The Returns to Knowledge Hierarchies*

Luis Garicano
London School of Economics

Thomas N. Hubbard
Northwestern University

October 21, 2009

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 an equilibrium model of knowledge hierarchies inspired by 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. Our estimates imply that hierarchical production leads to at least a 30% increase in productivity in this industry, relative to a situation where lawyers within the same office do not "vertically specialize." We further find that it amplifies earnings inequality, increasing the ratio between the 95th and 50th percentiles from 3.7 to 4.8. We conclude that the impact of hierarchy on productivity and earnings distributions in this industry is substantial but not dramatic, reflecting the fact that the problems lawyers face are diverse and that the solutions tend to be customized.

*We thank Pol Antras, Matthew Gentzkow, Lars Nesheim and Esteban Rossi-Hansberg and participants at various seminars, the 2006 Summer Meetings of the Econometric Society, and NBER Summer Institute for comments and suggestions. We are especially grateful to Steve Tadelis, who presented this paper and offered extensive comments at the latter. 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

Knowledge is an asset with increasing returns because acquiring it involves a fixed cost, independent of its subsequent utilization. But when knowledge is embodied in individuals, individuals must spend time applying it to each specific problem they face and possibly also communicating specific solutions to others. This makes it difficult for individuals to exploit these increasing returns, 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 a model inspired by 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.[1]

We proceed in two stages. We begin in Section II by developing an empirically estimable equilibrium model of hierarchical production, adapting Garicano and Rossi-Hansberg's (2006) framework to allow for hierarchies with more managers than workers and diminishing returns to managerial span or "leverage" (number of workers per manager). In this model, individuals have heterogeneous ability – some are more skilled than others – and hierarchical production allows more talented individuals to leverage their knowledge by applying it to others' time. Our production function captures the organization of the division of labor within hierarchies, which is derived from first principles in Garicano and Rossi-Hansberg (2006), through two assumptions. First, managers who increase their span of control must work with higher-skilled workers, since increasing span requires them to delegate tasks they previously did themselves. Second, working in a team involves coordination costs that do not appear when individuals work on their own. The "hierarchical production function" and equilibrium assignment that result contain two crucial features that facilitate estimation of the model's parameters: first, the productivity of a hierarchical team, per unit of time spent in production, is determined only by the

[1]See also Garicano and Hubbard (2007b) for empirical tests of Garicano (2000) that relate law offices' hierarchical structure to the degree to which lawyers field-specialize. Unlike this paper, our previous work neither examines earnings or assignment patterns nor estimates the returns to hierarchy.

manager's skill; second, the number of workers per manager or leverage is a sufficient statistic for worker skill. The first of these reduces the empirical problem to estimating the time cost of hierarchical production, and the second simplifies our application of hedonic techniques to estimate this cost.

We then estimate the parameters of this model using data from the U.S. Economic Census on thousands of law offices throughout the United States. These data contain law office-level information on, among other things, partners' earnings, associates' earnings, and associate-partner ratios. With these parameters in hand, we then use our framework to infer the "returns to hierarchy" – 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. We conclude that hierarchical production has a substantial effect on lawyers' productivity, raising lawyers' output by at least 30% relative to non-hierarchical production in which there is no vertical specialization within offices. We also find that hierarchies substantially expand earnings inequality, increasing the ratio between the 95th percentile and median earnings among lawyers from 3.7 to 4.8, 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 reasonably large, 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 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 (Lucas (1978), Rosen (1982)).[2] 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.[3] 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

[2] 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)

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

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 tractable 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 first estimate of the returns to hierarchical production.

II. ECONOMIC MODEL: HIERARCHIES, ASSIGNMENT, AND HETEROGENEITY

II.1. Tastes and Technology

Building on Lucas (1978), Rosen (1982), and Garicano and Rossi-Hansberg (2006) (hereafter, GRH), we develop a general hierarchical assignment model and characterize its equilibrium properties. Our model is based on a hierarchical production function where, like in these earlier papers, managerial skill raises the productivity of all the inputs to which it is applied.

We assume that demand for a service Z can be described by the demand of a mass of representative agents with semilinear preferences. We specify:

$$U = u(z) + y$$

where $u(z)$ is the utility attained from the consumption of quality $z \in [0, 1]$ of the service Z and y is the utility (in dollars) obtained from the rest of the goods in the agent's consumption bundle. We assume $u'(z) > 0$, and $u(0) = 0$. Each agent maximizes utility subject to a budget constraint $v(z) + y = Y$, where $v(z)$ is the cost (in dollars) of purchasing a level z of service Z.

Suppliers are endowed with a skill level $z \in [0, 1]$ and with one unit of time. That is they can supply up to quality level z costlessly and above that level at an infinite cost. The population of suppliers is described by a distribution of skill, $\Phi(z)$, with density function $\phi(z)$. Production of the service Z involves the application of individual suppliers' skill and time to production. An agent with skill z can provide a quality z of Z working on his own. This has a value in the market of $v(z)$; this value $v(z)$ is an equilibrium object that we characterize below.

Following GRH, z can be thought of in our empirical context as an index that reflects the share of client problems in a particular field that a lawyer working in this field can solve: more-skilled lawyers can solve a greater share of these problems than less-skilled lawyers, and the problems that a less-skilled lawyer can solve are a subset of those that a more-skilled lawyer can solve.

To fix ideas, consider the output of a set of suppliers who are each working on their own, n of whom have skill z_w, 1 of whom has skill z_m. Total output for these $n+1$ agents would be $v(z_m) + nv(z_w)$. Below we will compare output in this case with output when agents are allowed to organize hierarchically. This comparison will both illuminate the benefits and drawbacks of hierarchical production, and will convey the central challenge in estimating the returns to hierarchy.

Hierarchical Production.—

Agents may work on their own or as part of a hierarchical team. Following Lucas (1978), Rosen (1982), and GRH, we propose that hierarchical production allows one individual's skill to be applied to other individuals' time. The "skill" of a hierarchical team is therefore equal to the maximum skill available in the team, and output is equal to the value of this skill multiplied by the time the team spends in production. We thus specify the output of a team with one individual (the "manager") with skill z_m and a measure of n workers with skill z_w as:

$$F(z_m, z_w) = v(\max\{z_m, z_w\})g(n)$$

We make two key assumptions about this production function.

Assumption 1: Costly Coordination. Production is increasing in team size, $g' > 0$. Coordination costs imply that time spent in production is less than the team's time endowment, $g(n) < n + 1$.

Hierarchical production is thus costly in terms of time: it allows an agent's skill to be applied to others' time, but time is expended in the process. Thus $g(n)$ is the team's "effective time" in production, where $g(n)$ is a mapping from the team members' time endowment $n + 1$ to "effective time."

Our empirical analysis will assume that $g(n) = (n + 1)^\theta$, $0 < \theta < 1$, so that n units of worker time and 1 unit of manager time results in $(n + 1)^\theta$ units of time spent in production. θ is a returns to scale parameter, in this case capturing the returns to scale associated with managerial leverage. However, our analytical approach does not depend on this particular specification of $g(n)$.

Assumption 2. Span of Control and Worker Skill. The measure of workers n with whom a manager can work is positive and increasing in the skill of the workers z_w:

$$n(z_w) > 0 \text{ and } n'(z_w) > 0 \text{ for all } z_w.$$

This assumption captures the idea that managers' time is limited, but managers are able to delegate more difficult tasks to more-skilled workers. Time-constrained managers who wish to scale up must delegate to workers tasks that they used to do themselves, and this requires them to work with more-skilled workers. The greater the skill of the workers, the less help each worker needs, and the more workers the manager can have in his or her team.[4]

From Assumptions 1 and 2, it is immediate that it is never optimal to have $z_m < z_w$, and thus we can rewrite our hierarchical production function, without loss of generality, as:

$$F(z_m, z_w) = v(z_m)g(n(z_w))$$

The trade-off associated with hierarchical production is now evident. Figure 1 illustrates output under hierarchical and non-hierarchical production, in the problem solving variant, in which $g(n) = (n+1)^\theta$. The top of Figure 1 depicts nonhierarchical production, in which agents work on their own. The left side of this panel depicts the skill of $n+1$ agents. The lines depict these agents' time endowments, the shaded regions depict these agents' skill, or problem-solving ability. n of these agents have skill z_w, 1 has skill z_m. Assume that each of these agents confronts 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 by the thin bars on the right. Under nonhierarchical production, each of these agents simply handles the problems they themselves confront. The value of the output of each of the n lower-skilled agents would be $v(z_w)$ and value of the output of the higher-skilled agent would be $v(z_m)$. Total output would be $v(z_m) + nv(z_w)$.

The bottom part of this Figure depicts hierarchical production. Total output is $v(z_m)(n+1)^\theta$, the product of the value of the manager's skill and the time the $n+1$ agents spend directly in production. Output per unit of productive time is improved, relative to autarchic production, because managers can apply their knowledge to more than one set of problems. This improvement is the benefit of hierarchical production;

[4]Although we state this assumption in general terms, it is straightforward to generate it from first principles in a more specific framework. In Garicano (2000) and GRH, workers draw problems and ask for help from managers whenever they cannot solve it. Assuming that the probability that a worker needs help is $1 - z_w$, the number of workers a manager can help is determined by his time constraint $(1 - z_w)hn = 1$, where h is the per-problem time cost of helping, resulting in $n'(z_w) > 0$.

the drawback is that hierarchical production involves a loss in time spent in production.

The Figure also illustrates the empirical task we will confront when estimating the returns to hierarchy. Our goal is to compare realized production and earnings to what production and earnings would be, absent hierarchical production. Our data will contain information on $F(z_m, z_w)$, law offices' output, and n, law offices' associate-partner ratios. It follows that if θ were known, one could infer $v(z_m)$ – what partners would earn, absent hierarchical production – because $v(z_m) = F(z_m, z_w)/(n+1)^{\theta}$. Much of our empirical focus will therefore be aimed at estimating θ, or more broadly the time cost associated with hierarchical production.[5]

Production in most contexts, even human capital intensive contexts like ours, involves inputs other than individuals' knowledge. We allow for this by introducing "overhead" inputs into the model in the following way.

Assumption 3: Monotonic and convex overhead costs Overhead and other costs (e.g., for office space, support staff) are positive, increasing, and weakly convex in team size, $c(n+1) > 0$, $c'(n+1) > 0$, $c''(n+1) \geq 0$.

II.2. Equilibrium

Stochastic Elements.—

Our estimation framework will rely on two key equilibrium relationships: a first order condition characterizing managers' optimal choice of workers, and the equilibrium relationship between an associate's earnings w and the associate-partner ratio n at the associate's office. Utilizing these relationships empirically requires us to introduce stochastic elements into the model in a way that makes the structure of these relationships well-defined but not deterministic. We therefore make the following assumptions.

Assumption 4: Stochastic productivity. The productivity of teams is stochastic, so that the value of production at office i per unit of productive time is: $v_i(z_m) = v(z_m)\varepsilon_i$ where ε_i is i.i.d., $\varepsilon_i > 0$, $E(\varepsilon_i) = 1$. Productivity shocks are realized after organizational decisions are made, so that they affect partners' earnings but not the organizational equilibrium. These shocks affect the team's rate of output during

[5] $F(z_m, z_w)$ is closely related to production functions that have been applied elsewhere in the literature on hierarchical sorting. The production function in Lucas (1978) can be written as $F(z_m, z_w) = z_m g(n)$, and thus represents a special case of our model in which the skill of workers is irrelevant, and only the skill of managers matters. A two-layer version of the production function in GRH can be obtained from $F(z_m, z_w)$ by specifying $v(z) = z$, $n(z) = 1/(h(1-z))$, and $g(n) = n$ for $n > 1$. (Other elements of GRH are more general than our model; they allow for hierarchies with more than two layers, and allow the skill distribution among agents to be endogenous. We make fewer assumptions about the nature of the interaction between managerial and worker skill than GRH, but do not derive hierarchical production from first principles as they do.)

time in production.[6]

Assumption 5: Compensating differentials. Agents have preferences with respect to working as associates under different partners, so that the wage that a partner at office i must pay to compensate associates at the office by an amount w equals $w_i = w\xi_i$, where $\xi_i > 0$, $E(\xi_i) = 1$, and ξ_i is an absolutely continuous i.i.d. random variable, and is independent of all variables in the model.

An important element of Assumption 4 is that productivity is realized after the organizational equilibrium is obtained; we view this as reasonable in our empirical context, in which there is a distinct season for hiring associates and where some of the details of production (e.g., how time-intensive it is to communicate solutions to particular clients) are unknown at the time associates are hired.

Assumption 5 implies that any nonwage amenities of working as associates at office i are valued the same across individuals, and that this value is independent of the office's associate-partner ratio and, for that matter, the skill of the partners.[7] These conditions are necessary to keep the labor market equilibrium tractable because they will imply that any systematic sorting between agents is by skill and not other dimensions. This rules out multidimensional sorting, but is obviously stringent, in particular in this context because the independence assumption rules out the possibility that associates are willing to work for less under higher skilled partners than lower-skilled partners (perhaps for reasons having to do with training or client contacts). We will discuss the direction and magnitude of biases introduced by this assumption below.

Output Market.—

By the representative agents' first-order conditions, the price schedule in the output market is given by the value schedule of the representative consumer: $u'(z) = v'(z)$ subject to $u(z) \geq v(z)$. To simplify, we assume that the mass of demanders exceeds the capacity of suppliers. Under this assumption, the price schedule that solves this problem is given by the utility of the representative consumer, that is:

$$v(z) = u(z)$$

Trivially, this implies a price schedule where $v'(z) > 0$ and $v(0) = 0$.

[6] Note that this is distinct from the coordination cost θ, which reflects time loss from working with others. Individuals optimize knowing θ. Furthermore, productivity is stochastic at all offices, including offices where individuals work on their own.

[7] This is the standard compensating differentials assumption. With homogeneous agents, the equilibrium wage is such that individuals are indifferent between the different amenities: the wage "equalizes" utilities (see e.g. Rosen (1986)). Note also that individuals have the same preferences with respect to working as an associate at office i. This makes the equilibrium analysis simpler than in hedonic labor market models where workers differ in their preferences.

Labor Market.—

Like in GRH, the continuum of heterogeneous agents make occupational choices and team composition choices to maximize their compensation given the price schedule $v(z)$.[8] Each agent chooses whether to be a manager, to work on their own, or be a worker, and earn in expectation R, $v(z)$, or $w(z)$, respectively, where $w(z)$ relates the compensation an agent receives as a worker to the agent's skill. $w(z)$ is a hedonic wage function equates the supply and demand for skill at each skill level, and is thus an equilibrium object.

The labor market equilibrium involves solving a continuous assignment problem. The production function is continuous and involves complementarities between worker and manager skill, $\partial^2 F(z_m, z_w)/(\partial z_m \partial z_w) > 0$. Thus in general the assignment exists, is one to one in terms of skill, and is unique (Chiappori, McCann and Nesheim, forthcoming).[9] Thus there exist a matching function $z_m = m(z_w)$ derived from the equality of supply and demand for skill at each point that maps the skill of workers to the skill of their managers.

R, the residual earnings (or "rents") of a manager of skill z_m who has hired n workers of skill z_w, can be written as:

$$R = \varepsilon_i v(z_m)(n(z_w) + 1)^\theta - w_i n(z_w) - c(n(z_w) + 1) \tag{1}$$

The first term is the value of output; in our context, the revenues associated with a partner and his or her associates. The second term is the wage cost of hiring n associates of skill z_w. The third term is the overhead cost associated with the partner and associates.

Agents evaluate R at the point where they have chosen the skill level of their workers to maximize expected earnings, given the wage schedule they face $w_i = w(z_w)\xi_i$. This implies a first order condition:

$$v(z_m)\theta(n(z_w) + 1)^{\theta-1} n'(z_w) - w_i'(z_w)n(z_w) - w_i(z_w)n'(z_w) - c'(n+1)n'(z_w) = 0$$

Given $z_m = m(z_w)$, the above equation is a differential equation in $w(z)$; all the other objects are given. Assumption 2, combined with the fact that worker skill only enters the production function through managerial leverage, implies that managerial leverage n is an invertible function of worker skill z_w. Because of this, the labor market equilibrium can be equivalently characterized in terms of the supply and demand for leverage, n. The

[8] Compensation for agents who choose to be workers, w, includes any compensating differential.

[9] Specifically, Chiappori, McCann, and Nesheim (forthcoming) show that assignments in a class of problems including ours (e.g. including the possibility that some agents will not participate in the market) exist with great generality and that under a generalized Spence-Mirrlees condition (of which our positive cross-derivative is a special case), the assignments are unique and one-to-one outside a negligible set.

first order condition above can be restated as:

$$v(z_m)\theta(n+1)^{\theta-1} = w_i'(n)n + w_i(n) + c'(n+1) \tag{2}$$

and the hedonic wage function can be stated instead in terms of n, $w(n)$ and its derivative $w'(n)$.[10] This condition holds in equilibrium for all individuals who choose to be leveraged managers, and summarizes these agents' demand for leverage. As applied to our context, lawyers' choice of n is greater, the greater their skill, z_m, and the lower ξ_i: higher skill makes leverage more valuable, and lawyers with lower ξ_i can obtain it at lower cost. The fact that n is an invertible function of worker skill means that similar relationships hold when looking at worker skill: lawyers with higher z_m and lower ξ_i choose to work with more skilled, as well as more, associates. In equilibrium, there will be systematic sorting between lawyers on the basis of skill, but not on other characteristics given that lawyers, and in particular lawyers who choose to work as associates, are assumed to have the same preferences for nonwage amenities as captured by ξ_i.

An empirical advantage of reformulating the problem in terms of the supply and demand for leverage rather than the supply and demand for skill is that n is a variable we observe directly in the data – it is the number of associates per partner. This makes the first order condition more useful for estimation purposes, because we have eliminated an unobservable variable. It also helps with respect to utilizing hedonic techniques in recovering $w(n)$. A common problem that researchers encounter when utilizing hedonic techniques to recover supply or demand parameters is sorting on unobservables; absent this reformulation, we would face this problem as well, because we do not observe skill without error (in fact, we do not observe it at all). Given the assumptions of our model, this is not an issue here: because there is no systematic matching between partners and associates on characteristics other than skill, there is 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.[11] The invertibility property means that the quantities n and $w^{'}(n)$ summarize a lot of information in equilibrium. n is not only the number of workers per manager, but is also an error-free index of workers' skill.[12] Likewise $w^{'}(n)$ is not only the marginal price

[10] We are abusing notation slightly here, as $w(n)$ and $w(z)$ are different functions. $w_i'(n)$ in this equation is $\frac{dw_i(n)}{dn}$.

[11] Our exploitation of the invertibility of $n(z)$ 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.

[12] 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 $n(z)$. For example, if more skilled associates are willing to work for less for more skilled partners, n would continue to be a

of leverage, but is also a monotonic transformation of the marginal price of skill.

Using the definition of R and the production function $F(z_m, z_w) = \varepsilon_i v(z_m)(n+1)^{\theta}$, we can rewrite the first order condition as:

$$\begin{aligned} \underbrace{\frac{F(z_m, z_w)}{n+1}}_{} \theta \frac{1}{\varepsilon_i} &= \underbrace{w_i' n + w_i + c'}_{} \qquad (3) \\ \text{Average Revenues } \theta \frac{1}{\varepsilon_i} &= \text{ Marginal Cost} \end{aligned}$$

The left side of this equation is the marginal benefits of leverage, which are the average revenues per team member multiplied by θ/ε_i; the right is the marginal cost of leverage. This marginal cost contains three terms: the extra wage that needs to be paid to all team members w_i' (increasing leverage requires better as well as more workers), the wage of the additional agent, and the additional overhead cost. Thus the equilibrium relationship can be rewritten as:

$$\theta = \frac{MC}{AR} \varepsilon_i \qquad (4)$$

or:

$$\ln AR - \ln[w_i'(n)n + w_i + c'(n+1)] = -\ln\theta + \ln\varepsilon_i \qquad (5)$$

Our estimates of θ are based on this equation. Identification of θ is based on a straightforward idea that has been applied many times in the context of estimating returns to scale. In equilibrium, each manager chooses n such that the marginal benefits to leverage equal the marginal cost of leverage. If there are constant returns to leverage – $\theta = 1$ – then the average benefits of leverage equal the marginal benefits of leverage, and therefore the average benefits of leverage should equal the marginal cost of leverage. Finding that this is not the case – in particular, that the average benefits of leverage exceed the marginal cost of leverage – is evidence that there are decreasing returns to leverage, and under our specification the ratio between the marginal cost and the average benefits of leverage indicates the magnitude of decreasing returns.

The left side of this equation contains four terms: revenues per team member, the marginal price of leverage, worker pay, and the marginal overhead cost associated with workers. As we describe in further detail below, we obtain estimates for the marginal price of leverage from the coefficients in the hedonic wage regression implied by Assumption 5.[13]

sufficient statistic for associate skill, since this would merely reinforce positive sorting.

[13] We also explain below how we construct estimates of c', the marginal overhead cost. We defer this discussion because it relies on our data and institutional context, which we describe in the next section.

Dynamics and $w'(n)$.—

An important assumption in the model is that agents maximize their current-period compensation, given their skill. This specification of agents' objective, along with Assumption 5, ensures that agents face a schedule for the marginal price of leverage, $w_i'(n)$ that is independent of their skill and therefore the hedonic regression described above reveals this schedule. However, this 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 or contacts. Such dynamic aspects are very realistic in our context, but are difficult to incorporate directly in our model and estimation procedure. 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 place an additional value on working with higher-skilled partners.

If we have misspecified agents' tastes in this way, this will tend to bias *downward* our estimates of $w_i'(n)$, and thus the marginal cost of leverage $MC(n)$, and therefore ultimately θ, the coordination costs associated with hierarchy. If associates are willing to work for less for higher-skilled partners, then our estimate of $w_i'(n)$ will understate the marginal price of leverage faced by each individual partner; a partner 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 partner for a slightly-higher-skilled associate. Estimates of $w_i'(n)$ based on the hedonic regression described above 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 hierarchy, since a too-low θ implies that we overstate the extent to which the potential returns to hierarchy are eaten up by coordination costs. Our estimates should therefore be treated as a lower bound in this light.

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.

Relatedly, while our analysis largely abstracts from the details of internal labor markets, there is an issue of whether this abstraction leads us to misestimate the marginal cost of leverage, and hence θ. In particular, one can imagine a multi-period model inspired by tournament theory in which part of the marginal cost of an associate is a prize paid by incumbent partners to the most promising associates in the form of a transfer they receive from incumbent partners upon promotion. If this is the case, our analysis understates the marginal cost of leverage, and hence our estimates of the returns to hierarchy are a lower bound by the same logic as above. We think this perspective is incomplete, however, because characterizing promotions as a cost to incumbent partners ignores the prospective client-generation-related benefits of promoting promising lawyers. From incumbent partners' perspective, it is probably not a cost to promote lawyers who are expected to be at least as productive as existing partners.[14] If so, the promotion-related "prize" that accrues to the most promising associates should not be considered part of the marginal cost of leverage.

III. DATA AND ESTIMATION

III.1. 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 nonlawyers 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 across multiple fields. These variables allow us to control for the field composition of lawyers at various points in our analysis. Our main sample includes 9,283 law offices. This includes only observations in our sample that are legally organized as partnerships or proprietorships, because partners and associates are broken out separately only for these observations.[15] Throughout our analysis we use sampling weights supplied by the Census to account for the likelihood each was sampled.

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. Data that allows one to connect individuals' earnings with firm char-

[14] Levin and Tadelis (2005) propose that partnerships' profit-sharing aspects provide incumbent partners an incentive to only add new partners that raise the partnership's average skill level.

[15] Other offices are legally organized as "professional service organizations," or "PSOs." As we discuss at length in Garicano and Hubbard (2007b), it is unlikely that our use of only firms organized as partnerships or proprietorships in our main analysis leads to any important sample selection. States began to allow law firms to organize as PSOs mainly to allow partners the same tax-advantaged treatment of fringe benefits as employees of corporations, but this form had few other effects; it remained the case that shareholders consisted only of lawyers in the firm, and that there were no retained revenues. Selection of firms into this form as of 1992 largely reflected when the firm's state allowed for PSOs, and not differences in firm characteristics; PSOs were more much common in early-allowing states such as Florida than late-allowing states such as New York, but their prevalence varied little with their size. In 1992, PSOs made up about one-third of the industry in terms of lawyers, offices, and revenues.

acteristics 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: in particular, they do not directly report partners' earnings, and thus we have to estimate them from the data at hand.

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:

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

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

To do this, we utilize observations of law offices that are legally organized as "professional service organizations," or "PSOs." The data the Census collects for these offices differs from those the Census collects for partnerships. The Census collects data only on the total number of lawyers at these offices – and not separately the number of partners and associates. This makes these observations unusable in our main analysis, and thus they are not part of our main sample. However, it collects data on the total pay to lawyers – and not just to associates – which makes these observations useful for analyzing the determinants of law offices' overhead. The above identity 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 – revenues, lawyer pay, and nonlawyer pay – and thus allow us to infer overhead for each of these offices. We estimate the relationship between overhead and firm characteristics at PSOs, then use our estimates to obtain overhead estimates at the partnerships in our main sample; by the identity above, this implies estimates of partners' earnings.

Estimating Overhead.—

Our procedure for estimating overhead relies on knowledge of the structure of law firms' costs, derived mainly from reports on law offices from the Census' Operating Expenses Survey[16] and from Altman Weil's 1994 Survey of Law Firm Economics. In particular, our procedure is mindful of the following:

- "Non-payroll fringe benefits" such as health insurance and retirement plan contributions 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.

For example, rent increases with the number of people. Many of the expenses that increase with the amount of business, such as office supplies, communications, and expert services are "pass-through" expenses which are billed through to clients but will appear as both expenses and revenues in our data. This occurs, for example, when patent lawyers hire engineers.

- Some operating expenses such as rent tend to 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). The relationship between overhead and revenues might vary across fields because "pass-throughs" are more important in some than others (e.g., patent law).

We incorporate 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 incorporate the rest by specifying 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 allow the relationship between revenues and overhead to vary across fields.

We report the coefficient estimates from this specification in Table 1.[17] We allow the intercept term to vary with indicator variables that correspond to the employment size

[16] Bureau of the Census (1996).

[17] 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.

The error term in the OLS regression is heteroskedastic; the variance of the residual is higher for higher-revenue offices. We therefore use a GLS estimator to correct for this. The first stage regresses the logged square of the residual on a fourth-order polynomial of logged revenues. We use the predicted values of this regression as weights in the regression we report here.

of the county in which the office is located, and 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 \$28,500. The interactions suggest that the overhead associated with each additional individual is about \$2,900 in very small counties but this tends to be much greater in very large markets. We allow the coefficient on revenues to enter quadratically and to differ across fields. The estimates indicate that the relationship is concave for most fields, and strongest for patent, banking, and real estate law. The estimates imply that overhead increases by \$0.10-\$0.25 with each \$1.00 increase in revenues for most offices in our sample.

The R-squared for this regression, 0.70, is high. We found that more detailed specifications, including those that include county fixed effects instead of the market size dummies and that interact field shares 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]

We use these estimates in several ways. We use them to construct estimates of partner earnings at partnerships. This involves substituting in our estimate for oh_i, and deflating TR_i by our estimate of overhead's share of revenues at each office, $\partial oh_i/\partial(revenues_i)$ from the regression coefficients in Table 1. The latter accounts for the fact that some share of an offices revenues as reported in our data are pass-through expenses. In Table A1 in the Appendix, we show that the distribution of earnings that this procedure generates closely matches the distribution of earnings of privately-practicing lawyers from the 1990 PUMS data (up until the point at which earnings in the PUMS data are top-coded). To keep language simple, we will refer to the partner earnings generated by this procedure as "partner earnings" rather than "estimated partner earnings," even though the latter is more accurate.

We also use them to construct estimates of $c'(n+1)$ at each law office. With respect to the latter, we specify:

$$c'(n+1) = x_i + oh'_i/p_i$$

$c'(n+1)$ 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 per partner. This increase includes an increase 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 employ-

[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.

ment 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.

Summary Statistics and Patterns in the Data.—

Median earnings across all lawyers in our main sample are $77,000. The 25th and 75th percentiles are $44,000 and $141,000, respectively. The 95th percentile is about $350,000; there were about 435,000 privately-practicing lawyers in the U.S. in 1992, so this represents the earnings of roughly the 20,000th-ranked lawyer. About 40% of lawyers are associates, 25% are unleveraged partners (partners in offices with no associates), and 35% are leveraged partners. Among the latter, less than one-half work in offices with an associate-partner ratio greater than one.

Much of our analysis will be conducted from the perspective of partners' optimal choice of leverage; it is thus useful to report some statistics from the perspective of the average partner in our sample. The first column in Table 2 reports that average revenues per partner were $361,000, and average partner pay was $150,000. On average, partners had 0.6 associates, to whom they paid $36,000. The average partner worked in an office with 15 partners. In light of important ways in this industry are segmented (see Garicano and Hubbard (2007b)), we classify offices in the following way. We define "litigation" offices as those with at least 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. Table 2 indicates that the partners in our sample are evenly distributed across these three classes of offices.

The second and third columns, which report these averages separately according to whether offices have at least one associate, indicate that the averages in the first column mask a lot of variation in our sample. Offices with at least one associate are much larger in terms of the number of partners than those with no associates. Revenues per partner and partner pay are much higher as well. Our empirical analysis will revolve around relationships between earnings and offices' hierarchical organization; this table highlights the importance of accounting for differences in offices' scale and lawyers' fields (or their office's segment), both of which are correlated with both lawyers' earnings and hierarchical structure.

In other work (Garicano and Hubbard (2007a)), we have investigated earnings and organizational patterns in these data using a series of regressions. We found that controlling for lawyers fields, the size of the local market in which they work, and other variables, (a) partners' earnings and associates' earnings are positively correlated, (b) partners' earnings are higher in offices where associate-partner ratios are greater, and (c) associate earnings are greater in offices where associate-partner ratios are greater. These patterns are consistent with the hypothesis that manager skill, worker skill, and the worker-manager ratio should move together. We also investigate how earnings vary with lawyers' position. We found that, throughout most of our data and controlling for field and local market size, associates earn less than unleveraged lawyers, who in turn earn less than leveraged partners. Furthermore, we found that associates earn less than unleveraged partners, even when comparing associates at offices with high associate-partner ratios to unleveraged partners. The one segment of the labor market where these patterns do not hold is lawyers in very large cities serving business demands; there, unleveraged partners earn only as much or less than lawyers working as associates. In Garicano and Hubbard (2007a), we discuss how this exception reflects that the distribution of earnings among the small number of unleveraged partners in this market has a long lower tail, and thus may reflect that a disproportionate share of them work part time. Without evidence on the hours these lawyers work, however, we admit that this is a conjecture.

III.2. Estimation: w'(n) and the Marginal Cost of Leverage

We develop estimates of $w_i\prime(n)$ by specifying $\ln w(n)$ as a polynomial in n, controls for the field composition of lawyers in office i, and a full set of county fixed effects. We allow the polynomial to differ depending on whether the office is a "litigation," "business, non-litigation," or an "individual, non-litigation" office; allowing $w(n)$ to differ in this way accounts for the possibility that labor markets for lawyers are segmented along these lines. In practice, we found little additional explanatory power when adding terms in n beyond quadratic.[19]

Suppressing the controls, the first stage regression assumes:

$$\ln w_i(n) = \beta_0 + \beta_1 n + \beta_2 n^2 + \xi_i \tag{6}$$

Thus the marginal wage is:

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

We therefore estimate the wage-leverage surface by regressing the natural log of average associate earnings at office i on a quadratic in the associate/partner ratio at the office

[19] We have also estimated versions of this that interact the n terms with the market size dummies reported in Table 1; these allow the slope and curvature of the wage-leverage surface to differ across different market sizes. Unlike with the product market interactions, including these market size interactions did not significantly improve the fit of the model. One explanation for this result is that lawyers' mobility leads labor markets to be more segmented along the lines of product than geographic markets.

and a set of the above controls.[20] Our estimates allow us to construct an estimate of the marginal price of leverage, $\widehat{w_i'(n)}$, for partners at each office.

Table 3 reports our estimates of this equation, using offices in our main sample with at least one associate. Our estimates imply that $w'(n)$ is positive for the "business, non-litigation" offices, and that increasing the associate-partner ratio by one is associated with a \$7,750 increase in average associate pay. In the other segments, the wage-leverage surface is essentially flat, with none of the coefficients statistically different from zero. Drawing from the discussion in Section II, these results are consistent with a model in which the quality and quantity of workers' human capital are not perfect substitutes in the "business, non-litigation sector," but are perfect substitutes in the other two sectors.[21]

In Table 4, we report the mean and various quantiles of the marginal cost of leverage implied by these estimates. 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 \$139,000, though there is wide dispersion across offices. Our estimates imply that, on average, the pay the additional associate receives is only 45% 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, 28%), the cost of the associate and staff's fringe benefits (11%), and the cost of the additional overhead (13%). The combined fact that $w'(n)$ is generally not very high 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 these to our marginal cost estimates foreshadows our estimate of θ below, which is identified by the ratio of the marginal cost and average benefit of leverage. Average revenues per lawyer are about \$247,000, but the distribution of average revenues per lawyer is highly skewed across offices. Multiplying revenues per lawyer at each office by one minus our estimate of the overhead share of revenues (derived for each office from the coefficients on revenues in Table 1 – some revenues are "pass-through" expenses) gives an estimate of the average benefits of leverage. The quantiles of the average benefits distribution are 40-60% higher than our estimates of marginal cost, foreshadowing that there are not constant returns to leverage: θ will be less than one.

[20] In earlier versions of this paper, we used w_i rather than $\ln w_i$ as the dependent variable. Our main results are nearly identical when we did so; our estimates of the returns to vertical specialization and all quantiles of all earnings distributions are within \$1000 of those we report below.

[21] We have estimated $w(n)$ using various functional forms, and report results with w_i instead of $\ln w_i$ as the dependent variable in Garciano and Hubbard (2007a); unlike in the results we report here, $w'(n)$ is small and positive for the litigation offices in some of these specifications, but as reported above our estimates of the returns to hierarchy hardly change when we use these other functional forms.

III.3. Estimation: θ and the Coordination Cost of Hierarchy

Following equation (5), we derive an 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.[22] Including the field shares on the right side allows the coordination costs of hierarchy to vary across different fields of the law. We also include a polynomial of the number of partners in the office as a regressor. This accounts for the possibility that the coordination costs associated with leverage might be lower for larger offices, for example because larger offices might be able to more effectively utilize associates' time (or perhaps higher if coordination becomes more unwieldy as office size increases).[23] This OLS estimate, while easy to derive, is a biased estimate because $E(\ln \varepsilon_i) \neq 0$. However, the magnitude of this bias is very small relative to the estimates themselves, and we have found that accounting for it implies little change in the results from our counterfactual exercises.[24]

The right side of Table 3 reports our estimates. 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 0.71 with a standard error of 0.007: for a one-partner office consisting only of general practitioners, moving from $n = 0$ to $n = 1$ increases the time to which the partner's knowledge is applied by ($2^{0.71}$-1), or 64%. 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 18%. This estimate varies little with the number of partners in the office. Although the coefficients on the number of partners are jointly statistically significant, they are small in magnitude, and imply that θ decreases from 0.71 to 0.68 for a 50-partner office, then increases back to 0.70 for a 100-partner office.

[22] Revenues per lawyer, here and throughout, equal gross revenues times one minus our estimate of the overhead share of revenues, to account for the fact that gross revenues includes "pass-through" expenses.

[23] It has been suggested to us that we would obtain more efficient estimates by estimating our three equations jointly in a single step. In this case, however, the costs of doing so are high and the efficiency gains are limited because we must estimate the overhead equation using a different sample than the other equations; the covariance in the parameter estimates in the overhead and other equations equals zero. Our work with data from other years confirms that the efficiency gains are small even when the three equations can be estimated using the same sample. The 1977 version of these data include a variable that describes each office's operating expenses. This allows us to jointly estimate all three equations using only our partnership sample. We found that the estimates and standard errors when jointly estimating these equations were almost identical to those we obtained by estimating the system in multiple stages.

[24] The bias arises because, 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 $-\sigma^2/2$. 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.

In contrast we find larger differences in θ across fields. Our estimate of θ is lowest – about 0.50 – for an office with all negligence-plaintiff lawyers, and highest – about 0.87 – for an office with only specialists in corporate law, suggesting that the coordination costs associated with hierarchies tend to be high for the former and low for the latter.[25]

IV. THE RETURNS TO HIERARCHY

IV.1 Productivity

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, but the division of labor was constrained, so that partners and associates do not split work with each other optimally, but instead each works on a representative share of their office's problems, and no collaboration is allowed. What would be the value of the lost production?[26]

Consider this calculation for an office i one partner and n_i associates, referring again to Figure 1. This office's revenues, which are observed in the data, are $TR_i = \widehat{v(z_{mi})}(1+n_i)^{\theta}$. Absent the division labor, the office's revenues would equal $\widehat{v(z_{mi})} + n_i\widehat{v(z_{wi})}$, where $\widehat{v(z_{wi})} = \varepsilon_i v(z_{wi})$. In expectation this quantity is less than $v(z_{mi}) + n_i w_i$, because $w_i > v(z_{wi})$: from revealed preference, 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 $\widehat{v(z_{mi})}(((1+n_i)^{\theta} - 1) - n_i w_i)/(n_i + 1)$. We calculate this quantity for every office in our sample, exploiting the fact that $\widehat{v(z_{mi})} = TR_i/(1+n_i)^{\theta}$ and using our the coefficient estimates in Table 3 to construct an estimate of θ for each office. 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 5 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

[25] Predicted values of θ are within the unit interval for each of the offices in our sample, which indicates that the second order condition for optimization holds for each observation in our sample.

[26] We emphasize that our calculation does not compare equilibrium outcomes with and without hierarchies; if hierarchical production were banned, one would expect clients to adjust to this organizational change by improving their ability to match work to individual lawyers. The productivity effects of such changes in matching are not part of our analysis here, but would offset some of the loss that we calculate.

associates. Average revenues per lawyer in our sample equals \$227,000. We estimate that they would be \$175,000 if the division of labor were arbitrary. This estimate is robust in the sense that changing the estimate of θ by plus or minus one standard error implies changes in this estimate of less than 1%. From Table 5, vertical specialization associated with hierarchies increases productivity in the U.S. legal services industry by at least 30%. 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 58%; the median is 26%. The final column in Table 5 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 \$280,000, 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 30%. 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 4 that the marginal price of leverage is very low for nearly all partners, only \$5,000 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. Even assuming unrealistically that the marginal price of leverage is ten times what we estimate – \$50,000 rather than \$5,000 on average, our estimates imply that hierarchical organization increases productivity by 40% rather than 30%.

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 the estimated returns to leverage in an industry where most entities are low-leverage to begin with implies a very small change in the estimates of the returns to leverage that are in fact achieved.[27]

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

IV.2 Earnings

We next use our estimates of θ to derive estimates of $R_i(z_{mi}, 0) = \widehat{v(z_{mi})} - c_i(1)$ at offices with associates: this is what partners at office i would earn, absent hierarchical production. This differs from $\widehat{v(z_{mi})}$ because it accounts for the costs of operating a zero-associate office. We estimate $\widehat{v(z_{mi})}$ the same way we do in the previous subsection. We estimate $c_i(1) = x_i + oh_i/p_i$, using our data on nonlawyer pay per partner for x_i and the coefficients in the overhead equation to estimate oh_i. We compute quantiles of the distribution of $R_i(z_{mi}, 0)$ across the leveraged partners in our sample and compare them to quantiles associated with our observations of partner pay.

Figure 2 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 \$167,000. Our estimates imply that, absent hierarchical production, the median instead would be \$148,000, about 13% lower. Partner pay is 15-20% higher than $R_i(z_{mi}, 0)$ between the median and the 80th percentile, but is 35% and 50% higher at the 90th and 95th percentile, respectively. Considering only leveraged partners, lawyers' ability to leverage their knowledge through working with associates increases earnings inequality, producing a substantially more skewed earnings distribution. The difference between the 95th percentile and 50th percentile earnings increases from \$208,000 to \$364,000, and the ratio between these two percentiles increases from 2.4 to 3.2.

Figure 3 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 overall pay distribution.

The Figure indicates that, when looking across all lawyers, hierarchical production tends to make earnings distributions more skewed, but this effect is concentrated on the very upper parts of the earnings distribution. The difference between this and the previous Figure reflects the simple fact that well 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 earnings distributions. Our estimates indicate that hierarchical production leaves median earnings unchanged, but increases 95th percentile earnings by 31%. The ratio between the 95th percentile and median earnings increases from 3.7 to 4.8. Hierarchical production makes an already relatively skewed earnings distribution even more skewed. This is even more pronounced if the Figure extended to percentiles greater than the 95th.[28]

Finally, Figure 4 depicts these distributions separately for lawyers in the three classes of offices we defined earlier: "business, non-litigation," "individual, non-litigation," and "litigation" offices. The Figure indicates that hierarchical production has a similar 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 similar among these classes of lawyers,[29] and that hierarchical production amplifies this inequality similarly. In both cases, the 95th percentile of partner pay is close to 60% higher than $R_i(z_{mi}, 0)$, and hierarchical production has a broader-based impact on earnings than that in Figure 3. In contrast, lawyers in "individual, non-litigation" offices look much different; hierarchical production has a very small impact on the earnings distribution. Although lawyers in these offices tend to earn much less than lawyers in the other classes of offices, there is actually more earnings inequality by some measures. In part due to a long lower tail, the ratio between the 95th percentile and the median is 5.6. Absent hierarchical production, this would decrease only marginally, to 5.1. 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. The Figure 3 result that, overall, the impact of hierarchical production on earnings is concentrated on lawyers on the upper tail of the earnings distribution in part reflects that it has little effect on lawyers in this segment, who make up about 25% of privately-practicing lawyers in the U.S.

[28] Census disclosure regulations limit our ability to report results from very high percentiles, because these results would be based on a relatively small number of observations.

[29] 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.

V. 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 by 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 develop an equilibrium model of a hierarchy inspired by Garicano and Rossi-Hansberg (2006) 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 30%, 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 3.7 to 4.8, mostly by increasing substantially the earnings of the very highest percentile lawyers in business and litigation-related segments, and leaving other lawyers' earnings relatively unaffected.

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, 2008). We speculate that the complexity and customization of problem-solving in law firms limits the ability of agents to leverage their human capital: coordination costs are relatively high, as production requires some agent to spend time on each problem and communicating the specifics of an unsolved or new problem is costly. More work is necessary in order to uncover systematic differences in the return to knowledge hierarchies across sectors and to link such differences to the characteristics of the knowledge 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

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

Chiappori, Pierre A, Robert J. McCann and Lars Nesheim, "Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology, and Uniqueness." *Economic Theory,* Forthcoming.

Gabaix, Xavier and Augustin Landier. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, Vol 123, No. 1 (Feburary 2008): 49-100.

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. "The Returns to Knowledge Hierarchies." *NBER Working Paper 12815*, January 2007a.

Garicano, Luis and Thomas N. Hubbard. "Managerial Leverage Is Limited by the Extent of the Market: Theory and Evidence from the Legal Services Industry." *Journal of Law and Economics*, Vol. 50, No. 1 (February 2007b): 1-44.

Garicano, Luis and Thomas N. Hubbard. "Specialization, Firms, and Markets: The Division of Labor Within and Between Law Firms." *Journal of Law, Economics, and Organization*, Vol. 25, No. 2 (2009): 339-371.

Garicano, Luis and Esteban Rossi-Hansberg. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics*, No. 121, Vol. 4 (November 2006), 1383-1436.

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

Legros, Patrick and Andrew Newman. "Monotone Matching in Perfect and Imperfect Worlds." *The Review of Economic Studies*, vol. 69, No. 4 (Oct., 2002): 925-942.

Levin, Jonathan and Steven Tadelis. "Profit Sharing and the Role of Professional Partnerships." *Quarterly Journal of Economics*, Vol. 120, No. 1 (February 2005), 131-171.

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

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

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

Rosen, Sherwin "The Theory of Equalizing Differences." Ch. 12, p. 641-692 in Orley Ashenfelter and R. Layard, eds. *Handbook of Labor Economics*, vol.1, Elsevier, 1987.

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

Table 1

Overhead, Employment, and Revenues

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

Dependent Variable: (Revenues - Payroll)

C	**28.508**	Employment	**2.864**	Revenues	**0.213**	Revenues^2	**-7.61E-06**
	(2.508)		**(0.603)**		**(0.007)**		**(1.80E-06)**
Market Size Dummies		Market Size*Employment Interactions		Field*Revenues Interactions		Field*Revenues^2 Interactions	
20K-100K	-1.586	20K-100K*Employment	0.796	Share(Banking)*Rev	**0.148**	Share(Banking)*Rev^2	**1.26E-05**
	(3.023)		(0.662)		**(0.018)**		**(4.83E-06)**
100K-200K	4.089	100K-200K*Employment	0.984	Share(Corporate)*Rev	**-0.032**	Share(Corporate)*Rev^2	**9.86E-06**
	(3.319)		(0.701)		**(0.016)**		**(2.65E-06)**
200K-400K	**11.098**	200K-400K*Employment	**2.139**	Share(Insurance)*Rev	-0.016	Share(Insurance)*Rev^2	4.19E-06
	(2.809)		**(0.647)**		(0.012)		(2.49E-06)
400K-1M	**7.873**	400K-1M*Employment	**2.279**	Share(Negligence-Def)*Rev	-0.014	Share(Negligence-Def)*Rev^2	4.39E-06
	(2.756)		**(0.657)**		(0.013)		(2.88E-06)
More than 1M	**-20.181**	More than 1M*Employment	**13.896**	Share(Patent)*Rev	**0.121**	Share(Patent)*Rev^2	**5.84E-06**
	(3.032)		**(0.735)**		**(0.022)**		**(2.88E-06)**
				Share(Government)*Rev	-0.001	Share(Government)*Rev^2	9.12E-06
					(0.030)		(8.54E-06)
				Share(Environmental)*Rev	**-0.090**	Share(Environmental)*Rev^2	**1.16E-05**
					(0.039)		**(4.38E-06)**
				Share(Other Field)*Rev	**-0.026**	Share(Other Field)*Rev^2	**5.53E-06**
					(0.009)		**(2.00E-06)**
				Share(Real Estate)*Rev	**0.101**	Share(Real Estate)*Rev^2	1.48E-08
					(0.016)		(4.11E-06)
				Share(Tax)*Rev	**-0.085**	Share(Tax)*Rev^2	**2.17E-05**
					(0.021)		**(5.56E-06)**
				Share(Criminal)*Rev	0.036	Share(Criminal)*Rev^2	-1.20E-05
					(0.025)		(1.16E-05)
				Share(Domestic)*Rev	-0.009	Share(Domestic)*Rev^2	1.05E-05
					(0.024)		(1.26E-05)
				Share(Negligence-Plaintiff)*Rev	**0.038**	Share(Negligence-Plaintiff)*Rev^2	**-1.12E-05**
					(0.011)		**(2.69E-06)**
				Share(Probate)*Rev	-0.008	Share(Probate)*Rev^2	1.67E-05
					(0.028)		(1.19E-05)

R-Squared: 0.70

Specification also includes the (uninteracted) field shares of lawyers in the office. Omitted field category is "share(general practitioner)."

Market size dummies are defined in terms of total employment in the county in which the office is located.

Employment is the total number of individuals (lawyers and non-lawyers) working in the office, minus 2.

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

Table 2

Sample Averages

Partnerships and Proprietorships (N=9283)

	All Offices	Offices With Zero Associates	Offices With Associates
Revenues/Partner	361	142	518
Revenues/Lawyer	203	142	247
Partner Pay	150	57	216
Associate Pay/Partner	36	0	62
Associates/Partner	0.6	0	1.0
Nonlawyers/Lawyer	1.6	1.8	1.5
Nonlawyer Pay/Lawyer	33	24	39
Partners in Office	15	2	24
Business, Non-Litigation Office	0.33	0.23	0.41
Individual, Non-Litigation Office	0.33	0.61	0.14
Litigation Office	0.33	0.16	0.46
N	9283	5319	3964

All dollar figures are reported in thousands of 1992 dollars.
All calculations weight offices by the number of partners.

Table 3
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	**0.336** **(0.010)**
Associates/Partner -- "Business, Non-Litigation Offices"	**0.146** **(0.049)**	Number of Partners	**0.0013** **(0.0004)**
(Associates/Partner)**2 -- "Business, Non-Litigation Offices"	-0.021 (0.013)	Number of Partners**2	-9.30E-07 (2.82E-06)
Associates/Partner -- "Litigation Offices"	0.029 (0.043)		
(Associates/Partner)**2 -- "Litigation Offices"	0.007 (0.010)		
Associates/Partner -- "Individual, Non-Litigation Offices"	0.002 (0.060)		
(Associates/Partner)**2 -- "Individual, Non-Litigation Offices"	-0.026 (0.016)		
Share(Banking Law Specialist)	**0.193** **(0.062)**	Share(Banking Law Specialist)	-0.005 (0.037)
Share(Corporate Law Specialist)	**0.675** **(0.058)**	Share(Corporate Law Specialist)	**-0.201** **(0.029)**
Share(Insurance Law Specialist)	**0.232** **(0.046)**	Share(Insurance Law Specialist)	**-0.077** **(0.021)**
Share(Negligence-Defense Specialist)	**0.263** **(0.048)**	Share(Negligence-Defense Specialist)	**0.066** **(0.021)**
Share(Patent Law Specialist)	**0.413** **(0.055)**	Share(Patent Law Specialist)	**-0.075** **(0.029)**
Share(Government Law Specialist)	**0.548** **(0.070)**	Share(Government Law Specialist)	0.038 (0.044)
Share(Environmental Law Specialist)	**0.517** **(0.104)**	Share(Environmental Law Specialist)	**0.144** **(0.068)**
Share(Real Estate Law Specialist)	**0.375** **(0.049)**	Share(Real Estate Law Specialist)	**-0.141** **(0.035)**
Share(Tax Law Specialist)	**0.603** **(0.107)**	Share(Tax Law Specialist)	0.060 (0.063)
Share(Criminal Law Specialist)	**-0.265** **(0.057)**	Share(Criminal Law Specialist)	0.075 (0.046)
Share(Domestic Law Specialist)	0.082 (0.072)	Share(Domestic Law Specialist)	**0.139** **(0.053)**
Share(Negligence-Plaintiff Specialist)	**0.163** **(0.048)**	Share(Negligence-Plaintiff Specialist)	**0.366** **(0.026)**
Share(Probate Law Specialist)	**0.319** **(0.085)**	Share(Probate Law Specialist)	**-0.119** **(0.057)**
Share(Other Specialist)	**0.252** **(0.029)**	Share(Other Specialist)	0.007 (0.017)
R-Squared	0.61		0.07

The dependent variable in the wage-leverage surface regression is the natural log of average associate pay in the office. Offices with at least one lawyer specializing in insurance or negligence law are classified as "litigation" offices. All other offices are classified as "business" or "individual" depending on whether the majority of their revenues come from individuals. This regression includes county fixed effects as well as the variables above.

The dependent variable in the production function is ln(revenues/lawyer*(1-K))-ln(MC), where K is the derivative of overhead with respect to revenues in the overhead regression for the office, and MC is the estimated marginal cost of leverage for the office. The coefficients reported here correspond to -ln(theta) in the text. The 0.336 coefficient estimate for the constant implies an estimate of theta of 0.714 for an office of general practitioners (the omitted category).

Table 4

Revenues Per Lawyer and the Estimated Marginal Cost of Leverage

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

Percentile	Revenues/ Lawyer	Overhead Share of Revs.	Estimated Avg. Benefits	Estimated MC	Components of Estimated Marginal Cost: Associate Pay	Nonlawyer Pay	Benefits	Overhead	Marginal Price of Leverage
	TR/(n+1)	*(1-K)*	*TR/(n+1)*(1-K)*	*MC*	*w*	*x*	*0.15*(w+x)*	*oh*'/p*	*w'n*
10th	124	0.11	98	69	30	18	8	8	0
25th	173	0.17	137	98	44	26	11	10	1
50th	227	0.20	185	129	61	37	15	12	3
75th	297	0.21	244	173	77	49	19	16	7
90th	372	0.24	320	218	96	61	23	41	10
Mean	247	0.18	204	139	62	39	15	18	5
Relative to Estimated MC	1.78		1.47	1.00	0.45	0.28	0.11	0.13	0.04

Revenue and cost figures are in thousands of 1992 dollars.

Table 5
The Returns to Vertical Specialization
Partnerships and Proprietorships (N=9283)

	Revenues/Lawyer		
	Absent Vertical Specialization (upper bound)	Actual	Constant Returns to Leverage
	[zm+nw]/(n+1)	*TR/(n+1)*	*zm*
Percentile			
10th	76	83	89
25th	115	139	150
50th	165	209	249
75th	216	288	362
90th	273	374	494
Mean	175	227	280
Mean, Relative to Absent Returns to Specialization	1.00	1.30	1.60

All figures are reported in thousands of 1992 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
Comparison of Earnings Distributions Using Actual Data and Estimates

Variable	Actual Earnings	Estimated Earnings
Percentile		
10th	29	20
20th	40	39
30th	51	53
40th	61	66
50th	74	77
60th	88	94
70th		121
80th		168
90th		257
Sample	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 1992 dollars.

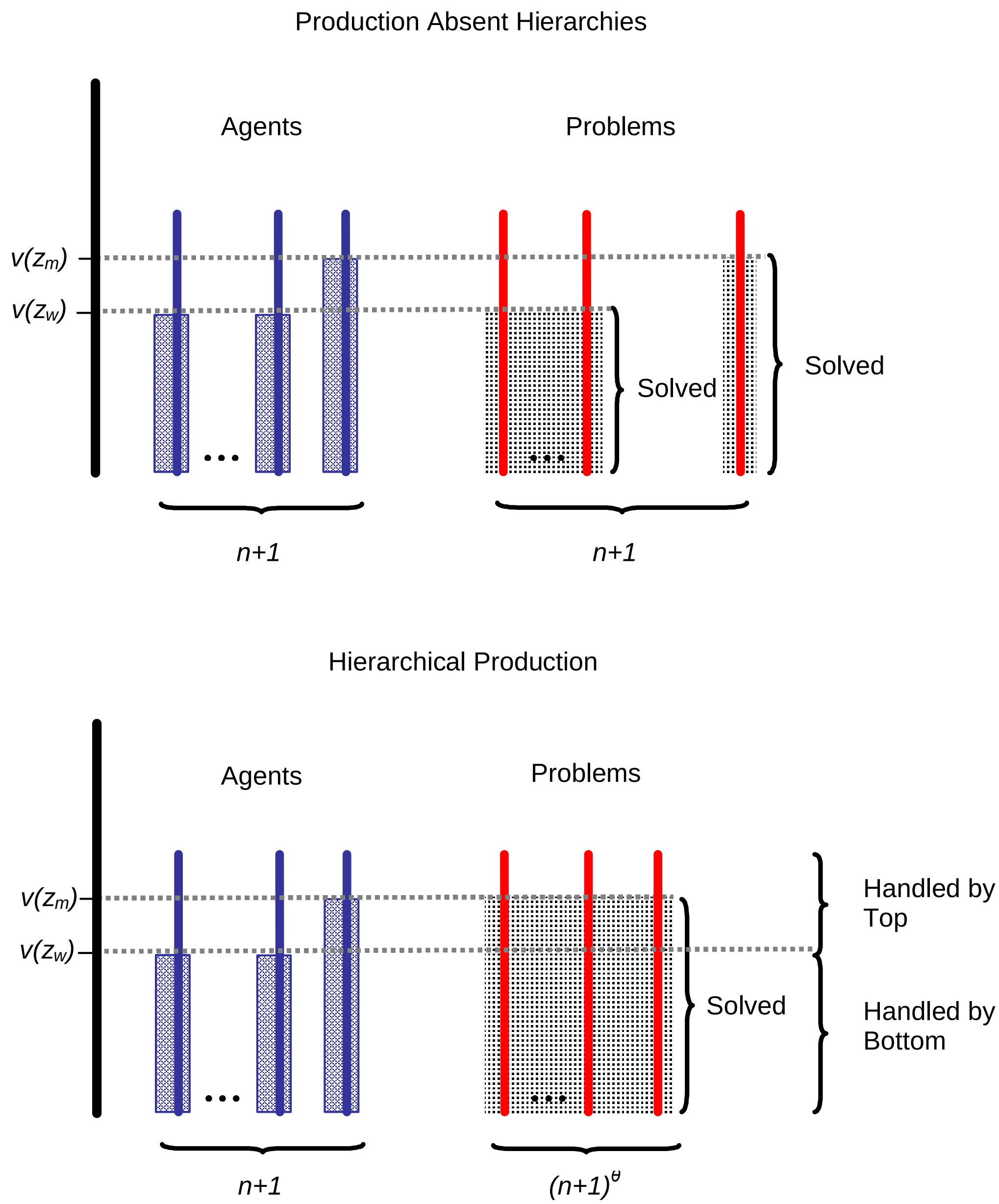

Figure 1. Non-Hierarchical and Hierarchical Production. The top panel depicts production absent hierarchies; sets of problems are allocated to lawyers arbitrarily and each lawyer applies their time and knowledge toward whatever set they confront. Output is $v(z_m) + nv(z_w)$. The bottom panel depicts 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 $v(z_m)(n + 1)^{\theta}$.

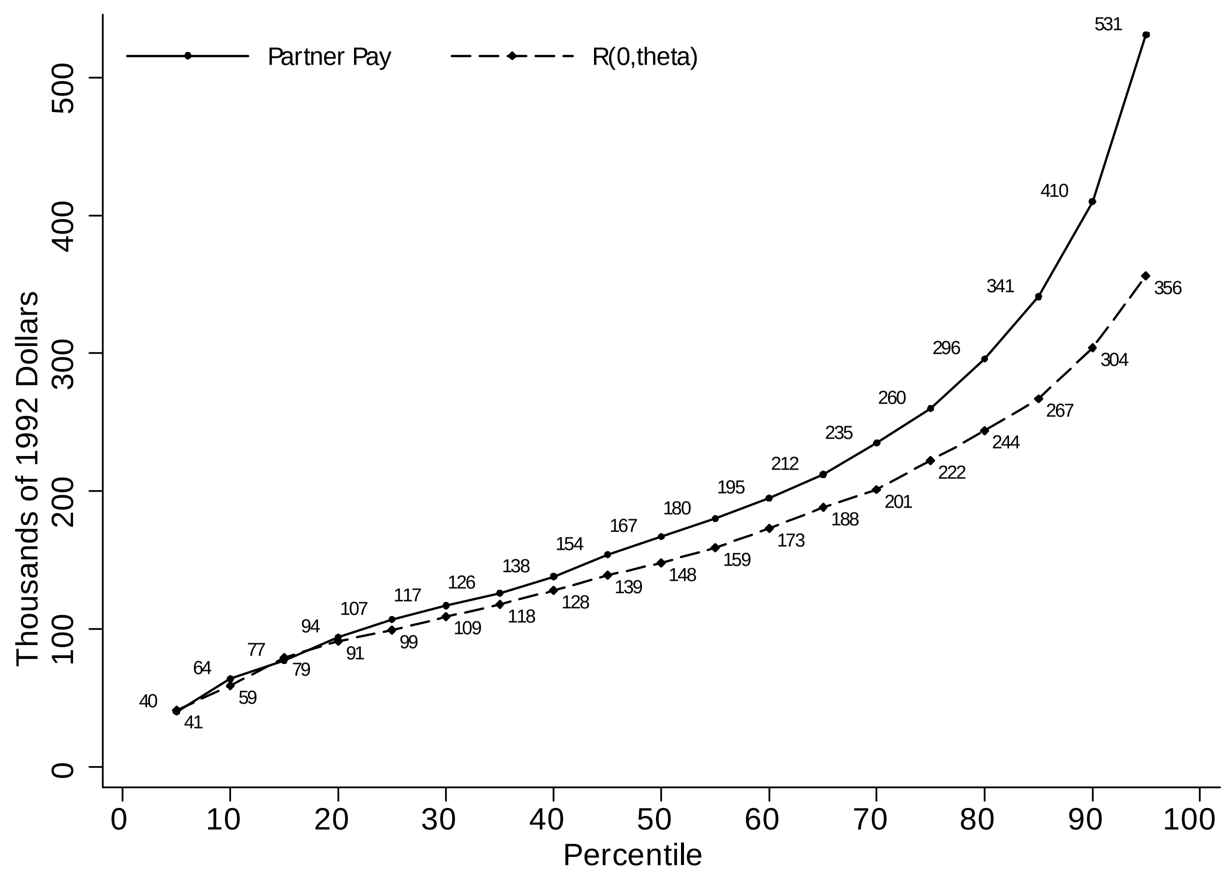

Figure 2. The Distribution of Partner Pay, Estimated Partner Pay Absent Hierarchies. This Figure reports 20 quantiles of the distribution of these quantities, looking only at partners at offices with associates. "Estimated pay absent hierarchies" is $R_i(z_{mi}, 0)$.

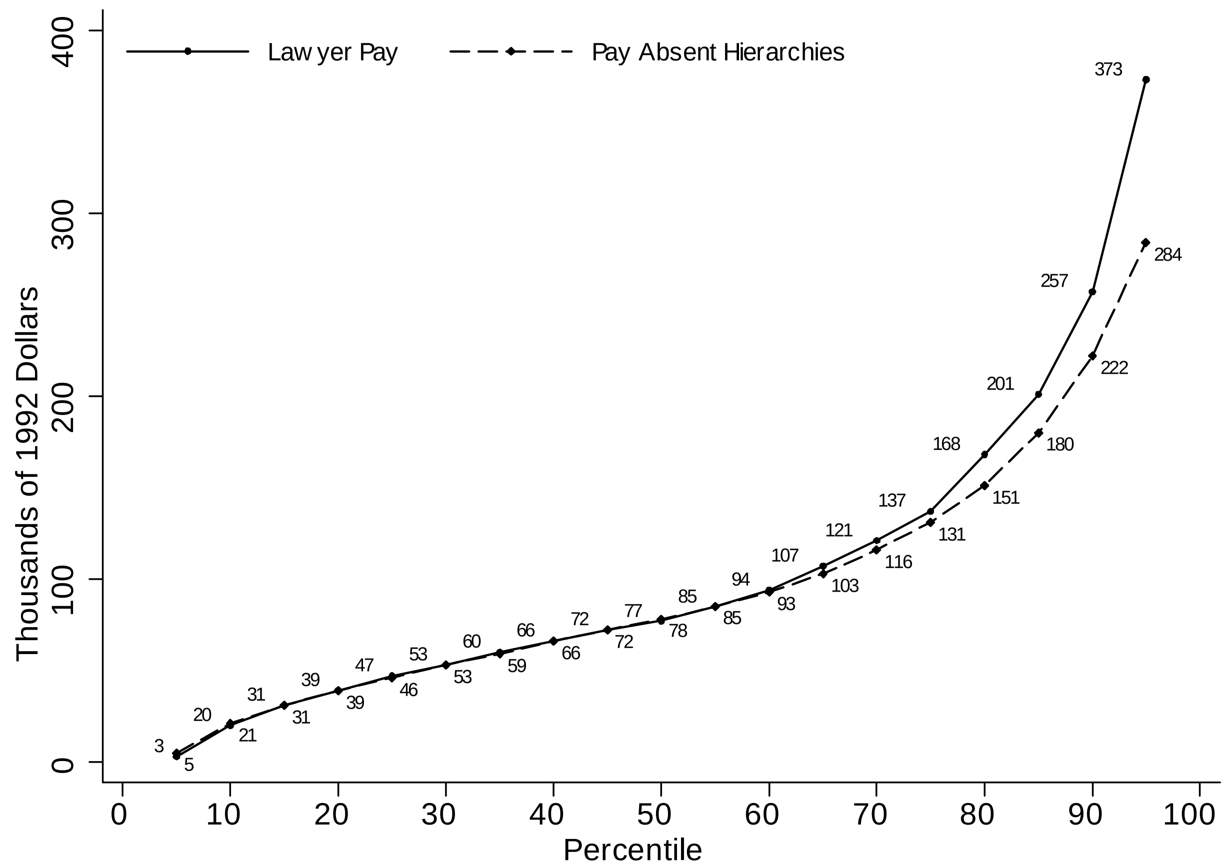

Figure 3. The Distribution of Lawyer Pay, Estimated Pay Absent Hierarchies. This Figure reports 20 quantiles of the distribution of these quantities. "Estimated pay absent hierarchies" is $R_i(z_m, 0)$ for partners at offices with associates. It is the same as lawyer pay for partners at offices without associates, as well as for associates. Because associates earn more as associates than they would absent hierarchies ($w(z) > R_i(z, 0)$), this overstates what these individuals would earn in this counterfactual. This upward bias primarily affects our estimates of lower quantiles.

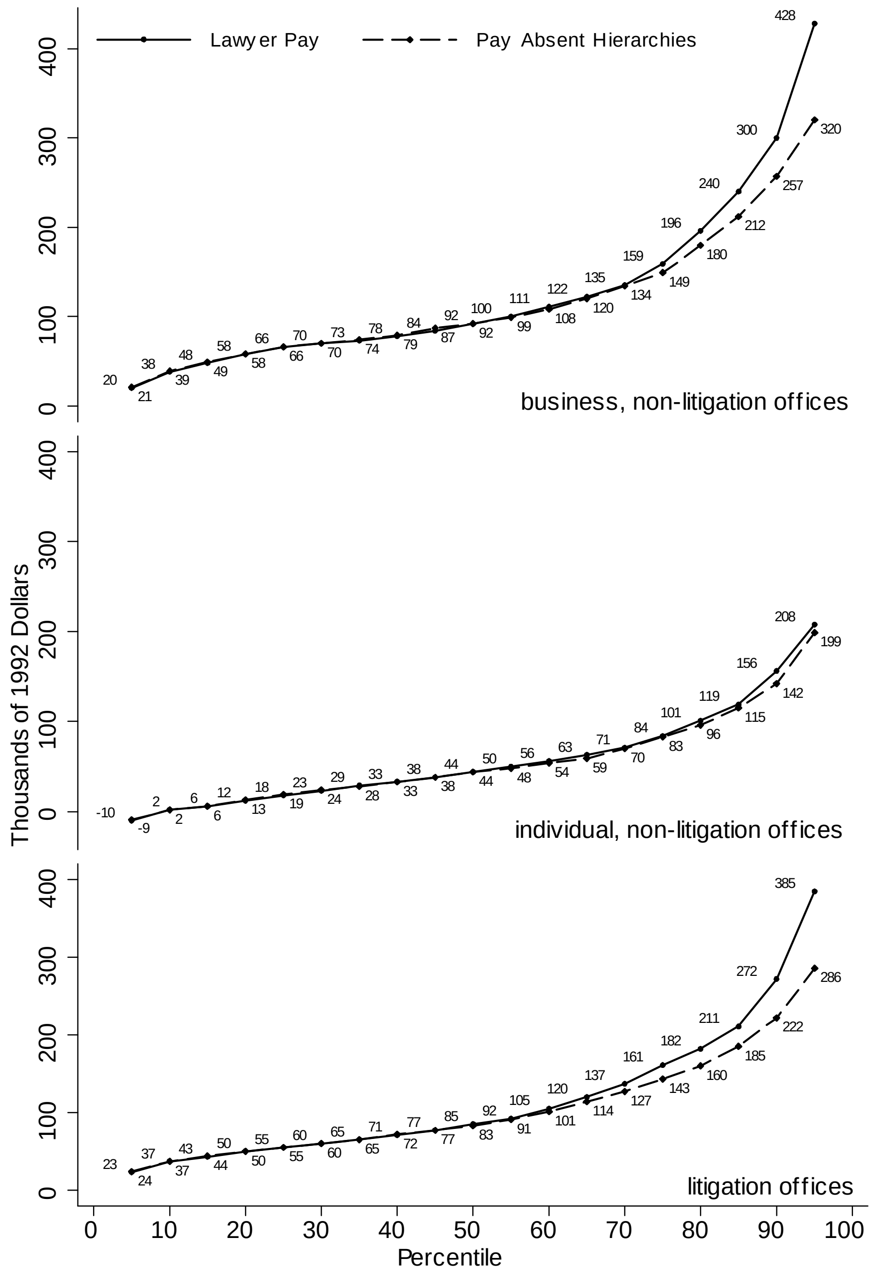

Figure 4. The Distribution of Lawyer Pay, Estimated Pay Absent Hierarchies, by Office Class. This Figure reports 20 quantiles of the distribution of these quantities for three classes of offices. "Estimated pay absent hierarchies" is $R_i(z_m, 0)$ for partners at offices with associates. It is the same 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 ($w(z) > R_i(z, 0)$), this overstates what these individuals would earn in this counterfactual. This upward bias primarily affects our estimates of lower quantiles.