Promotion, Turnover and Compensation in the Executive Market*

George-Levi Gayle, Limor Golan, Robert A. Miller
Tepper School of Business, Carnegie Mellon University

October 2009

Abstract

This paper is an empirical study of the market for managers, more specifically the effects of agency, human capital, and preferences on their promotion, tenure, turnover and compensation. From a large longitudinal data set compiled from observations on executives and their publicly listed firms, we construct a career hierarchy and report on its main features. Our summary results motivate a dynamic competitive equilibrium model, whose parameters we identify and estimate. Controlling for heterogeneity amongst firms, which differ by size and sector, and also managers, whose backgrounds vary by age, gender and education, our estimates are used to evaluate how important moral hazard and job experience are in jointly determining promotion rates, turnover and compensation.

1 Introduction

Chief executives are paid more than their subordinates, and internal promotions with the firm are positively correlated with wage growth.[1] Since high ranking executives are almost always drawn from the lower ranks, usually from within the firm, it is tempting to conclude that part of the reward from working hard in a low rank is the chance of promotion to earn rents. Theory provides several possible explanations, ranging from human capital acquired on lower level job, to superior ability being revealed with experience leading to wage dispersion, or as the prize in a tournament played by lower ranked executives to induce hard work.[2] The premise of all these explanations is the commonly held opinion that the CEO is better off than those he supervises. Yet several studies, conducted with data on executive compensation and returns from publicly traded firms, show quite conclusively that CEO compensation is more sensitive to the excess returns of firms than the

*We thank the participants of the 2008 World Congress on National Accounts and Economic Performance Measures for Nations 2008, Society of Labor Economists 2009, Econometric Society Summer meeting 2009, Cowles 2009, Society of Economic Dynamics 2009, NBER 2009 Summer Institute, CRES Applied Economics conference, the seminar participants in Carnegie Mellon, Stony Brook, Wisconsin, UCLA for comments and suggestions. This research is supported by the Center for Organizational Learning, Innovation and Performance in Carnegie Mellon University and National Science Foundation Grant Award SES0721098. Preliminary and Incomplete.

[1] See Lazear (1992), Baker, Gibbs and Holmstrom (1994a), McCue (1996)

[2] See Prendergast (1999), Gibbons and Waldman (1999) and Neal and Rosen (2000) for surveys.

compensation of lower ranked executives.[3] Thus at the upper levels of the career ladder, differently ranked jobs do not have the same characteristics. Whether one job is more desirable than another depends on the probability distribution of financial compensation that generates his income, as well as its nonpecuniary costs and benefits.

To the best of our knowledge, no one has attempted to quantify how much a CEO receives as a rent from human capital in management and leadership, and how much he is compensated for receiving a more volatile income. A small but growing literature on the structural estimation of moral hazard models investigates the empirical relationship between the principal's return and the agent's compensation, in order to quantify how incentives are used for inducing agents to work in the interests of their principals and truthfully revealing their hidden information.[4] These studies find that estimates of the higher risk premium necessary to compensate a CEO for a more uncertain income relative to the second in command are of the same order of magnitude as differences in expected compensation. Such findings do not resonate with common opinion, because they imply the CEO receives very little pecuniary rent from his promotion to that position. Published work does not, however, integrate human capital and its behavioral consequences into an optimal contracting framework, confounding any attempt to gauge the degree of on-the-job training provided at lower ranks relative to the nonpecuniary value of holding a job at any given rank. More generally, the empirical importance of human capital in the executive labor market, and the role of promotions in this process, is unclear.[5]

This paper is an empirical study of the effects of incentives, human capital, and preferences of managers, with goal of explaining the differences in the promotion, tenure, job turnover and compensation structure across managers using a dynamic competitive equilibrium model. Our data contain background information on executives, including age, gender, education, executive experience and the types of firms they work for, plus detailed information on their compensation and the financial returns of their firms and their rank within a career hierarchy. We identify and estimate a dynamic equilibrium model to analyze and disentangle the effects of competition in the market for managers using data on internal promotions, job turnover and the compensation of executives. estimate. Controlling for heterogeneity amongst firms and managers, our estimates are used to evaluate how important moral hazard and job experience are in jointly determining promotion rates, turnover and compensation.

The model is set up in the next two sections. The next section lays out the model. Executives choose job, firm and effort level every period. They have preferences over jobs, particularly, effort is costly. These taste parameters vary across jobs and firms. In addition, every period managers privately observe a firm-job specific taste shock. The effort level is private information as well. While working they accumulate firm-specific and general human capital. We assume human capital accumulation on a job is greater when the manager exerts effort. The rate of human capital accumulation varies across jobs and firm as well, therefore, working in some firms and jobs may increase the manager's stock of human capital. Firms offer contracts which provide incentives for managers to

[3] See Margiotta and Miller (2000) and Gayle and Miller (2008a, 2008b).

[4] Ferrall and Shearer (1999), Margiotta and Miller (2000), Dubois and Vukina (2005), Bajary and Khwaja (2006), Duflo, Hanna, and Ryan (2007), D'Haultfoeviller and Fevrier (2007), Einav, Finkelstein and Schrimpf (2007), Nekipelov (2007), Gayle and Miller (2008a,b,c).

[5] Frydman (2005) finds evidence on the increase importance of general skills in executive compensation.

exert effort. Because exerting effort increases the manager's stock of human capital, future promotion prospects provide incentives.[6] Thus, variation in compensation across firms and jobs partially reflect the different opportunities to accumulate human capital and different promotion prospects. In addition, managers' age and rank imply differences in career concerns affecting the optimal compensation schemes. Section 3 analyzes the optimization problems managers solve in equilibrium, namely their consumption, job choices and work effort each period, and analyzes the sequentially optimal contract we focus upon and the resulting equilibrium. The markets for executives is competitive, managers with different stocks of human capital and compensation adjusting to clear the market for each skill set.

Identification of the parameters of the model is analyzed in Section 4, and in that section we also lay out our estimation and testing strategy. Our data is described in Section 5, where we define the job hierarchy and wage compensation. Our measure of compensation is comprehensive, and includes salary and bonus, stock and option grants, retirement benefits, as well as income directly attributable to holding securities in the firm in lieu of a widely diversified portfolio. The compensation data is augmented with data on the titles of the executives, along with their professional and demographic background compiled from the Marquis "Who's Who" . Compensation of the executives are sensitive to fluctuations in the abnormal returns. In fact, the firm's excess return (over and above the market's return) is the most important determinant of managerial compensation, suggesting the importance of incentives and moral hazard. We find that in fact the higher the executive's rank in the firm, the more sensitive his compensation to the abnormal return. We also find that firm turnover is positively correlated with promotions and higher compensation.

Some preliminary estimates from the structural estimation are reported in the final section. We used four metrics to assess how much agency problems in executive markets are mitigated by their career concerns. Two of these measure the impact of an executive shirking rather than working, while the other two focus on the cost of eliminating the moral hazard problem. We find that firms are prepared to pay hardly anything to eliminate the moral hazard problem at the lower ranks, but that at the upper levels, the risk premium paid to executives for accepting an uncertain income stream that depends on the firm's abnormal returns, are considerably greater. Career concerns greatly ameliorate the moral hazard problem for lower level executives, but their importance declines monotonically with promotion through the ranks. Overall our empirical findings, based on a large sample of executives employed by a broad cross section of publicly traded firms, demonstrate that the design of the hierarchy and the promotion process are important tools, used in conjunction with compensation schemes, for disciplining employees and aligning their interests to the goals of the organization.

2 The Model

Our model analyzes promotion, turnover and executive compensation, where expected value maximizing shareholders are subject to moral hazard from choices made by their risk averse expected utility maximizing executives, who are more informed than their

[6] Gibbons and Murphy (1992) develop and empirically test a model of optimal contracts in the presence of career concerns in the market for CEOs.

employers about the value of their job matches. Executives earn returns from investment in human capital by gaining seniority within a position, from internal promotion, and turnover to other firms. These three factors, rooted in the technology of learning on the job, may induce them to trade off higher current income for better future prospects as their career opportunities unfold. Behavior on the job is also affected by these three factors, as well as the compensation schedule, which depends on signals shareholders receive about managerial performance. Designed to align the goals of the firm with the executive, this variability induces a risk premium. We derive a competitive equilibrium where the optimal contracts of shareholders guide managerial decisions on job choice and effort on the job.

At the beginning of every period, equity returns of firms from decisions made in the previous period are revealed to everyone, the human capital state variables of executives are updated, and each executive is compensated by following the schedule of the previous period's employment contract. Firms assess their demand for executives in the current period and advertise for executives internally and externally, by posting one-period contracts for positions within their firms. Then executives privately observe realizations of preference shocks and choose their consumption. They accept their most attractive employment offer, or quit management, and markets clear. Finally each executive chooses an effort level, a choice that is concealed from everyone else but nevertheless affect both his utility and the distribution of the returns of his firm realized at the beginning of the next period. Given the employment contracts offered by potential employers, executives sequentially maximize expected lifetime utility with respect to consumption, employment and effort level. This section develops the model.

2.1 Choices, Human Capital and Preferences

There are a finite number of firm types in the market indexed by $j \in \{1, ..., J\}$, with $j = 0$ representing retirement. There are K different types of positions within each firm type j, indexed by $k \in \{1, ..., K\}$ and ranked in hierarchical order. Let $t \in \{0, 1, ...\}$ denote the executive's age, let $d_{jkt} \in \{0, 1\}$ indicate the manger's job, his rank k in firm j at age t, and let d_{0t} denote the indicator variable for retirement, which is an absorbing state. The $JK + 1$ choices are mutually exclusive, implying:

$$d_{0t} + \sum_{j=1}^{J} \sum_{k=1}^{K} d_{jkt} = 1$$

for all ages $t \in \{0, 1, ...\}$ preceding retirement, which occurs upon reaching or before age $T < \infty$. Summarizing, $d_t \equiv (d_{0t}, d_{11t}, \ldots, d_{JKt})$ denotes the vector of job and rank choices an executive makes at age t. There are two activities within the firm, called working and shirking, denoted by $l_t \in \{0, 1\}$,where $l_t \equiv 0$ means the manager shirks at age t and $l_t \equiv 1$ means the manager works. Only the manager observes his own effort.

The background of the executive is defined by his age t, and his human capital, denoted by the vector h_t, sequentially determined by his choices. Given his age t choices of effort level l_t in the k^{th} rank at the j^{th} firm, his human capital at the beginning of the next period is determined by the mapping:

$$h_{t+1} \equiv H_{jk}(h_t)\, l_t + H'_{jk}(h_t)\,(1 - l_t)$$

where h_0 represents the initial endowment of the executive (such as fixed demographic characteristics such as gender and education). This specification encompasses three dimensions of how human capital is accumulated. The first relates to where it can be acquired, for example in lower ranks versus higher. The second dimension relates to where it might apply, such as to all firms versus only firms belonging to the same industry. Firm specific experience at rank k for example, $\sum_{s=0}^{t} d_{jkt-s}$, might increase productivity in firm j at rank k' more than elsewhere. The third dimension is who observes an executive's human capital. In our model some attributes can be only directly observed by the executive, such as accumulated effort $\sum_{k=1}^{K} \sum_{s=0}^{t} d_{jkt-s} l_{t-s}$, whereas other attributes are observed by everyone, such as total executive experience $\sum_{k=1}^{K} \sum_{s=0}^{t} d_{jkt-s}$.

Executives are infinitely lived, and their preferences are characterized by the discounted sum of a time additively separable constant absolute risk aversion utility function, which is multiplicative in consumption and nonpecuniary factors. Human capital affects both the productivity of the firm, as discussed below, and also enter preferences directly, through the ease with which tasks are accomplished. Thus the preference parameters of a manager depend on his employer and rank (j, k), his effort level l, and his background (t, h). Utility from consumption, which is exponential, is scaled by a weight that depends on the executive's work choices, that is relative to retirement which is normalized to one. Diligent work is scaled by the factor $\alpha_{jkt}(h)$ and shirking by $\beta_{jkt}(h)$. We assume there is more disutility from working than shirking or, noting that exponential utility is negative, $\alpha_{jkt}(h_t) > \beta_{jkt}(h_t)$ for all h_t. An individual taste shock indexed by firm, position and time, also affects current utility. Denote this shock by ε_{jkt} if workplace position (j, k) is selected, and by ε_{0t} if the executive retires. For notational convenience we assume, that if the executive retires in period t, then $\varepsilon_{0s} \equiv 0$ for all $s > t$. Thus life-time utility can be summarized as:

$$-\sum_{t=1}^{\infty} \delta^t \exp(-\rho c_t) \left\{ \begin{array}{l} d_{0t} \exp(-\varepsilon_{0t}) \\ + \sum_{j=1}^{J} \sum_{k=1}^{K} d_{jkt} \left[\alpha_{jkt}(h_t) l_t + \beta_{jkt}(h_t)(1 - l_t) \right] \exp(-\varepsilon_{jkt}) \end{array} \right\} \tag{1}$$

where δ is the subjective discount factor, ρ is the constant absolute risk aversion parameter, and if $d_{0t} = 1$ then $d_{0s} = 1$ for all $s > t$.

At date τ suppose the executive reaches age t. We assume there exists a complete set of markets for all publicly disclosed events relating to commodities with price measure Λ_τ and derivative λ_τ at date τ. This implies that consumption by the manager is limited by a lifetime budget constraint, which reflects the opportunities he faces as a trader and the expectations he has about future compensation. The lifetime wealth constraint is endogenously determined by the manager's work activities. By assuming markets exist for consumption contingent on any public event, we effectively attribute all deviations from the law of one price to the particular market imperfections under consideration. Let e_t denote his endowment at age t. We also measure $w_{jk,t+1}$, the manager's compensation for employment in position k at firm type j at the beginning of year $\tau + 1$, in units of current consumption. To indicate the dependence of the consumption possibility set on the set of contingent plans determining labor supply and effort, we define $E_t[\bullet | l_t, d_t, h_t]$ as the expectations operator conditional on work and effort level choices at age t, the subscript on the operator indicating shocks in the commodities market. The budget constraint can then be expressed as:

$$E_t\left[\lambda_{\tau+1}e_{t+1}|l_t, d_{jkt}, h_t\right] + \lambda_\tau c_t \leq \lambda_\tau e_t + E_t\left[\lambda_{\tau+1}w_{jkt+1}|l_t, d_{jkt}, h_t\right] \tag{2}$$

where $\tau(t)$ is the date when the executive reaches age t.[7]

2.2 Firms

We assume that the value of executive work to the firm is additive, an assumption of convenience that suppresses the role of teamwork and organizational capital. Specifically we assume that in period τ an executive aged $t(\tau)$

3 Optimization and Equilibrium

In our model each executive chooses his consumption stream as his income accrues from a sequence of lotteries with prizes of $w_{jk,t+1}$ at each age $t+1$ until retirement. During his working life he also receives nonpecuniary (scaled utility) benefits from participating of $\;\alpha_{jkt}(h_t)+\beta_{jkt}(h_t)\;\exp\;\;-\varepsilon^*_{jkt}\;\;$ where ε^*_{jkt} is the value of the disturbance when lottery (j,k) is selected at age t. Shareholders assess their demand for executives in the current period and post one-period contracts for positions within their firms to maximize their expected return from executive employment subject to two constraints. Driven by their demand for executive managers, firms set target acceptance rates for all positions and all executive backgrounds in the competitive equilibrium, thus satisfying a participation constraint. Every contract also satisfies an incentive compatibility constraint that induces each executive to work diligently. In competitive equilibrium each firm (type) $j \in \{1,\ldots,J\}$ fills its positions $k \in \{1,\ldots,K\}$ with managers with age/skill types (t,h_k) and at compensation levels $w_{jk,t+1}(h_k)$ that are at least as profitable for the firm as any alternative. Every executive choose his most desirable position and work effort given the menu of contracts offered by all firms to those endowed with his skills. Entry into the market for executives dissipates the rents from contracting with them to zero.

This section solves the stochastic sequence of executive's consumption and savings choices as a function of the compensation schemes offered by different firms, and lays out the dynamic discrete choice problem associated with workplace, rank and effort. Then we show what restrictions must be placed on contracts to induce executives to select their equilibrium work assignments and to prefer working diligently rather than shirking along the equilibrium path. Finally we derive the optimal one period contract, establish the existence of a unique competitive equilibrium and describe its properties.

3.1 Consumption and Saving

Consumption and savings at any given age depends on the manager's wealth and his career prospects, which encapsulate the value of future earnings. These in turn are determined by the opportunity cost of future consumption, compensation schedules for different jobs, and the probability of holding any given position. Writing as $\tau(t)$ the calendar date when the manager is t years old, we denote by $b_{\tau(t)}$ the period τ price of a infinitely lived bond, and define:

$$v_{jk,t+1} \equiv \exp\;\;-\rho w_{jk,t+1}/b_{\tau(t)+1} \tag{3}$$

as the risk adjusted utility weight for receiving compensation $w_{jk,t+1}$ at the beginning of period $\tau+1$ for working (j,k) in period τ if the manager is t years old. We denote by $p_{jkt}\;\;h, b_{\tau(t)}\;\;$ the probability that, if constrained to work shun shirking, the manager would optimally select job (j,k) at age t in period $\tau(t)$ given characteristics h and bond price $b_{\tau(t)}$. Similarly we denote by $p_{0t}\;\;h, b_{\tau(t)}\;\;$ the probability that the manager retires at age t, and note that $p_{0T}\;\;h, b_{\tau(T)}\;\;=1$.

We prove below that the manager's optimal consumption is additively separable in current wealth and human capital. To establish this result for the case in which the manager is always diligent at work, define the age of retirement $R \leq T$ by $d_{0R}=1$, set $A_t\;\;h, b_{\tau(t)}\;\;\equiv 1$ for all $t \geq R$, and for all $t \in \{1,\ldots,R-1\}$ recursively define $A_t\;\;h, b_{\tau(t)}$

for each (t,h) as:

$$
\begin{aligned}
A_t\left(h,b_{\tau(t)}\right) &= p_{0t}\left(h,b_{\tau(t)}\right) E\left[\exp\left(-\varepsilon_{0t}^{*}/b_{\tau(t)}\right)\right] \\
&\quad + \sum_{j=1}^{J}\sum_{k=1}^{K}\left\{\begin{array}{l} p_{jkt}\left(h,b_{\tau(t)}\right)\left[\alpha_{jkt}(h)\right]^{\frac{1}{b_{\tau(t)}}} E\left[\exp\left(-\varepsilon_{jkt}^{*}/b_{\tau(t)}\right)\right] \\ \times E_t\left[\upsilon_{jk,t+1}A_{t+1}\left(H_{jk}(h),b_{\tau(t)+1}\right)\right]^{1-\frac{1}{b_{\tau(t)}}} \end{array}\right\}
\end{aligned}
\tag{4}
$$

Let c_t^o denote the optimal consumption at age t, and define the value function as:

$$
\begin{aligned}
V_t\left(h,a_{\tau(t)},b_{\tau(t)}\right) &= -E_t\left\{\sum_{s=t+1}^{R}\sum_{j=1}^{J}\sum_{k=1}^{K}\delta^{s-t}\exp\left(-\rho c_s^o\right) p_{jks}\left(h_s,b_{\tau(s)}\right)\alpha_{jks}(h_s)\exp\left(-\varepsilon_{jks}\right)\right\} \\
&\quad -E_t\left\{\sum_{s=t+1}^{R}\delta^{s-t}\exp\left(-\rho c_s^o\right)p_{0s}\left(h_s,b_{\tau(s)}\right)\exp\left(-\varepsilon_{0s}\right)\right\} \\
&\quad -E_t\left\{\sum_{s=R+1}^{\infty}\delta^{s-t}\exp\left(-\rho c_s^o\right)\right\}
\end{aligned}
$$

In general, current consumption depends on all the state contingent prices, but for the exponential utility specialization, just two securities suffice, bond prices, and a_τ, the price of a security that pays off the (random) dividend $(\ln\lambda_{\tau+s} - (\tau+s)\ln\delta - \ln\lambda_{\tau+s})$. Lemma 1 solves the value function and optimal consumption in terms of the prices of financial securities $\left(a_{\tau(t)}, b_{\tau(t)}\right)$, the manager's wealth e_t, and the mapping $A_t\left(h,b_{\tau(t)}\right)$, which we now interpret as the scaled util value of human capital of a manager with background and skills (t,h).

Lemma 1

$$
V_t\left(h,a_{\tau(t)},b_{\tau(t)}\right) = -A_t\left(h,b_{\tau(t)}\right) b_{\tau(t)}\exp\left(-\frac{a_{\tau(t)}+\rho e_t}{b_{\tau(t)}}\right)
$$

If a manager with background and skills h selects position (j,k) when he is t years old, and is constrained to be diligent for the rest of his working life, then his optimal consumption is determined by:

$$
\begin{aligned}
c_t^o &= \frac{e_t}{b_{\tau(t)}} + \frac{t}{\rho}\ln\delta + \frac{a_{\tau(t)}}{\rho b_{\tau(t)}} - \left(\frac{b_{\tau(t)}-1}{\rho b_t}\right)\varepsilon_{jkt}^{*}\ln\alpha_{jkt}(h) \\
&\quad -\rho^{-1}\ln E_t\left[\upsilon_{jk,t+1}A_{t+1}\left(H_{jk}(h),b_{\tau(t)+1}\right)\right]
\end{aligned}
\tag{5}
$$

Note that $b_{\tau(t)}\exp\left[-\left(a_{\tau(t)}+\rho e_t\right)/b_{\tau(t)}\right]$ is the well known formula for the valuation function associated with exponential utility, that in this model applies to a retired manager. Thus $A_t\left(h,b_{\tau(t)}\right)$ is a weight, on the valuation function for a retiree, that values the working executive's career prospects given his human capital h, age t, and bond price $b_{\tau(t)}$. By inspection the index $A_t\left(h,b_{\tau(t)}\right)$ takes only strictly positive values. From the formula for $V_t\left(h,a_{\tau(t)},b_{\tau(t)}\right)$, lower values of $A_t\left(h,b_{\tau(t)}\right)$ are associated with a higher investment value and a higher valuation function.

The first three terms of Equation (5), the formula for optimal consumption, are familiar, spending the interest on the endowment, $e_t/b_{\tau(t)}$, discounting consumption over time due to impatience, $\rho^{-1}t\log\delta$, and adjusting for aggregate risk, $a_{\tau(t)}/\left(\rho b_{\tau(t)}\right)$. The next term depends on the effects of nonpecuniary features of the job on the marginal utility of consumption, that which is publicly observed, $\alpha_{jkt}(h)$, and the hidden component ε^*_{jkt}. Since both components are specific to rank and firm, so is optimal consumption. Finally current consumption also depends on the processes determining $\upsilon_{jk,t+1}$, compensation due next period, and the value of human capital at the beginning of next period, $A_{t+1}\left(H_{jk}(h), b_{\tau(t)+1}\right)$.

3.2 Job Choice and Work Effort

The supply of executives is determined by the job choices they make over their career. Given the optimal rule for consumption and diligent effort, the optimal position is found by selecting rank k in firm j at age t given human capital h_t and private value ε^*_{jkt} to sequentially maximize the sum of current utility, that is $\alpha_{jkt}(h)\exp(-\rho c^o_t-\varepsilon_{jkt})$ or $\exp(-\rho c^o_t-\varepsilon_{0t})$ in the case of retirement, plus the one period discounted expected value of future optimized utility, $\delta E_t\left[V_{t+1}\left(H_{jk}(h_t), a_{\tau(t)}, b_{\tau(t)}\right)\right]$ or $\delta E_t\left[V_{t+1}\left(1, a_{\tau(t)}, b_{\tau(t)}\right)\right]$. Substituting in the formulas for optimal consumption and the valuation function reduce the job choice problem to the following formulation.

Lemma 2 *If $d_{0t}=0$ for all $s\in\{0,\ldots,t-1\}$ and any $t\in\{0,\ldots,T\}$, then the optimal job choice indicators d_t are picked to maximize:*

$$\varepsilon_{0t}d_{0t}+\sum_{j=1}^{J}\sum_{k=1}^{K} d_{jkt}\left[\varepsilon_{jkt}-\ln\alpha_{jkt}(h)-\left(b_{\tau(t)}-1\right)\ln E_t[\upsilon_{jk,t+1}A_{t+1}\left(H_{jk}(h), b_{\tau(t)+1}\right)]\right]$$

In our framework the support of $\pi_{j\tau}$ does not depend on the effort choice, and executives work diligently in equilibrium. Consequently the action of shirking is not detected by shareholders, since all shirking outcomes that shareholders observe can be rationalized by executive working hard. Similarly, since ε_{jkt} is private information and has full support, job choices made at the beginning of each period can be rationalized by a history of always working diligently. So regardless of what the manager chooses, and more generally what outcomes shareholders observe, they update their beliefs of his human capital, which we denote by h'_t, as if he never strayed from the equilibrium path. Therefore the law of motion of h'_t is given by $h'_{t+1}\equiv H_{jk}(h'_t)$.

Consider an executive who, after deviating from equilibrium, accumulated human capital of h when shareholders believe he has h'. His conditional choice probabilities now depend on both h and h', because he knows he has h human capital, which stochastically determines his true productivity, yet he is paid as if he has h'. We denote them by $p_{jkt}\left(h, h', b_{\tau(t)}\right)$. Given job choice (j,k), the executive's state variables are then updated using the formula $h_{t+1}\equiv H_{jk}(h_t)$ if he works diligently and $H'_{jk}(h_t)$ if he shirks. Working diligently gives him $\alpha_{jkt}(h)$ disutility and abnormal firm returns are drawn from the $f_j(\pi)$ density, while shirking gives him $\beta_{jkt}(h)$ disutility but abnormal firm returns are drawn from the $f_j(\pi)g_{jkt}(\pi|h)$ density. Analogously to the definition of $A_t\left(h, b_{\tau(t)}\right)$ we define the recursion:

$$B_t\left(h,h',b_{\tau(t)}\right)=p_{0t}\left(h,h',b_{\tau(t)}\right)E\left[\exp\left(-\varepsilon^*_{0t}/b_{\tau(t)}\right)\right] \tag{6}$$
$$+\sum_{(j,k)}\left[\begin{array}{l} p_{jkt}\left(h,h',b_{\tau(t)}\right)E[\exp\left(-\varepsilon^*_{jkt}/b_{\tau(t)}\right)] \\ \times\max\left\{\begin{array}{l}\alpha_{jkt}(h)^{\frac{1}{b_{\tau(t)}}}E_t\left[\upsilon_{jk,t+1}B_{t+1}\left(H_{jk}(h),H_{jk}(h'),b_{\tau(t)+1}\right)\right]^{1-\frac{1}{b_{\tau(t)}}}, \\ \beta_{jkt}(h)^{\frac{1}{b_{\tau(t)}}}E_t\left[\upsilon_{jk,t+1}g\left(\pi_{j,\tau(t)+1},h\right)B_{t+1}\left(H'_{jk}(h),H_{jk}(h'),b_{\tau(t)+1}\right)\right]^{1-\frac{1}{b_{\tau(t)}}}\end{array}\right\}\end{array}\right];$$

This recursion yields the following generalization to the results we derived for equilibrium behavior, determining optimal job choice and effort selection off the equilibrium path.

Lemma 3 *The executive optimally selects his position and effort by choosing (d_t,l_t) to minimize:*

$$\varepsilon_{0t}d_{0t}+\sum_{j=1}^{J}\sum_{k=1}^{K}d_{jkt}\left[\varepsilon_{jkt}-l_t\ln\alpha_{jkt}(h)-(1-l_t)\ln\beta_{jkt}(h)\right]$$
$$-l_t\sum_{j=1}^{J}\sum_{k=1}^{K}d_{jkt}\left(b_{\tau(t)}-1\right)\ln E_t\left[\upsilon_{jk,t+1}B_{t+1}\left(H_{jk}(h),H_{jk}\left(h'\right),b_{\tau(t)+1}\right)\right]$$
$$-(1-l_t)\sum_{j=1}^{J}\sum_{k=1}^{K}d_{jkt}\left(b_{\tau(t)}-1\right)\ln E_t\left[\upsilon_{jk,t+1}g\left(\pi_{j,\tau(t)+1},h\right)B_{t+1}\left(H'_{jk}(h),H_{jk}\left(h'\right),b_{\tau(t)+1}\right)\right]$$

Whether the executive shirks or not depends on the relative benefits to his current utility, how it affects expected lifetime utility through compensation, and the differential investment value from working diligently versus shirking. A direct implication of Lemma 3 is that if position (j,k) is selected and diligence is optimal, meaning $d^o_{jkt},=l^o_t=1$, then:

$$\frac{E_t\left[\upsilon_{jk,t+1}B_{t+1}\left(H_{jk}(h),H_{jk}(h'),b_{\tau(t)+1}\right)\right]}{E_t\left[\upsilon_{jk,t+1}g\left(\pi_{j,\tau(t)+1},h\right)B_{t+1}\left(H'_{jk}(h),H_{jk}(h'),b_{\tau(t)+1}\right)\right]}\leq\left(\frac{\beta_{jkt}(h)}{\alpha_{jkt}(h)}\right)^{1/(b_t-1)}$$

3.3 Optimal Contracting

The demand for executives reflects their potential to add value to the firm, and is based on the perceptions firms have about their executive employees. As we remarked above, firms believe they can pinpoint the characteristics of every executive with the vector (t,h'). We express the demand by the j^{th} firm for an age t executive with perceived human capital of h' to fill the k^{th} position at time τ by a probability, denoted by $P_{jkt}(h',b_\tau)$. Later in this section we derive this demand probability as a function of the model's primitives in competitive equilibrium. But first we first derive the minimum expected cost of satisfying this probabilistic demand with executives who work diligently.

To achieve a success rate of $P_{jkt}(h',b_\tau)$, the firm must offer a sufficiently attractive compensation package to elicit this supply. Since firms have a point expectation of each manager's characteristics they treat the supply of managers as if they observed (t,h) rather than (t,h'). It follows that their success rate is achieved at each date τ if and only if:

$$P_{jkt}\left(h',b_\tau\right) \leq p_{jkt}\left(h',b_\tau\right)$$

for all executive types (t,h') and all positions (j,k), where the mapping $p_{jkt}(h',b_\tau)$ the conditional choice probabilities of a hard working executive with true characteristics (t,h'). Substituting the expressions we obtained from the supply side discrete choice problem to the right side of this inequality and rearranging, we obtain the expression:

$$P_{jkt}\left(h',b_\tau\right) \leq \Pr\left[(j,k) \in \underset{(j',k')}{\arg\max} \begin{array}{l} \varepsilon_{0t} - \varepsilon_{j'k't} + \ln \alpha_{j'k't}(h'), \\ (1-b_\tau)\ln E_t\left[\upsilon_{j'k',t+1} A_{t+1}\left(H_{j'k'}(h'), b_{\tau+1}\right)\right] \end{array}\right]$$

This inequality forms the basis for a participation constraint in the derivation of the optimal contract and the resulting compensation schedule.

Rather than working with the minimization operator associated with observing choices induced by the idiosyncratic private shocks

$$(\varepsilon_{11t} - \varepsilon_{0t}, \ldots, \varepsilon_{JKt} - \varepsilon_{0t})$$

it is less cumbersome to manipulate expressions that are based on the partition induced by the choice probabilities $P_{jkt}(h',b_\tau)$. Accordingly let:

$$P_t\left(h',b_\tau\right) \equiv \left(P_{11t}\left(h',b_\tau\right), \ldots, P_{JKt}\left(h',b_\tau\right)\right)$$

denote the conditional demand probability simplex, where the probability of retirement is simply $1 - \sum_{(j,k)} P_{jkt}(h',b_\tau)$. Noting that the logarithm of utility from retirement is ε_{0t}, by Proposition 1 of Hotz and Miller (1993), there exists a mapping $q(P)$ from the simplex to R^{JK} such that this inequality is met if and only if:

$$q_{jk}\left[P_t\left(h',b_\tau\right)\right] \leq \ln \alpha_{jkt}\left(h'\right) + (b_\tau - 1)\ln E_t\left[\upsilon_{jk,t+1} A_{t+1}\left(H_{jk}\left(h'\right), b_{\tau+1}\right)\right] \tag{7}$$

The left side of (7), $q_{jk}[P_t(h',b_\tau)]$, is the expected value of the disturbance $\varepsilon_{jkt} - \varepsilon_{0t}$ for an executive with characteristics (t,h') who is on the cusp of accepting the (j,k) position over all the other alternatives when, given the conditions of the job, summarized by compensation $\upsilon_{jk,t+1}$, human capital value $A_{t+1}[H_{jk}(h'), b_{\tau+1}]$, and nonpecuniary benefits $\alpha_{jkt}(h')$, the j^{th} firm offers a contract for the k^{th} position that is accepted with probability $P_{jkt}(h',b_\tau)$. We interpret the inequality above as a participation constraint firms are obliged to respect. Raising compensation, such as increasing $\upsilon_{jk,t+1}$ by a positive constant, better working conditions, represented by lower values of $\alpha_{1jk}(h_t)$, and higher investment utility, that is lower values of $A_{t+1}[H_{jk}(h'), b_{\tau+1}]$, reduce the right side of this inequality and thus help the firm attain any target demand probability $P_{jkt}(h',b_\tau)$.

Aside from the participation constraint described above, firms must also an incentive compatibility constraint that motivates their managers to work diligently. Following the same argument as before firms believe that every manager they hire has worked diligently up until that point. Thus to maintain incentives another period the inequality in Lemma 3 simplifies to:

$$\begin{aligned} & \alpha_{jkt}\left(h'\right)^{1/(b_\tau - 1)} E_t\left[\upsilon_{jk,t+1} A_{t+1}\left(H_{jk}\left(h'\right), b_{\tau+1}\right)\right] \\ \leq \; & \beta_{jkt}\left(h'\right)^{1/(b_\tau - 1)} E_t\left[\upsilon_{jk,t+1} g_{jkt}\left(\pi, h'\right) B_{t+1}\left(H'_{jk}\left(h'\right), H_{jk}\left(h'\right), b_{\tau+1}\right)\right] \end{aligned} \tag{8}$$

where we use the fact that by definition $A_t(h', b_\tau) = B_t(h', h', b_\tau)$. The incentive compatibility constraint is nonbinding if investment value from working diligently on the job dominates the disutility from working hard versus shirking and taking account of the investment value from shirking the job:

$$E_t\left[A_{t+1}\left(H_{jk}\left(h'\right), b_{\tau+1}\right)\right] \leq \left(\frac{\beta_{jkt}(h')}{\alpha_{jkt}(h')}\right)^{1/(b_\tau-1)} E_t\left[B_{t+1}\left(H'_{jk}\left(h'\right), H_{jk}\left(h'\right), b_{\tau+1}\right)\right] \tag{9}$$

In lower level jobs held at the beginning of a career this inequality might hold, obviating the need to tie remuneration to the abnormal returns of the firm and pay a risk premium, because the differential investment value is such an important motivator.

Following the approach Margiotta and Miller (2000), both constraints (7) and (8) can be expressed as linear in $\upsilon_{jk,t+1}$, and the objective function, the expected wage bill $E_t(w_{jk,t+1})$ can be expressed as a concave function of $\upsilon_{jk,t+1}$, namely $E_t(\ln \upsilon_{jk,t+1})$. Maximizing $E_t(\ln \upsilon_{jk,t+1})$ subject to the two constraints yields a unique stationary point from which the cost minimizing contract is derived.

Lemma 4 *If $h' = h$, and inequality (9) holds, then*

$$\overline{w}_{jk,t+1}(h, b_\tau, b_{\tau+1}) = \frac{b_{\tau+1}}{\rho}\left[\frac{1}{(b_\tau-1)}\ln \alpha_{jkt}(h) + \ln A_{t+1}\left[H_{jk}(h), b_{\tau+1}\right] + q_{jk}\left[P_t(h, b_\tau)\right]\right] \tag{10}$$

Otherwise the cost minimizing contract that elicits diligence and attracts a manager with experience h to the k^{th} position in the j^{th} firm at time t with probability $P_t(h)$ is:

$$\begin{aligned} w_{jk,t+1}(h, \pi, b_\tau, b_{\tau+1}) &= \overline{w}_{jk,t+1}(h, b_\tau, b_{\tau+1}) \qquad \textit{(OptCont)} \\ &+ \frac{b_{\tau+1}}{\rho}\left\{\ln\left[1 - \eta g_{jkt}(\pi, h) + \eta\left(\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right)^{1/(b_\tau-1)} \frac{A_{t+1}\left[H_{jk}(h), b_{\tau+1}\right]}{B_{t+1}\left[H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right]}\right]\right\} \end{aligned}$$

where η is the unique positive root to

$$\int\left[\frac{f_j(\pi)}{\eta^{-1} + \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{1/(b_\tau-1)} \frac{A_{t+1}\left[H_{jk}(h), b_{\tau+1}\right]}{B_{t+1}\left[H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right]} - g_{jkt}(\pi, h)}\right] d\pi = 1 \tag{11}$$

Equation ??, the compensation schedule characterizing the optimal contract, decomposes into four additive pieces. The first piece, $b_{\tau+1} \ln \alpha_{jkt}(h) / \rho(b_\tau - 1)$, is the amount that leaves a manager indifferent between retiring and accepting position (j, k) if the private values are the same across all the choices, there is no investment value from accepting the position, and the compensation is fixed. The second term:

$$b_{\tau+1} \ln\left[A_{t+1}\left[H_{jk}(h), b_{\tau+1}\right]\right] / \rho(b_\tau - 1)$$

is the investment value from the position, a wage discount that offsets higher future expected earnings. Next, $b_{\tau+1} q_{jk}\left[P_t(h, b_\tau)\right] / \rho$, sets compensation levels to make the position sufficiently attractive to the executive in the $P_{jkt}(h, b_\tau)$ fractal, a term that would arise in a static framework of job choice, such as the Roy model. The expected value of the last term is a risk premium for taking a position whose compensation depends on the firm's financial returns, and is therefore uncertain.

When human capital accumulation does not depend on effort so all human capital is public information, then $h_t = h'_t$ at every outcome, whether it is on the equilibrium path or not. Since $A_t(h, b_\tau) = B_t(h', h', b_\tau)$ the incentive compatibility constraint in this case reduces to the static model:

$$E[\upsilon_{jk,t+1}(\pi)\, g_{jkt}(\pi, h)\,|h_t] \leq \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{1/(b_\tau - 1)} E[\upsilon_{jk,t+1}(\pi)\,|h_t]$$

In this case the $A_{t+1}[H_{jk}(h), b_{\tau+1}]$ and $B_{t+1}\left[H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right]$ terms cancel each other in (??) and (11). Consequently the only difference in the optimal contract distinguishing a model with human capital that depends on past job choices, from a model of pure moral hazard without any human capital, is the investment cost component of compensation $b_{\tau+1}\ln[A_{t+1}[H_{jk}(h), b_{\tau+1}]]/\rho(b_\tau - 1)$. Regardless of whether human capital is observed or not, jobs associated with promotion prospects to higher paid jobs command an offsetting negative compensating differential, reflected in lower values of $A_{t+1}[H_{jk}(h), b_{\tau+1}]$. But when human capital is not observed, the prospect of promotion also ameliorates incentive problems, and would predict that lower level jobs on fast promotion tracks require less incentive pay.

Note that equilibrium compensation depends on the position within the firm (since tasks vary with the position), the applicant's known characteristics (since managers with different backgrounds bring different skills to the firm), the firm's random return next period (since this is a signal shareholders receive about unobserved managerial effort), but not the characteristics of the other executives in the management team (since, by our separability assumption, the manager's value to the firm is independent of the composition of the other executives on the management team). In our framework we assume firms cannot commit to long term multiperiod contracts with their executive staff. If human capital depends on employment history but not on effort, then the optimal long term contract decentralizes to a sequence of short term contracts, obviating the need to consider anything but one period contracts.[9] However if human capital is a function of unobserved effort, then there are benefits to shareholders from committing. In that scenario, the optimal long term compensation contract takes into account the signals a firm receives from abnormal returns about previous firm specific unobserved investments in human capital made by its workers. Absent a commitment device, firms engage in sequentially optimal short term one period contracts of this type.

3.4 Equilibrium Job Assignment

The demand for executive services is determined by a zero profit condition imposed in equilibrium:

$$E_t[w_{jk,t+1}(h, \pi, b_\tau, b_{\tau+1})] = F_{jk}(h)$$

[9] The proof of this statement follows arguments developed in Fudenberg, Holmstrom and Milgrom (1990).

Note that we are not assuming shareholders are unable to extract rent from hiring managers, merely that the expected surplus has already been impounded into the value of the firm before any contracts are written. We now demonstrate that combining this condition with the form of the optimal contract and the recursion for $B_{t+1}\left[H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right]$ sequentially yields the equilibrium solution as a mapping from the primitives of the model.

Define $W_{jkt}(h, b_\tau)$ as the expected compensation to be paid less the compensating differential associated with the unobserved component:

$$\begin{aligned}
& W_{jk,t+1}(h, \pi, b_\tau, b_{\tau+1}) \\
= \; & \frac{1}{(b_\tau - 1)} \left[\ln \alpha_{jkt}(h) + \ln A_{t+1}\left[H_{jk}(h), b_{\tau+1}\right] \right. \\
& \left. + \ln \left[1 - \eta g_{jkt}(\pi, h) + \eta \left(\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)} \right)^{1/(b_\tau - 1)} \frac{A_{t+1}\left[H_{jk}(h), b_{\tau+1}\right]}{B_{t+1}\left[H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right]} \right] \right]
\end{aligned}$$

Then from the optimal contract given in Equation (??):

$$E_t\left[w_{jk,t+1}(h, \pi, b_\tau, b_{\tau+1})\right] = \underbrace{E_t\left[\frac{b_{\tau+1}}{\rho} \left\{ W_{jk,t+1}(h, \pi, b_\tau, b_{\tau+1}) + q_{jk}\left[P_t(h, b_\tau)\right] \right\} \right]}_{F_{jk}(h)} = F_{jk}(h)$$

Form the JK dimensional vector $F(h)$ from arraying nonpecuniary benefits of shirking $\beta_{jkt}(h)$; and the probability density function of the idiosyncratic disturbance term to preferences $\varphi(\varepsilon)$.

Our model is estimated from cross sectional or longitudinal data on executive background, the job choices they make, their employer firms' abnormal returns, their compensation, and bond prices. Empirical researchers are not typically privy to the private information of their managers, but like the firms observe h', but not h. In equilibrium, though, $h' = h$, a condition we impose in estimation. Thus consistent estimates of the abnormal returns density, the choice probabilities, and the compensation schedule can obtained from cross sectional data on $t, j, k, h', \pi_{\tau(t)}, b_{\tau(t)}, b_{\tau(t)+1}$. Similarly the process governing bond prices can be estimated separately off time series data. So for the purpose of identification we assume the bond prices b_τ are observed and follow a known Markov process, that the probability density of abnormal returns, $f_j(\pi)$, is known, as is the compensation schedule $w_{jkt}(h', \pi, b_\tau, b_{\tau+1})$, and the conditional choice probabilities, $p_{jkt}(h, b_\tau)$ for every firm (type) $j \in \{1, \ldots, J\}$, managerial position $k \in \{1, \ldots, K\}$, and age $t \in \{1, \ldots, T\}$.

4.1 Identi...cation

The $f_j(\pi)$ density is identified directly from the data on abnormal returns, and the technology driving the demand for executive services, $F_{jkt}(h)$, is identified the zero profit condition imposed in equilibrium. Only the risk aversion parameter ρ, the parameter vectors defining tastes $\alpha_{jkt}(h)$ and $\beta_{jkt}(h)$, the likelihood ratio associated with diligence versus shirking $g_{jkt}(\pi|h)$ along with the density function for the unobserved utility components $\varphi(\varepsilon)$ remain to be identified.

We follow Gayle and Miller (2009) by imposing a regularity condition on the likelihood ratio, that shareholders are certain that all the executives have worked diligently during the period if firm performance at the end of the period is truly outstanding. Formally, we assume the likelihood ratio $g_{jkt}(\pi|h)$ converges to zero as π diverges to infinity, or that for all (j, k, h, t):

$$\lim_{\pi \to \infty} g_{jkt}(\pi|h) = 0 \tag{12}$$

Since variation in compensation is caused by changes in the value of the likelihood ratio of working diligently versus shirking, directly attributable to underlying changes in profitability, it follows that variation in compensation diminishes at very high levels of firm profitability.

An element in the parameter space, denoted by $\theta \in \Theta$, is defined as:

$$\theta \equiv \alpha_{jkt}(h), \beta_{jkt}(h), g_{jkt}(\pi|h), \varphi(\varepsilon), \rho$$

To study identification and develop an estimator, we develop notation that explicitly entertains any $\hat{\theta} \in \Theta$, while simultaneously acknowledging that the true model, parameterized by an unknown $\theta^o \in \Theta$ say, generated the data embodied in the density for abnormal returns, the compensation schedule and the choice probabilities. Put succinctly, apart from θ^o, what other $\hat{\theta} \in \Theta$ could have generated $f_j(\pi)$, $w_{jk,t+1}$ $h', \pi, b_{\tau(t)}, b_{\tau(t)+1}$, and $p_{jkt}(h, b_\tau)$?

To answer this question we develop mappings corresponding to the recursions $A_t(h, b_\tau)$ and $B_t(h, h', b_\tau)$, as well as the random variable $v_{jk,t+1}$, that reflect both any parameterization $\hat{\theta} \in \Theta$ being entertained, as well as the data generating process, summarized by the true model $\theta \in \Theta$. Let:

$$v_{jk,t+1}(\hat{\rho}, \pi) \equiv \exp\left(-\hat{\rho} w_{jk,t+1}\left(h, \pi, b_{\tau(t)}, b_{\tau(t)+1}\right)/b_{\tau(t)+1}\right)$$

and exploiting Equation (12) denote by:

$$\overline{v}_{jk,t+1}(\hat{\rho}) \equiv \lim_{\pi \to \infty} [v_{jk,t+1}(\hat{\rho}, \pi)]$$

We also define $\hat{A}_t(h, b_\tau)$ recursively, setting $\hat{A}_T\left(h, b_{\tau(t)}\right) = 1$, substituting $\hat{A}_{t+1}\left(h, b_{\tau(t)}\right)$ and $\hat{a}_{jkt}(h)$ into the decision problem of Lemma 2, solving the (essentially static) discrete choice problem, computing $\hat{E}\left[\exp\left(-\varepsilon^*_{0t}/b_{\tau(t)}\right)\right]$ and $\hat{E}\left[\exp\left(-\varepsilon^*_{jkt}/b_{\tau(t)}\right)\right]$ from $\hat{\varphi}(\varepsilon)$, and forming $\hat{A}_t\left(h, b_{\tau(t)}\right)$ as:

$$\begin{aligned} \hat{A}_t\left(h, b_{\tau(t)}\right) &= p_{0t}\left(h, b_{\tau(t)}\right) \hat{E}\left[\exp\left(-\varepsilon^*_{0t}/b_{\tau(t)}\right)\right] \\ &+ \sum_{j=1}^{J}\sum_{k=1}^{K} p_{jkt}\left(h, b_{\tau(t)}\right) \hat{E}[\exp\left(-\varepsilon^*_{jkt}/b_{\tau(t)}\right)] \exp\left(\hat{\alpha}_{jk} P_t\left(h, b_{\tau(t)}\right)\right)^{\frac{1}{b_{\tau(t)}}} \end{aligned} \tag{13}$$

We define the mapping $\hat{B}_t\left(h, h', b_{\tau(t)}\right)$ in a similar way. Letting $\hat{B}_T\left(h, h', b_{\tau(t)}\right) = 1$ and defining the expected value of the disturbances as above, we write:

$$\begin{aligned} \hat{B}_t\left(h, h', b_{\tau(t)}\right) &= p_{0t}\left(h, h', b_{\tau(t)}\right) E\left[\exp\left(-\varepsilon^*_{0t}/b_{\tau(t)}\right)\right] \\ &+ \sum_{(j,k)} \left(\begin{array}{l} p_{jkt}\left(h, h', b_{\tau(t)}\right) E[\exp\left(-\varepsilon^*_{jkt}/b_{\tau(t)}\right)] \\ \times \max\left(\exp\left\{q_{jk}\left[P_t(h', b_\tau)\right]\right\}^{1/b_\tau}, \beta_{jkt}(h)^{\frac{1}{b_{\tau(t)}}} \hat{B}_{jkt}\left(h, h', b_{\tau(t)}\right)^{1-\frac{1}{b_{\tau(t)}}}\right) \end{array}\right) \end{aligned} \tag{14}$$

where

$$\hat{B}_{jkt}\left(h, h', b_{\tau(t)}\right) = E_t\left[\frac{v_{jk,t+1}(\hat{\rho}, \pi) - \overline{v}_{jk,t+1}(\hat{\rho})}{1 - \overline{v}_{jk,t+1}(\hat{\rho})^{-1} E_t\left[v_{jk,t+1}(\hat{\rho}, \pi)^{-1}\right]}\right] E_t\left[B_{t+1}\left(H'_{jk}(h), H_{jk}(h), b_{\tau(t)+1}\right)\right] \tag{15}$$

In the proof of Theorem 1 below contained in the Appendix we establish that $\hat{A}_t\left(h, b_{\tau(t)}\right)$ and $\hat{B}_t\left(h, h', b_{\tau(t)}\right)$ respectively solve Equations (4) and (6), the recursions defining $A_t\left(h, b_{\tau(t)}\right)$ and $B_t\left(h, h', b_{\tau(t)}\right)$.

Within our framework identification can be reduced to three sets of equations indexed by age t, associated with the participation constraint, the first order condition for the optimal contract, and the incentive compatibility constraint. Our first result on identification are the necessary and sufficient conditions for observational equivalence.

Lemma 5 *Abbreviate by $E_t[\cdot]$ the conditional expectations operator $E\left[\cdot \mid h, b_{\tau(t)}\right]$. The parameter $\hat{\theta} \in \Theta$ is observationally equivalent to $\theta \in \Theta$ if and only if $\hat{\theta}$ solves the following equations for all firm types $j \in \{1, \ldots, J\}$, positions $k \in \{1, \ldots, K\}$, ages $t \in \{1, \ldots, T\}$ and backgrounds $h \in \{1, \ldots, K\}$ and bond prices b_τ:*

$$\hat{a}_{jkt}(h) = E_t\left[\hat{a}_{jk,t+1}\hat{A}_{t+1}\left(H_{jk}(h), b_{\tau(t)+1}\right)\right]^{(1-b_{\tau(t)})} \exp\left(\hat{\alpha}_{jk} P_t\left(h, b_{\tau(t)}\right)\right) \tag{16}$$

$$\hat{g}_{jkt}(\pi|h) = \frac{\overline{v}_{jk,t+1}(\hat{p})^{-1} - v_{jk,t+1}(\hat{p},\pi)^{-1}}{\overline{v}_{jk,t+1}(\hat{p})^{-1} - E_t[v_{jk,t+1}(\hat{p},\pi)]^{-1}} \times \frac{E_t\left\{\hat{B}_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]\right\}}{\hat{B}_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]} \tag{17}$$

$$\hat{\beta}_{jkt}(h) = \exp\left[\hat{\psi}_{jk}\left[P_t\left(h, b_{\tau(t)}\right)\right]\right]\left\{E_t\left[\hat{B}_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]\right]\right\}^{1-b_{\tau(t)}} \times \left(E_t[\hat{e}_{jk,t+1}(\hat{p})]\frac{\overline{v}_{jk,t+1}(\hat{p})^{-1} - E_t[v_{jk,t+1}(\hat{p},\pi)]^{-1}}{\overline{v}_{jk,t+1}(\hat{p})^{-1} - E_t[v_{jk,t+1}(\hat{p},\pi)^{-1}]}\right)^{1-b_{\tau(t)}} \tag{18}$$

Substituting $A_{t+1}\left[H_{jk}(h), b_{\tau(t)+1}\right]$ for $\hat{A}_{t+1}\left[H_{jk}(h), b_{\tau(t)+1}\right]$ in Equation (35) yields the participation constraint, derived from the inequality Inequality (7) which is satisfied with equality in equilibrium. Thus proving $\hat{\theta}$ must Equation (35) amounts to demonstrating $\hat{A}_t(h, b_\tau)$ satisfies the recursion (4). The expression for $\hat{\alpha}_{jkt}(h)$ essentially consists of three terms. In a standard labor supply model the disutility from work is offset by the utility equivalent of compensation, captured here by $\hat{e}_{jk,t+1}$. As in a Roy model, when alternative work is available, the attraction of taking (j,k) versus another job is reflected by a mapping of the choice probabilities, in this model $\exp\left[\hat{\psi}_{jk}\left[P_t\left(h, b_{\tau(t)}\right)\right]\right]$. Finally the greater human capital benefits, $\hat{A}_{t+1}\left[H_{jk}(h), b_{\tau(t)+1}\right]$, compensate for lower wages and worse working conditions.

Equation (36) is derived following the approach of Gayle and Miller (2009a) comes from manipulating the first order condition, and reflects the intuition that variation in utility arising from an optimally devised compensation schedule is ultimately driven by changes in the likelihood ratio for diligence versus shirking is driven by changes in profitability. When bond prices are fully anticipated the terms involving $\hat{B}_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]$ cancel out, then Equation (17) reduces to the formula derived by Gayle and Miller (2009a) for a model where there is both moral hazard and hidden information. Thus the introduction of dynamics through human capital accumulation only affects inferences about $g_{jkt}(\pi|h)$ to the extent that the interest rate is stochastic.

Equation (36) is derived by substituting Equations (35) and (36) into the incentive compatibility constraint (7) and simplifying. To establish sufficiency, the appendix proves that every $\hat{\theta}$ satisfying all three sets of conditions yields the data generating process as the outcome of the optimal contracting problem.

It is instructive to analyze the specialization where the bond price is a constant. When $b_\tau = b$ for all τ, we prove that $\alpha_{jkt}(h)$, $\beta_{jkt}(h)$, and $g_{jkt}(\pi|h)$ are nonparametrically identified if and only if both the risk aversion parameter ρ and the probability density function of the idiosyncratic disturbance term to preferences $\varphi(\varepsilon)$ are known.

Lemma 6 *Suppose $b_t = b$. For every $\rho > 0$ and all proper probability density functions $\varphi(\varepsilon)$ defined on the same support as $\varphi^*(\varepsilon)$, there exists a unique θ solving the equations in Lemma 1 which is observationally equivalent to θ^*.*

Variation in b_τ over time adds restrictions that aid identification. By inspection it is easy to verify, for example, that when the bond price is independently distributed, additional restrictions on the data emerge from these aggregate sources. In practice only a few additional restrictions arise from annual variation in bond prices over a relatively short panel. Restricting the effect of age on preferences, by specializing $\alpha_{jkt}(h)$ to $\alpha_{jk}(h)$ and $\beta_{jkt}(h)$ to $\beta_{jk}(h)$, coupled with variation in the optimal contract and conditional choice probabilities that arise from encroaching retirement in this multiperiod finite horizon setting, also helps achieve identification. Comparing the age $T-1$ problem, in which $A_T\left(h, b_{\tau(T)}\right) = 1$, with the age $T-2$ problem, for example, and noting $A_{T-1}\left(h, b_{\tau(T-1)}\right) \neq 1$, illustrates the horizon effects in providing identifying restrictions on $\varphi(\varepsilon)$ and ρ through the choice probabilities and the compensation contract when the utility and technology primitives of the model do not depend on t. Similarly other restrictions on the effects of executive age and background (t, h) on $\alpha_{jkt}(h)$, $\beta_{jkt}(h)$, and $g_{jkt}(\pi|h)$ are alteratives to making assumptions about the value of ρ and how to specify $\varphi(\varepsilon)$.

4.2 Parameterizing and Estimating the Model

In our empirical work we make two further assumptions of convenience, that $\varphi(\varepsilon)$ is distributed as Type 1 Extreme Value and that bond prices are known. The latter bond price assumption can be tested, as we demonstrate below. The parameterization of $\varphi(\varepsilon)$ considerably simplifies the estimation but can be relaxed providing the conditions of identification given in Theorem 1 are met by, if necessary imposing assumptions on $\alpha_{jkt}(h)$, $\beta_{jkt}(h)$ and $g_{jkt}(\pi|h)$ instead. The computational advantages of these assumptions are most evident by reviewing the estimation method.

The three sets of equalities of the Lemma used in identification form the basis for estimation. After estimating the reduced form of the conditional choice probabilities $P_t\left(h_t, b_{\tau(t)}\right)$, we estimated ρ and $\alpha_{jkt}(h)$ from Equation (35), derived from the participation constraint. Intuitively the estimation exploiting the idea that when risk averse managers make rational choices between different uncertain outcomes or lotteries they are revealing their attitude towards risk. Writing:

$$z \equiv p_{0,t+1}\left[H_{jk}(h), b_{\tau(t)+1}\right]^{\frac{1}{b_{\tau(t)+1}}} \Gamma\left(b_{\tau(t)}+1\right)/b_{\tau(t)} \left[\frac{p_{0t}\left(h, b_{\tau(t)}\right)}{p_{jkt}\left(h, b_{\tau(t)}\right)}\right]^{\frac{1}{\left(1-b_{\tau(t)}\right)}}$$

and estimating z with $\hat{z}$, we form a two stage GMM estimator for $(\hat{\rho}, \hat{\alpha}_{jkt})$ from the sample moments corresponding to:

$$E\left[\exp\left(\hat{\rho} w_{jkt}/b_{\tau(t)+1}\right) x - \hat{\alpha}_{jkt}(h)^{\frac{-1}{\left(1-b_{\tau(t)}\right)}} zx\right] = 0 \tag{19}$$

where x is a vector of instruments constructed from $(h, j, k, t, \tau(t))$ for each observation, and $\hat{z}$, formed using nonparametric estimators of the conditional choice probabilities $P_t\left(h_t, b_{\tau(t)}\right)$, is used in place of z. As we prove in the next lemma, Equation (19) can be derived from the identifying condition for ρ and $\alpha_{jkt}(h)$, Equation (35), by exploiting the formula for $A_t\left(h, b_{\tau(t)}\right)$ under the extreme value distribution assumption.

To estimate $g_{jkt}(\pi|h)$ we followed Gayle and Miller (2009b). By assumption bond prices are fully anticipated:

$$E_t\left[B_{t+1}\left(H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right)\right] = B_{t+1}\left(H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right) \tag{20}$$

so Equation (17), which identifies $g_{jkt}(\pi|h)$, reduces to:

$$\hat{g}_{jkt}(\pi|h) = \frac{\overline{v}_{jk,t+1}(\hat{\rho})^{-1} - v_{jk,t+1}(\hat{\rho},\pi)^{-1}}{\overline{v}_{jk,t+1}(\hat{\rho})^{-1} - E_t[v_{jk,t+1}(\hat{\rho},\pi)]^{-1}} \tag{21}$$

We formed $\hat{w}\left(h_t,\pi,b_{\tau(t)},b_{\tau(t)+1}\right)$, nonparametric estimates of the compensation schedule, using them in conjunction with our estimate of $\hat{\rho}$ obtained from the first stage, to compute estimates of $v_{jk,t+1}$ denoted by:

$$\hat{v}_{jk,t+1}(\hat{\rho},\pi) \equiv \exp\left(-\hat{\rho}\hat{w}\left(h_t,\pi,b_{\tau(t)},b_{\tau(t)+1}\right)\right)$$

We approximated the conditional expectation $E_t[v_{jk,t+1}(\hat{\rho},\pi)]$ by nonparametrically averaging $\hat{v}_{jk,t+1}(\hat{\rho},\pi)$ over the subsamples with similar $\left(j,k,t,h,b_{\tau(t)},b_{\tau(t)+1}\right)$, and computed $\overline{v}_{jk,t+1}(\hat{\rho})$ using Brunk's (1958) estimator. To be more specific, for each value of $\left(j,k,t,h,b_{\tau(t)},b_{\tau(t)+1}\right)$, we rank the observations on abnormal firm returns in increasing order by $\pi^{(q)}\left(j,k,t,h,b_{\tau(t)},b_{\tau(t)+1}\right)$, denoting by $v_{jk,t+1}\left(\hat{\rho},\pi^{(q)}\right)$ the corresponding (estimated) compensations, and estimate $\overline{v}_{jk,t+1}(\hat{\rho})$ with:

$$\overline{v}_{jk,t+1}(\hat{\rho}) \equiv \min_{q} \sum_{r=1}^{q} \frac{\hat{v}_{jk,t+1}\left(\hat{\rho},\pi^{(q)}\right)}{q} \tag{22}$$

Finally our estimate of $g_{jkt}(\pi|h)$ was obtained by substituting our estimates of $v_{jk,t+1}(\hat{\rho},\pi)$, $\overline{v}_{jk,t+1}(\hat{\rho})$ and $E_t[v_{jk,t+1}(\hat{\rho},\pi)]$ into Equation (21).

These procedures demonstrate that, when the managers in the sampled population can anticipate bond prices, $\alpha_{jkt}(h)$, $\beta_{jkt}(h)$ and $g_{jkt}(\pi|h)$ can be estimated without recourse to recursively computing $B_t\left(h',h,b_{\tau(t)}\right)$, in this avoiding the computational costs of nesting a fixed point algorithm within the estimation procedure. This still leaves the shirking parameter $\beta_{jkt}(h)$ to estimate. We solved $B_t\left(h,h',b_{\tau(t)}\right)$ and $\beta_{jkt}(h)$ recursively from Equations (14), (15) and (36) without imposing any restrictions, using the estimates of the compensation schedule and the other parameters obtained from the previous stages. Since $\beta_{jkt}(h)$ is exactly identified from $B_t\left(h,h',b_{\tau(t)}\right)$ no estimation is involved, and in the case of the Type 1 Extreme value distribution $B_t\left(h,h',b_{\tau(t)}\right)$ simplifies to a recursion that sequentially defines $\beta_{jkt}(h)$ in terms of the underlying probabilities and compensation, as indicated in the lemma below.

Lemma 7 *If ε_{jkt} is independently and identically distributed as Extreme Value Type I with location and scale parameters $(0,1)$, and bond prices are fully anticipated one period ahead, then:*

$$q_{jk}\left[P_t\left(h,b_{\tau(t)}\right)\right] = \ln p_{jkt}\left(h,b_{\tau(t)}\right) - \ln p_{0t}\left(h,b_{\tau(t)}\right) \tag{23}$$

and

$$E\left[\exp\left(-\varepsilon^{*}_{jkt}/b_{\tau(t)}\right)\right] = p_{jkt}\left(h,b_{\tau(t)}\right)^{1/b_{\tau(t)}}\Gamma\left[\left(b_{\tau(t)}+1\right)/b_{\tau(t)}\right] \tag{24}$$

Also:

$$A_t\left(h,b_{\tau(t)}\right) = p_{0t}\left(h,b_{\tau(t)}\right)^{1/b_{\tau(t)}}\Gamma\left[\left(b_{\tau(t)}+1\right)/b_{\tau(t)}\right] \tag{25}$$

and:

$$B_t\left(h,h',b_{\tau(t)}\right) \tag{26}$$
$$= \left(\sum_{(j,k)} \begin{array}{l} p_{jkt}\left(h,h',b_{\tau(t)}\right)^{1+1/b_{\tau(t)}}\Gamma\left(\left(b_{\tau(t)}+1\right)/b_{\tau(t)}\right) \\ \times \max\left\{ \frac{p_{0t}\left(h,h',b_{\tau(t)}\right)}{p_{jkt}\left(h,h',b_{\tau(t)}\right)}, \beta_{jkt}(h)\, E_t\left[\begin{array}{l} \frac{v_{jk,t+1}-\overline{v}_{jk,t+1}}{1-\overline{v}_{jk,t+1}^{-1}E_t\left[v_{jk,t+1}(\pi)^{-1}\right]} \\ \times B_{t+1}\left(H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right)\end{array}\right]^{b_{\tau(t)}-1}\right\}^{\frac{1}{b_{\tau(t)}}} \end{array}\right)$$
$$+p_{0t}\left(h,h',b_{\tau(t)}\right)^{1+1/b_{\tau(t)}}\Gamma\left(\left(b_{\tau(t)}+1\right)/b_{\tau(t)}\right)$$

where:

$$p_{jkt}\left(h,h',b_{\tau(t)}\right) = M_{jkt}\left(h,h',b_{\tau(t)}\right)^{\left(b_{\tau(t)}-1\right)}\left\{1+\sum_{j'=1}^{J}\sum_{k'=1}^{K} M_{j'k't}\left(h,h',b_{\tau(t)}\right)^{\left(b_{\tau(t)}-1\right)}\right\}^{-1} \tag{27}$$

and:

$$M_{jkt}\left(h,h',b_{\tau(t)}\right) \equiv \max\left\{\begin{array}{c} \alpha_{jkt}(h)^{\frac{1}{b_{\tau(t)}-1}} E_t\left[v_{jk,t+1}\right] B_{t+1}\left(H_{jk}(h), H'_{jk}(h'), b_{\tau(t)+1}\right), \\ \beta_{jkt}(h)^{\frac{1}{b_{\tau(t)}-1}} \frac{E_t\left[v_{jk,t+1}\right]-\overline{v}_{jk,t+1}}{1-\overline{v}_{jk,t+1}E_t\left[v_{jk,t+1}\right]^{-1}} B_{t+1}\left(H'_{jk}(h), H_{jk}(h'), b_{\tau(t)+1}\right) \end{array}\right\}$$

4.3 Testing the Model

Our empirical framework lends itself for investigating two questions to about the nature of human capital, the role education and previous working experience plays in determining compensation levels, and whether certain kinds of backgrounds also ameliorates incentive problems, at different points in the career of a manager. Our starting point is to observe that the bond price should not enter preferences directly. Comparing two managers of different ages who have the difference between our estimates of their preferences at a given age, say 50, should not be time dependent if they had the same background and experience at 50. Since the value of human capital does depend on the bond price, it follows that if we incorrectly specify the model by assuming human capital matters when it does not, or vice versa, then our estimates of their preference will be time dependent, which is revealed by differencing across successive cohorts.

We compute the value of tests statistics for the dynamic framework with hidden information about human capital that we analyze, and also compare them with test statistics for the simpler models, in which there is no human capital, and where all human capital is public knowledge. When there is no human capital $A_t\left(h,b_{\tau(t)}\right)=1$ and the preference parameter for working diligently, estimated from the participation constraint modeled in Equation (35), reduces to:

$$\alpha_{jkt}(h) = E\left[v_{jk,t+1}\left|h,b_{\tau(t)}\right.\right]^{\left(1-b_{\tau(t)}\right)}\exp\left\{q_{jk}\left[P_t\left(h,b_{\tau(t)}\right)\right]\right\} \tag{28}$$

Differencing across cohorts, it immediately follows that for all τ:

$$E\left[v_{jk,t+1}\left|h,b_{\tau}\right.\right]^{(1-b_{\tau})}\exp\left\{q_{jk}\left[P_t\left(h,b_{\tau}\right)\right]\right\} = E\left[v_{jk,t+1}\left|h,b_{\tau+1}\right.\right]^{(1-b_{\tau+1})}\exp\left\{q_{jk}\left[P_t\left(h,b_{\tau+1}\right)\right]\right\} \tag{29}$$

If human capital affects career choices, meaning $A_t\left(h, b_{\tau(t)}\right) \neq 1$ for $t < T$, but is public knowledge (perhaps because it only depends on past positions held but not on work effort), then $B_{t+1}\left(h, h', b_{\tau(t)}\right) = 1$ and the preference parameter for shirking, estimated from the participation constraint modeled in Equation (36), reduces to:

$$\beta_{jkt}(h) = \exp\left(q_{jk} P_t\left(h, b_{\tau(t)}\right)\right) \left(E_t[v_{jk,t+1}(\rho)] \frac{\overline{v}_{jk,t+1}(\rho)^{-1} - E_t[v_{jk,t+1}(\rho)]^{-1}}{\overline{v}_{jk,t+1}(\rho)^{-1} - E_t[v_{jk,t+1}(\rho)^{-1}]}\right)^{1-b_{\tau(t)}} \tag{30}$$

Differencing estimates of $\beta_{jkt}(h)$ obtained from different cohorts then indicates whether there are aggregate affects that are not accounted for, but in the event of a rejection is not informative about why. However repeating the same tests for the estimated dynamic model with hidden human capital directly addresses that issue.

Bond prices are a good measure of the value of market so eminently appropriate in thinking about human capital, as a statistical matter there is nothing intrinsic about focusing on bond prices not entering preferences, as opposed to some other variable, say h^{int}, which under the null hypothesis affects compensation but does not enter preferences. Testing these exclusion restrictions amounts to checking their significance levels.

5 The Data

The data for our empirical study was compiled from three sources. From Standard & Poor's ExecuComp database we extracted records on the job title and compensation of the eight highest paid executives in the S&P 500, Midcap, and Smallcap firms for the years 1992 through 2006 inclusive. Data on the employer firms were supplemented by the S&P COMPUSTAT North America database and monthly stock price data from the Center for Securities Research (CRSP) database. We matched the names, birth dates and gender of 16,300 executives from 1800 firms with information in Who's Who to augmented their records with biographical data. The resulting data set gives us unprecedented access to detailed firm characteristics, including accounting and financial data, along with their managers' characteristics, namely the main components of their compensation, including pension, salary, bonus, option and stock grants plus holdings, their socio-demographic characteristics, including age, gender, education, and a description of their career history through the five ranks and firms.

This section summarizes the aggregate features of our data set. We present estimates of the distribution of abnormal returns and show how they vary with executive characteristics. We estimate elasticities of compensation with respect to returns and measures of executive experience. Finally we investigate, empirically, how experience and other background variables affect job transitions and thus define the career paths of executives. In this way we describe the variation in the data that supports the identification and estimation of our model of executive compensation and career choice.

5.1 Summary Statistics

Most of the characteristics of the executives and firms in our sample require no explanation, but the construction of several variables merit comment. The sample of firms was initially partitioned into three industrial sectors by GICS code. Sector 1, called primary, includes firms in energy (GICS:1010), materials (1510), industrials (2010,2020,2030), and utilities (5510). Sector 2, consumer goods, comprises firms from consumer discretionary (2510,2520,2530,2540,2550) and consumer staples (3010,3020,3030). Firms in health care (3510,3520), financial services (4010,4020,4030,4040), information technology and telecommunication services (410, 4520, 4030, 4040, 5010) comprise Sector 3, which we call services. In our sample 37 percent of the firms belong to the primary sector, 28 percent to the consumer goods sector, and the remaining 35 percent to the services sector. Firm size was categorized by total employees and total assets, the median firm in each size category determining whether the other firms are called large or small. The sample mean value of total assets is $18.2 billion (2000 US) with standard deviation $76.2 billion, while the sample mean number of employees is 23,659 with standard deviation 65,702.

Table 1 describes the characteristics of management by sector and firm size. Jobs were assigned to a rank using the hierarchy ordering we developed in our work on gender discrimination.[10] At 27 percent, Rank 2 is the most commonly observed rank, which reflects the diversity of promotion schemes across firms. By way of contrast, the top and bottom ranks each only contribute 6 percent to the sample population. The distribution of ranks across the three sectors is roughly independent but small firms, as measured by either assets of employment, have a greater proportion of their executives congregating in the lower ranks, with 30 percent versus 20 in the bottom two ranks. Four measures of experience were included to capture the potential of on-the-job training. Executive experience is the number of years elapsed since the manager was first recorded as one of the top eight paid executives in the sample. Tenure is years spent working at the employee's current firm. We also tracked the number of moves the manager made throughout his career in different jobs and ranks, as well as the number of moves since becoming an executive. Promotion is a indicator variable for whether the manager was promoted recently or not.

The mean age of executives is almost 54 years with a standard deviation of about 9. Only 4 percent of the sample are female, ranging between 3 percent in the primary sector and 5 percent in the consumer sector. Roughly speaking, formal education is uniformly distributed evenly between bachelor degree or less, professional certification (in accounting or law for example), MBA, some other Master's degree, and Ph.D. The distribution is approximately independent of firm size and sector, ranging from 15 percent with an MS/MA in the consumer sector to 27 percent in small firms by employee for professionally certified executives.

Tenure in the firm averages about 14 years, about 40 years less than age, with standard deviation of about 11, two years more. The sectors are ranked the same way with respect to age and tenure; similarly firms with small assets have both the oldest executives and the longest tenure. In these respects average age, firm sector and size are almost sufficient statistics for average tenure, giving the deceptive appearance at this level of aggregation that executives within firms follow a well defined career track. Averaging across the sample, there are two rank and/or firm turnover moves per observation, one of which has occurred since acquiring executive status. About one third of executives have been promoted within the last two years.

The most important differences between the executives across firm size and sector

[10] See Gayle, Golan and Miller (2008).

relate to their compensation. Regardless of which measure is used, the mean salary and bonus in small firms is about two thirds the mean in large firms, about half the total compensation, with standard deviations about one third smaller.[11] This suggests that similarly named positions in small firms are not comparable to their analogues in large firms and may help explain differences between internal and external transitions.

Summarizing differences across firm type, the consumer sector has the lowest percent of executives with advance degrees and the highest percent of female executives, while the service sector has the lowest average tenure and the highest promotion rate and highest total compensation. Total compensation is roughly twice as large in large firms (using both measures), promotion and turnover rates are greater, tenure is lower, and there are more executives holding MBA degrees.

Table 2 describes the characteristics of executives by rank. The average age between Rank 1 and 3 declines from 60 to 52, but is more or less constant as rank falls off further. Similarly average tenure is roughly constant in the lower and middle ranks at 14 but rises to 15 and 17 for Ranks 2 and 1 respectively. The average gap between Ranks 1 and 3 in executive experience is 6 years. To summarize, relative to the lower ranks, Ranks 1 and 2 are 8 years older, with only 6 years more executive experience and just 2 years more tenure, late bloomers hired by the firm late in their career. Not that they are likely to move more than those who do not reach the top levels; although 8 years older the they average the same number of past moves, before and after becoming an executive.

Females form a very small fraction of the executive sample, and they are not uniformly distributed by rank. By a factor of two to three, females congregate in the lower executive ranks relative to males; 2 percent of the top two ranks are females, while 6 percent of Ranks 5 and 6 are female. With regard to the education background variables, the two most striking features are that there is higher percent (out of total executives in the rank) of executives with MBA degrees in the top 4 ranks, the percent of executive with another Masters degree or a Ph.D. is greater in the bottom there ranks, and there is a larger percent of executives with professional certification in the bottom 4 ranks.

Average total compensation and the salary components rise from Rank 7, are maximized at Rank 2, at levels that are more than twice as high as the corresponding figures for Rank 7, and decline. The salary component for Rank 1 is only eclipsed by Rank 2, but it is an open question whether the total financial compensation package offered for a Rank 1 position is more or less desirable than the offer for a Rank 5 position. Although the average compensation \$2.7 million for Rank 2 exceeds the Rank 5 mean by almost \$400,000, the standard deviation for the former is more than twice that of the latter. For example, if all compensation variation observed in the data was resolved before an executive accepted a position, implying the standard deviation simply reflects heterogeneity in fixed pay contracts, then there would be many Rank 5 positions that pay better than many Rank 2 positions. Alternatively if all the variation in compensation was resolved after the executive accepted his job, implying the standard deviation is a measure of the income uncertainty, the executive would prefer Rank 5 to Rank 1 position if he was sufficiently

[11] We followed Antle and Smith (1985, 1986), Hall and Liebman (1998), Margiotta and Miller (2000) and Gayle and Miller (2008a, 2008b) by using total compensation to measure executive compensation. Total compensation is the sum of salary and bonus, the value of restricted stocks and options granted, the value of retirement and long term compensation schemes, plus changes in wealth from holding firm options, and changes in wealth from holding firm stock relative to a well diversified market portfolio instead.

risk averse.

5.2 Abnormal returns

We defined the abnormal returns of the firm as the residual component of returns that cannot be priced by aggregate factors the manager does not control. In an optimal contract compensation to the manager might depend on this residual in order to provide him with appropriate incentives, but it should not depend on changes in stochastic factors that originate outside the firm, which in any event can be neutralized by adjustments within his wealth portfolio through the other stocks and bonds he holds. More specifically, letting ϑ_{jt} denote the value of the j^{th} firm at time t, the gross abnormal return attributable to all the executives' actions is the residual

$$\pi_{jt} \equiv \left(\vartheta_{jt} + D_{jt} + \sum_{k=1}^{K} w_{jkt} \right) / \vartheta_{jt-1} - x_t \tag{31}$$

where x_t is the return on the market portfolio in period t and D_{jt} is the dividend. This study assumes that π_t is a random variable that depends on the managers' efforts in the previous period but, conditional on the effort vector of the executive branch $\{l_{jkt}\}_{k=1}^{K}$, is independently and identically distributed across both firms and periods.[12]

We now show how abnormal returns depends on the experience and the other characteristics of the executives.

5.3 Compensation

We estimated annual excess returns for firms in equation 31 from the data, and then computed, conditional on the state variables, a nonparametric estimator of total compensation from our imputed values compiled from the data, which we assume is the sum of true compensation and independent measurement error. We used Kernel methods to nonparametrically estimate $w_{2mk}^{o}(\pi, h)$, the compensation schedule for diligent work, for each (m, k, h) as:

$$w_{2mk}^{(N)}(\pi, h) = \frac{\sum_{s=1, s\neq n}^{N} \sum_{t=1}^{T} w_{st} I\{d_{mkst} = 1, h_{st} = h,\} K\left(\frac{d_{mt}-\pi}{\delta_{xN}}\right)}{\sum_{s=1, s\neq n}^{N} \sum_{t=1}^{T} I\{d_{mkst} = 1, h_{st} = h,\} K\left(\frac{d_{mt}-\pi}{\delta_{xN}}\right)}$$

Table 3 reports OLS and LAD results from regressing how compensation varies with firms' and executives' characteristics. The (conditional) level effects are given in the first two columns of estimates, their interactions with abnormal returns in the second two. Controlling for background demographics and tenure more or less leaves intact the qualitative rank ordering on total compensation we found in Table 3. Total compensation to Ranks 6 and 7 differ by a statistically insignificant amount, and then rises with promotion, spiking at Rank 2, compensation to Rank 1 falling between Ranks 3 and 4. In contrast the unconditional means and standard deviations reported in Table 3, however, the results from the regression analysis separate the effects of excess return, which induces

[12] In our sample the mean abnormal return is -0.005 with standard deviation 0.6, and we do not reject the null hypothesis that it is uncorrelated with the stock market.

uncertainty to manager's total compensation, from the background variables that determine observed heterogeneity. Note that Rank 1 is more affected by excess returns than every rank except 2. Thus Rank 1 has a lower (OLS) or the same (LAD) estimated mean and more dependence on abnormal returns than Rank 3, while Rank 2 has a higher mean but more dependence than Rank 3. Therefore Rank 3 offers a superior total compensation package to Rank 1, and for sufficiently risk averse executives, a more attractive compensation package than the Rank 2. Continuing in this vein, dependence on excess returns is declining in the remaining middle or lower ranks.

All the firm size and sector variables have significant coefficients except the OLS estimator of the level effect distinguishing the consumer from service sector. None of the background variables for executives interact significantly in the OLS regression, but almost all have significant level effects irrespective of estimator. A notable exception are the coefficients relating to gender. The OLS estimator indicates that gender has no effect on compensation level or its dependence on abnormal returns, whereas the LAD estimator implies there is a small positive level effect of $91,731 and significantly reduced dependence on abnormal returns, both factors making an executive positions more attractive to females relative to males.

With respect to education the OLS results show, that after controlling for the other observed differences, Ph.D. and MBA graduates earn more than $300,000 in excess of executives with undergraduate degrees only, who earn $386,793 more than those with professional certification only. Compensation is quadratic in age as is the case in wage regressions for many occupations. Tenure, executive experience and the number of past moves have statistically significant effects on compensation but are small and inconsequential in magnitude. More noteworthy is the large estimated sign-on bonus associated with turnover, $551,859 for LAD and $994,989 for OLS.

Overall our results suggest that after controlling for rank and firm type, there are significant returns from acquiring general human capital in formal education, but little from firm specific capital that is measured in terms of tenure within any one job and/or experience acquired at a variety of jobs. Similarly gender is not a useful predictor of wages given the other executive's and other characteristics and the nature of the job. To summarize, aside from formal education, job transitions and the abnormal returns of their own firms are the main drivers determining how wealthy executives become.

5.4 State Variables and Conditional Choice Probabilities

We denote the state variables relevant for the n^{th} manager at the time t by h_{nt}, one of $h < \infty$ possible characteristics, the ranks by $r \in \{1, \ldots, R\}$ and the firm types by $s \in \{1, \ldots, S\}$. In our model $h_{n,t+1}$, the n^{th} manager's state variables in the period $t+1$, are fully determined by h_{nt}, the type of firm he transitions to, denoted s_{nt}, and his rank next period, r_{nt}, by a mapping $h_{n,t+1} \equiv f(h_{nt}, r_{nt}, s_{nt})$, which we define in the next section. Our theory models the transition of h_{nt} to $h_{n,t+1}$ through the competitive equilibrium choices of (r_{nt}, s_{nt}), a stochastic process that generates the data. The structural estimation of our theoretical framework uses as input reduced form estimates of $P(r_{nt}, s_{nt} | h_{nt})$, the probability of (r_{nt}, s_{nt}) conditional on h_{nt}.

We report our estimates for the reduced form of our model. Since R and S are finite, and we assume H is a finite set, it follows that in principle cell estimators could be used to recover $P(r_{nt}, s_{nt} | h_{nt})$. Although our sample size, 59,066, is very large compared with all previous studies of this market, the comprehensive detail that accompanies each observation also greatly magnifies the total number of cells RSH,needed to estimate the model, so this procedure is not feasible. For example only 5 percent of the observations in our sample are female, and none of them have doctorates and head small firms. Many smoothing algorithms are asymptotically equivalent. We used multinomial logits to estimate the reduced form, because of their computational tractability in recovering the structural parameters, because the logit estimates are easy to interpret, and because they illustrate how the variation in our data is used to estimate the underlying structure. For expositional convenience we decomposed $P(r_{nt}, s_{nt} | h_{nt})$ into

$$P(r_{nt}, s_{nt} | h_{nt}) \equiv P(r_{nt} | h_{nt}, s_{nt}) P(s_{nt} | h_{nt})$$

and separately estimated $P(s_{nt} | h_{nt})$, the probability of firm type selected as a function of the state variables, from $P(r_{nt} | h_{nt}, s_{nt})$, the selection of rank conditional on both the state variables and also the firm selected.

Table 4 presents our estimates of $P(s_{nt} | h_{nt})$. The columns refer to the type of firm chosen conditional on moving from the current employer, and the state variables are defined by the rows. The omitted (column) choice is to remain employed with the current firm one more period, and the base line (row) category is a college educated Rank 1 executive employed in a firm of type 1.

MBAs go to 7. MSMAs and Ph.D.'s don't transit as much, as we saw in the previous table. controlling for other state variables we now also see that no degree executives also do not move as much as the college educated group. Female behave the same as males. Similarly tenure and male have no significant effects on the probability of an external move. Older execs are more likely to leave and conditional on leaving are less likely to go 3 than the other types.

Perhaps the most striking feature of this table is that when executives move they join firms similar to the ones they left, that is defined in terms of sector and size. Furthermore conditional on moving to a firm of different size, they are more likely to join a firm in the same sector as the one they left. Broadly speaking, the bottom rows, referring to the rank of the executive at the beginning of the period, show that highly ranked executives are less likely to move than the lower ranked ones, evident form the fact that the estimated coefficients increase in each row.

The final column of Table 4 reports on the probability of leaving the sample for at least two years and never returning, a condition we call retirement. The higher the rank the less likely the probability of retirement, indicated by the decreasing sequence of coefficients on rank. Possibly for very different reasons, executives and those without formal qualifications are less likely to exit this sample than groups with other formal education. The indicator variable for gender has a far bigger impact than any of the education variables. Mirroring female labor supply more generally, women in this highly select and lucrative market are more likely to withdraw from it than their male colleagues and competitors. Finally there are significant sector differences.

Finally our estimates of $P(r_{nt} | h_{nt}, s_{nt})$ are presented in Table 5. It shows female executives with a doctorate are more likely to select into the bottom rank. The conditional choice probability estimates shed light on the effects of tenure and age. Here we see that, controlling for all other state variables, last period employer, and this year's employer as well, Rank 2 executives are in fact older than Rank 1 executives, signified by the higher coefficient estimate. Given values of the other observed factors, lower ranked employees have more tenure. The highest coefficients invariably show staying in the same rank is the most likely outcome, and an executive in the lowest rank is more likely to move to Rank i than Rank $i + 1$. Similarly Rank 4 executives are more likely to be demoted than be promoted to Rank 3, evident from the estimated coefficients in Table 4. The results in Table 4 show that relative to other executives, turnover for a Rank 2 manager is more likely than external promotion.

6 Investment versus Moral Hazard

In the concluding section to this paper we assess how much agency problems in executive markets are mitigated by their career concerns. Two of the four metrics we use measure the impact of an executive shirking rather than working. We estimated how much abnormal returns would fall if shareholders failed to incentivize one of its executives but continued to pay the other according to the optimal schedule. This is one measure of how much a firm stands to lose by ignoring the moral hazard problem. The executive, on the other hand, is much more concerned with the compensating differential between diligence and shirking. We computed the compensating differential to an executive from following his interests (shirking) rather than acting according to the interests of the shareholders (working diligently). The other two metrics focus on the cost of eliminating the moral hazard problem. We report on how much the firm pays to induce diligence in the presence of human capital investment, a risk premium for eliminating the moral hazard problem. Finally we calculate how much more a firm would have to pay if executives were not motivated by career concerns, ambition that helps to internalize what would otherwise be a more substantial moral hazard problem.

Each metric was computed using the structural estimates obtained from the previous section, by executive rank, averaged over firm type and executive background. Thus successive rows in Table 6 report a sample average for the rank and its standard deviation, conditional on optimal behavior by the rest of the management team. For the purposes of comparisons with other studies in this literature we also report the estimated risk aversion parameter, the top entry. Quite plausible, and comparable to previous estimates found, we note that an executive with exponential utility and risk aversion parameter of 0.45 would be willing to pay \$217, 790 to insure against an actuarially fair gamble that offers a loss of \$1 million with probability one half and a gain of \$1 million with probability one half.

The first metric is an average over $\tau_{1mk}(h)$, the expected gross loss in the value of the firm of type m in percentage terms if a rank k executive with background h tends his own interests for one year, instead of maximizing the expected value of the firm, that is before netting out the decline in expected compensation all executives would incur from the deteriorating financial performance of the firm. When all executives work diligently, by definition abnormal returns have mean zero, meaning $E[\pi] = 0$. Thus $\tau_{1mk}(h)$ is found by integrating abnormal returns conditional on the executive in question shirking, when every other executive works diligently:

$$\tau_{1mk}(h) \equiv E\left\{\pi\left[1 - g_{mk}(\pi, h)\right)\right]\} = -E\left[\pi g_{mk}(\pi, h)\right]$$

We interpret $\tau_{1mk}(h)$ as a measure of the executive's span of control, because it indicates his potential impact on the firm from behaving irresponsibly. Not surprisingly we find Rank 2 executives exercise the greatest span of control; at 11 percent per year, a chief executives can drive the value of firm equity down to less than half its current value in 8 years, shareholders willing. Similarly, the result that the estimated span of control declines through the middle and lower ranks, confirms our intuition. More remarkable is our finding that executives in Ranks 2 and 3 have a greater span of control than those in Rank 1, as do many in Rank 4.

Taking the manager's perspective rather than the firm's, the compensating differential between working hard and shirking, which we denote by $\tau_{2mk}(h)$, is measured by differencing $w^0_{1mk}(h)$, the manager's reservation certainty equivalent wage to shirk, from $w^0_{2mk}(h)$, the manager's reservation certainty equivalent wage to work diligently under perfect monitoring. Derived from the participation constraint, these certainty equivalents can be expressed as:

$$w^0_{1mk}(h) = \frac{b_{t+1}}{\rho}\log(\alpha^{t+1,1}_{mkt}(h)) + \frac{b_{t+1}}{\rho(b_t - 1)}\log(\alpha_{1mk}/U^E_{mk}(h_m))$$

and

$$w^0_{2mk}(h) = \frac{b_{t+1}}{\rho}\log(\alpha^{t+1,2}_{mkt}(h)) + \frac{b_{t+1}}{\rho(b_t - 1)}\log(\alpha_{2mk}/U^E_{mk}(h_m))$$

Thus

$$\begin{aligned}\tau_{2mk}(h) &\equiv w^0_{2mk}(h) - w^0_{1mk}(h) \\ &= \frac{b_{t+1}}{\rho}\log(\alpha^{t+1,2}_{mkt}(h)/\alpha^{t+1,1}_{mkt}(h)) + \frac{b_{t+1}}{\rho(b_t - 1)}\log\left(\alpha_{2mk}/\alpha_{1mk}\right)\end{aligned}$$

If a manager does not maximize the value of the firm, he gains utility from the nonpecuniary benefits of pursuing his own interests, but does not acquire so much human capital, and thus reduces his chances of higher wages and better positions in the future.

The first factor would also arise in a static model of pure moral hazard where there are no career concerns, and in our formulation does not depend on the executives background characteristics:

$$\tau^{PM}_{2mk} \equiv \frac{b_{t+1}}{\rho(b_t - 1)}\log\left(\alpha_{2mk}/\alpha_{1mk}\right)$$

Our estimates in Table 6 show that contemporaneous nonpecuniary shirking/working benefit differential associated with the Rank 2 position, at \$2.48 million, exceed those associated with any of the other ranks, but that the annual differential from the Rank 1 position is the next highest. Thus Rank 1 has a lesser span of control than Rank 3, but more nonpecuniary benefits. Again these benefits decline through the middle and lower ranks.

The second factor determining $\tau_{2mk}(h)$ reflects those dynamic features of our framework relating to career concerns

$$\tau^{H}_{2mk}(h) \equiv \frac{b_{t+1}}{\rho}\log(\alpha^{t+1,2}_{mkt}(h)/\alpha^{t+1,1}_{mkt}(h))$$

Here we find that, on average, the benefits of human capital accumulation decline monotonically with rank, and that compared with τ_{2mk}^{PM}, are much less dispersed throughout the population of firm types and executive backgrounds. At the lower ranks these benefits are quite considerable. On average a Rank 5 executive is willing to forego \$1.88 million per year because of the greater opportunities working diligently versus shirking affords him, while a Rank 1 executive only values the human capital component of the compensating differential at \$400,000 million per year.

By inspection the compensating differential $\tau_{2mk}(h)$ is the sum of these two factors

$$\tau_{2mk}(h) = \tau_{2mk}^{H}(h) + \tau_{2mk}^{PM}$$

Our estimates imply the compensating differential for every rank except the second is about \$2 million per year, but exceeds \$3 million per year for Rank 2 executives.

How much a firm would be willing to eliminate moral hazard is measured by $\tau_{3mk}(h)$. Under a perfect monitoring scheme shareholders would pay a manager the fixed wage of $w_{2mk}^{0}(h)$, and thus eliminate the risk premium they pay him in the form of a favorable lottery over the outcome of abnormal returns to induce diligent work. Hence the expected value of a perfect monitor to shareholders, denoted $\tau_{3mk}(h)$, is the difference between expected compensation under the current optimal scheme and $w_{2mk}^{0}(h)$, or:

$$\begin{aligned}\tau_3 &\equiv E\left[w_{mk}\left(\pi\right)|h\right] - w_{2mk}^{0}(h) \\ &= E\left[w_{mk}\left(\pi\right)|h\right] - \frac{b_{t+1}}{\rho}\log(\alpha_{mkt}^{t+1,2}(h)) - \frac{b_{t+1}}{\rho(b_t-1)}\log(\alpha_{2mk}/U_{mk}^{E}(h_m))\end{aligned}$$

Our findings in Table 6 show that the firms are prepared to pay hardly anything to eliminate the moral hazard problem at the lower ranks, but that at the Ranks 1 and 3, the benefits of a perfect monitor are considerably more. Curiously, the average risk premium paid to Ranks 1 and 3, \$1.6 million and \$1.7 million respectively, are quite close, despite the fact that the other measures of moral hazard are not.

As one final check on the relevance of human capital to resolving moral hazard problems in the executive market, we estimated the extra premium shareholders would pay to eliminate the moral hazard problem if the benefits of acquiring human capital was ignored by an executive, say because neither the organizational structure nor the market rewarded his diligence. In our model this is represented by:

$$\tau_{4mk}(h) \equiv \frac{b_{t+1}}{\rho}\log(\alpha_{mkt}^{t+1,2}(h))$$

The estimates in Table 6 show that career concerns greatly ameliorate the moral hazard problem for lower level executives but their importance declines monotonically with promotion through the ranks, bordering on irrelevance for many Rank 1 executives.

7 Appendix (incomplete)

Proof of Lemma 1. For all ages $t \in \{1, \ldots, T\}$ we set $A_T(h, b_\tau) \equiv 1$ and recursively define $A_t(h, b_\tau)$ as:

$$\begin{aligned} A_t(h, b_\tau) &= p_{0t}(h, b_\tau) E\left[\exp\left(-\varepsilon^*_{0t}/b_\tau\right)\right] \\ &+ \sum_{j=1}^{J} \sum_{k=1}^{K} p_{jkt}(h, b_\tau)\, \alpha_{jkt}(h)^{\frac{1}{b_\tau}} E \left[\exp\left(-\varepsilon^*_{jkt}/b_\tau\right) E_t[\upsilon_{jk,t+1} A_{t+1}(H_{jk}(h, b_{\tau+1}))]^{1-\frac{1}{b_\tau}}\right] \end{aligned}$$

where ε^*_{jkt} is the value of the disturbance when (j, k) is selected at age t, b_τ is the current price of a perpetual bond at time τ and $H_{jk}(h)$ is the value of the state variables at age $t+1$ induced by the [illegible] averaging over job choices (j, k) using the choice probabilities yields:

$$\begin{aligned}
& V_T\left(h, e_T, a_\tau, b_\tau\right) \\
\equiv\ & -b_\tau \left[p_{0t}\left(h, b_\tau\right) V_{0T}\left(h_T, b_\tau\right) + \sum_{j=1}^{J} \sum_{k=1}^{K} p_{jkt}\left(h, b_\tau\right) V_{jkT}\left(h_T\right) \right] \\
=\ & -b_\tau \exp\left(-\frac{a_\tau + \rho e_T}{b_\tau}\right) \\
& \times \left(\begin{array}{l} p_{0t}\left(h, b_\tau\right) \alpha_0\left(h\right)^{1/b_\tau} E\left[\exp\left(-\varepsilon_{0T}^*/b_\tau\right)\right] + \\ \sum_{j=1}^{J} \sum_{k=1}^{K} p_{jkt}\left(h, b_\tau\right) \alpha_{jkT}\left(h\right)^{1/b_\tau} E\left[\exp\left(-\varepsilon_{jkT}^*/b_\tau\right) E_T\left[\upsilon_{jk,T+1}\right]^{1-\frac{1}{b_\tau}}\right] \end{array} \right) \\
=\ & -b_\tau \exp\left(-\frac{a_\tau + \rho e_T}{b_\tau}\right) A_T\left(h_T, b_\tau\right)
\end{aligned}$$

2. The proof is completed with an induction showing that for all ages $t \in \{1, \ldots, T-1\}$:

$$\begin{aligned}
V_{jkt}\left(h, e_t, a_\tau, b_\tau, \varepsilon_{jkt}^*\right) \equiv\ & -\alpha_{jkt}\left(h\right)^{1/b_\tau} \exp\left(-\varepsilon_{jkt}^*/b_t\right) \\
& \times E_t\left[\upsilon_{jk,t+1} A_{t+1}\left(H_{jk}\left(h\right), b_{\tau+1}\right)\right]^{1-\frac{1}{b_\tau}} b_\tau \exp\left(-\frac{a_\tau + \rho e_t}{b_\tau}\right)
\end{aligned} \tag{33}$$

and

$$V_t\left(h, e_t, a_\tau, b_\tau\right) = -b_\tau \exp\left(-\frac{a_\tau + \rho e_t}{b_\tau}\right) A_t\left(h, b_\tau\right)$$

Suppose both equations are true for all ages $s \in \{t+1, \ldots, T\}$. Given job selection (j, k) the solution to the consumption savings decision at age t is found by maximizing:

$$\begin{aligned}
& -\alpha_{jkt}\left(h\right) \exp\left(-\varepsilon_{jkt}^*\right) \exp\left(-\rho c_t\right) - E_t\left[V_{t+1}\left(H_{jk}\left(h\right), e_{t+1}, a_{\tau+1}, b_{\tau+1}\right)\right] \\
=\ & -\alpha_{jkt}\left(h\right) \exp\left(-\varepsilon_{jkt}^*\right) \exp\left(-\rho c_t\right) - E_t\left[b_{\tau+1} \exp\left(-\frac{a_{\tau+1} + \rho e_{t+1}}{b_{\tau+1}}\right) \upsilon_{jk,t+1} A_{t+1}\left[H_{jk}\left(h_t\right), b_{\tau+1}\right]\right]
\end{aligned}$$

with respect to (c_t, e_{t+1}). Substituting t for T and $\upsilon_{jk,t+1} A_{t+1}\left[H_{jk}\left(h_t\right), b_{\tau+1}\right]$ for $\upsilon_{jk,T+1}$ in Expression (32) above, Expression (??) follows directly. Integrating over $(\varepsilon_{0t}, \varepsilon_{11t}, \ldots, \varepsilon_{JKt})$, the idiosyncratic disturbance vector that is revealed at the beginning of the period, and averaging over the (j, k) job selections yields:

$$\begin{aligned}
& V_t\left(h, e_t, a_\tau, b_\tau\right) \\
=\ & p_{0t}\left(h, b_\tau\right) E_t\left[V_{0t}\left(h, e_t, a_\tau, b_\tau, \varepsilon_{0t}^*\right)\right] + \sum_{j=1}^{J} \sum_{k=1}^{K} p_{jkt}\left(h, b_\tau\right) E_t\left[V_{jkt}\left(h, e_t, a_\tau, b_\tau, \varepsilon_{jkt}^*\right)\right] \\
=\ & -b_\tau \exp\left(\frac{a_\tau + \rho e_t}{b_\tau}\right) \left\{ \begin{array}{l} p_{0t}\left(h, b_\tau\right) E\left[\exp\left(-\varepsilon_{0t}^*/b_t\right)\right] \\ + \sum_{j=1}^{J} \sum_{k=1}^{K} p_{jkt}\left(h, b_\tau\right) \alpha_{jkt}\left(h\right)^{1/b_\tau} E\left[\exp\left(-\varepsilon_{jkt}^*/b_\tau\right) \right. \\ \qquad \left. \times E_t\left[\upsilon_{jk,t+1} A_{t+1}\left(H_{jk}\left(h\right), b_{\tau+1}\right)\right]^{1-\frac{1}{b_\tau}}\right] \end{array} \right\} \\
=\ & -b_\tau \exp\left(\frac{a_\tau + \rho e_t}{b_\tau}\right) A_t\left(h, b_\tau\right)
\end{aligned}$$

the third equality following from the recursive definition of $A_t\left(h, b_\tau\right)$.

3. To prove Equation (5) in the text, we differentiate the right side of Equation (??) to obtain the optimal consumption as the unique stationary point for this concave programming maximization problem.

■

Proof of Lemma 2. From Expression (??), the conditional valuation functions take the form of:

$$V_{jkt}\left(h, e_t, a_\tau, b_\tau, \varepsilon^*_{jkt}\right) \equiv -\alpha_{jkt}(h)^{1/b_\tau} \exp\left(-\varepsilon^*_{jkt}/b_t\right) \times E_t\left[\upsilon_{jk,t+1} A_{t+1}\left(H_{jk}(h), b_{\tau+1}\right)\right]^{1-\frac{1}{b_\tau}} b_\tau \exp\left(-\frac{a_\tau + \rho e_t}{b_\tau}\right)$$

The manager optimizes his expected lifetime utility at age t by choosing the highest valued conditional valuation function, given by Expression (??) over (j,k), or to retire. Noting the solution to this problem is invariant to monotone transformations, he picks (j,k) or retires to minimize the negative of the logarithm of Expression (??):

$$\ln b_\tau + \frac{1}{b_\tau}\ln \alpha_{jkt}(h) - \frac{\varepsilon_{jkt}}{b_\tau} + \left(1 - \frac{1}{b_\tau}\right) \ln E_t\left[\upsilon_{jk,t+1} A_{t+1}\left(H_{jk}(h), b_{\tau+1}\right)\right] - \frac{a_t + \rho e_t}{b_\tau}$$

where the logarithm of $-V_{0t}(h_t)$, the conditional valuation function from retiring, specializes to $[\ln b_t - (\varepsilon_{0t} + a_t + \rho e_t)/b_t]$. Subtracting $[\ln b_t - (a_t + \rho e_t)/b_t]$ from each conditional valuation function, and multiplying by $-b_t$ yields the objects in the maximization problem defined in the Lemma, as required. ■

Proof of Lemma 3. The proof of this lemma essentially follows the proofs of Lemma 1 and 2 by substituting B for A. ■

Proof of Lemma 4. Since the expectations operator preserves linearity, both the participation constraint (7) and the incentive compatibility constrain (8) can be expressed as linear in $\upsilon_{jk,t+1}$, namely:

$$\exp\left\{q_{jk}\left[P_t(h_t, b_\tau)\right]\right\}^{1/(b_\tau - 1)} \le \alpha_{jkt}(h)^{1/(b_\tau-1)} E_t\left\{\upsilon_{jk,t+1} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right]\right\}$$

and:

$$\begin{aligned} & \alpha_{jkt}(h)^{1/(b_\tau-1)} E_t\left\{\upsilon_{jk,t+1} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right]\right\} \\ \le \; & \beta_{jkt}(h)^{1/(b_\tau-1)} E_t\left[\upsilon_{jk,t+1} g_{jkt}(\pi, h) B_{t+1}\left(H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right)\right] \end{aligned}$$

The objective function, the expected wage bill $E_t(w_{jk,t+1})$ can be expressed as a concave function of $\upsilon_{jk,t+1}$, namely $E_t(\ln \upsilon_{jk,t+1})$. Therefore the Kuhn Tucker Theorem applies, and the Lagrangian for the problem in which the j^{th} firm elicits diligent work for the k^{th} rank can be written as:

$$\begin{aligned} & E_t[\ln(\upsilon_{jk,t+1})] + \eta_0 E_t\left[\begin{array}{r} \upsilon_{jk,t+1}\alpha_{jkt}(h)^{1/(b_\tau-1)} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right] \\ - \exp\left\{q_{jk}\left[P_t(h_t, b_\tau)\right]\right\}^{1/(b_\tau-1)} \end{array}\right] \\ & +\eta E_t\left[\upsilon_{jk,t+1}\left(\begin{array}{r} \beta_{jkt}(h)^{1/(b_\tau-1)} g_{jkt}(\pi, h) B_{t+1}\left[H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right] \\ -\alpha_{jkt}(h)^{1/(b_\tau-1)} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right] \end{array}\right)\right] \end{aligned}$$

The first order condition is:

$$
\begin{aligned}
\upsilon_{j,k,t+1}^{-1} &= -\eta_0 \alpha_{jkt}(h)^{1/(b_\tau-1)} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right] \\
&\quad -\eta \left\{ \begin{array}{l} \beta_{jkt}(h)^{1/(b_\tau-1)} g_{jkt}(\pi, h) B_{t+1}\left[H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right] \\ \qquad -\alpha_{jkt}(h)^{1/(b_\tau-1)} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right] \end{array} \right\}
\end{aligned} \tag{34}
$$

We multiply this equation by $\upsilon_{j,k,t+1}$, add $\eta_0 \exp\{q_{jk}[P_t(h_t, b_\tau)]\}^{1/(b_\tau-1)}$ to both sides, and take expectations to solve for η_0 obtaining:

$$
\eta_0 = \exp\{q_{jk}[P_t(h_t, b_\tau)]\}^{1/(b_\tau-1)}
$$

Substituting for η_0 back into the first order condition yields:

$$
\begin{aligned}
\upsilon_{j,k,t+1}^{-1} &= -\alpha_{jkt}(h)^{1/(b_\tau-1)} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right] \exp\{q_{jk}[P_t(h_t, b_\tau)]\}^{1/(b_\tau-1)} \\
&\quad -\eta \left\{ \begin{array}{l} \beta_{jkt}(h)^{1/(b_\tau-1)} g_{jkt}(\pi, h) B_{t+1}\left[H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right] \\ \qquad -\alpha_{jkt}(h)^{1/(b_\tau-1)} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right] \end{array} \right\}
\end{aligned}
$$

and substituting for $\upsilon_{j,k,t+1}$ from Equation (3), upon taking logarithms of both sides, we obtain:

$$
\begin{aligned}
\frac{\rho w_{jk,t+1}}{b_{\tau(t)+1}} &= -\alpha_{jkt}(h)^{1/(b_\tau-1)} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right] \exp\{q_{jk}[P_t(h_t, b_\tau)]\}^{1/(b_\tau-1)} \\
&\quad -\eta \left\{ \begin{array}{l} \beta_{jkt}(h)^{1/(b_\tau-1)} g_{jkt}(\pi, h) B_{t+1}\left[H'_{jk}(h), H_{jk}(h), b_{\tau+1}\right] \\ \qquad -\alpha_{jkt}(h)^{1/(b_\tau-1)} A_{t+1}\left[H_{jk}(h_t), b_{\tau+1}\right] \end{array} \right\}
\end{aligned}
$$

from which the compensation schedule given in the lemma's statement follows by substituting in the the formula. Finally the solution for η is obtained by substituting the equation above into the incentive compatibility constraint and simplifying. ■

Proof of Theorem 1. The proof comprises five steps and a concluding lemma. First we show that $\tilde{A}_t\left(h, b_{\tau(t)}\right) = A_t\left(h, b_{\tau(t)}\right)$. Then we prove:

$$
\begin{aligned}
\tilde{g}_{jkt}(\pi|h) &= \frac{\overline{\upsilon}_{jk,t+1}(\tilde{p})^{-1} - \upsilon_{jk,t+1}(\tilde{p}, \pi)^{-1}}{\overline{\upsilon}_{jk,t+1}(\tilde{p})^{-1} - E_t[\upsilon_{jk,t+1}(\tilde{p}, \pi)]^{-1}} \\
&\quad \times \frac{E_t\left\{B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]\right\}}{B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]}
\end{aligned}
$$

The third step shows $\tilde{B}_t\left(h, h', b_{\tau(t)}\right) = B_t\left(h, h', b_{\tau(t)}\right)$. Then we prove:

$$
\tilde{\alpha}_{jkt}(h) = E_t\left\{\upsilon_{jk,t+1}(\tilde{p}, \pi) A_{t+1}\left[H_{jk}(h), b_{\tau(t)+1}\right]\right\}^{(1-b_{\tau(t)})} \exp\left\{\tilde{q}_{jk}\left[P_t\left(h, b_{\tau(t)}\right)\right]\right\} \tag{35}
$$

while the fifth step is to show:

$$
\begin{aligned}
\tilde{\beta}_{jkt}(h) &= \exp\left\{\tilde{q}_{jk}\left[P_t\left(h, b_{\tau(t)}\right)\right]\right\} E_t\left\{B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]\right\}^{1-b_{\tau(t)}} \\
&\quad \times \left\{ E_t[\tilde{v}_{jk,t+1}(\tilde{p})] \frac{\overline{\upsilon}_{jk,t+1}(\tilde{p})^{-1} - E_t[\upsilon_{jk,t+1}(\tilde{p}, \pi)]^{-1}}{\overline{\upsilon}_{jk,t+1}(\tilde{p})^{-1} - E_t[\upsilon_{jk,t+1}(\tilde{p}, \pi)^{-1}]} \right\}^{1-b_{\tau(t)}}
\end{aligned} \tag{36}
$$

Combining these five steps demonstrates that $\tilde{\theta}$ satisfies Equations (35), (17) and (36). Lemma 7 proves that if any $\hat{\theta} \in \Theta$ satisfies those equations, it is observationally equivalent to $\tilde{\theta}$.

1. Comparing the definition of $A_t\left(h, b_{\tau(t)}\right)$, given by Equation (4), with the definition of $\hat{A}_t\left(h, b_{\tau(t)}\right)$, given in Equation (13) by inspection that $\tilde{A}_t\left(h, b_{\tau(t)}\right) = A_t\left(h, b_{\tau(t)}\right)$.

2. The second equation is derived from the first order condition of the optimization problem for the contract. Following the approach of Gayle and Miller (2009b) we substitute the solution for η_{0t} into (34) the first order condition for the optimization problem to obtain:

$$v_{jk,t+1}(\pi)^{-1} = E_t\left[v_{jk,t+1}(\pi)\right]^{-1} + \eta \left\{ \begin{array}{c} \left[\alpha_{jkt}(h)/\beta_{jkt}(h)\right]^{1/(b-1)} A_{t+1}\left[H_{jk}(h), b_{\tau(t)+1}\right] \\ -g_{jkt}(\pi, h)\, B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right] \end{array} \right\}$$

Taking expectations yields:

$$E_t\left[v_{jk,t+1}(\pi)^{-1}\right] = E_t\left[v_{jk,t+1}(\pi)\right]^{-1} + \eta \left\{ \begin{array}{c} \left[\alpha_{jkt}(h)/\beta_{jkt}(h)\right]^{1/(b-1)} A_{t+1}\left[H_{jk}(h), b_{\tau(t)+1}\right] \\ -E_t\left\{B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]\right\} \end{array} \right\}$$

Also:

$$\overline{v}_{jk,t+1}^{-1} = E\left[v_{jk,t+1}(\pi)\right]^{-1} + \eta\left\{\left[\alpha_{jkt}(h)/\beta_{jkt}(h)\right]^{1/(b-1)} A_{t+1}\left[H_{jk}(h), b_{\tau(t)+1}\right]\right\}$$

Making η the subject of the difference of the second two equations we obtain:

$$\eta = \frac{\overline{v}_{jk,t+1}^{-1} - E_t\left[v_{jk,t+1}(\pi)^{-1}\right]}{E_t\left\{B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]\right\}}$$

Subtracting $v_{jk,t+1}(\pi)^{-1}$ from $\overline{v}_{2t}^{-1}$ gives:

$$\overline{v}_{2t}^{-1} - v_{jk,t+1}(\pi)^{-1} = \eta g_{jkt}(\pi, h)\, B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]$$

or

$$g_{jkt}(\pi, h) = \frac{\overline{v}_{2t}^{-1} - v_{jk,t+1}(\pi)^{-1}}{\eta B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]}$$

so substituting for η we obtain upon rearrangement:

$$g_{jkt}(\pi, h) = \frac{\overline{v}_{2t}^{-1} - v_{jk,t+1}(\pi)^{-1}}{\overline{v}_{jk,t+1}^{-1} - E_t\left[v_{jk,t+1}(\pi)^{-1}\right]} \frac{E_t\left\{B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]\right\}}{B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]} \tag{37}$$

as required.

3. Finally:

$$
\begin{aligned}
B_t\left(h,h',b_{\tau(t)}\right) &= p_{0t}\left(h,h',b_{\tau(t)}\right) E\left[\exp\left(-\varepsilon^*_{0t}/b_{\tau(t)}\right)\right] \\
&+ \sum_{(j,k)} \left[\begin{array}{l} p_{jkt}\left(h,h',b_{\tau(t)}\right) E[\exp\left(-\varepsilon^*_{jkt}/b_{\tau(t)}\right)] \\ \times \max \left\{\begin{array}{l} \alpha_{jkt}(h)^{\frac{1}{b_{\tau(t)}}} E_t\left[\upsilon_{jk,t+1} B_{t+1}\left(H_{jk}(h), H_{jk}(h'), b_{\tau(t)+1}\right)\right]^{1-\frac{1}{b_{\tau(t)}}}, \\ \beta_{jkt}(h)^{\frac{1}{b_{\tau(t)}}} E_t\left[\upsilon_{jk,t+1} g\left(\pi_{j,\tau(t)+1}, h\right) B_{t+1}\left(H'_{jk}(h), H_{jk}(h'), b_{\tau(t)+1}\right)\right]^{1-\frac{1}{b_{\tau(t)}}} \end{array}\right\} \end{array}\right]
\end{aligned}
$$

A $\widehat{B}_t\left(h,h',b_{\tau(t)}\right)$ mapping is derived in a similar way. Letting $\widehat{B}_T\left(h,h',b_{\tau(t)}\right) = 1$ we write:

$$
\exp\left(q_{jk}\left(P_t\left(h',b_\tau\right)\right)\right)^{1/b_\tau} = \alpha_{jkt}\left(h'\right)^{1/b_\tau} E_t\left(\upsilon_{jk,t+1} A_{t+1}\left(H_{jk}\left(h'\right), b_{\tau+1}\right)\right)^{1-1/b_\tau}
$$

$$
\begin{aligned}
\widehat{B}_t\left(h,h',b_{\tau(t)}\right) &= p_{0t}\left(h,h',b_{\tau(t)}\right) E\left[\exp\left(-\varepsilon^*_{0t}/b_{\tau(t)}\right)\right] \\
&+ \sum_{(j,k)} \left[\begin{array}{l} p_{jkt}\left(h,h',b_{\tau(t)}\right) E[\exp\left(-\varepsilon^*_{jkt}/b_{\tau(t)}\right)] \\ \times \max \left\{\begin{array}{l} \alpha_{jkt}(h)^{\frac{1}{b_{\tau(t)}}} E_t\left[\upsilon_{jk,t+1} B_{t+1}\left(H_{jk}(h), H_{jk}(h'), b_{\tau(t)+1}\right)\right]^{1-\frac{1}{b_{\tau(t)}}}, \\ \beta_{jkt}(h)^{\frac{1}{b_{\tau(t)}}} E_t\left[\upsilon_{jk,t+1} g\left(\pi_{j,\tau(t)+1}, h\right) B_{t+1}\left(H'_{jk}(h), H_{jk}(h'), b_{\tau(t)+1}\right. \end{array}\right. \end{array}\right.
\end{aligned}
$$

Substituting for:

$$
\alpha_{jkt}\left(h'\right)^{1/b_\tau} E_t\left(\upsilon_{jk,t+1} A_{t+1}\left(H_{jk}\left(h'\right), b_{\tau+1}\right)\right)^{1-1/b_\tau}
$$

we get:

$$
\begin{aligned}
\widehat{B}_t\left(h,h',b_{\tau(t)}\right) &= p_{0t}\left(h,h',b_{\tau(t)}\right) E\left[\exp\left(-\varepsilon^*_{0t}/b_{\tau(t)}\right)\right] \\
&+ \sum_{(j,k)} \left[\begin{array}{l} p_{jkt}\left(h,h',b_{\tau(t)}\right) E[\exp\left(-\varepsilon^*_{jkt}/b_{\tau(t)}\right)] \\ \times \max \left\{\begin{array}{l} \exp\left\{q_{jk}\left[P_t\left(h',b_\tau\right)\right]\right\}^{1/b_\tau}, \\ \beta_{jkt}(h)^{\frac{1}{b_{\tau(t)}}} E_t\left[\upsilon_{jk,t+1} g\left(\pi_{j,\tau(t)+1}, h\right) B_{t+1}\left(H'_{jk}(h), H_{jk}(h'), b_{\tau(t)}\right.\right. \end{array}\right. \end{array}\right.
\end{aligned}
$$

Also:

$$
g_{jkt}(\pi,h) = \frac{\overline{v}_{2t}^{-1} - \upsilon_{jk,t+1}(\pi)^{-1}}{\overline{v}_{jk,t+1}^{-1} - E_t\left[\upsilon_{jk,t+1}(\pi)^{-1}\right]} \frac{E_t\left\{B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]\right\}}{B_{t+1}\left[H'_{jk}(h_t), H_{jk}(h_t), b_{\tau(t)+1}\right]} \tag{38}
$$

so we put in:

$$
\begin{aligned}
\widehat{B}_t\left(h,h',b_{\tau(t)}\right) &= p_{0t}\left(h,h',b_{\tau(t)}\right) E\left[\exp\left(-\varepsilon^*_{0t}/b_{\tau(t)}\right)\right] \\
&+ \sum_{(j,k)} \left[\begin{array}{l} p_{jkt}\left(h,h',b_{\tau(t)}\right) E[\exp\left(-\varepsilon^*_{jkt}/b_{\tau(t)}\right)] \\ \times \max \left\{\begin{array}{l} \exp\left\{q_{jk}\left[P_t\left(h',b_\tau\right)\right]\right\}^{1/b_\tau}, \\ \beta_{jkt}(h)^{\frac{1}{b_{\tau(t)}}} E_t\left(\frac{\upsilon_{jk,t+1} - \overline{v}_{2t}}{1-\overline{v}_{2t}^{-1} E_t\left[\upsilon_{jk,t+1}(\pi)^{-1}\right]}\right)^{1-\frac{1}{b_{\tau(t)}}} E_t\left\{B_{t+1}\left[H'_{jk}(h_t), H\right.\right. \end{array}\right. \end{array}\right.
\end{aligned}
$$

4. Since the participation constraint, the inequality (7), is binding in the optimal contract:

$$\exp\{q_{jk}[P_t(h_t,b_\tau)]\}^{1/(b_\tau-1)} = \alpha_{jkt}(h)^{1/(b_\tau-1)} E_t\{\upsilon_{jk,t+1}A_{t+1}[H_{jk}(h_t),b_{\tau+1}]\}$$

Making $\alpha_{jkt}(h')$ the subject of the equation yields:

$$\alpha_{jkt}(h') = \exp\{q_{jk}[P_t(h',b_\tau)]\} E_t\{\upsilon_{jk,t+1}A_{t+1}[H_{jk}(h'),b_{\tau+1}]\}^{(1-b_\tau)} \tag{39}$$

which is the equation that obtains by setting $\theta = \hat{\theta}$ in Equation (35).

5. The incentive compatibility constraint (8) is also satisfied with equality implying that in equilibrium:

$$\begin{aligned} & \beta_{jkt}(h) E_t\{\upsilon_{jk,t+1}g_{jkt}(\pi,h) B_{t+1}[H'_{jk}(h),H_{jk}(h),b_{\tau+1}]\}^{(b_\tau-1)} \\ = \; & \alpha_{jkt}(h) E_t\{\upsilon_{jk,t+1}A_{t+1}[H_{jk}(h),b_{\tau+1}]\}^{(b_\tau-1)} \end{aligned}$$

Substituting for $\alpha_{jkt}(h)$ using Equation (39) and making $\beta_{jkt}(h)$ the subject of the equation yields:

$$\beta_{jkt}(h) = \exp\{q_{jk}[P_t(h',b_\tau)]\} E_t\{\upsilon_{jk,t+1}g_{jkt}(\pi,h) B_{t+1}[H'_{jk}(h),H_{jk}(h),b_{\tau+1}]\}^{(1-b_\tau)}$$

Finally substituting for $g_{jkt}(\pi,h)$ from Equation (37) into the equation above for $\beta_{jkt}(h)$ returns Equation (36) evaluated at $\theta = \hat{\theta}$:

$$\beta_{jkt}(h) = \exp\{q_{jk}[P_t(h',b_\tau)]\} \left\{ \frac{E_t[\upsilon_{jk,t+1}] - \overline{\upsilon}_{jk,t+1}}{1 - E_t[\upsilon_{jk,t+1}(\pi)^{-1}]\,\overline{\upsilon}_{jk,t+1}} \right\}^{(1-b_\tau)} E_t\{B_{t+1}[H'_{jk}(h_t),H_{jk}(h_t),b_{\tau(t}$$

■

Proof of Lemma 6. Given the probability distribution of choices and wages, Lemma 5 uniquely define $\alpha_{1jk}(h)$, $\alpha_{1jk}(h)$ and $g_{jk}(\pi,h)$ for any ρ and $\varphi(\varepsilon)$. We now show that if the equations in Lemma 5 are satisfied, then the compensation contract observed in the data is optimal for the parameterization. Abbreviating the probability

1. From the first equation in Lemma 6 it follows that

$$\widehat{\alpha}_{1jk}\left(h_t\right) = E_t\left[\upsilon_{jk,t+1}\right]\exp\left\{-\widehat{q}_{jk}\left[P_t\left(h_t\right)\right] - \left(b-1\right)\widehat{A}_{s-1}\left[H_{jk}\left(h_t\right)\right]\right\}$$

   implying the participation constraint is met with equality, as required by the solution to the optimization problem.

2. From the definitions of $\alpha_{0jk}\left(h_t\right)$, $\alpha_{1jk}\left(h_t\right)$ and $\widehat{g}_{jk}\left(\pi, h\right)$ given in the third equation of Lemma 6:

$$\frac{A_{s-1}\left[H_{jk}\left(h_t\right)\right]}{B_{s-1}\left[H'_{jk}\left(h_t\right), H_{jk}\left(h_t\right)\right]}\left(\frac{\alpha_{1jk}\left(h_t\right)}{\alpha_{0jk}\left(h_t\right)}\right)^{\frac{1}{b_t-1}} = \frac{E_t\left[\upsilon_{jk,t+1}g\left(\pi, h_t\right)\right]}{E_t\left[\upsilon_{jk,t+1}\right]}$$

   Substituting in the definition of $\widehat{g}_{jk}\left(\pi, h\right)$ given in the second equation of Lemma 6 we obtain:

$$\frac{A_{s-1}\left[H_{jk}\left(h_t\right)\right]}{B_{s-1}\left[H'_{jk}\left(h_t\right), H_{jk}\left(h_t\right)\right]}\left(\frac{\alpha_{1jk}\left(h_t\right)}{\alpha_{0jk}\left(h_t\right)}\right)^{\frac{1}{b_t-1}} = \frac{\overline{\upsilon}_t^{-1} - E_t\left[\upsilon_{jk,t+1}\right]^{-1}}{\overline{\upsilon}_t^{-1} - E_t\left[\upsilon_{jk,t+1}\left(\pi\right)^{-1}\right]}$$

   Hence:

$$\widehat{g}_{jk}\left(\pi, h\right) - \frac{A_{s-1}\left[H_{jk}\left(h_t\right)\right]}{B_{s-1}\left[H'_{jk}\left(h_t\right), H_{jk}\left(h_t\right)\right]}\left(\frac{\alpha_{1jk}\left(h_t\right)}{\alpha_{0jk}\left(h_t\right)}\right)^{\frac{1}{b_t-1}} = \frac{\overline{\upsilon}_{2t}^{-1} - \upsilon_{jk,t+1}\left(\pi\right)^{-1}}{\overline{\upsilon}_{2t}^{-1} - E_2\left[\upsilon_{jk,t+1}\left(\pi\right)^{-1}\right]} - \frac{\overline{\upsilon}_t^{-1} - E_t\left[\upsilon_{jk,t+1}\right]^{-1}}{\overline{\upsilon}_t^{-1} - E_t\left[\upsilon_{jk,t+1}\left(\pi\right)^{-1}\right]}$$

   Multiplying both sides by:

$$\overline{\upsilon}_{2t}^{-1} - E_2\left[\upsilon_{jk,t+1}\left(\pi\right)^{-1}\right] = \widehat{\eta}_t B_{s-1}\left[H'_{jk}\left(h_t\right), H_{jk}\left(h_t\right)\right]$$

   yields:

$$\widehat{\eta}_t\left\{\widehat{g}_{jk}\left(\pi, h\right)B_{s-1}\left[H'_{jk}\left(h_t\right), H_{jk}\left(h_t\right)\right] - A_{s-1}\left[H_{jk}\left(h_t\right)\right]\left(\frac{\alpha_{1jk}\left(h_t\right)}{\alpha_{0jk}\left(h_t\right)}\right)^{\frac{1}{b_t-1}}\right\} = E_t\left[\upsilon_{jk,t+1}\right]^{-1} - \upsilon_{jk,t+1}\left(\pi\right)^{-1} \tag{40}$$

   By definition

$$\eta' = \exp\left\{-q_{jk}\left[P_t\left(h_t\right)\right]\right\}^{1/(b-1)}$$

   Substituting η' into the definition of $\alpha_{1jk}\left(h_t\right)$ given in Lemma 6 yields

$$\begin{aligned}\alpha_{1jk}\left(h_t\right) &= E_t\left[\upsilon_{jk,t+1}\right]^{(b-1)}A_{s-1}\left[H_{jk}\left(h_t\right)\right]^{(b-1)}\exp\left\{-q_{jk}\left[P_t\left(h_t\right)\right]\right\} \\ &= E_t\left[\upsilon_{jk,t+1}\right]^{(b-1)}A_{s-1}\left[H_{jk}\left(h_t\right)\right]^{(b-1)}\eta'^{(b-1)}\end{aligned}$$

   Making $E_t\left[\upsilon_{jk,t+1}\right]$ the subject of the equation:

$$E_t\left[\upsilon_{jk,t+1}\right] = \alpha_{1jk}\left(h_t\right)^{1/(b-1)}A_{s-1}\left[H_{jk}\left(h_t\right)\right]\eta'$$

   Inverting and substituting the resulting expression for $E_t\left[\upsilon_{jk,t+1}\right]^{-1}$ now yields the first order condition in for $\upsilon_{jk,t+1}\left(\pi\right)$.

3. From its definition $\tilde{\eta}_t > 0$. Also the equation above implies:

$$\tilde{y}_{jk}(\pi,h) - \frac{A_{s-1}\left[H_{jk}(h_t)\right]}{B_{s-1}\left[H'_{jk}(h_t), H_{jk}(h_t)\right]}\left(\frac{\alpha_{1jk}(h_t)}{\alpha_{0jk}(h_t)}\right)^{\frac{1}{b_t-1}} = \frac{E_t\left[v_{jk,t+1}\right]^{-1} - v_{jk,t+1}(\pi)^{-1}}{\overline{v}_{2t}^{-1} - E_2\left[v_{jk,t+1}(\pi)^{-1}\right]}$$

   Multiplying through by $v_{jk,t+1}(\pi)$ and taking the expectation with respect to π conditional (h,j,k) state yields:

$$\begin{aligned}
& E_t\left\{\left[\tilde{y}_{jk}(\pi,h) - \frac{A_{s-1}\left[H_{jk}(h_t)\right]}{B_{s-1}\left[H'_{jk}(h_t), H_{jk}(h_t)\right]}\left(\frac{\alpha_{1jk}(h_t)}{\alpha_{0jk}(h_t)}\right)^{\frac{1}{b_t-1}}\right] v_{jk,t+1}(\pi)\right\} && (41)\\
= \; & \tilde{\eta}_t^{-1} E_t\left\{\left[E_t\left[v_t(x,\tilde{\eta})\right]^{-1} - v_t(x,\tilde{\eta})^{-1}\right] v_t(x,\tilde{\eta})\right\} && (42)\\
= \; & \tilde{\eta}_t^{-1} E_t\left\{v_t(x,\tilde{\eta}) / E_t\left[v_t(x,\tilde{\eta})\right] - 1\right\} \\
= \; & 0
\end{aligned}$$

   Therefore the incentive compatibility condition is satisfied with equality.

4. By construction $\eta'_t(\tilde{\eta})$ is the Lagrange multiplier for the participation constraint. From the Equation (41):

$$E_t\left\{v_{st}(x,\tilde{\eta})\left[\tilde{y}_{jk}(\pi,h) - \frac{A_{s-1}\left[H_{jk}(h_t)\right]}{B_{s-1}\left[H'_{jk}(h_t), H_{jk}(h_t)\right]}\left(\frac{\alpha_{1jk}(h_t)}{\alpha_{0jk}(h_t)}\right)^{\frac{1}{b_t-1}}\right]\right\} = 0$$

   Using the first order condition as expressed by Equation (40) to substitute out $v_t(x,\tilde{\eta})$ yields $\tilde{\eta}_t$ as a solution to the equation:

$$E_2\left\{\frac{B_{s-1}\left[H'_{jk}(h_t), H_{jk}(h_t)\right]\tilde{y}_{jk}(\pi,h) - A_{s-1}\left[H_{jk}(h_t)\right]\left[\frac{\alpha_{1jk}(h_t)}{\alpha_{0jk}(h_t)}\right]^{\frac{1}{b_t-1}}}{\eta\tilde{y}_s(x) - \eta\left(\tilde{\eta}_2/\tilde{\eta}_1\right)^{1/(b_t-1)} - E_s\left[v_{st}(x,\tilde{\eta})\right]^{-1}}\right\} = 0$$

   in η which defines the Kuhn Tucker multiplier associated with the incentive compatibility constraint.

■

Proof of Lemma 7. For notational convenience define:

$$W_{jklt} \equiv \ln\alpha_{jklt} + (b_t-1)\ln A_{s-1}\left[H_{jk}(h_t)\right] + (b_t-1)\log\left\{E[v_{jk,t+1}|h_t,l_t]\right\}$$

Then (j,k) is chosen if:

$$\varepsilon_{jkt} + W_{jklt} \geq \varepsilon_{j'k't} + W_{j'k'lt}$$

for $l = l_t$. Let $G(\varepsilon_{11t},\ldots,\varepsilon_{JKt})$ denote the probability distribution function for $(\varepsilon_{11t},\ldots,\varepsilon_{JKt})$ and $G_{jk}(\varepsilon_{11t},\ldots,\varepsilon_{JKt})$ its derivative with respect to ε_{jkt}. Since $G(\varepsilon_{11t},\ldots,\varepsilon_{JKt})$ is the product of independently distributed standard Type 1 Extreme value probability distributions in our model :

$$G_{jk}(\varepsilon_{11t},\ldots,\varepsilon_{JKt}) = \exp(-\varepsilon_{jkt})\prod_{(j',k')}\exp\left[-\exp\left(-\varepsilon_{j'k't}\right)\right]$$

Using the well known fact that:

$$W_{jklt} - W_{j'k'lt} = \log p_{jkt} - \log p_{j'k't}$$

it now follows that :

$$\begin{aligned}
& G_{jk}\left(\varepsilon_{jkt} + W_{jklt} - W_{11lt}, \ldots, \varepsilon_{jkt} + W_{jklt} + W_{JKlt}\right) \\
= \; & \exp\left(-\varepsilon_{jkt}\right) \prod_{(j',k')} \exp\left[-\exp\left(-\varepsilon_{jkt} + W_{j'k'lt} - W_{jklt}\right)\right] \\
= \; & \exp\left[-\varepsilon_{jkt} - \sum_{(j',k')} \exp\left(-\varepsilon_{jkt} + W_{j'k'lt} - W_{jklt}\right)\right] \\
= \; & \exp\left[-\varepsilon_{jkt} - \exp\left(-\varepsilon_{jkt}\right) \sum_{(j',k')} \exp\left(\log p_{j'k't} - \log p_{jkt}\right)\right] \\
= \; & \exp\left[-\varepsilon_{jkt} - \exp\left(-\varepsilon_{jkt} - \log p_{jkt}\right) \sum_{(j',k')} \exp\left(\log p_{j'k't}\right)\right] \\
= \; & \exp\left[-\varepsilon_{jkt} - \exp\left(-\varepsilon_{jkt} - \log p_{jkt}\right)\right]
\end{aligned}$$

From Equation the conditional choice probability for (j, k) can be expressed as

$$p_{jkt} = \int_{-\infty}^{\infty} G_{jk}\left(\varepsilon_{jkt} + W_{jklt} - W_{11lt}, \ldots, \varepsilon_{jkt} + W_{jklt} + W_{JKlt}\right) d\varepsilon_{jkt}$$

Hence the probability density function of $\varepsilon^*_{jkt} \equiv d_{jk}\varepsilon_{jkt}$ is Type 1 extreme value with location parameter $-\log p_{jkt}$ and unit scale parameter since:

$$\begin{aligned}
\overline{\varphi}\left(\varepsilon^*_{jkt}\right) &= p^{-1}_{jkt} \frac{\partial}{\partial \varepsilon^*_{jkt}} \left[\int_{-\infty}^{\varepsilon^*_{jkt}} G_{jk}\left(\varepsilon_{jkt} + W_{jklt} - W_{11lt}, \ldots, \varepsilon_{jkt} + W_{jklt} + W_{JKlt}\right) d\varepsilon_{jkt}\right] \\
&= p^{-1}_{jkt} \exp\left[-\varepsilon^*_{jkt} - \exp\left(-\varepsilon^*_{jkt} - \log p_{jkt}\right)\right] \\
&= \exp\left[-\varepsilon^*_{jkt} - \log p_{jklt} - \exp\left(-\varepsilon^*_{jkt} - \log p_{jkt}\right)\right]
\end{aligned}$$

To derive:

$$E\left[\exp\left(-\varepsilon^*_{jkt}/b_t\right)\right]$$

we draw from Equations (15) and (17) of Chapter 21 of Johnston and Kotz (1970, pages 277 - 278) proving that the moment generating function for ε^*_{jkt} is:

$$E\left[\exp\left(t\varepsilon^*_{jkt}\right)\right] = \exp\left(-t \log p_{jkt}\left(h, b_\tau\right)^{1/b_t}\right) \Gamma\left(1 - t\right)$$

Setting $t = -b^{-1}_t$ this simplifies to:

$$E\left[\exp\left(\varepsilon^*_{jkt}/b_t\right)\right] = \exp\left(\log p_{jkt}\left(h, b_\tau\right)^{1/b_t}\right) \Gamma\left[\left(b_t + 1\right)/b_t\right] = p_{jkt}\left(h, b_\tau\right)^{1/b_t} \Gamma\left[\left(b_t + 1\right)/b_t\right]$$

1. Finally we prove the formula for $A_t\left(h, b_{\tau(t)}\right)$. When ε is distributed Type 1 Extreme Value then:

$$q_{jk}\left[P_t\left(h, b_\tau\right)\right] = \ln p_{0t}\left(h, b_\tau\right) - \ln p_{jkt}\left(h, b_\tau\right)$$

and the proof to the optimal contract shows that the participation constraint is met with equality, which implies:

$$\alpha_{jkt}(h)^{1/(b_\tau-1)} E_t\left\{\upsilon_{jk,t+1}A_{t+1}\left[H_{jk}(h), b_{\tau+1}\right]\right\} = \left[p_{0t}(h,b_\tau)/p_{jkt}(h,b_\tau)\right]^{1/(b_\tau-1)}$$

Rearranging this equality yields:

$$\alpha_{jkt}(h)^{1/b_\tau} E_t\left\{\upsilon_{jk,t+1}A_{t+1}\left[H_{jk}(h), b_{\tau+1}\right]\right\}^{\frac{b_\tau-1}{b_\tau}} = \left[p_{jkt}(h,b_\tau)/p_{0t}(h,b_\tau)\right]^{1/b_\tau}$$

Substituting for:

$$\alpha_{jkt}(h)^{1/b_\tau} E_t\left\{\upsilon_{jk,t+1}A_{t+1}\left[H_{jk}(h), b_{\tau+1}\right]\right\}^{\frac{b_\tau-1}{b_\tau}}$$

in the recursion for $A_{t+1}(h,b_\tau)$ given in (4) we now obtain:

$$\begin{aligned}
A_t(h,b_\tau) &= p_{0t}(h,b_\tau)^{1+\frac{1}{b_\tau}}\Gamma\left[(b_\tau+1)/b_\tau\right] \\
&\quad + \sum_{j=1}\sum_{k=1}\left(p_{jkt}(h,b_\tau)^{1+\frac{1}{b_\tau}}\Gamma\left[(b_\tau+1)/b_\tau\right]\left[\frac{p_{0t}(h,b_\tau)}{p_{jkt}(h,b_\tau)}\right]^{1/b_\tau}\right) \\
&= p_{0t}(h,b_\tau)^{\frac{1}{b_\tau}}\Gamma\left[(b_\tau+1)/b_\tau\right]\left\{p_{0t}(h,b_\tau)+\sum_{j=1}\sum_{k=1}p_{jkt}(h,b_\tau)\right\} \\
&= p_{0t}(h,b_\tau)^{\frac{1}{b_\tau}}\Gamma\left[(b_\tau+1)/b_\tau\right]
\end{aligned}$$

■

Proof of Lemma 9. If bond prices are fully anticipated:

$$E_t\left[\upsilon_{jk,t+1}A_{t+1}\left(H_{jk}(h), b_{\tau(t)+1}\right)\right] = E_t\left[\upsilon_{jk,t+1}\right]A_{t+1}\left(H_{jk}(h), b_{\tau(t)+1}\right)$$

So when $\varphi(\varepsilon)$ is distributed as Type 1 Extreme Value, substituting for $\hat{q}_{jk}\left(P_t\left(h,b_{\tau(t)}\right)\right)$ and $\hat{\upsilon}_{jk,t+1}$, and rearranging Equation (35), we obtain:

$$\begin{aligned}
\alpha_{jkt}(h)^{\frac{1}{\left(1-b_{\tau(t)}\right)}} &= E_t\left[\upsilon_{jk,t+1}\right]A_{t+1}\left(H_{jk}(h), b_{\tau(t)+1}\right)\exp\left(q_{jk}\left(P_t\left(h,b_{\tau(t)}\right)\right)\right)^{\frac{1}{\left(1-b_{\tau(t)}\right)}} \\
&= E_t\left[\upsilon_{jk,t+1}\right]p_{0,t+1}\left(H_{jk}(h), b_{\tau(t)+1}\right)^{\frac{1}{b_{\tau(t)+1}}} \\
&\quad \times\Gamma\left(b_{\tau(t)}+1\right)/b_{\tau(t)}\left[\frac{p_{0t}\left(h,b_{\tau(t)}\right)}{p_{jkt}\left(h,b_{\tau(t)}\right)}\right]^{\frac{1}{\left(1-b_{\tau(t)}\right)}}
\end{aligned}$$

or

$$\begin{aligned}
E_t\left[\exp\left(\rho w_{jkt}/b_{\tau(t)+1}\right)\right] &= \alpha_{jkt}(h)^{\frac{-1}{\left(1-b_{\tau(t)}\right)}}p_{0,t+1}\left(H_{jk}(h), b_{\tau(t)+1}\right)^{\frac{1}{b_{\tau(t)+1}}} \\
&\quad \times\Gamma\left(b_{\tau(t)}+1\right)/b_{\tau(t)}\left[\frac{p_{0t}\left(h,b_{\tau(t)}\right)}{p_{jkt}\left(h,b_{\tau(t)}\right)}\right]^{\frac{1}{\left(1-b_{\tau(t)}\right)}}
\end{aligned}$$

■

References

[1] Antle, R. and A. Smith "An Empirical Investigation of the Relative performance Evaluation of Corporate Executives," *Journal of Accounting Research*, 24 pp. 1-39, 1986.

[2] Antle, R. and A. Smith "Measuring Executive Compensation: Methods and as Application, " *Journal of Accounting Research*, 23 pp. 296-325, 1985.

[3] Bajari, P. and A. Khwaja, "Moral Hazard, Adverse Selection and Health Expenditures: A Semiparametric Analysis," NBER Working Paper No. W12445, August 2006.

[4] D'Haultfoeviller X. and P. Fevrier, "Identification and Estimation of Incentive Problems: Adverse Selection," Working paper, September 2007.

[5] P. Dubois and, T. Vukina, "Optimal Incentives under Moral Hazard and Heterogeneous Agents: Evidence from Production Contracts Data", Working paper, December 2005.

[6] E. Du‡o, R. Hanna, and S. Ryan, "Monitoring Works: Getting Teachers to Come to School", Working paper, MIT, November 2007.

[7] L. Einav, A. Finkelstein and P. Schrimpf, "The Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market," NBER Working Paper No. 13228, July 2007.

[8] Ferrall C. and Shearer B. "Incentives and Transactions Costs Within the Firm: Estimating an Agency Model Using Payroll Records," *Review of Economic Studies*, 66, 2, 309-338, 1999.

[9] Frydman, Carola. "Rising Through the Ranks: The Evolution of the Market for Corporate Executives, 1936-2003." Columbia University, 2005.

[10] Fudenberg, Drew, Bengt Holmstrom and Paul Milgrom. 1990. " Short-Term Contracts and Long-Term Agency Relationships." *Journal of Economic Theory*, Vol. 50, pp. 1-31.

[11] Gayle, George-Levi and Miller, Robert A. 2008a. "Has Moral Hazard become a More Important Factor in Managerial Compensation?" forthcoming, *American Economic Review*.

[12] Gayle, George-Levi and Miller, Robert A. 2008b. " The Paradox of Insider Information and Performance Pay" forthcoming, *CESifo Economic Studies*.

[13] Gayle, George-Levi and Robert A. Miller. 2008c "Identifying and testing models of managerial compensation" Tepper school of Business, Carnegie Mellon University.

[14] Gayle, George-Levi, Limor Golan and Robert A. Miller. 2008 "Are There Glass Ceilings for Female Executives?" Tepper school of Business, Carnegie Mellon University.

[15] Gibbons R. and K. J. Murphy. "Optimal Incentive Contract in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy, 1992, vol. 100 (3), pp 468-505*

[16] Gibbons, R. and M. Waldman "Careers in Organizations: Theory and Evidence," *Handbook of Labor Economics* Vol. 3b. pp 2373–2437, 1999

[17] Hall, Brian J. and Je¤rey B. Liebman. 1998. "Are CEOS Really Paid Like Bureaucrats?" *The Quarterly Journal of Economics,* August 1998, CXIII pp. 653-680.

[18] Johnson, Norman L. and Samuel Kotz 1970. *Continuous Univariate Distriburtions-1,* John Wiley and Sons, New York.

[19] Lazear E. 1992. "The job as a concept," in W. Bruns, ed., *Performance Measurement, Evaluations and Incentives.* Harvard University Press, Boston, MA, pg.183-215

[20] Margiotta, Marry M. and Robert A. Miller. 2000. "Managerial Compensation and The Cost of Moral Hazard." *International Economic Review,* 41 (3) pp. 669-719.

[21] Masson, R. "Executive Motivations, Earnings, and Consequent Equity Performance,"

Table 1: Executives Characteristics by Sector and Firm Size
Compensation and Salary are measured in Thousand of 2006US$

Variable	Service	Primary	Consumer	Asset Small	Asset Large	Employee Small	Employee Large
Rank 1	0.04	0.05	0.07	0.04	0.06	0.04	0.06
Rank 2	0.21	0.27	0.26	0.28	0.26	0.28	0.26
Rank 3	0.07	0.06	0.09	0.05	0.08	0.05	0.08
Rank 4	0.22	0.20	0.22	0.18	0.22	0.18	0.22
Rank 5	0.20	0.17	0.18	0.15	0.18	0.15	0.18
Rank 6	0.18	0.18	0.14	0.21	0.15	0.22	0.15
Rank 7	0.08	0.06	0.04	0.09	0.05	0.08	0.06
Age	52.7 (9.5)	54.8 (9.2)	53.6 (9.4)	53.9 (10.3)	53.7 (9.3)	53.7 (11.2)	53.8 (9.3)
Female	0.056	0.03	0.06	0.06	0.04	0.05	0.04
No Degree	0.20	0.18	0.26	0.23	0.21	0.21	0.21
Bachelor	0.82	0.81	0.73	0.77	0.79	0.78	0.78
MBA	0.23	0.24	0.22	0.19	0.23	0.18	0.23
MS/MA	0.22	0.19	0.15	0.24	0.18	0.23	0.19
Ph.D.	0.18	0.20	0.15	0.18	0.18	0.21	0.17
Prof. Certification	0.21	0.24	0.21	0.26	0.21	0.27	0.21
Executive Experience	18.28 (53.3)	18.7 (49.8)	17.9 (18.7)	20.6 (12.3)	17.1 (11.3)	19.4 (12.1)	17.2 (11.3)
Tenure	13.62 (10.93)	15.0 (11.5)	14.28 (11.5)	16.2 (12.07)	14.1 (11.4)	15.7 (12.1)	14.1 (11.4)
# of past moves	2.11 (1.98)	2.02 (2.01)	2.00 (2.00)	2.5 (2.2)	2.0 (2.0)	2.3 (2.1)	2.0 (2.0)
# of executive moves	0.82 (1.32)	0.82 (1.34)	0.846 (1.39)	0.93 (1.5)	0.81 (1.3)	0.86 (1.4)	0.82 (1.33)
Promotion	0.085 (0.28)	0.34 (0.47)	0.34 (0.475)	0.33 (0.47)	0.36 (0.47)	0.34 (0.47)	0.36 (0.47)
Salary	442 (271)	496 (296)	584 (392)	327 (185)	544 (334)	361 (233)	546 (334)
Total Compensation	3,270 (14,435)	1,841 (8461)	2,041 (12,153)	1,350 (10,188)	3,022 (13,858)	1,538 (11,311)	3,056 (13,753)

*Standard Deviation in Parenthesis

Table 2: Executives Characteristics

Compensation and Salary are measured in Thousand of 2006 US$

Variable	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6	Rank7
Age	59.6	55.7	52.4	52.0	52.8	52.4	52.2
	(9.8)	(7.6)	(8.0)	(8.8)	(10)	(10.3)	(11.2)
Female	0.02	0.02	0.03	0.05	0.06	0.06	0.05
	(0.13)	(0.12)	(0.16)	(0.23)	(0.24)	(0.24)	(0.21)
No Degree	0.25	0.21	0.25	0.21	0.21	0.17	0.21
	(0.43)	(0.41)	(0.43)	(0.40)	(0.41)	(0.37)	(0.41)
MBA	0.24	0.26	0.23	0.27	0.19	0.18	0.22
	(0.42)	(0.44)	(0.42)	(0.44)	(0.39)	(0.39)	(0.41)
MS/MA	0.16	0.17	0.17	0.19	0.21	0.21	0.21
	(0.37)	(0.37)	(0.37)	(0.39)	(0.41)	(0.40)	(0.40)
Ph.D.	0.15	0.15	0.14	0.13	0.21	0.27	0.17
	(0.37)	(0.35)	(0.34)	(0.33)	(0.41)	(0.44)	(0.38)
Prof. Certification	0.15	0.14	0.15	0.22	0.24	0.37	0.30
	(0.36)	(0.34)	(0.35)	(0.42)	(0.43)	(0.47)	(0.45)
Executive Experience	22.3	19.8	16.1	15.9	16.6	16.5	16.9
	(13.0)	(10.5)	(10.7)	(11.0)	(12)	(11.7)	(11.7)
Tenure	17.1	15.1	13.7	13.8	14.1	13.7	14.2
	(13.5)	(11.7)	(11.4)	(11.2)	(12)	(11.0)	(10.8)
# of past moves	1.9	1.9	1.7	1.9	2.2	2.3	2.3
	(2.0)	(1.9)	(1.9)	(1.9)	(2.0)	(2.1)	(2.1)
# of Executive Moves	0.9	0.93	0.73	0.76	0.77	0.80	0.84
	(1.4)	(1.38)	(1.3)	(0.13)	(1.32)	(1.3)	(1.4)
Salary	640	767	591	438	408	323	340
	(375)	(398)	(320)	(197)	(190)	(141)	(217)
Total Compensation	2682	4199	4055	2587	2311	1598	1867
	(18229)	(20198)	(14892)	(8536)	(7319)	(5539)	(6634)

Table 3: Compensation Regressions

Level	OLS	LAD	Slope	OLS	LAD
Constant	964.053	1,222	Excess Return	11,636.76	8,478.87
	(1,417)	(191.9)**		(967.506)**	(129.384)**
			Excess Return Square	-908.68	-238.373
				(27.210)**	(3.649)**
Consumer	-4.737	83.106	Excess Return×Consumer	2,246.78	334.718
	(161.543)	(21.863)**		(353.561)**	(47.699)**
Service	965.097	519.103	Excess Return×Service	2,694.64	1,427.43
	(149.900)**	(20.291)**		(288.870)**	(39.047)**
Assets	0.029	0.03	Excess Return×Asset	0.115	0.086
	(0.001)**	(0.000)**		(0.006)**	(0.001)**
Employees	16.82	16.613	Excess Return×Employees	34.181	32.124
	(1.346)**	(0.182)**		(4.481)**	(0.606)**
Rank 2	2,090.11	1,388.09	Excess Return×Rank 2	-388.042	1,423.73
	(289.289)**	(39.143)**		(655.597)	(88.196)**
Rank 3	896.515	65.889	Excess Return×Rank 3	-7,142.15	-5,254.64
	(352.374)*	-47.683		(745.473)**	(100.422)**
Rank 4	-197.024	-767.392	Excess Return×Rank 4	-12,219.21	-8,068.44
	(302.908)	(40.986)**		(665.071)**	(89.477)**
Rank 5	-484.074	-932.005	Excess Return×Rank 5	-14,409.11	-8,921.51
	(308.492)	(41.736)**		(675.818)**	(90.755)**
Rank 6	-998.282	-1,139.54	Excess Return×Rank 6	-14,047.82	-9,188.51
	(313.464)**	(42.411)**		(670.508)**	(90.146)**
Rank 7	-783.61	-1,109.86	Excess Return×Rank 7	-13,148.96	-9,227.35
	(379.645)*	(51.357)**		(748.188)**	(100.593)**

Table 3(cont.): Compensation Regressions

Level	OLS	LAD	Slope	OLS	LAD
Age	75.732	20.155	Excess Return× Age	136.767	29.214
	(47.603)	(6.444)**		(12.835)**	(1.711)**
Age Square	-0.879	-0.155			
	(0.411)*	(0.056)**			
Female	355.209	91.731	Excess Return×Female	-377.221	-286.293
	(339.929)	(45.917)*		(607.244)	(75.045)**
No. Degree	136.194	12.363	Excess Return×No. Degree	-622.6	-68.224
	(189.753)	(25.679)		(328.146)	(44.118)
MBA	367.872	130.474	Excess Return×MBA	-249.712	234.566
	(162.991)*	(22.060)**		(314.901)	(42.495)**
MS/MA	-79.861	-74.731	Excess Return×MS/MA	-64.16	-355.654
	(165.083)	(22.344)**		(299.351)	(40.481)**
Ph.D.	309.473	32.827	Excess Return×Ph.D.	-22.42	100.848
	(172.953)	(23.409)		(312.742)	(42.259)*
Prof. Cert.	-385.793	-101.85	Excess Return×Prof. Cert.	-1,478.81	-199.566
	(160.076)*	(21.665)**			
Exec. Experience	-0.977	-0.078	Excess Return×Exec. Experience	-2.464	-1.086
	(1.582)	(0.203)		(1.891)	(0.151)**
Tenure	-17.339	-4.573	Excess Return×Tenure	15.764	9.271
	(6.709)**	(0.906)**		(11.078)	(1.469)**
# of past moves	-32.503	-31.781	Excess Return×# of past moves	-392.886	-80.655
	(48.569)	(6.574)**		(84.423)**	(11.360)**
# of Executive Moves	52.739	21.603	Excess Return×# of Exec. moves	153.524	10.868
	(65.354)	(8.839)*		(114.343)	(15.297)
First Year with firm	994.989	551.859	Excess Return× first year in firm	-579.266	-513.588
	(464.134)*	(62.789)**		(854.534)	(115.601)**

Table 4: Multinominal Logit of Firm Choice
(Staying with your Current Firm in the Based)

Variables	1	2	3	4	5	6	Retirement
MBA	-0.026	0.205	0.146	0.167	0.413	0.353	-0.049
	(0.200)	(0.181)	(0.140)	(0.230)	(0.280)	(0.161)*	(0.036)
MS/MA	-0.467	-0.727	-0.335	-0.145	-0.107	-0.207	-0.014
	(0.225)*	(0.238)**	(0.164)*	(0.240)	(0.314)	(0.192)	(0.035)
PhD	-0.787	-0.338	-0.316	-0.281	-0.371	-0.151	-0.080
	(0.248)**	(0.217)	(0.168)	(0.270)	(0.363)	(0.205)	(0.037)*
No Degree	-0.319	-0.436	-0.298	0.435	0.184	0.113	-0.118
	(0.246)	(0.242)	(0.184)	(0.254)	(0.332)	(0.204)	(0.041)**
Moves befere Exec.	-0.141	-0.265	-0.202	-0.046	-0.315	-0.377	0.045
	(0.063)*	(0.075)**	(0.055)**	(0.066)	(0.107)**	(0.073)**	(0.010)**
Female	0.198	0.127	-0.242	-0.173	-1.410	-0.226	0.342
	(0.365)	(0.349)	(0.328)	(0.482)	(1.021)	(0.344)	(0.073)**
Tenure	-32.248	-32.149	-32.277	-31.894	-32.262	-31.935	0.010
	(1.09e+6)	(9.9e+5)	(7.8e+5)	(9.3e+5)	(1.4e+5)	(6.8e+5)	(0.002)**
Moves after Exec.	-0.024	-0.021	0.061	-0.108	-0.123	0.003	0.062
	(0.052)	(0.050)	(0.035)	(0.067)	(0.086)	(0.044)	(0.010)**
Age	0.340	0.165	0.360	0.270	0.340	0.321	0.039
	(0.105)**	(0.075)*	(0.083)**	(0.130)*	(0.173)*	(0.101)**	(0.009)**
Age square	-0.003	-0.001	-0.003	-0.002	-0.003	-0.003	-0.000
	(0.001)**	(0.001)	(0.001)**	(0.001)	(0.002)	(0.001)**	(0.000)*
Firm Type : 2	-0.197	0.650	0.463	-0.781	-0.303	-1.182	0.291
	(0.219)	(0.230)**	(0.200)*	(0.457)	(0.473)	(0.474)*	(0.044)**
Firm Type : 3	-0.932	0.049	0.640	-1.097	-1.378	-0.262	0.232
	(0.210)**	(0.223)	(0.175)**	(0.407)**	(0.516)**	(0.298)	(0.038)**
Firm Type : 4	-1.500	-1.058	-1.096	2.048	1.587	1.452	0.673
	(0.476)**	(0.538)*	(0.441)*	(0.293)**	(0.388)**	(0.304)**	(0.048)**
Firm Type : 5	-1.954	-1.316	-2.072	0.859	1.286	1.317	0.440
	(0.603)**	(0.613)*	(0.728)**	(0.383)*	(0.426)**	(0.319)**	(0.060)**
Firm Type : 6	-1.743	-1.323	-0.729	0.846	0.573	1.828	0.339
	(0.340)**	(0.370)**	(0.254)**	(0.304)**	(0.379)	(0.254)**	(0.044)**
Previous Rank :2	-1.064	0.083	0.059	-0.176	0.239	-0.277	-1.060
	(0.422)*	(0.455)	(0.277)	(0.649)	(0.768)	(0.278)	(0.054)**
Previous Rank :3	0.186	0.810	0.535	1.170	1.478	0.065	-0.560
	(0.454)	(0.503)	(0.308)	(0.662)	(0.802)	(0.331)	(0.069)**
Previous Rank :4	0.677	1.382	0.633	1.310	1.426	0.293	-0.340
	(0.373)	(0.435)**	(0.267)*	(0.606)*	(0.742)	(0.265)	(0.048)**
Previous Rank : 5	0.857	1.134	0.391	1.746	1.329	-0.255	-0.340
	(0.391)*	(0.460)*	(0.295)	(0.611)**	(0.765)	(0.313)	(0.052)**
Constant	-12.389	-8.882	-12.618	-11.794	-14.162	-11.705	-2.918
	(2.794)**	(2.086)**	(2.208)**	(3.325)**	(4.471)**	(2.603)**	(0.281)**
Observations	59066	59066	59066	59066	59066	59066	35019

Table 5 : Multinominal Logit of Rank Choice (Rank 4 is excluded)

variables	1	2	3	5
MBA	0.232	0.232	0.011	-0.021
	(0.082)**	(0.067)**	(0.069)	(0.062)
MS/MA	-0.011	-0.131	-0.117	0.014
	(0.089)	(0.073)	(0.075)	(0.061)
PhD	-0.117	-0.094	-0.147	0.187
	(0.094)	(0.076)	(0.079)	(0.060)**
No Degree	0.198	0.142	0.144	-0.086
	(0.091)*	(0.075)	(0.075)	(0.070)
Moves befere Exec.	-0.144	-0.169	-0.117	0.038
	(0.028)**	(0.023)**	(0.023)**	(0.017)*
Female	-0.749	-0.608	-0.435	0.220
	(0.214)**	(0.162)**	(0.152)**	(0.106)*
Tenure	-0.002	-0.008	-0.006	0.001
	(0.004)	(0.003)**	(0.003)*	(0.003)
Moves after Exec.	-0.008	-0.019	-0.048	0.013
	(0.026)	(0.022)	(0.023)*	(0.019)
Age	0.156	0.226	0.060	-0.009
	(0.025)**	(0.024)**	(0.022)**	(0.015)
Age square	-0.001	-0.002	-0.001	0.000
	(0.000)**	(0.000)**	(0.000)**	(0.000)
Firm Type : 2	0.077	0.193	0.084	-0.224
	(0.104)	(0.086)*	(0.088)	(0.073)**
Firm Type : 3	0.283	0.352	0.216	-0.374
	(0.089)**	(0.075)**	(0.076)**	(0.067)**
Firm Type : 4	-0.585	-0.388	-0.324	0.020
	(0.133)**	(0.104)**	(0.110)**	(0.079)
Firm Type : 5	-0.262	-0.115	0.013	-0.152
	(0.148)	(0.118)	(0.118)	(0.099)
Firm Type : 6	0.239	0.195	0.191	-0.262
	(0.103)*	(0.086)*	(0.087)*	(0.077)**
Previous Rank :2	-2.196	3.745	-0.413	0.209
	(0.132)**	(0.144)**	(0.177)*	(0.296)
Previous Rank :3	-3.544	0.652	3.031	0.265
	(0.159)**	(0.154)**	(0.162)**	(0.309)
Previous Rank :4	-7.890	-4.656	-3.662	-1.951
	(0.124)**	(0.134)**	(0.145)**	(0.255)**
Previous Rank : 5	-7.181	-3.512	-2.402	3.922
	(0.232)**	(0.170)**	(0.168)**	(0.253)**

Table 6: Structural Estimates and Simulations

τ_2, τ_3 and τ_4 are measured in US100,000 of dollars

τ_1 is measured in percentage per year

Measure	Rank	Estimates	Standard Deviation.
ρ		0.45	
	1	5.2	3.4
	2	10.9	14
	3	8.3	2.9
τ_1	4	4.2	2.7
	5	1.6	1.2
	1	4.0	0.2
	2	9.0	0.5
τ_2^H	3	11.8	0.9
	4	16.4	1.3
	5	18.8	2.2
	1	18.6	34.7
	2	24.8	56.6
τ_2^{PM}	3	8.3	14.2
	4	2.5	8.6
	5	.9	1.2
	1	17.3	34.0
τ_3	2	32.5	45.6
	3	16.03	24.8
	4	1.2	2.5
	5	0.8	1.3
	1	0.5	1.4
	2	2.6	3.9
	3	12.0	14.3
τ_4	4	14.0	18.9
	5	18.2	22.7