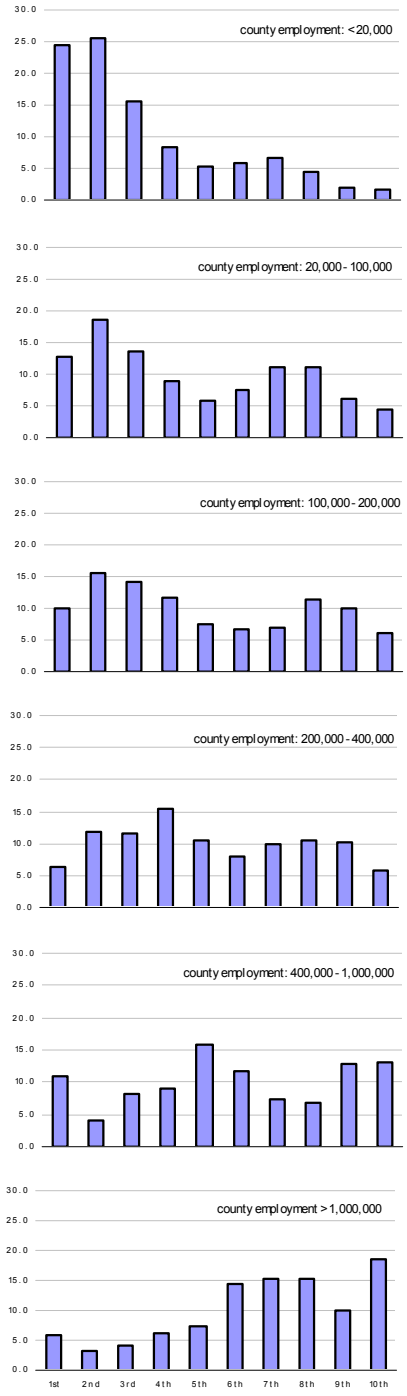


Figure 2. Equilibrium Earnings in Knowledge-Based Production. The three curves in this Figure represent agents' earnings as a function of their skill, z, if they are (a) managers with others working under them ($R(z,n)$), (b) working on their own ($R(z,0)$) and (c) workers working under managers ($w(z)$). The equilibrium wage-skill surface is the outer envelope of these curves, which is in bold. A goal of the empirical work is to estimate how much this outer envelope differes from $R(z,0)$.



This Figure depicts the how the distribution of lawyers across earnings deciles varies across local markets of different sizes.

We developed this Figure in the following way. First, we computed earnings deciles across all markets, and assigned associates and partners within each office to earnings deciles accordingly. Then, we computed and plotted frequency distributions of lawyers across these deciles within market size categories. The first bar of the top panel indicates that 24.5% of lawyers in counties with less than 20,000 employees have earnings that put them in the 1st decile, when earnings deciles are calculated across all markets. If earnings distributions are identical across differently-sized local markets, these frequency plots would depict uniform distributions.

These plots indicate that, although higher-earning lawyers tend to work in larger markets, earnings and market size do not appear to be positively associated throughout their domains. Instead, these frequency distributions tend to be bimodal, with both modes increasing as one moves from smaller to larger local markets.

The earnings deciles are:

10th:	13,110	60th:	94,875
20th:	36,333	70th:	121,679
30th:	52,777	80th:	168,201
40th:	65,993	90th:	256,388
50th:	76,657		

Figure 3. The Distribution of Lawyers Across Earnings Deciles by Market Size.

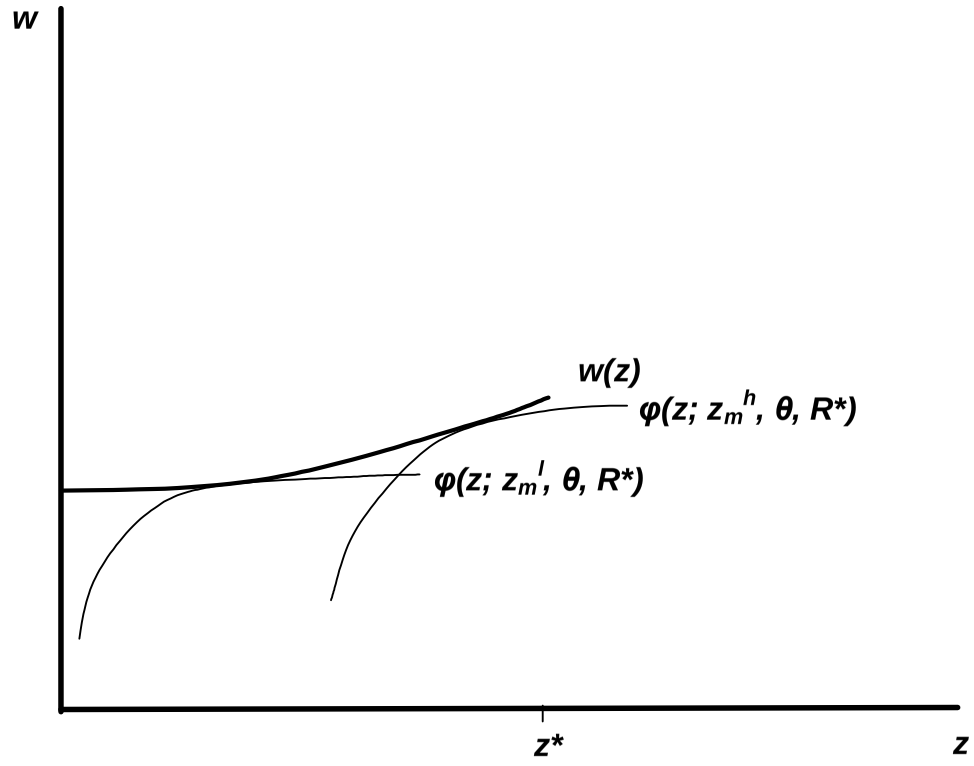


Figure 4. The Wage-Skill Surface and Bid Curves. The section below z^* in Figure 2 can be interpreted as a hedonic wage surface that relates associates' wage to their skill. This surface is the locus of tangency points between associates' offer curves and partners' bid curves. Here we show the wage surface and the bid curves of two partners, with skill z_m^l and z_m^h, respectively.

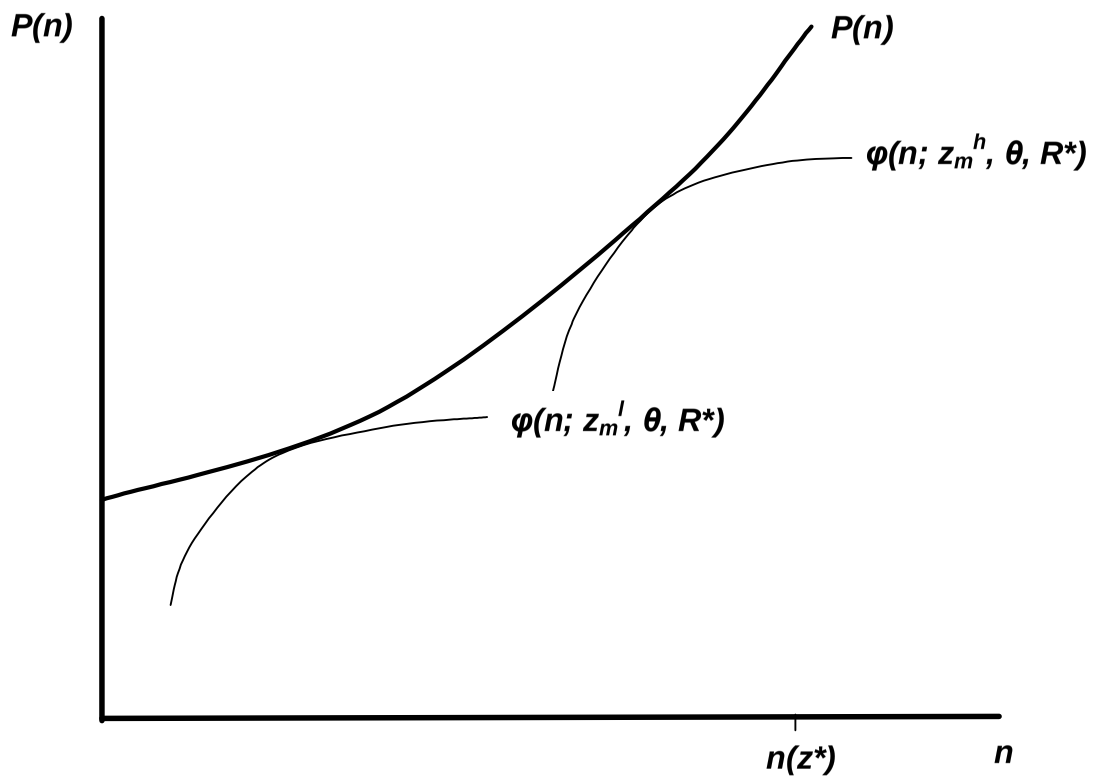


Figure 5. The Wage-Leverage Surface and Bid Curves. The hedonic equilibrium in Figure 3 can be recast in terms of the supply and demand for leverage because $n(z)$ is invertible. $P(n) = w(n)n$ is a surface relating the market price of leverage faced by partners to n. Here we show the wage-leverage surface and the bid curves of two partners, with skill z_m^l and z_m^h, respectively.

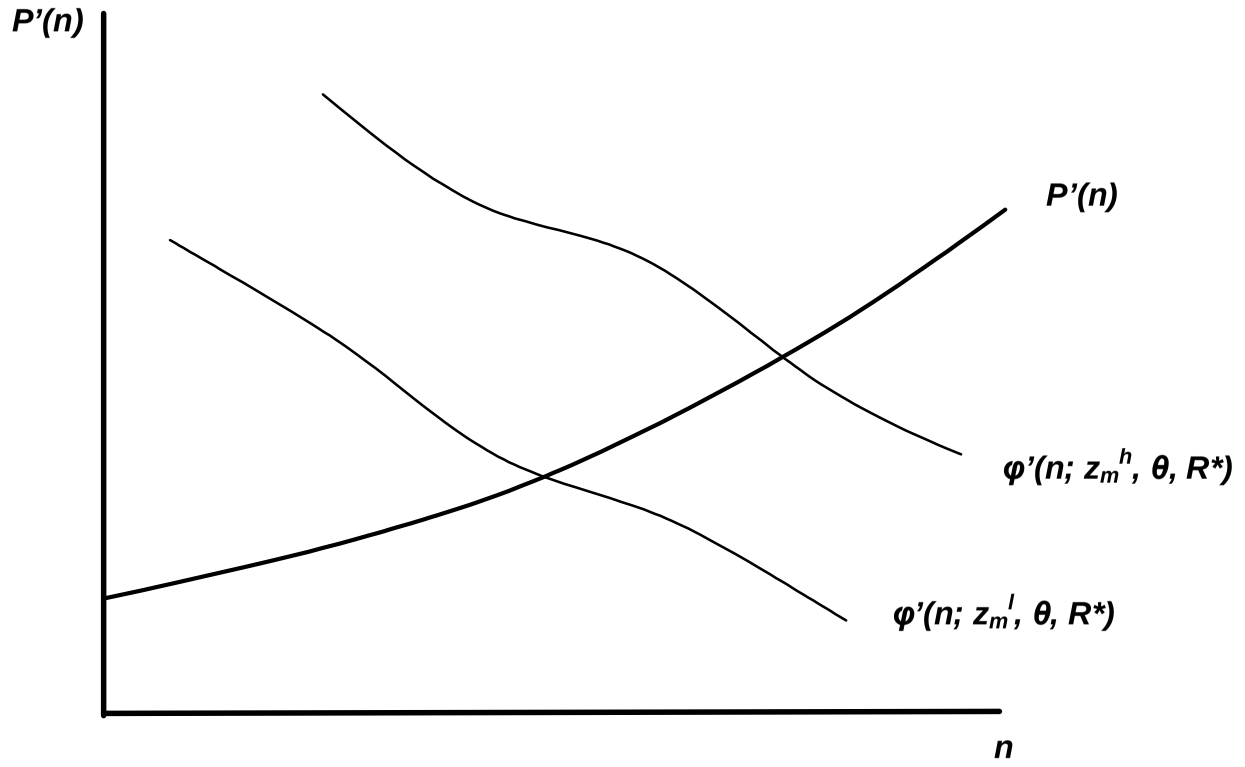


Figure 6. The Marginal Price of Leverage and Partners' Marginal Benefit of Leverage. The first derivative of partners' offer curves is their marginal benefit of leverage. The leverage a partner acquires is determined by the intersection between their marginal benefit curve and the marginal price surface. The parameter vector θ determines the shape of partners' marginal benefit curve; an empirical goal is to estimate its value. It will be identified by the ratio between our estimates of partners' average benefit of leverage and their marginal cost of leverage; the higher this ratio, the steeper the marginal benefits curves depicted here.

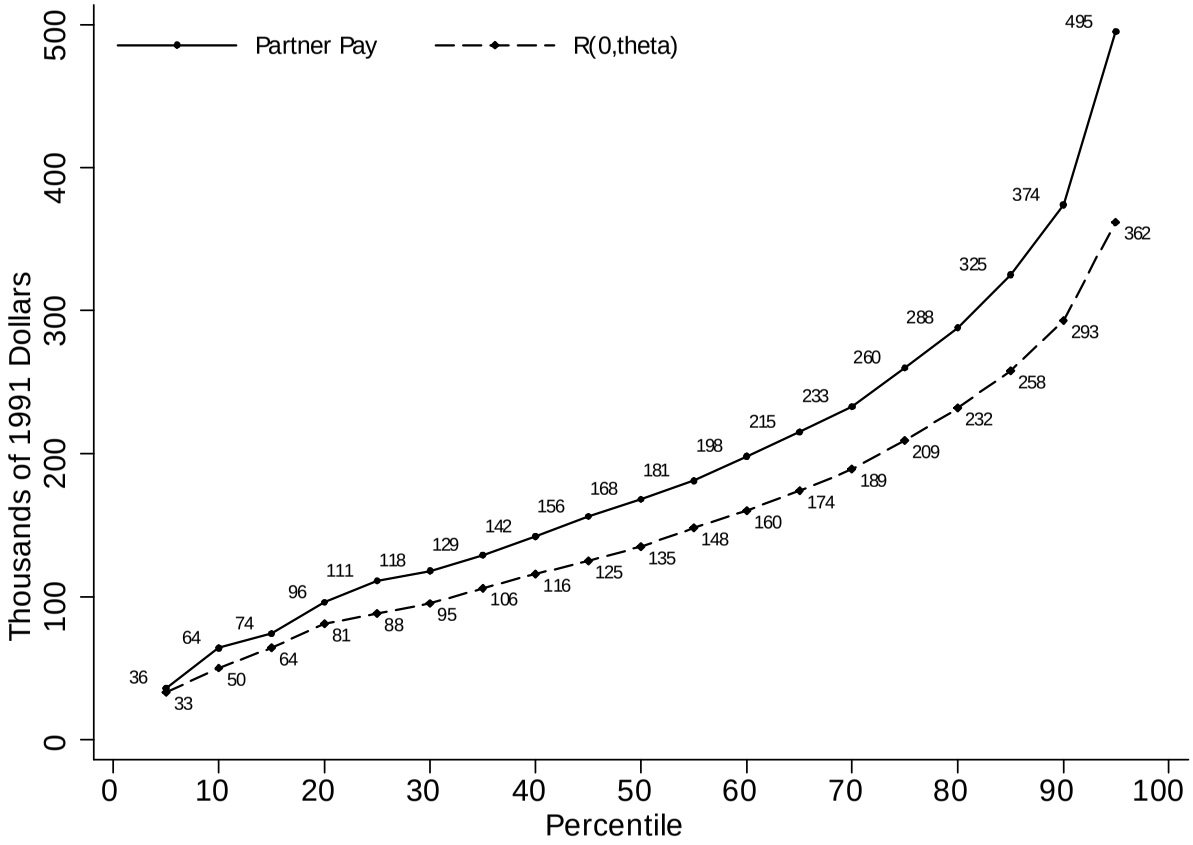


Figure 7. The Distribution of Partner Pay, Estimated Partner Pay Absent Hierarchies. This Figure reports 20 quantiles of the distribution of these quantities, looking only at partners at offices with associates. "Estimated pay absent hierarchies" is $R_i(z_{mi}, 0)$.

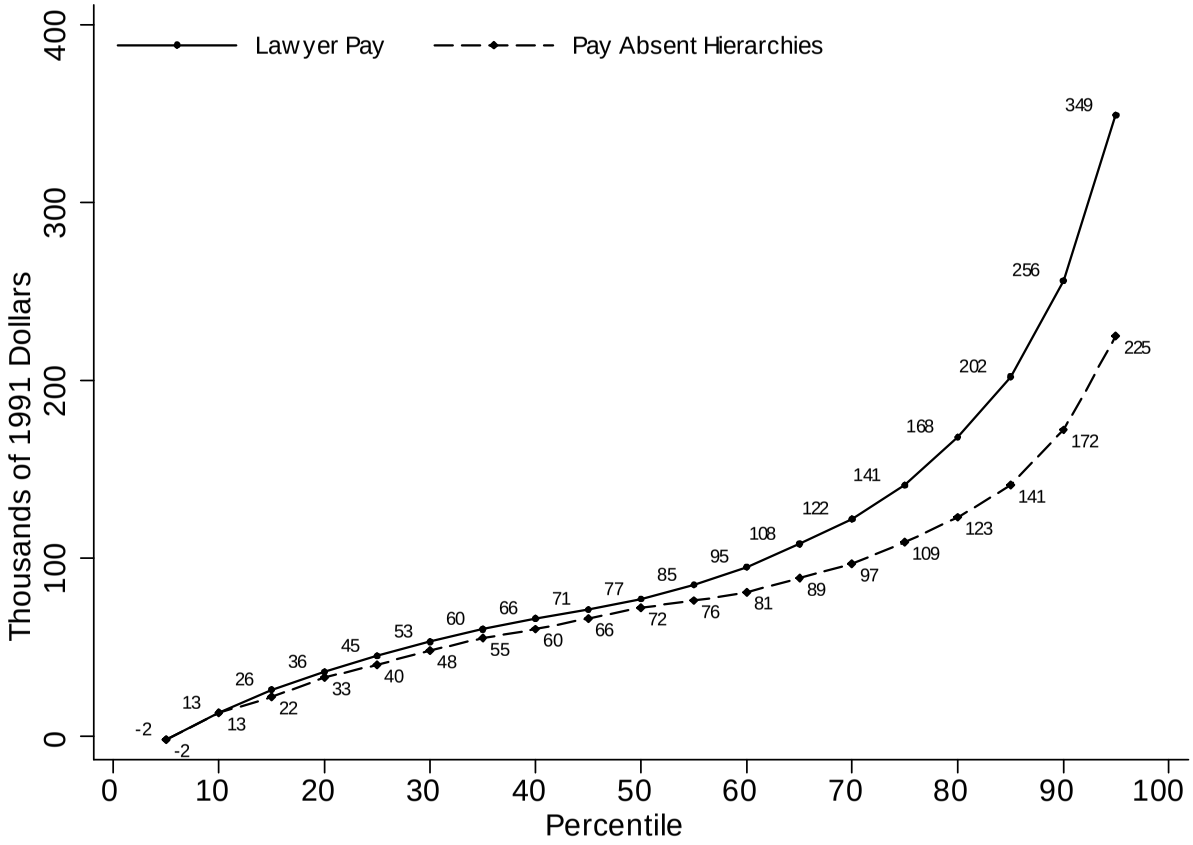


Figure 8. The Distribution of Lawyer Pay, Estimated Pay Absent Hierarchies. This Figure reports 20 quantiles of the distribution of these quantities. "Estimated pay absent hierarchies" is $R_i(z_m, 0)$ for partners at offices with associates. It is the same as lawyer pay for partners at offices without associates, as well as for associates. Because associates earn more as associates than they would absent hierarchies ($w_i > z_{wi}$), this overstates what these individuals would earn in this counterfactual. This upward bias primarily affects our estimates of lower quantiles.

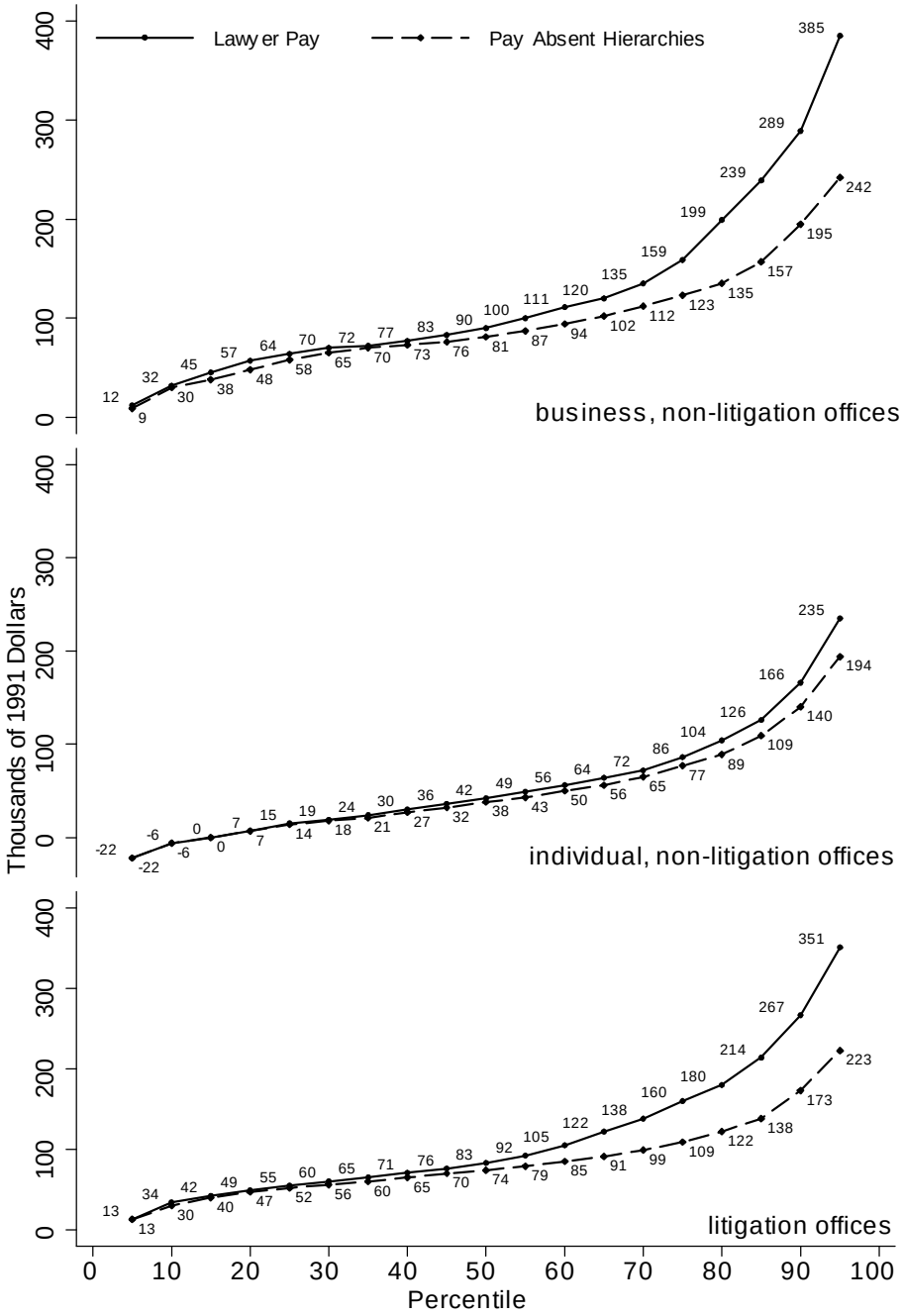


Figure 9. The Distribution of Lawyer Pay, Estimated Pay Absent Hierarchies, by Office Class. This Figure reports 20 quantiles of the distribution of these quantities for three classes of offices. "Estimated pay absent hierarchies" is $R_i(z_m, 0)$ for partners at offices with associates. It is the same as lawyer pay for partners at offices without associates, as well as for associates. Because associates earn more as associates than they would absent hierarchies ($w_i > z_{wi}$), this overstates what these individuals would earn in this counterfactual. This upward bias primarily affects our estimates of lower quantiles.