

Promotion, Turnover and Compensation in the Executive Labor Market*

George-Levi Gayle

Department of Economics, Washington University in St. Louis

Limor Golan

Department of Economics, Washington University in St. Louis

Robert A. Miller

Tepper School of Business, Carnegie Mellon University

June 3, 2014

Abstract

This paper constructs and estimates a generalized Roy model with human-capital accumulation and moral hazard to control for self-selection by executives across firms and jobs. It uses the estimated model to decompose the firm-size pay gap into five sources, namely: (1) utility from working, (2) divergence in incentives, (3) monitoring, (4) demand, and (5) human-capital accumulation. It finds no support for (1), because although total compensation and incentive pay increase with firm size, the certainty-equivalent wage decreases with firm size. The risk premium paid to correct inefficiencies caused by (2) and (3) accounts for roughly more than 80 percent of firm-size total-compensation gap, while (4) accounts for the rest. The quality of the signal about effort is unambiguously poorer in larger firms, which completely explains the larger risk premium. While human capital from formal education and experience gained from different firms are individually significant, collectively their effect on the firm-size pay differentials nets out.

1 Introduction

One of the most robust empirical findings in labor economics is that pay increases with firm size (Oi and Idson, 1999). This is also true in the executive labor market: executives in large firms are paid 2.7 times as much as their counterparts in small firms. Recently, a number of papers have used this relationship between firm size and compensation to justify the increasing trend in executive compensation (Gabaix and Landier, 2008; Terviö, 2008; Gayle and Miller, 2009b). The literature on the firm-size pay premium has proposed three major behavioral reasons for the relationship between firm size and pay: monitoring cost, shirking and efficiency wages, and demand for entrepreneurial talent. However, none

*We thank the editor of this journal and three anonymous referees for their comments. We also thank the participants at the World Congress on National Accounts and Economic Performance Measures for Nations 2008, the Society of Labor Economists 2009, the Econometric Society summer 2009 meeting, the Cowles Microeconometrics Conference 2009, the Society of Economic Dynamics 2009, the NBER 2009 Summer Institute, the CRES Applied Economics conference 2009, the conference in honor of James Heckman at the Milton Friedman Institute 2010, and seminar participants at Carnegie Mellon, Columbia, Duke, Federal Reserve Bank of Richmond, Harvard, IDB, NYU, Minnesota, Stony Brook, Wisconsin-Madison, Queens, Rochester, UCLA, UNC-Chapel Hill, and Virginia for comments and suggestions. This research is supported by the Center for Organizational Learning, Innovation and Performance in Carnegie Mellon University and the National Science Foundation's Grant SES0721098.

of the papers on the firm-size pay premium in the executive labor market include all these possible explanations for the firm-size pay premium, nor do they assess their relative importance.¹ In labor markets differences in compensation arise from differences in job characteristics, such as employment stability, the nature of the tasks, promotion opportunities, and the work environment. This paper develops a framework encompassing these job features to investigate the reasons for the firm-size pay premium in the executive labor market. This is the first paper to explicitly analyze the problem in the context of a dynamic model of executive careers. It delivers several new empirical findings relating firm size to compensation and interprets them within a unified conceptual framework.

We control for self selection by executives across firms and ranks, by extending the sorting model of Roy (1951) to incorporate nonpecuniary job utilities, agency issues and human capital. These three extensions are motivated by empirical regularities found in our data, a matched sample of over 30,000 executives and 2,500 firms spanning fourteen years. The stylized fact that larger firms pay more compensation than smaller firms might be attributable to inferior working conditions in the former. Second, top executives are paid a significant portion of their total compensation in stock and options, raising their expected total compensation by a risk premium. Third, our data show that previous executive experience in other firms raises executive compensation at higher ranks in the hierarchy. Forward-looking managers accumulate this form of human capital when choosing between jobs.

Our data also show that the composition of firm-denominated securities varies substantially across ranks and executives at different points in their lifecycles: for example executives closer to their retirement position or age receive substantially more incentive pay, increasing total expected compensation through a higher risk premium. This regularity gives additional empirical motivation for including both agency and dynamic considerations. Incorporating a theory of career concerns allows us to account for this empirical regularity, and investigate whether it varies with firm size.

In the model, executives make sequential job and effort choices taking into account the compensation, nonpecuniary benefits from working and future value of accumulated human capital. Their effort choices are private information and ultimately the source of moral hazard. We incorporate career concerns by allowing human capital accumulated on the job to depend on effort. The other dimensions of human capital are defined by formal education, plus previously held executive positions and their durations. Thus each job choice has an investment component. At the beginning of every period, the equity returns of firms from decisions made in the previous period are revealed to everyone, the executives' human-capital state variables are updated, and each executive is compensated following the schedule of the previous period's employment contract. Firms assess their demand for executives in the current period and post one-period contracts for positions within their firms. The one-period equilibrium spot contracts are sequentially optimal. The contract aligns executive goals with those of shareholders, by making compensation depend on the executive's characteristics: both the nonpecuniary benefits and the amelioration of monetary incentives due to career concerns vary with executive characteristics, which change over the lifecycle.

The structural econometric model we estimate comes from two equations that hold in the sequential equilibrium we analyze. The first equation applies to a manager who is indifferent between taking any job match and exiting in equilibrium. It equates the systematic portion of the manager's expected utility (the sum of current utility, the certainty equivalent of compensation and the investment value of human capital), conditional on human capital and job-match choice, with the net value of the disturbance from exiting. The net value of the marginal disturbance and the value of human capital can be written as functions of the conditional-choice probabilities.

The second equation is derived from the wage schedule for the optimal contract. We show that,

¹ Gabaix and Landier's (2008) and Terviö's (2008) models were based on the demand for entrepreneurial talent while Gayle and Miller's (2009b) model is based on shirking.

up to a factor of proportionality, the slope of the contract identifies the likelihood ratio of abnormal returns for different effort choices. This fact provides the means for estimating the model's remaining parameters. We also show the extent to which our model is nonparametrically identified. We prove an observational equivalence holds between long-term optimal contracts when career concerns are absent, and equilibrium spot contracts when career concerns are present. We then show that all the elements of the pay-differential decomposition are independent of this distinction except one, career concerns: thus the identification of the costs and benefits of shirking does not hinge on whether there are career concerns or not. Finally the extent to which career concerns ameliorate the agency problem requires either exclusion restrictions or functional-form assumptions on the evolution of human capital when managers shirk.

Empirically, we document a sizable firm-size pay premium for executives in both total compensation and incentive pay. The paper shows that this firm-size pay gap is robust to controls for industry, executive rank, human capital and individual characteristics. Average pay increases as executives are promoted, and executive experience accumulated in different firms increases human capital, raising the chance of becoming a CEO, empirical regularities that are consistent with Fox's (2009) model of hierarchy matching. To assess sources of the firm-size pay premium, we control for sorting and risk aversion by calculating the certainty-equivalent wage by firm size. We control for risk aversion because over two thirds of executive compensation is paid in the form of firm-denominated securities.

An important finding of this paper is that the certainty-equivalent wage declines with firm size. To understand why, we further decompose the certainty-equivalent wage into four components: the compensating differential for the disutility of working, the compensating differential for human-capital accumulation, the agency-risk premium, and the demand for executive talent. The compensating differential for the disutility from work would explain the firm-size pay gap if larger firms had negative job attributes. However, we find that the nonpecuniary costs of working are larger in smaller firms. This is the main reason the certainty-equivalent wage is decreasing in firm size.

We find that human capital accumulation does not decline through the ranks, but peaks at the rank just below and at the CEO level, primarily because attaining either position promises a longer future tenure with the firm than the others. Similarly we find that to counteract declining career concerns as an executive approaches retirement, explicit incentives increase with age and dead end positions. In net, the compensating differential for human-capital accumulation does not vary much with firm size.

How then, do we explain the sizable firm-size pay premium observed in the executive labor market? A risk premium, rationalized in our model by incentive contracts to deter shirking, accounts for approximately 80 percent of the firm-size pay premium. More specifically, the estimated risk premium is \$1.6 million for small firms, \$2.6 million for medium-size firms and \$4.9 million for large firms. Loosely interpreted these findings are consistent with explanations that suggest large firms pay large efficiency wages to prevent shirking (Doeringer and Piore,1971; Raff and Summers,1987; Katz and Summers,1989). They also corroborate findings in Gayle and Miller (2009b) that the increase in firm size, through its effect on the moral-hazard problem, can explain the growth of CEO compensation over the past 50 years.

Since the average equity value is \$322 million for small firms, \$1,071 million for medium-size firms and \$6,022 million for large firms, the risk premium is concave increasing in firm size. Moreover, we find that opportunities to invest in human capital do not vary appreciably with firm size, and as noted above, large firms provide more nonpecuniary benefits than small firms. Consequently these three factors cannot explain why further amalgamation does not occur. Our estimates attribute the remaining 20 percent of the firm-size pay premium to a higher demand for executives from larger firms that attract and retain executives who would otherwise exit the occupation. These results on the relationship and importance of agency costs to firm size provide some of the first empirical evidence

that speaks to the theoretical predictions of Lucas (1978) and Aron (1988).

We also explore what drives differentials in the risk premium. The risk premium arises from the agency problem, and its severity depends on three factors. First, the more executives value shirking versus working, the greater the risk premium in the equilibrium contract. We find the utility from shirking versus working is higher in small firms than in large firms. Therefore this factor cannot explain the firm-size risk-premium gap. Second, career concerns ameliorate the agency problem and reduce the risk premium, because in the extended version of our model, working provides human capital. Empirically we find that this does not vary by firm size. Third, the quality of the signal about effort, which in our model is the likelihood ratio of the density of excess returns from shirking versus working, affects the cost of moral hazard, that is the risk premium. We find that signal quality is unambiguously poorer in larger firms, overwhelming the other two effects. On reflection this is not surprising: the hierarchy of ranks varies significantly across size, larger firms having more supervisory positions, accountability is more difficult, leading to greater reliance on incentive pay.

Finally, a coherent interpretation of how management teams function within corporations can be gleaned from the estimated model. We find that the equity lost from an executive shirking declines with his rank, contradicting conventional wisdom. Since those lower in the ranks and closer to operations can most affect excess return to the firm, a CEO is clearly not paid more because of his power to create or destroy shareholder value! Furthermore we do not find support for another traditional view that high level executives have more discretion than low level managers to seize upon opportunities they value at the expense of shareholders; although the estimated benefits from shirking modestly increase with rank, we cannot reject the null hypothesis of equality. The effects of weaker signals at higher ranks that translates to a higher risk premium explains most of the differences total compensation across ranks, a finding that is broadly consistent with the monitoring paradigm of McNulty's (1984). More generally, highly ranked executives are paid more than lower ranked executives for largely the same reasons that executives are paid more in large firms than in small firms: they are further from operations, can do less damage to the firm, so the signal they give shareholders is less informative, inducing in equilibrium a more incentivized contract supported by a much bigger risk premium.

The rest of the paper is organized as follows. Section 2 discusses the related literature. Section 3 introduces the data and documents the stylized facts on the firm-size pay gaps in the executive labor market. Section 4 presents the basic model without implicit incentives. Section 5 extends the model to include implicit incentives. Section 6 analyses the identification of the model. Section 7 outlines the estimation strategy. Section 8 presents the estimates and the decomposition of the firm-size pay gaps. Section 9 concludes. The proofs of all the main results are in the appendix at the end of the paper. More details on the data construction, additional results, and derivations of examples used in the paper are collected in an online appendix.

2 Related Literature

Several papers (Lucas, 1978; Rosen, 1983; Gabaix and Landier, 2008; Terviö, 2008) have used assignment and sorting to model the executive labor market; none combine assignment and sorting with moral hazard and human-capital accumulation to study how information frictions affect the equilibrium assignment and pay of managers to firms. This paper also allows human-capital accumulation to be a function of hidden actions that have direct consequences for shareholders, giving rise to a dynamic moral-hazard problem in a nonstationary environment where current actions have future consequences. Moral hazard models, built on the assumption of hidden actions, are the principal paradigm for rationalizing incentive pay in the executive labor market. Letting hidden actions also determine human-capital accumulation induces career concerns without adding a second source of pri-

vate information; previous theoretical work (Gibbons and Murphy, 1992; Chevalier and Ellison, 1999; Dewatripont, Jewitt, and Tirole, 1999; Holmström, 1999) relies on an additional source of private information to generate career concerns. Our more parsimonious approach aids identification and hence the interpretation of empirical results.

Only a handful of recent theoretical papers have studied dynamic contracting with moral hazards in nonstationary environments where current actions have consequences over a long horizon (Garrett and Pavan, 2012; Li, 2014; Sannikov, 2014). The closest to ours is Garrett and Pavan’s (2012). In both models match quality between a firm and its managers changes stochastically over time, shocks to managerial productivity are anticipated at the time of contracting, but only privately observed by the managers. In our model, the match changes endogenously over time through human-capital accumulation, but not in theirs. Providing appropriate incentives to managers becomes less onerous over time in their model, not more onerous, as in our model; empirically, we find executives are more expensive to motivate in the twilight of their careers.

The theoretical apparatus used to model job assignment, sorting, and human capital is based on a vast literature that dates back to Roy (1951), Becker (1964) and Ben-Porath (1967). In the standard general human capital framework diversity of experience has no value. Our model adds an additional dimension to this literature by allowing current compensation to directly depend on the range of jobs the executive has held in the past, which creates a trade off between firm specific tenure and this form of general human capital. This creates an incentive for younger executives to gain experience in different work environments. Similar predictions apply to younger workers in the experimentation human-capital literature (Miller, 1984; Antonovics and Golan, 2012; Sanders, 2013). We find that obtaining experience in different jobs is indeed statistically significant and quantitatively important.

A number of papers have studied identification and inference in the generalized Roy model (Bayer, Khan, and Timmins, 2011; D’Haultfoeuille and Maurel, 2013); ours is the first to analyze identification and estimate a generalized Roy model with moral hazard and human-capital accumulation. Additionally, this paper establishes the identification of a sequential-equilibrium signaling game, which, to the best of our knowledge, has never been analyzed before. The identification results are also related to Gayle and Miller (2013), who show that the static and repeated moral-hazard models are only set identified. This paper extends that work by exploiting the equilibrium-sorting and assignment equations to achieve point identification.

Several papers have estimated equilibrium models and used them to decompose pay differences in labor markets. Some papers use worker employment data in an equilibrium framework (Altuğ and Miller, 1998; Lee and Wolpin, 2006; Gayle and Golan, 2012), while others, like ours, use matched firm–manager data, which allows the incorporation of firm and worker heterogeneity (Postel-Vinay and Robin, 2002; Cahuc, Postel-Vinay, and Robin, 2006; Taber and Vejlin, 2010). Of these papers, only Gayle and Golan (2012) motivate turnover and wages with information asymmetries between workers and firms, but they do not use data on firms. Our paper contributes to this literature by providing a unified framework for investigating information asymmetries and career concerns with data on both the suppliers and the demanders for labor. Finally, our empirical results also add to the empirical literature on the firm-size pay premium (Brown and Medoff, 1989; Oi and Idson, 1999; Winter-Ebmer and Zweimüller, 1999, among others). Our finding, that workers get significant nonpecuniary benefits from working in larger firms, contradicts the belief that large firms offer inferior working conditions, and corroborates similar empirical results in Brown and Medoff (1989).

3 Data

The data for our empirical study come from three sources. The main data source is Standard & Poor’s ExecuComp database, which contains annual records on 30,614 individual executives, itemizing their compensation and describing their titles. Each executive worked for one of the 2,818 firms comprising the (composite) S&P 500, Midcap, and Smallcap indices for at least one year spanning the period 1992 to 2006, which covers about 85% of the U.S. equities market; in the years for which we have observations, the executive was one of the eight top-paid employees in the firm whose compensation was reported to the SEC. Data on the 2,818 firms for the ExecuComp database were supplemented by the S&P COMPUSTAT North America database and monthly stock-price data from the Center for Securities Research database. We also gathered background history for a subsample of 16,300 executives, recovered by matching the 30,614 executives from our COMPUSTAT database using their full name, year of birth and gender with the records in Who’s Who, which contains biographies of about 350,000 executives. The matched data 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 sociodemographic characteristics, including age, gender, and education; and a comprehensive description of their career path sequence described by their annual transitions through the possible positions and firms.

We construct a hierarchy consisting of five ranks using a rational ordering over a set of job titles based on transition independent of compensation. (See Gayle, Golan, and Miller, 2012, for a detailed description of the titles and the construction of the hierarchy.) Following is a rough description of the titles on each rank: (*Rank 1*) chairman of the board of the company or chairman of a subsidiary who does not have any other executive positions in the firm; (*Rank 2*) CEO of the company; (*Rank 3*) COO, CFO and chairman of board of the company who holds some other executive position in the company other than CEO; (*Rank 4*) other high-level corporate executives and heads of subsidiaries or regional chiefs; (*Rank 5*) other lower-level executives. The first observation is that CEOs are not in Rank 1 but instead in Rank 2: Since this hierarchy is based on transitions, this reflects a lifecycle consideration more than control. However, it collaborates institutional use of the term Rank, which emphasizes the supervisory roles of managers over their subordinates. For example, the chairman of the board of directors monitors the CEO of the firm. We retain a position for executives that serve on the board of directors of their own company (Exedir). This is because directors are of special interest to good corporate governance and span of control of a position in the hierarchy. The literature on the firm-size pay premium has highlighted this as important. We also classify firms into three industrial sectors, primary, consumer and service.

Firms are further classified into three sizes, large, medium and small, based on the value of their assets and number of employees over the sample period. A firm is classified as large if both its value of assets and its number of employees are above the median for its sector over the sample period and as small if both its value of assets and number employees are below the median for its sector over the sample. All other firms are classified as medium. We also classified firms according to the number of “insiders” on their board relative to the industrial norm. That is, a company is classified as having a large insider board if the number of insiders on its board is above the median for its sector and firm size. Finally, we classified a firm from the perspective of an executive as “new” if this is the first year the executive is working in the firm and as “old” if the executive has worked in the firm for more than one year. This variable allows us to capture executive turnover. In all, there are 36 firm types in our analysis characterized by firm size, industrial sector, size of insider board, and whether the firm is a new firm for the executive.

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.² Hence, the change in wealth from holding their firms' stock is the value of the stock at the beginning of the period multiplied by the abnormal return, defined as the residual component of returns that cannot be priced by aggregate factors the manager does not control.

Individual characteristics consist mainly of labor-market experience, which has several dimensions: years of tenure in the firm, years worked as top executive, number of firms an executive worked in before becoming an executive, and the number of firms an executive worked in after becoming an executive. In addition, we observe education (such as MBA, M.Sc., Ph.D., etc.), gender, age, and interlocked.³ The price of a console bond will play a central role for consumption smoothing in our analysis; hence, we construct a bond price series from the Federal Reserve's Economic Research Database. A full description of the construction and summary of the data is contained in Online Appendix B.

3.1 Preliminary Analysis of the Data

This section documents the firm-size differences in compensation, hierarchy, education, experience and mobility patterns in the executive labor market. The previous literature on pay and firm size has focused mostly on other labor markets. Those that looked at the executive labor market (Gabaix and Landier, 2008; Gayle and Miller, 2009b) did not have data on hierarchy, education, work experience and mobility. Therefore, documenting the basic empirical regularities is a worthwhile exercise.

The first empirical regularity is evident from Figure 1a, which shows that both total compensation and the fixed component, salary, increases with firm size. However, total compensation increases significantly more than salary. For example, the average total compensation for an executive in a large firm is 2.7 times that of an executive in a small firm, but the average salary for an executive in a large firm is only 1.7 times that of an executive in a small firm. Thus, not only is compensation increasing with firm size, but so too is incentive pay. Turning to Figure 1b, we see that hierarchy also varies with firm size. For example, large firms are more likely than small firms to separate the jobs of CEO (Rank 2) from Chairman of the Board (Rank 1). This could be evidence that larger firms have a monitoring problem as proposed by the literature as a reason for the firm-size pay premium. Also, Rank 5 is more likely in a small firm than a large firm while the opposite is true for Rank 4.

Figure 2a shows that executives in large firms have more formal education than executives in small firms. However, even among the executives with formal education there are also differences by firm size. While executives with a Ph.D. degree are equally distributed across firm size, large firms have a higher concentration of executives with an MBA, but a lower concentration of nonbusiness masters degrees. This might suggest that large firms have a higher demand for talent. However Figure 2b gives reason to pause, as both tenure and years of executive experience decrease with firm size. On the other hand, age increases with firm size. Taken together, Figures 2a and 2b follow Mincer's (1974) arguments about the value of schooling: executives in large firms have less job experience and

² Changes in wealth from holding firm stock and options reflect the cost a manager incurs from not being able to fully diversify her wealth portfolio because of restrictions on stock and option sales. When forming their portfolio of real and financial assets, managers recognize that part of the return from their firm-denominated securities should be attributed to aggregate factors, so they reduce their holdings of other stocks to neutralize those factors. See Antle and Smith (1985, 1986), Hall and Liebman (1998), Margiotta and Miller (2000), and Gayle and Miller (2009a,b) for other papers using this measure of total executive compensation.

³ An executive is classified as interlocked if at least one of the following is true: (a) The executive serves on the board committee that makes her compensation decisions. (b) The executive serves on the board of another company that has an executive officer serving on the compensation committee of the indicated executive's company. (c) The executive serves on the compensation committee of another company that has an executive officer serving on the board of the indicated executive's company.

are older because they acquired more formal education. Our data comes from a truncated upper management in publicly held companies, so we cannot infer much about the lengthy incubation phase that characterizes executive selection. However, we can nevertheless infer something about the value of human capital acquired through experience on the job, by investigating the movement and decisions through the hierarchy and their subsequent careers conditional on their human capital upon entering management.

Given the interaction between firm size, hierarchy, and human capital, Table 1 presents the main characteristics of our sample by executive rank. Rank 1 has the highest exit rate while Rank 2 has the lowest exit rate and the highest turnover rate. Average age, tenure and executive experience increase with rank. Rank 2 executives have the most experience in other firms since becoming an executive, but the least experience with other firms before becoming an executive. Those with no college are more likely to fill the upper ranks, while those with a Ph.D. are most likely to be found in Ranks 4 and 5. Thus, Rank 5 is the most educated by every measure except MBA while a Rank 2 executive is more likely to have an MBA than an executive in any other rank. Salary, total compensation, and the likelihood of being on the board rise with advancing rank, peak at Rank 2, and then decline at Rank 1.

None of the results on compensation and mobility documented in Table 1 and Figure 1a account for interactions between firm size and: hierarchy (Figure 1b), education (Figure 2a) and experience (Figure 1c). Table 2 shows the effects, on compensation and three indicators of job mobility, of using conditioning information in four regressions. The first is a second-order polynomial compensation regression that specifically breaks out compensation in terms of its fixed and variable components; the first three columns of Table 2 report the coefficients and their estimated standard errors from this one regression on rank (in Panel A), firm type (Panel B) and human capital plus individual heterogeneity (Panel C). The second is a multivariable logit that summarizes promotion. The third and fourth are logit regressions that summarize the probability of changing firms and retirement, respectively. Panel A of Table 2 demonstrates that the empirical regularities in the firm-size pay premium are robust to controlling for these interactions. Panel B of Table 2 shows three empirical regularities with regard to compensation and firm type: (i) Larger firms compensate executives with more fixed pay, as is usually found in labor markets, and on average more incentive pay as well. (ii) Firms with a larger number of insiders on their board of directors have more incentive pay but the same fixed pay. (iii) The service sector pays the most in fixed pay and offers the most incentive pay, while the primary sector pays the least in fixed pay and offers the least incentive pay. Note that (i) does not imply that the certainty equivalent wage is higher in large firms than small firms. Answering this question requires us to estimate the risk parameter of executives from an identified behavioral model we assume generates the data.

Panel C of Table 2 demonstrates three empirical regularities with regards to compensation and human capital: (i) The effect of tenure is highly nonlinear and varies by rank. (ii) Tenure in a given rank does not affect the fixed component of pay, but does affect the variable component. (iii) Years of executive experience affects the variable but not the fixed component of executive pay. These empirical regularities demonstrate the significance of human capital in determining compensation. The last seven columns of Table 2 show that firm size does not seem to affect promotion, turnover or exit, but human capital does.

In summary, with the notable exception that there is less mobility between firms in the primary sector, which could well be due to technological considerations and specialized training, firm size and sector differences affect only compensation—not promotion, turnover or exit—suggesting that a static model of compensating differentials might account for them. Exit is convex increasing in age, older executives are more likely to be found in the highest paid ranks, and are, moreover, paid a premium for any rank they hold. In addition, they have substantially more incentive pay. This begs a nonstationary dynamic model with career concerns in which aging executives become increasingly

productive but less willing (and ultimately unable) to remain employed with the firm. Job turnover complicates the picture because newly hired executives at Ranks 2 and 3 receive a substantial sign-on bonus, reinforced by declining compensation with increased tenure. Similarly, newly hired executives at all ranks are not subject to the same performance-pay criteria as executives with more tenure. This could be construed as evidence for an orientation phase in which new hires are initially given less responsibility so they can familiarize themselves with their working environment. Consequently, they are not held as accountable for firm performance in their first year. However, the distribution of ranks and human capital varies by firm size, suggesting that sorting out the determinants of the firm-size pay premium requires a model that simultaneously incorporates all of these factors.

4 The Basic Model

The building blocks of the model are moral hazard, sorting, nonpecuniary benefits from jobs, human capital and career concerns. These building blocks are parsimoniously combined to facilitate estimation of the underlying technology and utility parameters rationalizing the observed compensation schedule for the different executives and firms, as well as its evolution with age, tenure and experience. This section presents the model without career concerns because it is easier to understand. In the basic model, expected-value-maximizing shareholders are subject to moral hazard from choices made by risk-averse executives who have private information about their own effort levels. Executives invest in firm-specific and general human capital through experience on the job. They sequentially choose employment, bargain with firms about their compensation, and choose their effort levels, which determine the probability distribution of the returns to the firms. Through this process, executives extract all the rent from their job matches.

4.1 Executives and Firms

A finite number of firms in the executive market are indexed by $j \in \{1, \dots, J\}$, with $j = 0$ representing retirement. There are K positions within each firm j , indexed by $k \in \{1, \dots, K\}$ and ranked in hierarchical order. Different combinations of firms and ranks capture heterogeneity of jobs in the economy. Firms belong to different industries and have different sizes of capital and employment. Thus, the position of a CEO in a large firm in the manufacturing industry, for example, may be different from a CEO position in a small firm in the service industry, in terms of the tasks performed, skill requirements and nonpecuniary benefits and costs. Let $t \in \{0, 1, \dots\}$ denote each executive's age, with retirement upon reaching or before age $T < \infty$. To simplify the notation, we assume that executives are infinitely lived. Each executive's background is defined by age t and a vector of human capital, h_t , which includes fixed demographic characteristics and indexes work experience.

4.2 Choices

At the beginning of period t , which denotes age, an executive chooses her consumption, c_t , and, for any $t \leq T$, makes her employment choices. She negotiates her compensation and signs an employment contract determining how she will be paid. She then chooses her effort, which is unobserved by the shareholders. Let $d_{jkt} \in \{0, 1\}$ indicate the executive's choice of rank k in firm j at age t , and let d_{0t} denote the indicator variable for retirement. The $JK + 1$ choices are mutually exclusive, implying

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

Summarizing, $d_t \equiv (d_{0t}, d_{11t}, \dots, d_{JKt})$ denotes the vector of job matches from which an executive chooses at any age t preceding retirement.

There are two effort levels, working diligently and shirking, denoted by $l_t \in \{0, 1\}$, where $l_t = 0$ means the executive shirks at age t and $l_t = 1$ means the executive works. Effort affects the distribution of the firm's returns and the executive's current-period nonpecuniary utility. As in standard moral hazard models, the goals of executives and shareholders are not aligned. Therefore, the term shirk refers to activities that benefit her but not shareholders, and working describes effort and activities undertaken to achieve shareholder goals.

4.3 Preferences

The executive's preferences depend on her consumption and nonpecuniary utility associated with labor-supply choices. Preferences are characterized by the discounted sum of a time-additively separable, constant absolute risk-aversion (CARA) utility function. The utility function is decomposed into utility from consumption and a nonpecuniary cost-of-working. The nonpecuniary costs of working and shirking are allowed to be different in each rank and firm, and are further decomposed into systematic and nonsystematic components. The nonsystematic component captures the executive's firm- and rank-specific idiosyncratic-taste shock, which does not depend on effort. The taste-shock vector in period t is denoted by $\varepsilon_t \equiv (\varepsilon_{0t}, \varepsilon_{11t}, \dots, \varepsilon_{JKt})$, where ε_{0t} is the shock from choosing retirement, and the taste shock from working in firm j at rank k is ε_{jkt} . The systematic component of the nonpecuniary utility from working depends on the executive's effort, characteristics and experience h_t , as well as the firm and rank. When the executive works (setting $l_t = 1$), her nonpecuniary utility is $\alpha_{jkt}(h_t)$; when she shirks it is $\beta_{jkt}(h_t)$. The executive's lifetime utility can thus be summarized as

$$-\sum_{t=1}^{\infty} \delta^t \exp(-\rho c_t) \left[d_{0t} \exp(-\varepsilon_{0t}) + \sum_{j=1}^J \sum_{k=1}^K d_{jkt} [\alpha_{jkt}(h_t) l_t + \beta_{jkt}(h_t) (1 - l_t)] \exp(-\varepsilon_{jkt}) \right], \quad (2)$$

where δ denotes the subjective discount factor and ρ is the constant absolute risk-aversion parameter. The systematic component of nonpecuniary benefits from retiring are normalized to one. We assume there is more disutility from working than from shirking, so $\alpha_{jkt}(h) > \beta_{jkt}(h)$. The difference between $\beta_{jkt}(h_t)$ and $\alpha_{jkt}(h_t)$ captures the divergence between the shareholder and executive goals. This formulation of the utility function captures differences across rank-firm nonpecuniary costs. This allows the model to account for different levels of moral hazard between large and small firms and among ranks and industries. The formulation also allows executives with different characteristics to have different disutilities from firm-rank and effort choices. CARA utility is commonly assumed because the lack of wealth effects makes the dynamic problem more tractable. More specifically, the log of indirect utility is linear in outside wealth, and additively separable in the taste shocks and shifters. Consequently outside wealth drops out of contention when comparing different employment options, which is a particularly attractive feature in applications of executive compensation, where data sets rarely, if ever, include detailed information on outside wealth.

4.4 Human Capital

Human capital is multidimensional and includes skills that depend on education and work experience. We define a vector of time-invariant characteristics and skills, h_1 , that captures gender and education dummies. We further define a vector to capture the individual's history of rank-firm choices, including retirement, as $h_{2t} = (h_{211t}, \dots, h_{2JKt})$. Thus, the vector that captures all human capital is $h_t = (h_1, h_{2t})$. We also define a transition function, $\bar{H}_{jk}(h_{2t})$, to capture the evolution of human capital; we assume the function is deterministic and that human capital follows the law of motion:

$$h_{2t+1} = \sum_{j=1}^J \sum_{k=1}^K d_{jkt} \bar{H}_{jk}(h_{2t}). \quad (3)$$

Our specification of human-capital accumulation, captured by h_{2t} , encompasses two dimensions. First, the model captures information about where (firm and rank) human capital is acquired; therefore, it contains information about industry and firm size. Second, the specification captures the applicability of the human capital. That is, it captures by how much experience in the j^{th} firm at rank k increases productivity in each firm and rank, and allows for increments to differ by firm and rank.

To illustrate how human capital accumulation works in our model, suppose that h_{2t} is a three-dimensional vector, $h_{2t} \equiv (h_{2t}^{(1)}, h_{2t}^{(2)}, h_{2t}^{(3)})$, and that firm is the cross between two sets of indices: The first, $j_1 \in \{0, 1\}$, denotes whether this is a new firm, $j_1 = 0$, or the executive worked for this firm last period, $j_1 = 1$. The second, $j_2 \in \{1, 2, \dots, J_2\}$, denotes firm size and industrial sector; hence, $j = j_1 \otimes j_2 \in \{0, 1\} \otimes \{1, 2, \dots, J_2\}$. Therefore, let $h_{2t}^{(1)}$ measure tenure of the executive in the current firm, and hence capture firm-specific capital. Let $h_{2t}^{(2)}$ measure the number of years of executive experience, and let $h_{2t}^{(3)}$ measure the number of different firms the executive has worked in since becoming an executive. The last two, $h_{2t}^{(2)}$ and $h_{2t}^{(3)}$, are meant to capture the years of general human capital. The second, $h_{2t}^{(2)}$, is standard in the learning-by-doing human-capital-accumulation literature; however, the third, $h_{2t}^{(3)}$, is meant to capture the idea that management may require many different skills, and the greater the number of firms an executive worked in the better she may be when she becomes, say, CEO. We then specify the transition function of human capital as

$$\bar{H}_{jk}(h_{2t}) = h_{2t} + \Delta_{jk}, \quad (4)$$

where $\Delta_{jk} \equiv (\Delta_{jk}^{(1)}, \Delta_{jk}^{(2)}, \Delta_{jk}^{(3)})$. For example, if the executive does not retire but chooses a new firm, then $\Delta_{jk}^{(1)} = -h_{2t}^{(1)}$, $\Delta_{jk}^{(2)} = 1$ and $\Delta_{jk}^{(3)} = 1$. This means she would lose all her firm-specific capital, gain an additional year of executive experience, and increase the number of firms she worked in. On the other hand, if she remains with her current firm, then $\Delta_{jk}^{(1)} = 1$, $\Delta_{jk}^{(2)} = 1$ and $\Delta_{jk}^{(3)} = 0$. With this formulation of human capital, the executive is also gaining firm-specific capital. In this simple way we capture the standard formulation of specific and general human-capital accumulation, as represented by $h_{2t}^{(1)}$ and $h_{2t}^{(2)}$ in this example. Since $h_{2t}^{(2)}$ is portable across firms, older workers would also have better outside options and hence can have compensation increasing with tenure and experience. However greater firm-specific capital makes the executive appear less versatile.

The additional element we add an to the standard model of human capital accumulation, captured by $h_{2t}^{(3)}$ in the illustration, implies that younger executives may change firms more often than otherwise to gain this dimension of human capital and increase the chance of advancing to a high rank in the future. This element is similar to the prediction of the experimentation literature on human capital (Miller, 1984; Antonovics and Golan, 2012; Sanders, 2013) except that the experimentation literature requires learning about skills unknown to the executives, whereas we get the same prediction if the upper level of the hierarchy values a combination of skills that can be more efficiently acquired by working in multiple firms.

4.5 Firm Technology

In this subsection alone, it is necessary to identify the executive pool explicitly, since each firm may employ more than one executive, and distinguish their different ages from calendar time, because aggregate technological shocks are dated by the latter. Accordingly we now suppose there are N executives who sort themselves into positions, denote by $t(\tau, n)$ the age of the n^{th} executive at calendar time τ , and her human capital at τ by $h_{t(\tau, n)}$. Let $F_{jk}(h_{t(\tau, n)})$ denote the executive's contribution to the j^{th} firm's output in τ if she chooses the k^{th} job with that firm by setting $d_{jkt(\tau, n)} = 1$. Let $\pi_{\tau+1}$

denote a return from an exogenous aggregate productivity shock that affects every firm, and $\pi_{j,\tau+1}$ denote the (net) excess return to the j^{th} firm. Let $e_{j\tau}$ denote the value of firm j at the beginning of calendar time τ . Finally, denote by $w_{jk\tau+1}^{(n)}$ the firm's compensation to executive n if she worked at rank k in period τ . We assume the production of firm j at τ is then defined as the sum of three components:

- (i) $\sum_{n=1}^N \sum_{k=1}^K d_{jkt(\tau,n)} F_{jk}(h_{t(\tau,n)})$ is the contribution to output from all the firm's executives.
- (ii) $e_{j\tau}(\pi_{\tau+1} - 1)$ is the value of output attributable to the aggregate productivity shock.
- (iii) $e_{j\tau}\pi_{j,\tau+1}$ is the value from the excess return to the firm, $\pi_{j,\tau+1}$, which depends on effort of all the executives.

The first component of the output is additively separable in the productivity of each executive; it depends on the characteristics of executives in the firm but not on their efforts and captures each executive's contribution to the firm which depends on her human capital, $h_{t(\tau,n)}$. The individual factor is deterministic, has a level effect on the executive's marginal product, and is independent of the individual's effort and other executives' characteristics and efforts. The second component, $\pi_{\tau+1}$, captures the effect of aggregate factors on the firm's equity.⁴ In standard moral-hazard models, the optimal contract does not depend on the market portfolio or aggregate factors the executive cannot affect, because they are risk averse and a contract depending on such factors imposes additional risk on them without providing any additional incentive. Assuming all dividends are paid when the firm is liquidated, the equity of the firm evolves according to the law of motion⁵

$$e_{j,\tau+1} \equiv \sum_{n=1}^N \sum_{k=1}^K d_{jkt(\tau,n)} \left[F_{jk}(h_{t(\tau,n)}) - w_{jk\tau+1}^{(n)} \right] + e_{j\tau}(\pi_{\tau+1} + \pi_{j,\tau+1}). \quad (5)$$

Rearranging Equation (5) to make $\pi_{j,\tau+1}$ the subject reveals the standard definition of excess returns in the asset-pricing literature: the rate of increase in the value of the firm above and beyond the net return on the market portfolio.

The efforts of executives only affect the firm through $\pi_{j,\tau+1}$, which is determined stochastically. If all the executives in the firm work diligently, the value of $\pi_{j,\tau+1}$ is drawn from a distribution with probability density function $f_j(\pi)$. If everyone except the k^{th} ranked executive works, conditional on any level of human capital h , the value of $\pi_{j,\tau+1}$ is drawn from the distribution $f_j(\pi)g_{jk}(\pi | h)$.⁶ If two or more executives shirk, we further that assume the distribution for π does not depend on how many others shirk, and denote its density by $f_{0j}(\pi)$. The potential for moral hazard arises from the assumption that the greater the number of executives who shirk, their preferred action because $\alpha_{jkt}(h_t) > \beta_{jkt}(h_t)$, the lower the expected value of excess returns: formally for all (h, j, k) ,

$$\int \pi f_j(\pi) d\pi > \int \pi f_j(\pi)g_{jk}(\pi | h) d\pi > \int \pi f_{0j}(\pi) d\pi. \quad (6)$$

The left inequality in (6) shows that shareholders have higher expected payoffs when all executives work, yet executives prefer to shirk rather than work, creating a conflict between the executives' and shareholders' goals. We assume the likelihood ratio, $g_{jk}(\pi | h_t)$, is bounded; this prevents shareholders

⁴ Here, we are abstracting from other costs faced by the firm, such as the wage bill for the nonexecutive work force, and assuming that the market index already impounds these costs for a given firm size.

⁵ This formula can be easily modified to allow for dividends to be distributed throughout the life of the firm, but the firm's dividend policy does not affect the compensation paid to managers in our model.

⁶ Thus $g_{jk}(\pi | h)$ is the ratio of the density when the executive with h in position k shirks while all other executives work, and the density when all executives work diligently.

from achieving a first best contract with a (constant) wage, accomplished by deterring each executive from shirking with a sufficiently harsh punishment administered in states that could not have occurred if everyone worked. We also impose the regularity condition

$$\lim_{\pi \rightarrow \infty} g_{jk}(\pi | h) = 0. \quad (7)$$

Intuitively this condition states that if firm performance at the end of the period is truly outstanding, then shareholders are almost certain that all the executives have worked during the period. Our assumptions ensure the existence of an optimal contract with bounded compensation (Mirrlees, 1975), and are clearly weaker than the common monotonicity assumption requiring $g_{jk}(\pi | h)$ to decline in π . Given Nash responses by the executives, the second inequality in (6) implies there are only three basic cases to consider: either everybody works, everyone shirks, or all but one executive works. Under these assumptions, if more than one executive shirks, the shareholders can pay lower compensation to the remaining executives by encouraging them to shirk as well. The justification for simplifying the analysis this way when more than one executive shirks is that all the executives in our data are incentivized, so in the Nash equilibrium supporting the optimal contract only single deviations are considered.

Executives, therefore, have two separable components of productivity. We assume that when all executives work, the equilibrium outcome in our model, the ex-ante expected value of excess returns is zero, which is consistent with standard asset-pricing theory. The distortion caused by an executive who shirks does depend on firm and executive characteristics; therefore our model captures differences in the degree of divergence of executives' and shareholders' goals across firm and executive characteristics. It allows the severity of the agency problem to vary by firm and executive characteristic through technology and preferences. However the likelihood ratio, $g_{jk}(\pi | h)$, does not depend on the number of executives in their firm or their human capital; relaxing this assumption would endogenize the optimal number of executives and the configuration of human capital within the team, a challenge for future research. Since $g_{jk}(\pi | h)$ measures the degree to which executive effort affects the firm's returns, it can be interpreted as a measure of span of control. The estimates of $g_{jk}(\pi | h)$ provide some insight regarding the role of rank in the firm. For example, we can test whether our measure of span of control declines with rank, which would be consistent with Williamson (1967).⁷ Although the assumption of constant returns to scale precludes us from making predictions about the size distribution of firms, we can however whether the span of control increases with firm size. If so, then using the utility-function estimates of the costs of shirking, we can calculate whether the costs of agency increase in firm size. This might provide one justification to a diminishing returns to scale in firm size as postulated in Lucas (1978).

4.6 Information, Timing and Capital Markets

Information: Each executive is privy to her taste shocks, effort level and outside wealth. Similarly, consumption choices by executives are not public. All other information is symmetric. Everyone observes human capital, executive rank and firm choices of all executives plus their compensation for the previous period's employment. Although $F_{jk}(h_t)$ cannot be separately observed, it is also public knowledge. To summarize, at the beginning of each period τ , the market observes $(h_{t(\tau,n)}, d_{t(\tau,n)})$ and $\sum_{j=1}^J \sum_{k=1}^K d_{jk,t(\tau-1,n)} w_{jk\tau}^{(n)}$ for all N executives, plus the aggregate return π_τ , and the initial equity

⁷ In Williamson's (1967) hierarchical model of firms, there are decreasing returns to scale for labor as a manager moves up the hierarchy as a result of cumulative loss of "compliance" across the ranks. In our formulation, $g_{jk}(\pi | h)$ varies across the ranks of the hierarchy. Therefore, we can test whether managers' shirking causes larger distortions in higher ranks. In contrast, Mirrlees (1975) offers an alternative view of a firm as a decentralized contractual organization.

$e_{j\tau}$ and excess returns $\pi_{j\tau}$ of all J firms, while every executive also observes her own outside wealth $\xi_\tau^{(n)}$, and her idiosyncratic-taste shocks $\varepsilon_\tau^{(n)}$, and in addition recalls her own effort history $\{l_s^{(n)}\}_{s=0}^{t-1}$ as well.

Timeline: At the beginning of each period, executives are compensated according to their contracts. After observing her own taste shock vector, each executive privately chooses her consumption and makes her asset portfolio choice. She simultaneously decides whether to retire or not; and if she decides not to retire, which firm to be employed in, and at what rank and effort level. She approaches the firm, and begins negotiating with the shareholders. We assume the executive makes an ultimatum offer that the shareholders can only accept or reject. If no agreement is reached, the executive does not work during that period, and there is no additional hiring by the firm.

Capital Markets: Following Margiotta and Miller (2000), we assume that executives have sufficient access to financial markets to smooth their outside wealth without using their firm as a bank. In our model this means there exists a complete contingent-claims market for consumption, including all publicly disclosed events relating to commodities with price measure Λ_τ and derivative λ_τ at date τ . Thus for each $\tau \in \{0, 1, 2, \dots\}$, the term Λ_τ is the price at time 0 of contingent claims to consumption delivered at date τ . For example, $E[\lambda_\tau]$, is the number of consumption units forgone in date 0 to obtain a sure-consumption unit in date τ and $\{E[\lambda_\tau]\}^{-1} - 1$ is the τ -period interest rate. We measure $w_{jkt(\tau)+1}$, the executive's compensation for employment in position k at firm type j at the beginning of age $t + 1$, in units of current consumption. Since the executive's wealth is endogenously determined by her compensation, it cannot be fully insured if it depends on the firm's returns $\pi_{j,\tau+1}$. Naturally value maximizing banks would not voluntarily insure executives against volatile excess returns in their own firm, because the executive might then find it optimal to shirk, generating expected losses to the bank: public disclosure laws require top executives to declare their financial holdings in securities issued by their own firm, so given our technology, it is easy for banks to protect themselves against this form of insider trading.

4.7 Intertemporal Consumption and Employment Choices

The separability of preferences, the executives' absolute risk aversion and the completeness of the capital market allow us to focus on the executives' indirect utility function, which maps their expected utility as a function of the relevant security prices, the portion of their wealth that can be fully diversified, the distribution of any unanticipated changes in their wealth induced by the undiversifiable component of their contingent compensation and the option value of their stock of human capital. The indirect utility will be characterized in two steps. First, we derive the indirect utility function for an executive entering retirement at age $t + 1$. Upon retirement, the executive faces a single budget constraint for perpetuity. Second, we recursively solve the problem for the same executive at age t using Bellman's (1957) principle, the age- $t + 1$ utility is just the indirect utility function derived in the first step.

The consumer choice problem solved in the first step is standard in the asset-pricing literature (Debreu, 1959, Chap 7). As Rubinstein (1981) shown, the CARA assumption implies very few securities are required to characterize the optimal financial portfolio. In particular, let b_τ denote the price of a bond that, contingent on the history through date τ , pays a unit of consumption from period τ in perpetuity in period- τ prices:

$$b_\tau \equiv E_\tau \left[\sum_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau} \right]. \quad (8)$$

Similarly, let a_τ denote the price of a security that pays the random quantity $(\ln \lambda_s - s \ln \delta)$ of consumption from period τ in perpetuity in period- τ prices:

$$a_\tau \equiv E_\tau \left[\sum_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau} (\ln \lambda_s - s \ln \delta) \right]. \quad (9)$$

The executives have another asset to be priced: their human-capital stock. Before doing so, we introduce some additional notation. We denote the utility of the present value of compensation by

$$v_{jkt(\tau)+1} \equiv \exp \left(-\rho w_{jkt(\tau)+1}(h_t, \pi) / b_{\tau+1} \right). \quad (10)$$

Let $p_{jkt}(h)$ denote the probability of choosing (j, k) at age t conditional on h . Similarly, we denote the retirement probability by $p_{0t}(h)$. Finally, denote by ε_{jkt}^* the value of ε_{jkt} conditional on choosing (j, k) at t . Thus, $\varepsilon_{jkt}^* = \varepsilon_{jkt}$ if $d_{jkt} = 1$ and is not defined if $d_{jkt} = 0$. We define an index of human capital for a t -year-old executive with characteristics h who always works as

$$A_t(h) = p_{0t}(h) E \left[\exp \left(-\frac{\varepsilon_{0t}^*}{b_\tau} \right) \right] + \sum_{j=1}^J \sum_{k=1}^K p_{jkt}(h) \alpha_{jkt}(h)^{\frac{1}{b_\tau}} E \left[\exp \left(-\frac{\varepsilon_{jkt}^*}{b_\tau} \right) \right] \{A_{t+1} [\bar{H}_{jk}(h)] E_t[v_{jkt(\tau)+1}]\}^{1-\frac{1}{b_\tau}}. \quad (11)$$

The index $A_t(h)$ is a choice-probability-weighted average of expected outcomes from making different (j, k) choices, including retirement. By inspection, the index is strictly positive, and lower values of $A_t(h)$ are associated with higher values of human capital. Thus, increasing expected compensation reduces $E_t[v_{jkt(\tau)+1}]$ and $A_t(h)$. Similarly, $A_t(h)$ is monotonically increasing in $\alpha_{jkt}(h)$, the utility-weighted nonpecuniary losses of job characteristics. Combining the first and second steps gives us the rationale for defining a human-capital index in this way, as shown in Lemma 4.1.

Lemma 4.1 *Let $V_t(h, \xi_t, a_t)$ denote the discounted sum of expected utility from age $t < R$ onwards. For an executive with characteristics h and wealth ξ_t who has not yet observed ε_t and will make optimal consumption and job-match choices thereafter, subject to never shirking,*

$$V_t(h, \xi_t, a_t) = -b_\tau \exp \left(-\frac{a_t + \rho \xi_t}{b_\tau} \right) A_t(h). \quad (12)$$

The term $-b_\tau \exp[-(a_t + \rho \xi_t)/b_\tau]$ is the value function for a retiree defined above. Thus, Equation (12) shows that the optimized lifetime expected utility is the product of utility from financial wealth and human capital. This simplifies the maximization problem faced by executives: They can use the indirect utility from Lemma 4.1 in the lifetime utility formulation, Equation (2), to solve for their employment choice. This is summarized in Theorem 4.1.

Theorem 4.1 *If $t \leq R$ and $l_s = 1$ for all $s \in \{t, \dots, R\}$, then job choices d_t are picked to sequentially maximize*

$$d_{0t} \varepsilon_{0t} + \sum_{j=1}^J \sum_{k=1}^K d_{jkt} \left\{ \varepsilon_{jkt} - \ln \alpha_{jkt}(h) - (b_\tau - 1) \ln A_{t+1} [\bar{H}_{jk}(h)] - (b_\tau - 1) \ln E_t[v_{jkt(\tau)+1}] \right\}. \quad (13)$$

The formulation above encompasses several models of labor market sorting. First, it is a generalized Roy model: Based on human capital, executives sort into jobs and firms. The generalized form includes a Roy model and a compensating-differentials model: As in Roy models, the current compensation is given as $v_{jkt(\tau)+1}$, and as in hedonic-price models the nonpecuniary benefits from the job are given by $\alpha_{jkt}(h)$ and ε_{jkt} . The generalized dynamic Roy model is augmented by an additional component,

$A_{t+1}[\overline{H}_{jk}(h)]$, the future expected utility attached to a job. This index of utility includes pecuniary compensation, expected growth of earnings, as well as expected nonpecuniary utility from future jobs and ranks and can be interpreted as the value of human capital acquired in different ranks and jobs. Thus, working and acquiring experience in some ranks and firms is associated with assignments to firms, industries and ranks in which the pay for skills is higher, whereas some job choices provide human capital associated with lower expected earnings growth and lower promotion probability over the executive's career.

Next, we characterize the firm- and rank-choice probabilities and how they change over the lifecycle in an equilibrium in which all executives work diligently. These choice probabilities will map the model's parameters and the observed choice probabilities in the data, and therefore play an important role in the estimation strategies. The vector of conditional-choice probability functions, $p_t(h) \equiv (p_{11t}(h), \dots, p_{JKt}(h))$, that the executive uses to compute $A_t(h)$ in Equation (11) are precisely the probability functions that characterize her choices when solving the optimization function described by (13). We appeal to Proposition 1 of Hotz and Miller (1993): a mapping exists, $q(p) \equiv (q_{11}[p_t(h)], \dots, q_{JK}[p_t(h)])$, from the simplex to R^{JK} such that

$$q_{jk}[p_t(h)] = \ln \alpha_{jkt}(h) + (b_\tau - 1) \ln A_{t+1}[\overline{H}_{jk}(h)] + (b_\tau - 1) \ln E_t[v_{jkt(\tau)+1}]. \quad (14)$$

Given h , the solution to the optimization problem in Equation (13) depends only on the vector of differences $(\varepsilon_{11t} - \varepsilon_{0t}, \dots, \varepsilon_{JKt} - \varepsilon_{0t})$ rather than their levels, ε_t . This becomes apparent from substituting out $d_{0t} = 1 - \sum_{j=1}^J \sum_{k=1}^K d_{jkt}$ in Equation (13), collecting terms involving d_{jkt} , and noting that the additive constant, ε_{0t} , has no effect on the optimal choices. Substituting Equation (14) into (13), we see that if position (j, k) is the optimal employment choice, then $\varepsilon_{jkt} - \varepsilon_{0t} > q_{jk}[p_t(h)]$ and

$$(j, k) = \arg \max_{(j', k')} \{ \varepsilon_{j'k't} - q_{j'k'}[p_t(h)] \}. \quad (15)$$

Given (t, h) , the executive is indifferent between all positions if ε_t satisfies the condition

$$(\varepsilon_{11t} - \varepsilon_{0t}, \dots, \varepsilon_{JKt} - \varepsilon_{0t}) \equiv q[p_t(h)] \equiv (q_{11t}, \dots, q_{JKt}). \quad (16)$$

It now follows that $(\varepsilon_{0t}, q_{11t} + \varepsilon_{0t}, \dots, q_{JKt} + \varepsilon_{0t})$ defines, for all ε_{0t} , the set of idiosyncratic shocks, ε_t , for a executive who would marginally accept any of the JK positions or retire.

We end this section by showing the form Equation (14) takes if the distribution of ε_t is Type I extreme value and the transition function of human capital is as defined in Equation (4). The formal derivation is in the online appendix.

$$\ln \left(\frac{p_{jkt}(h)}{p_{0t}(h)} \right) = - \ln \alpha_{jkt}(h) - (b_\tau - 1) \left(\frac{1}{b_{\tau+1}} \ln p_{0,t+1}(h + \Delta_{jk}) + \ln \Gamma \left[1 + \frac{1}{b_{\tau+1}} \right] + \ln E_t[v_{jkt(\tau)+1}] \right), \quad (17)$$

where $\Gamma[\cdot]$ is the complete gamma function. In this particular case, the model delivers a log-linear equilibrium sorting function in the log-odds ratio. Equation (17) characterizes the supply for any given market rental rates of human capital in different ranks and firms, $v_{jkt+1}(h)$. While there is no wealth effect due to the absolute-risk-aversion-preference assumption, there is a clear intertemporal substitution effect captured by the relative price of a bond today versus tomorrow, b_τ and $b_{\tau+1}$. If the probability of retirement next period increases for some exogenous reason, this would decrease the probability of choosing any job today.

Equation (17) shows that executives trade off jobs based on three dimensions: nonpecuniary benefit, $\alpha_{jkt}(h)$; human-capital accumulation, Δ_{jk} ; and expected utility from compensation, $E_t[v_{jkt(\tau)+1}]$. For illustrative purposes, let us assume h is one dimensional and that Δ_{jk} is ordinal. Consider three scenarios:

- (i) Two jobs, say j and j' in the same rank, with different nonpecuniary benefits, say $\alpha_{jkt}(h) > \alpha_{j'kt}(h)$, the same level of human-capital accumulation, say $\Delta_{jk} = \Delta_{j'k}$, and different monetary compensation, say $w_{jkt(\tau)+1}(h_t, \pi) > w_{j'k,t(\tau)+1}(h_t, \pi)$ for all (h_t, π) . This implies $E_t[v_{jkt(\tau)+1}] < E_t[v_{j'k,t(\tau)+1}]$.
- (ii) Two jobs with different nonpecuniary benefits and human-capital accumulation, but the same compensation. In this case $\alpha_{jkt}(h) > \alpha_{j'kt}(h)$, $\Delta_{jk} > \Delta_{j'k}$, and $w_{jkt(\tau)+1}(h_t, \pi) = w_{j'kt(\tau)+1}(h_t, \pi)$.
- (iii) Two jobs with the same nonpecuniary benefits but different human capital and expected compensation, say $\alpha_{jkt}(h) = \alpha_{j'kt}(h)$, $\Delta_{jk} > \Delta_{j'k}$, and $w_{jkt(\tau)+1}(h_t, \pi) < w_{j'kt(\tau)+1}(h_t, \pi)$.

According to Equation (17), the executive cares about $E_t[v_{jkt(\tau)+1}]$ but not necessarily $w_{jkt(\tau)+1}(h_t, \pi)$. However, we simplify the discussion here by looking at $w_{jkt(\tau)+1}(h_t, \pi)$ in relation to $w_{j'kt(\tau)+1}(h_t, \pi)$ since that immediately allows us to find the sign of the difference between $E_t[v_{jkt(\tau)+1}]$ and $E_t[v_{j'kt(\tau)+1}]$. In (i) a higher value of $\alpha_{jkt}(h)$ relative to $\alpha_{j'kt}(h)$ will decrease the probability of choosing firm j relative to firm j' . On the other hand a higher $w_{jkt(\tau)+1}(h_t, \pi)$ relative to $w_{j'kt(\tau)+1}(h_t, \pi)$ will yield a lower $E_t[v_{jkt(\tau)+1}]$ relative to $E_t[v_{j'kt(\tau)+1}]$ and this will increase the probability of choosing firm j relative to firm j' . This highlights the trade-off between nonpecuniary and pecuniary benefits embedded in the model. A similar comparison in (ii) and (iii) is not possible because the probability of retiring next period, $p_{0,t+1}(\cdot)$, is an equilibrium object and we cannot find the sign of its derivative with respect to h unambiguously outside of the general equilibrium. If $p_{0,t+1}(\cdot)$ is decreasing in human capital, then in (ii) we get a similar trade-off as in (i) and in (iii) the executive is trading off higher human capital for lower expected compensation today. However, if $p_{0,t+1}(\cdot)$ is increasing in human capital, then in both (ii) and (iii) the effect from higher human capital tomorrow reinforces the effect from lower wages today. While the assumption that $p_{0,t+1}(\cdot)$ is decreasing in human capital may seem reasonable, this predicate relies on the assumption that human capital increases with expected compensation. We are not able to assert that at this stage as compensation is determined in equilibrium, and we have not yet analyzed the demand side of the executive market.

4.8 Labor Demand and Optimal Contract

Shirking by just one executive is disguised because every firm outcome that might occur when one executive shirks could also occur when every executive works; technically, the likelihood ratio, $g_{jk}(\pi | h)$, is bounded. In the equilibrium, every job history has strictly positive mass even though no shirking occurs along the equilibrium path. Underlying this result is our assumption that ε_{jkt} has full support and is privately known to only the executive. Now we consider that the executive has the choice of shirking. This does not affect the state variables' deterministic effect on the next period's human capital, but it does give the executive another combination of nonpecuniary and financial packages to choose from. As a consequence of the assumptions in Equation (6), we only need to look at a one-period optimal contract in which the shareholders guard against one executive shirking given that the rest of the team works diligently. We will later show that, in this version of the model, the one-period spot contract is the T -period replication of the optimal long-term contract.

The shareholders' objective will be to minimize the executive team's expected aggregate compensation bill because they are concerned about only a single executive shirking conditional on the others working diligently. This is equivalent to minimizing $E_t[w_{jkt(\tau)+1}(h, \pi) | h]$ or, equivalently, $E_t[\ln v_{jkt(\tau)+1} | h]$. The shirking contract minimizes $E_t[\ln v_{jkt(\tau)+1} | h]$ subject to a market-participation constraint characterized by the executive-employment-decision rule summarized in Equa-

tion (14). We define the certainty-equivalent compensation as

$$w_{jkt(\tau)+1}^*(h) = \rho^{-1} b_{\tau(t)+1} \left\{ (b_{\tau(t)} - 1)^{-1} \ln \alpha_{jkt}(h) + \ln A_{t+1} [\bar{H}_{jk}(h)] - (b_{\tau(t)} - 1)^{-1} q_{jk} [p_t(h)] \right\}. \quad (18)$$

If effort could be costlessly monitored and demand existed for executives giving effort, it follows from Equations (10) and (14) that a cohort of executives aged t all with human capital h confronted with job opportunities across K ranks in J firms offering $w_{jkt(\tau)+1}^*(h)$ would sort into the jobs following the probability distribution $p_t(h)$. The equilibrium compensation schedule must conform with the participation constraint implied by the certainty-equivalent wage to attract the mix of executives dictated by the conditional-choice probabilities and does not depend on the firm's excess return. There is little reason to presume that a contract subject to the participation constraint induces working over shirking. Because firms do not observe effort expended, they resort to voluntarily inducing working through the use of incentives in their employment contracts.

To elicit diligence from the executives, the shareholders must offer a contract that gives a higher utility than the outside option, Equation (14), and a higher expected utility than shirking provides. In this version of the model, the incentive-compatibility constraint that guarantees that the executives get a higher expected utility than shirking provides simplifies to

$$\alpha_{jkt}(h)^{1/(b_\tau-1)} E_t[v_{jkt(\tau)+1}] \leq \beta_{jkt}(h)^{1/(b_\tau-1)} E_t[v_{jkt+1} g_{jk}(\pi | h)]. \quad (19)$$

The compensation schedule minimizes expected wage payments from employment subject to the market-participation and incentive-compatibility constraints, which can be decomposed into two parts, a fixed and a variable component. We first define the variable part of compensation by

$$r_{jkt(\tau)+1}(h, \pi) \equiv \rho^{-1} b_{\tau(t+1)} \ln \left[1 - \eta(h, b_\tau) \left\{ g_{jk}(\pi | h) - [\alpha_{jkt}(h)/\beta_{jkt}(h)]^{1/(b_\tau-1)} \right\} \right], \quad (20)$$

where $\eta(h, b_{\tau(t)})$ is the unique positive root to

$$\int \left[\eta^{-1} + [\alpha_{jkt}(h)/\beta_{jkt}(h)]^{1/(b_{\tau(t)}-1)} - g_{jk}(\pi | h) \right]^{-1} f_j(\pi) d\pi = 1. \quad (21)$$

It is evident from (20) that a greater $g_{jk}(\pi | h)$, which implies that the outcome π is relatively more likely to occur when there is shirking, leads to a lower $r_{jkt(\tau)+1}(h, \pi)$. Contracting to pay less in states that are relatively more likely to occur when there is shirking encourages the executive to work. Since $g_{jk}(\pi | h) \rightarrow 0$ as $\pi \rightarrow \infty$, it follows that $r_{jkt(\tau)+1}(h, \pi)$ has a finite upper bound of

$$\bar{r}_{jkt(\tau)+1}(h) \equiv \rho^{-1} b_{\tau(t+1)} \ln \left[1 + \eta(h, b_\tau) [\alpha_{jkt}(h)/\beta_{jkt}(h)]^{1/(b_\tau-1)} \right]. \quad (22)$$

The higher the firm's returns, the less likely they could have been generated by shirking, and hence the lower the slope of the variable component to compensation. Theorem 4.2 states that the optimal contract is the sum of the compensating-equivalent wage and the variable component defined in the optimal contract.

Theorem 4.2 *The cost-minimizing one-period contract that attracts a executive of age t with experience h to select the k^{th} position in the j^{th} firm with probability $p_t(h)$ and work is*

$$w_{jkt(\tau)+1}(h, \pi) = w_{jkt(\tau)+1}^*(h) + r_{jkt(\tau)+1}(h, \pi). \quad (23)$$

The optimal long-term contract can be implemented by a sequence of the one-period contracts defined in (23). Intuitively, if the firm is not serving a banking function for wealth the executive has already accumulated and if the firm does not receive any further information about a shirking deviation after the period in which it occurs, then any punishment the firm might wish to administer for poor performance can be administered immediately (Malcomson and Spinnewyn, 1988; Fudenberg, Holmström, and Milgrom, 1990; Rey and Salanie, 1990).

We end this section by showing the form Equation (18) takes if the distribution of ε_t is Type I extreme value and the transition function of human capital as defined in Equation (4):

$$w_{jkt(\tau)+1}^*(h) = \frac{b_{\tau+1}}{\rho} \left\{ \frac{\ln \alpha_{jkt}(h)}{b_{\tau} - 1} + \frac{\ln \left(p_{0,t+1}(h + \Delta_{jk}) \Gamma \left[1 + \frac{1}{b_{\tau+1}} \right] \right)}{b_{\tau+1}} + \frac{1}{b_{\tau} - 1} \ln \left(\frac{p_{jkt}(h)}{p_{0t}(h)} \right) \right\}. \quad (24)$$

There are three sources of pay differentials required to attract a executive of characteristics h with probability $p_t(h)$. Differentials in the certainty-equivalent wage arise because jobs differ in the value and nonpecuniary costs of working, $\alpha_{jkt}(h)$, and in the value of human capital provided by the job, $p_{0,t+1}(h + \Delta_{jk})$. The lower the probability of retirement, the greater the future opportunities for extracting rent in the executive market, and hence the lower the certainty equivalent wage. In addition, jobs are different in the agency–risk premium, which is determined by the likelihood ratio, and by the relative disutility of working versus shirking in a particular job.

4.9 Equilibrium

Next, we complete the characterization of the equilibrium contract that clears the market. The executive supply, the choice probabilities of the different rank–firm combinations and retirement, is characterized by Equation (14) relating the compensation ($E_t [v_{jkt(\tau)+1}|h]$) and the choice probabilities. Theorem 4.2 characterizes the cost-minimizing contract that satisfies the market-participation constraint and the incentive-compatibility constraint. The market-participation constraint relates the certainty-equivalent wage required to attract any type of executive with characteristics h at a certain probability for each job. In equilibrium, the perceived probability of attracting an executive is the choice probability derived from the executive’s utility-maximization problem; the market-participation constraint derives from the supply equation ensuring this condition. Additionally, the incentive-compatibility constraint is satisfied so the executive works diligently. To close the model, we must pin down the demand (i.e., the implied probability of hiring a executive for a given contract). We assume free entry into the market for firms, implying a zero expected profit from hiring a manger in equilibrium.

Theorem 4.3 *In the equilibrium with a one-period incentive-compatible contract, the expected compensation equals the executive’s marginal productivity:*

$$E_t[w_{jkt(\tau)+1}(h, \pi)|h] = F_{jk}(h). \quad (25)$$

The above theorem states that each executive earns her expected marginal productivity in each period, conditional on accepting a contract in which the executive works diligently; the firm makes zero expected profit, as the expected return on the net equity value is zero, also conditional on all executives working diligently. Since the contract is incentive compatible, all executives work diligently. Solving backwards to the negotiation stage, given that all other executives work diligently, the executive extracts all the rents given that she has an incentive to work diligently. A executive cannot extract additional rents resulting from the distortion she causes by shirking instead of working diligently

because, given the incentive-compatibility constraint, the threat to shirk is not credible. Therefore, any offer higher than the expected productivity conditional on diligent work is rejected as the expected value of a vacancy, if a firm rejects the offer, is zero.

As in Rosen (1974), the market-clearing condition is achieved in a labor market in which there are no frictions in hiring or finding jobs. There is no scarcity of positions of any kind, and no costs associated with a vacant position, and the output is additively separable in executives. The marginal cost of attracting a executive at a given rate, however, is increasing in the nonpecuniary costs and risk; it also depends on the dynamic component, the continuation value of human capital. Prices adjust to attract the marginal executive with a taste shock that makes her indifferent between choosing the job and retirement with a given probability determined by the assumption on rent sharing. The equilibrium prices determine the fraction of executives for every h assigned to each position (j, k) . However, ex-post, a single firm may have zero, one, or more executives hired for a given rank in the firm.⁸ The strategies described above are optimal as we assume that there are no costs to hiring an additional executive to a position except for the compensation; since vacancies cost nothing and output is additively separable in each executive's output, we have constant returns to scale.

The value of accepting a job is firm specific because of firm-specific skills and because executives have independently distributed taste shocks, which are private information. Thus, in equilibrium, there is a surplus above the market outside option. In competitive models with match-specific surplus, firms make zero expected profit at the time of hiring, and the first-period wage adjusts to include the expected future profits firms make (Becker, 1964; Harris and Holmström, 1982; Thomas and Worrall, 1988; Felli and Harris, 1996). Here, executives earn their expected marginal product every period, so firms make zero expected profit from hiring an executive in each period. Our modeling choice is similar to those of Jovanovic (1979) and Miller (1984), among others. We assume that executives can make take-it-or-leave-it offers in order to achieve the per-period zero expected profit as an equilibrium outcome. Other mechanisms of surplus sharing in which the executives and shareholders share the surplus may be more realistic. Cahuc, Postel-Vinay, and Robin (2006) show that labor-market competition allows skilled workers to extract more surplus than unskilled workers, some of the rents skilled workers extract cannot be explained by competition over workers. Our sample, however, is of the very top level of the U.S. executive market, where talent is scarce. Thus, it is reasonable that these executives extract more rents than skilled workers in representative samples of the population. Our modeling choice of the ultimatum game simplifies the empirical implementation. In contrast to previous work, which estimates different rent-sharing mechanisms, we model internal promotions and ranks as opposed to modeling an employer as one job, and we characterize assignment within firms. If we allow heterogeneity in bargaining power for executives, the firm's problem will involve choosing an optimal size of management and configuration of employment (Stole and Zwiebel, 1996a, 1996b). These are important issues we leave for future research.

4.10 Example

We end this section by showing how one would calculate the equilibrium in our model if the distribution of ε_t is Type I extreme value and the transition function of human capital is as defined

⁸ In the data, we observe similar firms employing different numbers of managers in a given rank.

in Equation (4). Define

$$W_{jkt}(h, b_\tau) = -\ln \alpha_{jkt}(h) - (b_\tau - 1) \left(\frac{1}{b_{\tau+1}} \ln p_{0,t+1}(h + \Delta_{jk}) + \ln \Gamma \left[1 + \frac{1}{b_{\tau+1}} \right] \right) + (b_\tau - 1) \times \left[\frac{\rho}{b_{\tau+1}} F_{jk}(h) - E \ln \left((1 - \eta(h, b_\tau)) \left\{ g_{jk}(\pi | h) - [\alpha_{jkt}(h) / \beta_{jkt}(h)]^{\frac{1}{b_\tau - 1}} \right\} \right) \right]. \quad (26)$$

Then, the equilibrium ex-ante choice probabilities are

$$p_{jkt}(h) = \frac{\exp[W_{jkt}(h, b_\tau)]}{1 + \sum_{j=1}^J \sum_{k=1}^K \exp[W_{jkt}(h, b_\tau)]} \quad \text{if } j = 1, \dots, J \quad (27)$$

$$p_{0t}(h) = \frac{1}{1 + \sum_{j=1}^J \sum_{k=1}^K \exp[W_{jkt}(h, b_\tau)]} \quad (28)$$

Examining the right-hand side of Equation (26), we see that the first three components are the same as in Equation (17). Therefore, the final two components are the log of the expected utility from compensation. This is the value, measured in units of the difference between the marginal product produced by the executive and the risk premium needed for the variable part of the executive's compensation. Therefore, an executive's higher marginal product in that firm and rank is associated with a higher probability of choosing that firm–rank pair, holding everything else constant. In this way, the assignment to a firm and rank is efficient in our model. However, more moral hazard in that firm and rank leads to a lower probability of choosing that rank, holding everything else constant. Hence, the agency problem introduces inefficiencies into the sorting and assignment problem.

The optimal contract simplifies to

$$w_{jkt+1}(h, \pi) = F_{jkt}(h) - \frac{b_{\tau+1}}{\rho} E_t \ln \left((1 - \eta(h, b_\tau)) \left\{ g_{jk}(\pi | h) - [\alpha_{jkt}(h) / \beta_{jkt}(h)]^{\frac{1}{b_\tau - 1}} \right\} \right) + \frac{b_{\tau+1}}{\rho} \ln \left((1 - \eta(h, b_\tau)) \left\{ g_{jk}(\pi | h) - [\alpha_{jkt}(h) / \beta_{jkt}(h)]^{\frac{1}{b_\tau - 1}} \right\} \right). \quad (29)$$

Note that $w_{jkt+1}(h, \pi)$ depends only on primitives of the model, but the equilibrium sorting probabilities also depend on the next-period retirement probabilities—which is an equilibrium object—along with the model's primitives. Therefore, the equilibrium can be calculated in the four steps:

- (i) Solve for $\eta(h, b_\tau)$ using Equation (21) and use it to compute $w_{jkt+1}(h, \pi)$ using Equation (29).
- (ii) For each executive, set $t = R - 1$ and compute $W_{jkR-1}(h, b_{\tau(R-1)})$ and $p_{0R-1}(h)$, which will be a function of only the primitives of the model and $\eta(h, b_{\tau(R-1)})$ calculated in Step 1.
- (iii) Form $W_{jkR-2}(h, b_{\tau(R-2)})$ using the primitive of the model, $p_{0R-1}(h)$ from Step 2 and $\eta(h, b_{\tau(R-2)})$ calculated in Step 1.
- (iv) Recursively repeat Step 3 for $R - 3, \dots, t$.

5 The Extended Model

This section extends the basic model to account for career concerns' incentive effect. This is done by relaxing the assumption that human capital evolves independently of the executives' effort. This implies that human capital is now the executives' private information and is unobserved by the firms

and markets. Executives have full knowledge of their productivity, which evolves deterministically according to their choices. Current effort affects productivity, providing implicit incentives because current effort may impact future employment choices, promotions and pay. The private–human-capital model nests the public–human-capital model.⁹ As such, the notations and assumptions of the basic model on executives and firms, choices, and preferences will be common to both models and hence will not be modified. Below are the parts that will be modified.

5.1 Human-Capital Accumulation and Effort

Human capital is still multidimensional and the dichotomy between h_1 and h_{2t} remains the same as in the basic model. We assume that if an executive in rank k of the j^{th} firm works diligently, her human capital is augmented according to the transition function $\overline{H}_{jk}(h_{2t})$. The same transition function as in the basic model. However, if she shirks, then her human capital evolves according to another transition function, $\underline{H}_{jk}(h_{2t})$. Human capital is now private information to the executive because her effort choice is observed by neither the firm nor the market. Therefore, the law of motion of human capital is now

$$h_{2t+1} = \sum_{j=1}^J \sum_{k=1}^K d_{jkt} [l_t \overline{H}_{jk}(h_{2t}) + (1 - l_t) \underline{H}_{jk}(h_{2t})]. \quad (30)$$

If $l_t = 1$, human capital evolves according to the same function as in the basic version of the model, Equation (3). However, if $l_t = 0$, human capital evolves according to the law of motion $h_{2t+1} = \sum_{j=1}^J \sum_{k=1}^K d_{jkt} \underline{H}_{jk}(h_{2t})$. If $\overline{H}_{jk}(h_{2t}) = \underline{H}_{jk}(h_{2t})$ for all (j, k, h) , the effort choice l_t drops out of Equation (30). Therefore, Equation (30) nests Equation (3). Consider the example at the end of Section 4.4. Equation (4) still hold and we specify a similar equation for $\underline{H}_{jk}(h_{2t})$,

$$\underline{H}_{jk}(h_{2t}) = h_{2t} + \underline{\Delta}_{jk}, \quad (31)$$

where $\underline{\Delta}_{jk} \equiv (\underline{\Delta}_{jk}^{(1)}, \underline{\Delta}_{jk}^{(2)}, \underline{\Delta}_{jk}^{(3)})$. For example, if j is such that $j_1 = 0$ and $d_{0t} = 0$, then $\underline{\Delta}_{jk}^{(1)} = -h_{2t}^{(1)}$, which is the same as if the executive had worked diligently. However, $\underline{\Delta}_{jk}^{(2)} = 0$ and $\underline{\Delta}_{jk}^{(3)} = 0$, meaning that the executive would lose all her firm-specific capital and does not gain an additional year of executive experience or an increase in the number of firms she worked in if she shirks. On the other hand, if she chooses a j with $j_1 = 1$ and $d_{0t} = 0$, then $\underline{\Delta}_{jk}^{(1)} = 0$, $\underline{\Delta}_{jk}^{(2)} = 0$ and $\underline{\Delta}_{jk}^{(3)} = 0$ if she shirks.

5.2 Firm Technology and Effort

Productivity still has three components: the probability distributions, $g_{jk}(\pi | h) f_j(\pi)$ and $f_j(\pi)$, of excess returns and the individual marginal product, $F_{jk}(h)$. Since effort and human capital are linked, $g_{jk}(\pi | h)$ and $F_{jk}(h)$ now depend on the manager's past effort. In order to simplify the equilibrium characterization, we place some basic structure on these two objects:

- (i) Human capital effects individual output, $F_{jk}(h)$, more than joint output, $g_{jk}(\pi | h)$:

$$F_{jk}(\overline{H}_{jk}(h_{2t-1})) - F_{jk}(\underline{H}_{jk}(h_{2t-1})) > e_{j\tau} \int \pi [g_{jk}(\pi | \overline{H}_{jk}(h_{2t-1})) - g_{jk}(\pi | \underline{H}_{jk}(h_{2t-1}))] f_j(\pi) d\pi, \text{ for all } (j, k, t, h_{2t-1}).$$

⁹ There are other ways to introduce career concerns into the basic model. One is to add symmetric learning about executives' productivities, but this involves additional sources of uncertainty and is empirically less parsimonious. Another is to assume that each executive has a different cost of effort, which is known to the executive but unknown to firms. This involves a substantial extension to current model that could have empirical relevance on the equilibrium path. This would be a dynamic model with adverse selection and ratchet effects, but it is beyond the scope of this paper.

(ii) If $l_0 = 0$, then $F_{jk}(h) \equiv \underline{F}$.

Assumption (i) ensures managers' incentive to overreport their human capital. Assumption (ii) is an initial condition that places an upper bound on output and simplifies the notation and off-equilibrium-path analysis since it ensures a firm does not benefit from employing an executive who has shirked in the initial period.¹⁰

5.3 Capital Markets, Timing and Information

The capital market and timing assumptions of the extended and basic models are the same; however, the information structure of the extended model is a bit more complicated. Since $F_{jk}(h_t)$ cannot be separately observed and given that human capital is the executives' private information, $F_{jk}(h_t)$ is private information. We assume that all accepted and rejected contracts and employment histories are observed by all firms. This simplifies the off-equilibrium-path analysis by reducing the number of observed histories.

5.4 Intertemporal Consumption and Employment Choices

The managers' intertemporal consumption choices are unchanged from the basic model, but employment choice needs some additional notations and concepts. Let $h'_t = (h'_1, h'_{2t})$ denote shareholders' belief about a manager's human capital—that is, the manager's reputation—while $h_t = (h_1, h_{2t})$ continues to denote the manager's actual human capital. The contract is based on the manager's reputation, h'_t , not the manager's actual human capital, h_t . However, if the executive shirks, firm returns are related to the manager's actual human capital and drawn from $g_{jk}(\pi | h)f_j(\pi)$, not $g_{jk}(\pi | h')f_j(\pi)$. Consequently, the conditional-choice probabilities depend on both the manager's actual human capital, h_t , and the manager's reputation, h'_t . Shirking by just one manager is disguised because every firm return outcome that might occur when one manager shirks could also occur when every manager works. Similarly, firms cannot definitively recognize past shirking because individual productivity, $F_{jk}(h_t)$, is not observed separately from the executive team's aggregate output. The initially shirking executive's choices of job-match profiles do not reveal past shirking either. In the equilibrium, every job history has a strictly positive mass even if no shirking occurs along the equilibrium path. Underlying this result is the assumption that ε_{jkt} has full support and is private information. Therefore, when contracts are only offered for diligent work, shareholders believe that h'_t follows the law of motion $h'_{t+1} = \bar{H}_{jk}(h'_t)$ in any given history. In truth, if a manager deviates and shirks at age t , her next-period human capital is $h_{t+1} = \underline{H}_{jk}(h_t)$.

To complete the description of the manager's choice problem, we formulate the value of job matches to the manager when $h'_t \neq h_t$. We then describe the manager's optimal labor-supply choices, on and off the equilibrium path, and the cost-minimizing contract, assuming shareholders' beliefs are as described above. Later, we show that these shareholder beliefs are correct in equilibrium. Denote the manager's choice probabilities over positions in firms by $p_{jkt}(h, h')$. As compensation payments are based on firms' perception of human capital, in place of $v_{jkt(\tau)+1}$, the risk-adjusted utility from compensation is

$$v'_{jkt(\tau)+1} \equiv \exp(-\rho w_{jkt(\tau)+1}(h'_t, \pi)/b_{\tau(t+1)}). \quad (32)$$

Analogous to the definition of $A_t(h)$, we define the recursion

$$B_t(h, h') = p_{0t}(h, h') E_t \left[\exp\left(\frac{-\varepsilon_{0t}^*}{b_\tau}\right) \right] + \sum_{j=1}^J \sum_{k=1}^K \left\{ p_{jkt}(h, h') E_t \left[\exp\left(\frac{-\varepsilon_{jkt}^*}{b_\tau}\right) \right] V'_{jkt}(h, h') \right\}, \quad (33)$$

¹⁰ The human capital of a manager who did not shirk in the first period, but shirks later, evolves according to $\underline{H}_{jk}(h_{2t})$.

where

$$V'_{jkt}(h, h') \equiv \min \left[\alpha_{jkt}(h)^{\frac{1}{b_\tau}} \left\{ B_{t+1} \left[\overline{H}_{jk}(h), \overline{H}_{jk}(h') \right] E_t \left[v'_{jkt(\tau)+1} \right] \right\}^{1-\frac{1}{b_\tau}}, \right. \\ \left. \beta_{jkt}(h)^{\frac{1}{b_\tau}} \left\{ B_{t+1} \left[\underline{H}_{jk}(h), \overline{H}_{jk}(h') \right] E_t \left[v'_{jkt+1} g_{jk}(\pi | h) \right] \right\}^{1-\frac{1}{b_\tau}} \right]. \quad (34)$$

The difference between $A_t(h)$ and $B_t(h, h')$ stems from the minimization used to define $V'_{jkt}(h, h')$, the conditional valuation function of match (j, k) for a manager with demographics (t, h) and reputation h' . The first element of the minimization operator in Equation (34) is the manager's conditional valuation function, net of lifetime utility conferred by endowment wealth, at age t in position (j, k) with human capital h and reputation h' from choosing to work. If the manager could not choose to shirk, human capital and reputation would always be equated, and $B_t(h, h')$ would simplify to $A_t(h)$. Thus, the second element is a conditional-valuation function for a similarly placed manager to choose shirking: She reaps the immediate benefit from shirking since $\beta_{jkt}(h) < \alpha_{jkt}(h)$, but firm returns are drawn from $g_{jk}(\pi | h)f(\pi)$ rather than $f_j(\pi)$, affecting the probability distribution of her compensation; her reputation subsequently diverges further from her true human capital. Theorem 5.1 now extends the job-match problem from Equation (13) to include the choice of effort.

Theorem 5.1 *If $h'_{t+1} \equiv \overline{H}_{jk}(h'_t)$, then job matches d_t and effort levels l_t are picked to sequentially maximize*

$$\varepsilon_{0t} d_{0t} + \sum_{j=1}^J \sum_{k=1}^K d_{jkt} \left[\varepsilon_{jkt} - \ln V'_{jkt}(h, h') \right]. \quad (35)$$

An omitted induction can be used to prove that $B_t(h, h') \leq A_t(h)$ for any given compensation schedule because adding the option to shirk unambiguously increases the opportunity set. Consequently, the value from solving Equation (35) exceeds the value from solving (13). Recalling that the case of private information of human capital nests the public-information case, it immediately follows that if $\underline{H}_{jk}(h) = \overline{H}_{jk}(h)$ for all (j, k, t, h) , then $B_t(h, h) = A_t(h)$ for all (t, h) . So, if $t \leq R$ and $l_s = 1$ for all $s \in \{t, \dots, R\}$, we obtain the same characterization of the conditional probabilities as in the basic model with public information, Equation (14). Furthermore, the rest of the employment-choice analysis carries through to the private-information model.

Suppose that ε_{jkt} is independently and identically distributed as a Type I extreme value with location and scale parameters $(0, 1)$. Denote the probability of retirement for a manager with demographics (t, h) and reputation h' as $p_{0t}(h, h')$, then $B_t(h, h')$ simplifies to

$$B_t(h, h') = \Gamma \left(\frac{b_t + 1}{b_t} \right) p_{0t}(h, h')^{\frac{1}{b_t}}, \quad (36)$$

where $p_{0t}(h, h') = \left[1 + \sum_{j=1}^J \sum_{k=1}^K V'_{jkt}(h, h')^{-b_t} \right]^{-1}$. Equation (36) has the same form as Equation (17), the definition of $A_t(h)$, except it depends on $p_{0t}(h, h')$ instead of $p_{0t}(h)$ to reflect the role of the executives' reputation versus their actual human capital.

5.5 Labor Demand and Optimal Contract

The main difference between the labor demand and contracts in the basic and extended models is that career concerns may ameliorate the divergence of incentives between managers and shareholders in

the extended model. The definition of $V'_{jkt}(h, h')$ given in Equation (34) shows that the compensation schedule must satisfy the incentive-compatibility constraint:

$$\alpha_{jkt}(h)^{1/(b_\tau-1)} E_t[v_{jkt(\tau)+1}] A_{t+1} [\overline{H}_{jk}(h)] \leq \beta_{jkt}(h)^{1/(b_\tau-1)} E_t[v_{jkt(\tau)+1} g_{jk}(\pi | h)] B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]. \quad (37)$$

Thus, whenever $A_{t+1} [\overline{H}_{jk}(h)] < B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$, career concerns ameliorate the agency problem. Therefore, for any constant compensation, Equation (37) is satisfied if and only if

$$\ln \alpha_{jkt}(h) + (b_\tau - 1) \ln A_{t+1} [\overline{H}_{jk}(h)] \leq \ln \beta_{jkt}(h) + (b_\tau - 1) \ln B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]. \quad (38)$$

So, when the investment value of human capital is large enough relative to the increase in disutility from working, the incentive-compatibility constraint does not bind, obviating the need to tie remuneration to the firm's abnormal returns and pay a risk premium. Thus, career concerns provide implicit incentives that substitute explicit incentives provided by incentive contracts; since implicit incentives are larger when executives are young, explicit incentives increase as managers get closer to retirement age.

As before, the compensation schedule minimizes expected wage payments from employment subject to the participation and incentive-compatibility constraints decomposed into fixed and variable components. Define the variable component by

$$r_{jkt(\tau)+1}(h, \pi) \equiv \frac{b_{\tau+1}}{\rho} \ln \left[1 - \eta(h, b_\tau) \left[g_{jk}(\pi | h) - \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)} \right]^{1/(b_{\tau(t)}-1)} \frac{A_{t+1} [\overline{H}_{jk}(h)]}{B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]} \right] \right], \quad (39)$$

where $\eta(h, b_\tau)$ is the unique positive root to

$$\int \left[\eta^{-1} + \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)} \right]^{1/(b_\tau-1)} \left[\frac{A_{t+1} [\overline{H}_{jk}(h)]}{B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]} - g_{jk}(\pi | h) \right]^{-1} f_j(\pi) d\pi = 1. \quad (40)$$

For a manager who worked diligently up to period $t - 1$, the difference between the risk premiums in the basic and the extended models is the value of human capital attained by diligent work relative to the value of human capital attained if the manager had shirked. Theorem 5.2 states that the optimal contract is the sum of the compensating-equivalent wage and the variable component defined in the optimal contract.

Theorem 5.2 *If $h' = h$, then the cost-minimizing one-period contract that attracts a manager of age t with experience h to select the k^{th} position in the j^{th} firm with probability $p_t(h)$ and work is*

$$w_{jkt(\tau)+1}(h, \pi) = w_{jkt(\tau)+1}^*(h) + r_{jkt(\tau)+1}(h, \pi). \quad (41)$$

The difference between the cost-minimizing contracts in the basic and extended models is the risk premium, which is weakly smaller when there are career concerns. In addition, the result that the optimal long-term contracts can be implemented as a sequence of short-term contracts does not apply in the extended model. In the extended model shirking executives affect the firm's future returns, both directly through F_{jk} , and also, since $h \neq h'$ for shirking executives, indirectly through the cost of achieving incentive compatibility. Thus, a long-term contract that promises to punish managers for poor firm performance several periods from now has a current deterrent effect, and when used in conjunction with immediate punishment is potentially cheaper to implement because more than one signal is used to achieve incentive compatibility in any given period. We interpret the optimal one-period contract in the extended model as an economically meaningful departure from the null

hypothesis that the data can be rationalized by a sequence of short-term contracts replicating an optimal long-term contract.

As $w_{jkt(\tau)+1}^*(h)$ is similar under the basic and the extended models, the main difference is in the risk premium, $r_{jkt(\tau)+1}(h, \pi)$. There, the main inclusion is $A_{t+1} [\overline{H}_{jk}(h)] / B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$ which simplifies in our extreme-value example to

$$\frac{A_{t+1} [\overline{H}_{jk}(h)]}{B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]} = \left(\frac{p_{0t+1}[h + \Delta_{jk}, h + \Delta_{jk}]}{p_{0t+1}[h + \underline{\Delta}_{jk}, h + \Delta_{jk}]} \right)^{\frac{1}{b_{\tau+1}}}. \quad (42)$$

The incentive compatibility constraint would automatically be satisfied, so $\eta(h, b_{\tau}) = 0$ if

$$(b_{\tau} - 1) / b_{\tau+1} \ln(p_{0t+1}[h + \Delta_{jk}, h + \Delta_{jk}] - p_{0t+1}[h + \underline{\Delta}_{jk}, h + \Delta_{jk}]) \leq \ln[\beta_{jkt}(h) - \alpha_{jkt}(h)]. \quad (43)$$

This would mean that the optimal contract that elicits diligent effort is given by

$$w_{jkt(\tau)+1}(h, \pi) = \frac{b_{\tau+1}}{\rho} \left\{ \frac{\ln \alpha_{jkt}(h)}{b_{\tau} - 1} + \frac{\ln p_{0t+1}[h + \underline{\Delta}_{jk}, h + \Delta_{jk}] \Gamma \left[1 + \frac{1}{b_{\tau+1}} \right]}{b_{\tau+1}} + \frac{1}{b_{\tau} - 1} \ln \left(\frac{p_{jkt}(h, h)}{p_{0t}(h, h)} \right) \right\}, \quad (44)$$

which is independent of π . So in contrast to the basic model, an executive compensation independent of the firm's excess return does not necessarily mean the shareholders are demanding shirking.

5.6 Equilibrium

In contrast to the basic model, the game in the extended model is a signaling game. Given the support of the realization of output and the support of the taste shock, all outcomes and job-rank choices are consistent with the beliefs that no manager has shirked. Thus, job-rank choices and output realizations do not serve as a signal. However, the contracts executives offer may serve to signal their level of human capital. We use the sequential-equilibrium refinement because, after the first period, the entire game consists of one subgame.

Theorem 5.3 *A sequential equilibrium with one-period contracts exists where expected compensation equals the worker's marginal productivity:*

$$E_t [w_{jkt(\tau)+1}(h, \pi) | h] = F_{jk}(h). \quad (45)$$

At the offer stage, a manager with any level of human capital h offers the cost-minimizing contract specified in Equation (44) for the beliefs h' . These offers are accepted if the manager has never deviated from making these equilibrium offers in the past. Any other offer is rejected. Firms believe that all managers making offers deviating from the above contract have shirked in all periods. In equilibrium no executive shirks and $h' = h$.

The full description of strategies and beliefs on and off the equilibrium path and a proof is in the appendix. We establish by construction the existence of a sequential equilibrium in which managers sequentially expropriate all the rent that can be extracted from one-period contracts. Along the equilibrium path, managers work every period, so $h = h'$ for all t . If the manager shirks, $h \neq h'$, and the variable pay components, designed for reputation h' , do not necessarily align the incentives of shareholders with those of the manager who is off the equilibrium path. Having deviated from the equilibrium path by shirking once, it may be optimal for a manager to shirk at some future time, as Equation (33) indicates. One possibility not accommodated by the construction of $B_t(h, h')$ is

a manager who has always shirked attempting to confess during her negotiations with shareholders. What happens if she offers a contract in the ultimatum game that differs from $w_{jkt(\tau)+1}(h', \pi)$, such as $w_{jkt(\tau)+1}(h, \pi)$? In the equilibrium we construct, shareholders interpret any deviation from $w_{jkt(\tau)+1}(h', \pi)$ as proof the manager has shirked initially and is, therefore, a liability to the firm. In particular, it is straightforward (but not instructive) to write down an upper bound for \underline{F} in terms of the model's primitives that ensures no manager who has shirked in all t periods has accumulated sufficient wealth to compensate the firm for expected losses that the firm will incur by retaining such a manager. This assumption effectively truncates behavior off the equilibrium path because, given the shareholders' beliefs, it is a best response of the manager who has optimally selected (j, k) to demand $w_{jkt(\tau)+1}(h', \pi)$ and follow the continuation path implied by $B_t(h, h')$.¹¹

5.7 Example

We end this section by showing how one would calculate the equilibrium in the extended model if the distribution of ε_t is Type I extreme value and the transition function of human capital is as defined in Equations (4) and (31). We assume that $\underline{\Delta}_{jk} = 0$ and, when $h = h'$, we compress the double argument to one for illustrative purposes. Redefine $W_{jkt}(h, b_\tau)$ to be inclusive of both the basic and extended models:

$$W_{jkt}(h, b_\tau) = -\ln \alpha_{jkt}(h) - (b_\tau - 1) \left(\frac{1}{b_{\tau+1}} \ln p_{0t+1}[h + \Delta_{jk}] + \ln \Gamma \left[1 + \frac{1}{b_{\tau+1}} \right] \right) + (b_\tau - 1) \times \left[\frac{\rho}{b_{\tau+1}} F_{jk}(h) - E \ln \left(1 - \eta(h, b_\tau) \left\{ g_{jk}(\pi | h) - \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)} \right]^{\frac{1}{b_\tau-1}} \left(\frac{p_{0t+1}[h + \Delta_{jk}]}{p_{0t+1}[h, h + \Delta_{jk}]} \right)^{\frac{1}{b_{\tau+1}}} \right\} \right) \right]. \quad (46)$$

The equilibrium ex-ante choice probabilities have the same form as in Equation (27) using the new definition for $W_{jkt}(h, b_\tau)$ in Equation (46). The optimal contract simplifies to

$$w_{jk,+1}(h, \pi) = F_{jk}(h) - \frac{b_{\tau+1}}{\rho} E_t \ln \left(1 - \eta(h, b_\tau) \left\{ g_{jk}(\pi | h) - \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)} \right]^{\frac{1}{b_\tau-1}} \left(\frac{p_{0t+1}[h + \Delta_{jk}]}{p_{0t+1}[h, h + \Delta_{jk}]} \right)^{\frac{1}{b_{\tau+1}}} \right\} \right) + \frac{b_{\tau+1}}{\rho} \ln \left(1 - \eta(h, b_\tau) \left\{ g_{jk}(\pi | h) - \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)} \right]^{\frac{1}{b_\tau-1}} \left(\frac{p_{0t+1}[h + \Delta_{jk}]}{p_{0t+1}[h, h + \Delta_{jk}]} \right)^{\frac{1}{b_{\tau+1}}} \right\} \right). \quad (47)$$

There is one major difference between the sorting probabilities in the extended and basic models: Human capital has two different effects on the sorting patterns in equilibrium. The first, the human-capital motive captured by $p_{0t+1}[h + \Delta_{jk}]$ in Equation (46), is already in the basic model; the second comes from the career concerns' incentive effect, captured by $p_{0t+1}[h + \Delta_{jk}]/p_{0t+1}[h, h + \Delta_{jk}]$ in Equation (46). This is because career concerns reduce the risk premium that must be paid to an executive. Therefore, if the executive is comparing two jobs with the same productivity technology (i.e., $F_{jk}(h)$, $f_j \pi$) and $g_{jk}(\pi | h)$, nonpecuniary benefits (i.e., $\alpha_{jkt}(h)$ and $\beta_{jkt}(h)$) and human-capital accumulation potential (i.e., Δ_{jk}), but different career concerns, the executive has a higher probability of choosing the job with the greater career concerns because the certainty-equivalent wage would be higher there. Therefore, career concerns ameliorate the inefficiencies introduced into the sorting and assignment problem by the agency problem. Also, $w_{jkt(\tau)+1}(h, \pi)$ depends not only on primitives as in the basic model, but also on the next-period retirement probability, which is an equilibrium object. Therefore,

¹¹ We can make other assumptions and construct off-equilibrium-path behavior in which no manager truthfully reveals her type and that no contracts eliciting shirking behavior are offered. There might be other equilibria consistent with the estimation. However, since the out game is elaborate, the off-equilibrium path becomes less tractable.

unlike the basic model, the objects of the equilibrium must be calculated recursively. Note that in period $t = R - 1$ the contract is the same as in the basic model and therefore the equilibrium can be calculated with the following steps.

- (i) For each executive, set $t = R - 1$.
 - (a) Solve for $\eta(h, b_{\tau(R-1)})$ using Equation (21) and use it to compute $w_{jk,R-1}(h, \pi)$ using Equation (29).
 - (b) Compute $W_{jkR-1}(h, b_{\tau(R-1)})$ and $p_{0R-1}(h, h')$, which will be a function of only the model's primitives and $\eta(h, b_{\tau(R-1)})$ calculated in Step 1(a).
- (ii) For each executive, set $t = R - 2$.
 - (a) Solve for $\eta(h, b_{\tau(R-2)})$ using Equation (40) and use it to compute $w_{jk,R-2}(h, \pi)$ using Equation (47) with $p_{0R-1}(h, h')$ calculated in Step 1(b).
 - (b) Compute $W_{jkR-2}(h, b_{\tau(R-2)})$ and $p_{0R-2}(h, h')$ using the primitives of the model, $p_{0R-1}(h, h')$ from Step 1(a) and $\eta(h, b_{\tau(R-2)})$ calculated in Step 2(a).
- (iii) Recursively repeat Step 3 for $R - 3, \dots, t$.

6 Identification

The extended model nests the basic model, so it suffices to analyze identification in the extended model. The model is characterized by the preference parameters, ρ , $\alpha_{jkt}(h_t)$, $\beta_{jkt}(h_t)$ and $G(\varepsilon_{11t}, \dots, \varepsilon_{JKt})$; the technology parameters, $F_{jk}(h)$, $f_j(\pi)$ and $g_{jk}(\pi | h_t)$; and the human-capital transition functions, $\underline{H}_{jk}(h)$ and $\overline{H}_{jk}(h)$. Our data consist of matched panel data on firms and their managers in different time periods, $(w_{ijk\tau}, d_{ijk\tau}, \pi_{j\tau}, h_{i\tau}, t_{i\tau}, b_{\tau})$ where $i = 1, \dots, I$ indexes the individual executives, $j = 0, \dots, J$ indexes the firms, $k = 1, \dots, K$ indexes the rank and $\tau = 1, \dots, \Upsilon$ indexes the time periods. There are two cases to investigate: When is it optimal for managers to shirk in equilibrium? When is it optimal for managers to work in equilibrium? In the first case, managers' compensation, $w_{ijk\tau}$, should be independent of $\pi_{j\tau}$. Since this is never the case in our data, we focus on what can be identified when it is optimal for managers to work in equilibrium, and when the incentive compatibility constraint is met with equality.

When the data are generated by an equilibrium where managers work in equilibrium, $F_{jk}(h)$, $f_j(\pi)$ and $\overline{H}_{jk}(h)$ are immediately identified from the data: $F_{jk}(h)$ is identified from the conditional expectation of $w_{ijk\tau}$ on $h_{i\tau}$, $t_{i\tau}$ and $d_{ijk\tau}$ using the rent-extraction condition in Equation (45); $f_j(\pi)$ is identified from observations on $\pi_{j\tau}$; while $\overline{H}_{jk}(h)$ is identified from the empirical distribution of $h_{i\tau+1}$ at $t_{i\tau}+1$ conditional on $d_{ijk\tau}$ and $h_{i\tau}$ at $t_{i\tau}$. As shown in Magnac and Thesmar (2002), $G(\varepsilon_{11t}, \dots, \varepsilon_{JKt})$ is not identified nonparametrically. For this we assume the econometrician knows $G(\varepsilon_{11t}, \dots, \varepsilon_{JKt})$, and analyze the identification of ρ , plus the semiparametric identification of $\alpha_{jkt}(h_t)$, $\beta_{jkt}(h_t)$, $g_{jk}(\pi | h_t)$ and $\underline{H}_{jk}(h)$. It is instructive to highlight the differences between the basic and extended models, by letting $1_{\{\text{private}\}}$ denote an indicator function taking a value of one if human capital is private and zero if not, and defining a virtual shirking parameter as:

$$\beta_{jkt}^*(h) \equiv \beta_{jkt}(h) \left\{ \frac{B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]}{A_{t+1} [\overline{H}_{jk}(h)]} \right\}^{1_{\{\text{private}\}}(b_t-1)}. \quad (48)$$

We proceed in three steps: first the identification of $\alpha_{jkt}(h_t)$, $\beta_{jkt}^*(h_t)$ and $g_{jk}(\pi | h_t)$ is considered when ρ is known. Then we explore conditions under which ρ is identified. The third step establishes conditions under which $\beta_{jkt}(h)$ and $\underline{H}_{jk}(h)$ are identified from the knowledge of $\beta_{jkt}^*(h)$.

Step 1: The finite-upper-bound property of $r_{jkt(\tau)+1}(h, \pi)$ as established in Equation (22) and the optimal compensation schedule in Equation (41) imply that compensation is bounded and the manager's maximum compensation is

$$\lim_{\pi \rightarrow \infty} w_{jkt(\tau)+1}(h, \pi) = w_{jkt(\tau)+1}^*(h) + \bar{r}_{jkt(\tau)+1}(h) \equiv \bar{w}_{jkt(\tau)+1}(h). \quad (49)$$

Suppose ρ is known and define the mappings for $g_{jk}(\pi | h_t, \rho)$, $\alpha_{jkt}(h_t, \rho)$ and $\beta_{jkt}^*(h_t, \rho)$ as

$$g_{jk}(\pi | h_t, \rho) = \frac{e^{\rho \bar{w}_{jkt(\tau)+1}(h_t)/b_{\tau+1}} - e^{\rho w_{jkt+1}(h_t, \pi)/b_{\tau+1}}}{e^{\rho \bar{w}_{jkt+1}(h_t)/b_{\tau+1}} - E[e^{\rho w_{jkt(\tau)+1}(h, \pi)/b_{\tau+1}} | h_t, j]} \quad (50)$$

$$\alpha_{jkt}(h_t, \rho) = \frac{\exp(q_{jk}[p_t(h)])}{A_{t+1} [\bar{H}_{jk}(h)]^{b_{\tau}-1}} E \left[e^{-\rho w_{jkt(\tau)+1}(h, \pi)/b_{\tau+1}} | h_t, j \right]^{1-b_{\tau}} \quad (51)$$

$$\beta_{jkt}^*(h_t, \rho) = \frac{\exp(q_{jk}[p_t(h)])}{A_{t+1} [\bar{H}_{jk}(h)]^{b_{\tau}-1}} E \left[e^{\rho w_{jkt(\tau)+1}(h, \pi)/b_{\tau+1}} g_{jk}(\pi | h, \rho) | h_t, j \right]^{1-b_{\tau}}. \quad (52)$$

These mappings are derived from Equations (41), (14) and (37). They are similar to the mappings derived in Gayle and Miller (2013), except that $g_{jk}(\pi | h_t, \rho)$ is conditional on h_t , while $\alpha_{jkt}(h_t, \rho)$ and $\beta_{jkt}^*(h_t, \rho)$ are scaled by $\exp(q_{jk}[p_t(h)])A_{t+1} [\bar{H}_{jk}(h)]^{1-b_{\tau}}$ to reflect the equilibrium sort and the dynamic human-capital accumulation. Additionally, instead of $\beta_{jkt}(h)$, Equation (52) refers to $\beta_{jkt}^*(h)$ which is not a primitive of the model, but instead an equilibrium object. The equilibrium compensation schedule, $w_{jkt+1}(h_t, \pi)$, is identified by the conditional expectation of $w_{ijk\tau}$ on $(d_{ijk\tau}, \pi_{j\tau}, h_{i\tau}, t_{i\tau}, b_{\tau})$; therefore, $\bar{w}_{jkt(\tau)+1}(h_t)$ is also identified by the maximum of $w_{ijk\tau}$ conditional on $(d_{ijk\tau}, h_{i\tau}, t_{i\tau}, b_{\tau})$. Therefore, the likelihood ratio, $g_{jk}(\pi | h_t, \rho)$, is identified if ρ is known.

Appealing to Proposition 1 of Hotz and Miller (1993), a mapping $q_{jk}[\cdot]$ exists conditional on $G(\varepsilon_{11t}, \dots, \varepsilon_{JKt})$. Consider the exponent version of Equation (14) raised to the power of $1/b_{\tau}$

$$\alpha_{jkt}(h)^{\frac{1}{b_{\tau}}} \{E_t[v_{jkt+1}]A_{t+1} [\bar{H}_{jk}(h)]\}^{1-\frac{1}{b_{\tau}}} = \exp[q_{jk}(p_t[h])/b_{\tau}]. \quad (53)$$

Substituting Equation (53) into Equation (11) gives

$$A_t(h) = p_{0t}(h)E \left[\exp \left(\frac{-\varepsilon_{0t}^*}{b_{\tau(t)}} \right) \right] + \sum_{j=1}^J \sum_{k=1}^K p_{jkt}(h)E \left[\exp \left(\frac{-\varepsilon_{jkt}^*}{b_{\tau(t)}} \right) \right] \exp[q_{jk}(p_t[h])/b_{\tau}]. \quad (54)$$

Hotz and Miller (1993) show that if $G(\varepsilon_{11t}, \dots, \varepsilon_{JKt})$ is known, then $E[\exp(-\varepsilon_{jkt}^*/b_{\tau(t)})]$ can be written as a known function of the conditional-choice probabilities. Therefore, $A_t(h)$ can be written as

$$A_t(h) = \varphi(p_t[h], h, b_{\tau}), \quad (55)$$

where $\varphi(\cdot)$ is a known function. The choice probability, $p_t(h)$, is identified by the conditional expectation of $d_{ijk\tau}$, on $(h_{i\tau}, t_{i\tau}, b_{\tau})$ and therefore $A_t(h)$ is identified. It follows immediately from Equations (51) and (52) that $\alpha_{jkt}(h_t, \rho)$ and $\beta_{jkt}^*(h_t, \rho)$ are identified up to ρ , since $\exp(q_{jk}[p_t(h)])$, $A_{t+1} [\bar{H}_{jk}(h)]^{1-b_{\tau}}$ and $g_{jk}(\pi | h_t, \rho)$ are identified. To summarize, if the risk-aversion parameter is known, then $\alpha_{jkt}(h)$, $g_{jk}(\pi|h)$ and $\beta_{jkt}^*(h)$ are semiparametrically identified.

Step 2: Gayle and Miller (2013) show that in a model of moral hazard with neither turnover, promotion nor human-capital accumulation, the risk-aversion parameter, ρ , is only set identified. Their analysis exploits conditions derived from both cost minimization and profit maximization—that

in equilibrium the (shareholder) principal only offers work contracts if it is more profitable than paying (executive) agents to shirk. Their analysis proves that any positive value of the risk aversion parameter can be rationalized by the cost minimization conditions; the profit maximization condition is necessary to obtain an inequality that defines an interval for the identified set of ρ . In our model, the introduction of turnover, promotion and human capital yield additional moments for identification. Viewing the compensation schedule offered in different ranks and firms as a lottery, we use the equilibrium sorting condition over ranks and firm types to point identify ρ .

The equilibrium sorting condition identifies ρ when exclusion restrictions exist that limit the dependence of the taste parameters on variables the help determine the contract. Substituting for $A_{t+1} [\overline{H}_{jk}(h)]$ in (53) using (55), and rearranging, we obtain

$$\alpha_{jkt}(h)^{\frac{1}{b_\tau-1}} E_t[v_{jkt(\tau)+1}] = \frac{\exp[q_{jk}(p_t[h])/(b_\tau-1)]}{\varphi(p_{t+1}[\overline{H}_{jk}(h)], \overline{H}_{jk}(h), b_{\tau+1})} \equiv z_{jkt}(h, b_\tau, b_{\tau+1}), \quad (56)$$

where $z_{jkt}(h, b_\tau, b_{\tau+1})$ is a known function of the data. Identification then follows from assumptions that some components of (j, k, t, h, b_τ) affect $z_{jkt}(h, b_\tau, b_{\tau+1})$ but neither ρ nor $\alpha_{jkt}(h)$: all the elements in (j, k, t, h, b_τ) belong to the information set of the executive at the beginning of each age period t (by the assumptions of the model), affect her choices (which can be ascertained by checking for variation in the conditional choice probabilities), and are therefore qualify as valid instruments if they do not affect preferences as well. For example, human capital provides a natural source of exclusion restrictions. In this paper we assume that ρ is independent of the executives' level of human capital, and that the nonpecuniary cost of switching firms or ranks does not depend on some dimension of human capital accumulation: in estimation, we use previous ranks. Similarly b_τ is a valid instrument if, as we assume, ρ and $\alpha_{jkt}(h)$ are independent of the aggregate state of the economy.

Let x denote a vector of instruments constructed from (h, j, k, b_τ) for each observation, and define the unconditional density of π as $f(\pi)$. Applying the law of iterated expectations to (56) implies

$$E[z_{jkt}(h, b_\tau, b_{\tau+1}) | x] = E \left[\alpha_{jkt}(h)^{\frac{1}{b_\tau-1}} \exp \left(\frac{-\rho w_{jkt(\tau)+1}(\pi, h)}{b_{\tau+1}} \right) \frac{f_j(\pi)}{f(\pi)} | x \right]. \quad (57)$$

Thus ρ and $\alpha_{jkt}(h)$ are identified off Equation (57).

Step 3: Using Equation (48), we rewrite Equation (52) as

$$\beta_{jkt}(h) B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]^{b_\tau-1} = \beta_{jkt}^*(h) A_{t+1} [\overline{H}_{jk}(h)]^{b_\tau-1} E \left[e^{-\rho w_{jkt(\tau)+1}(h, \pi)} | h_t, j \right]^{1-b_\tau}. \quad (58)$$

The product $\beta_{jkt}(h) B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$ is identified from (58) because the right side that equation is identified from the previous steps. However no further headway can be made without adding restrictions to the model. Imagine the data is generated by the extended model and substitute the virtual parameter $\beta_{jkt}^*(h)$ defined in Equation (48), the incentive-compatibility constraint for the extended model, into Equation (37). This gives the incentive-compatibility constraint for the basic model, (19), with $\beta_{jkt}^*(h)$ replacing $\beta_{jkt}(h)$. Neither (13) nor (14) depend on $\beta_{jkt}(h)$ or the information structure because the manager works in the equilibrium of both models. Therefore the solution to the optimal-contract problem given by Equations (20), (21) and (23) for the private-information model is obtained by replacing $\beta_{jkt}(h)$ with $\beta_{jkt}^*(h)$.

These arguments suggest that models with private information about human capital, and induce career concerns, are observationally equivalent models with public information about human capital, where there are none. Specifically, $\beta_{jkt}^*(h)$ indexes observationally equivalent models that differ only

in their specification of $\underline{H}_{jk}(h)$ and $\beta_{jkt}(h)$. We formally state this result as follows, for the case where bond prices are constant over time.¹²

Theorem 6.1 *Let Θ denote the class of models under consideration, consisting of elements*

$$\theta \equiv (\alpha_{jkt}(h), \beta_{jkt}^*(h), \rho, f_j(\pi), g_{jk}(\pi | h), G(\varepsilon)).$$

Suppose $b_\tau = b$ for all τ and $(w_{ijk}, d_{ijk}, \pi_j, h_i, t_i)$ is generated by $\tilde{\theta}$. For every $\hat{\rho} > 0$ and all proper probability distribution functions $\tilde{G}(\varepsilon)$ defined on the same support as $G(\varepsilon)$, there exists a unique $\hat{\theta}$ solving Equations (35), (41), (45), (50), (51) and (52) that is observationally equivalent to $\tilde{\theta}$.

Nevertheless the models of hidden information but no screening can be distinguished from models with career concerns with the aid of additional restrictions. For the purposes of decomposition it suffices to identify $B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$, the continuation value for an executive shirks for the first time. This section concludes by showing three such restrictions:

- We could specialize $\beta_{jkt}(h)$, by assuming it does not depend on the executive's age: $\beta_{jkt}(h) = \beta_{jk}(h)$ for all t , and assume there is a maximal age of retirement R . Recall $A_{t+1} [\overline{H}_{jk}(h)] = B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)] = 1$ for all $t \geq R$, because there is no incentive effect from career concerns in the period immediately preceding retirement. Consequently the shirking parameter is identified for executives at age $R - 1$ from (48) as

$$\beta_{jk}(h) = \beta_{jkR-1}^*(h) E \left[e^{-\rho w_{jk,R}(h,\pi)} | h_t, j \right]^{1-b_\tau}.$$

Having identified $\beta_{jkt}(h)$ the continuation value associated with shirking the first time is then identified off (52) for all $t \leq R - 2$ as

$$B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)] = [\beta_{jkt}^*(h) / \beta_{jk}(h)]^{\frac{1}{b_\tau-1}} A_{t+1} [\overline{H}_{jk}(h)] / E \left[e^{-\rho w_{jk,t(\tau)+1}(h,\pi)} | h_t, j \right]$$

Intuitively the incentive effect from career concerns at younger ages can be identified by comparing the aggregate incentive at younger ages to the aggregate incentive effect one year before retirement.

- Similarly, suppose $\beta_{jkt}(h)$ is independent of the aggregate prices in the economy, as summarized in our model by b_τ . For simplicity, assume two distinct bond prices b_τ and $b_{\tau'}$: From Equation (58), we can show that $B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_\tau]$ ¹³ is identified relative to a normalization that $B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau'}] = 1$.
- Finally, if we assume that the off equilibrium belief about the law of motion of human-capital accumulation, $\underline{H}_{jk}(h)$, is known, we can calculate $B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_\tau]$ numerically from $t = R$ backward using Equation (33).

¹² A more general result holds when b_τ varies over time, providing the parameters are also permitted to vary with calendar time.

¹³ Here, we make explicit the dependence of $B_{t+1} [\underline{H}_{jkt}(h), \overline{H}_{jkt}(h)]$ on b_τ , which was suppressed for notional simplicity.

7 Estimation

In our empirical framework we assume throughout that ε_t is distributed as a Type 1 extreme value. The computational advantages of parameterizing $G(\varepsilon)$ this way are evident from the formulas for $A_t(h)$ and $B_t(h, h')$ in Equation (36) and the expression for $q_{jk}[p_t(h)]$ in Equation (17). On and off the equilibrium path, the human-capital transition functions are deterministic; see Equations (4) and (31) for $\overline{H}_{jk}(h)$ and $\underline{H}_{jk}(h)$, respectively.

We employ a four-step procedure to estimate and test our models, which is directly the identification conditions:

- (i) Flexibly estimate $w_{jkt}(\pi, h)$, $\overline{w}_{jkt}(h)$, $f_j(\pi)$, $f(\pi)$, $\overline{H}_{jk}(h)$, and $p_{jkt}(h)$.
- (ii) Estimate ρ and $\alpha_{jkt}(h)$ off sample moments formed from population moments implied by (57), replacing $w_{jkt}(\pi, h)$, $\overline{w}_{jkt}(h)$, $f_j(\pi)$, $f(\pi)$, $\overline{H}_{jk}(h)$ and $p_{jkt}(h)$ with their estimates obtained from Step 1.
- (iii) Use the formulas from Equations (50) and (52) to estimate $g_{jk}(\pi | h)$ and $\beta_{jkt}^*(h)$ by replacing ρ with its estimate from Step 2 and $w_{jkt}(\pi, h)$, $\overline{w}_{jkt}(h)$, $f_j(\pi)$, $\overline{H}_{jk}(h)$ and $p_{jkt}(h)$ with their estimates from Step 1.
- (iv) Numerically calculate $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$ recursively, Assuming that $\beta_{jkt}(h)$ is independent of b_τ and that $\underline{H}_{jk}(h)$ is known, we and test the implied overidentifying restrictions.

An alternative estimation strategy is to exploit the equilibrium-computation algorithm outlined in Section 5.7. It involves computing a nested fixed-point algorithm to calculate $\eta(h, b_\tau)$ and $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$ for different values of the primitives in an inner loop, and using the results from the inner loop to estimate the primitives of the model in an outer loop. This alternative strategy is not only computationally burdensome, and in practical applications somewhat obscure. It also requires a fully parametric specification of $f_j(\pi)$, $g_{jk}(\pi | h)$ and $F_{jk}(\pi | h)$. In our paper, all these parameters are nonparametrically estimated. Another advantage of the estimation strategy used above is that it allows us to impose the different identification restrictions only when needed: for example, the restrictions needed to identify $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$ are only imposed when we estimate the effects of career concerns.

Step 1: The state space for the dynamic system is the Cartesian product of the manager's age, t , and personal background, $h_t \in \{1, \dots, H\}$ at the beginning of each period, as well as a vector that includes her employer firm last period, $j_{t-1} \in \{1, \dots, 36\}$, management rank last period, $k_{t-1} \in \{0, 1, \dots, 5\}$, fixed components (such as cohort, gender and education) and other variable components (such as measures of executive experience). Job matches in our model follow a stochastic law of motion, $p_{jkt}(h_t)$ and $p_{0t}(h_t)$. We estimate a multinomial logit model of firm type and position transitions with some (but not all) interactions for exit, promotions, and turnover. In estimation, we exploit Bayes' rule: Given background h , the (joint) probability, $p_{jkt}(h_t)$, is the product of the probability of choosing the j^{th} firm conditional on choosing the k^{th} rank, and the (marginal) probability of choosing Rank k . The compensation schedule, $w_{jkt(\tau)}(\pi, h)$, is estimated using a polynomial, and the boundary condition, $\overline{w}_{jkt(\tau)}(h)$, is estimated using the maximum of $w_{jkt(\tau)}(\pi, h)$ over π . Finally, $f_j(\pi)$ and $f(\pi)$ are estimated using kernel-density estimators with normal kernel and the Silverman rule of thumb for the bandwidth.

Step 2: To estimate ρ and $\alpha_{jkt}(h)$ we exploit the exclusion restrictions discussed in the identification section by forming population moments from (57)

$$E[z_{jkt}(h, b_\tau, b_{\tau+1})x] = E \left[\alpha_{jkt}(h)^{\frac{1}{b_\tau-1}} \exp \left(\frac{-\rho w_{jkt(\tau)+1}(\pi, h)}{b_{\tau+1}} \right) \frac{f_j(\pi)}{f(\pi)} x \right]. \quad (59)$$

Upon substituting (14) and (36) into (56) yields, $z_{jkt}(h)$ simplifies to

$$z_{jkt}(h) \equiv \Gamma \left[\frac{b_{t(\tau)+1} + 1}{b_{t(\tau)+1}} \right]^{-1} p_{0,t+1} (\overline{H}_{jk}(h))^{\frac{-1}{b_{t(\tau)+1}}} \left[\frac{p_{0t}(h)}{p_{jkt}(h)} \right]^{\frac{1}{(b_{t(\tau)-1})}}. \quad (60)$$

We approximate $z_{jkt}(h)$ by substituting the Step 1 estimates of the conditional-choice probabilities, $p_{0t}(h)$, $p_{jkt}(h)$ and $p_{0,t+1}(\overline{H}_{jk}(h))$ into (60). Sample analogs for the conditional-choice probabilities, compensation schedule, and conditional and unconditional densities of the abnormal return from Step 1 are substituted into Equation (59). Consistent estimates of ρ and $\alpha_{jkt}(h)$ are then obtained from the approximate sample moments along with (consistently estimates of their) standard errors adjusted for the pre-estimation.

We specify $\alpha_{jkt}(h)$ as a log-linear function of age, age squared, tenure, tenure squared, executive experience, executive experience squared, number of employers before becoming an executive, number of employers after becoming an executive, and indicators for board membership, interlocked, no college degree, MBA, MS/MA, Ph.D., and gender. We estimate an unrestricted version of the model that allows $\alpha_{jkt}(h)$ and ρ to be fully interacted with rank and firm type. This allows us to test whether ρ is a function of firm size, a possibility that might arise if our absolute-risk-aversion assumption is violated (Baker and Hall, 2004). We interact these 16 variables with rank and firm type to form $\alpha_{jkt}(h)$. We also permit the risk-aversion parameter to vary by the 36 firm types, but not by rank. In total, there are $(16 \times 5 + 1) \times 36 = 2,916$ parameters to be estimated. Equation (59) yields an orthogonal condition for each rank and firm combination, giving $5 \times 36 = 180$ moment conditions. In addition to the variables affecting $\alpha_{jkt}(h)$, we use bond prices and the lag of Ranks 1 through 4 as instruments, adding another $5 \times 20 \times 36 = 3,600$ moment conditions. After rejecting the null hypothesis that ρ varies with firm size, we impose these and other non-rejected restrictions on the results and reestimate the model. These restrictions are a common ρ for all firm types and that the effect of rank and firm type in $\alpha_{jkt}(h)$ is additive. We now have $(16 \times 36 + 5 \times 16 + 1) = 657$ parameters to estimate. We obtain similar results from both the restricted and unrestricted versions and hence only the restricted version is reported.

Step 3: We form $\widehat{w}(h_t, \pi)$, the nonparametric estimates of the compensation schedule, as a polynomial expansion from Step 1, using them in conjunction with our estimate of the risk-aversion parameter obtained from Step 2. We approximate the conditional expectation, $E_t[\exp(-\widehat{\rho}\widehat{w}(h_t, \pi)/b_{\tau(t)+1})]$, by integration using the nonparametrically estimated density of π for a given j , from Step 1, and compute $\overline{w}_{jkt(\tau)+1}(h)$ using the maximum $\widehat{w}(h_t, \pi)$ for each value of (j, k, t, h) . Finally, our estimate of $g_{jkt(\tau)}(\pi | h)$ is obtained by substituting our estimates of $\overline{w}_{jkt(\tau)+1}(h)$, ρ and $E_t[v_{jkt(\tau)+1}(\rho, \pi)]$ into Equation (50). A similar procedure is used in the estimation of $\beta_{jkt}^*(h)$ using Equation (52).

Step 4: In the extended model, substituting the Type I extreme-value functional form of $q_{jk}[p_t(h)]$ into Equation (52) and rearranging gives

$$\beta_{jkt}(h) \equiv \frac{p_{0t}(h)}{p_{jkt}(h)} B_{t+1} [\underline{H}_{jk}(h), \overline{H}_{jk}(h)]^{1-b_{t(\tau)}} \left\{ \frac{E_t[v_{jkt(\tau)+1}] - \overline{v}_{jkt(\tau)+1}^{-1}}{1 - \overline{v}_{jkt(\tau)+1} E_t[v_{jkt(\tau)+1}^{-1}]} \right\}^{1-b_{t(\tau)}} \quad (61)$$

for all (j, k, t, h) . Estimates of $\beta_{jkt}(h)$ and $B_t(h, h')$ are obtained recursively. Noting that $B_{T+1}(h, h') \equiv 1$ and substituting our estimated risk-aversion parameter and conditional-choice probabilities into Equation (61) yields $\beta_{jkT}(h)$. Substituting $\beta_{jkT}(h)$ into Equation (34) yields $V'_{jkT}(h, h')$ and hence $B_T(h, h')$, using Equation (36). More generally, given $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$, $\beta_{jkt}(h)$ is obtained from Equation (61), and hence estimates of $V'_{jkt}(h, h')$ and $B_t(h, h')$ are produced from Equations (34) and (36), respectively.

8 Pay Differential in the Executive Labor Market

This section presents estimates of the different components of pay that explain the sources of pay differential across ranks and firms in the executive labor market. It decomposes the differential into compensating variation in utility, investment value in human capital, and a risk premium. To understand the differentials in risk premium, we further analyze the variation in the net benefit and costs of shirking across firms, ranks and executives, which depend on the technology and preference parameters.

8.1 Expected Compensation Decomposition

We first present the estimates of the components of Equation (18). Expected compensation can be decomposed into two additive components: certainty equivalent pay and a compensating differential to risk averse executives for bearing risk in the form of firm-denominated securities, called a risk premium. The certainty equivalent wage factors into three additive components:

$$w_{jkt(\tau)+1}^*(h) = \frac{b_{t(\tau)+1}}{\rho(b_{t(\tau)} - 1)} \ln \alpha_{jkt}(h) + \frac{b_{t(\tau)+1}}{\rho} \ln A_{t+1}[\overline{H}_{jk}(h)] - \frac{b_{t(\tau)+1}}{\rho(b_{t(\tau)} - 1)} q_{jk}[p_t(h)]. \quad (62)$$

The first component of Equation (62), denoted $\Delta_{jkt}^\alpha(h)$, is the compensating differential due to the nonpecuniary utility gain or loss incurred by working in (j, k) relative to the outside option; it could even arise in a static model. The second component of (62), denoted $\Delta_{jkt}^A(h)$, is the investment value of (j, k) from accumulating human capital. The third component, denoted $\Delta_{jkt}^q(h)$, is a compensating differential due the idiosyncratic preference shocks (Rosen, 1974). It measures the compensating differential due to demand for labor in the firm and rank (j, k) : $q_{jk}[p_t(h, b_\tau)]$ is the value of the disturbance $\varepsilon_{jkt} - \varepsilon_{0t}$ that makes the marginal executive in (j, k) indifferent between that position and her outside option. The only structural parameters needed to estimate the certainty equivalent and its decomposition which cannot be estimated nonparametrically are $\alpha_{jkt}(h)$ and ρ . The other ingredients, the choice probabilities, the compensation schedule, and the distribution of abnormal return, are all estimated nonparametrically.

We then present estimates of the risk premium, defined as the difference between expected compensation and its certainty equivalent in equilibrium. From (25) expected compensation is simply the expected value of the executive's marginal product:

$$\Delta_{jkt}^r(h) \equiv E_t[r_{jkt(\tau)+1}(h, \pi)] = F_{jk}(h) - w_{jkt(\tau)+1}^*(h). \quad (63)$$

Since the executive does not shirk in equilibrium, $\Delta_{jkt}^r(h)$ does not directly depend on $\underline{H}_{jk}(h)$, $\beta_{jkt}(h)$, or $B_{t+1}(\cdot)$, that is what happens to human capital, utility, or the continuation value if the executive shirks.

8.1.1 Firm Size and Ranks

Figure 3 presents the components of expected pay decomposition by firm size and rank. Figure 1b shows that executive expected pay is greater in large firms and in higher ranks (up to Rank 2). The most striking result in Figure 3a is that the certainty-equivalent wage decreases with firm size. The average certainty equivalent wage of an executive in a small firm is \$780,000, falling to \$430,000 for a medium-size firm, and to \$390,000 for a large firm. The discount for the value of human capital accumulation does not vary appreciably with firm size, and larger firms have a higher demand for executives reflected in greater compensating differentials for the marginal executive hired to meet demand. However these two factors are overwhelmed by a third one: small firms inflict greater nonpecuniary losses on executives than large firms.

There is a positive relationship between firm size and the variance of compensation.¹⁴ In principle, the higher variability of compensation in large firms could be due to volatility in abnormal returns that factors into compensation packages and are accounted for by the risk premium, or to other forms of heterogeneity, both observed and unobserved. Figure 3b shows that a risk premium designed to solve agency problems can reconcile expected compensation that increases in firm size with certainty equivalent wages that decrease in firm size.

In addition to the negative relationship between firm size and nonpecuniary benefit from working, the distribution of ranks across firm size, as demonstrated in Figure 1b, contributed to the difference between the average compensation and the certainty equivalent by firm size. Figure 3c shows that the certainty-equivalent wage is concave over ranks, lowest in Rank 5, \$570,000, increasing monotonically to \$900,000 in Rank 2, before declining to \$690,000 in Rank 1. It is instructive to note that Rank 3 executives have a higher certainty equivalent compensation, \$730,000, than Rank 1 executives, but that Rank 1 executives have a slightly higher certainty equivalent compensation than Rank 4 executives, \$660,000. This ordering follows that of the average total compensation by executive rank reported in Table 1, which ranges from \$1,269,000 (for Rank 5) to \$4,794,000 (for Rank 2). However, the compression of the certainty equivalent pay is due to the risk premium.

The lifecycle theory of human capital predict that as executives age, human-capital investment becomes less important. In support of the theory, Table 1 shows higher ranks are held by older executives with more executive experience, and the value of human-capital investment decreases with all measures of experience. However, Figure 3b also shows that executives give up more compensation for human capital investment as they progress through the ranks right up until Rank 1, where the trend falls off. In our model, the investment value of human capital is inversely related to the probability of exit. Thus, this pattern reflects the exit probability, which from Table 2 is lowest in Rank 2 and highest in Rank 1 and is lower in larger firms. Intuitively, the effective discount factor used to compute the value of human capital, in terms of summed future increased earnings within the occupation, must account for the probability of exit. Consequently, standard models of human capital where everybody retires at the same rank would overpredict human-capital investment in the lower ranks and underpredict the level of investment in higher ranks. As a fraction of their certainty equivalent wage, the value of human capital is bracketed between approximately one quarter and one half of total compensation, remarkably high given the distribution of ages, positions, and the lengths of future careers. The major new finding on human capital investment is that even late in the career cycle, variety in job experience adds to human capital, and that the value of human capital is higher in large firms. This finding suggests that in the top ranks of the executive occupation there might be general human capital accumulated while executives accumulate management experience in different environments.¹⁵ Although human capital accumulation is important, the risk premium is the largest

¹⁴ See Table 2A in the online appendix.

¹⁵ Table 9A in the online appendix showing that the value of human capital increases with turnover by roughly \$13K

component explaining variation in pay across firm sizes and ranks. We further discuss the sources of this variation and the nature of the agency problem in different ranks and firms in the next subsection.

8.1.2 The Risk Premium

As indicated by the above results, risk-aversion measures account for much of the variation in pay across ranks and firms with different sizes. The parameters in our model for estimation of the risk premium are the risk-aversion parameters; hence, we first examine the robustness of their estimates. We initially specified the risk-aversion parameter as a function of gender and firm size, but at the one percent level could not reject the null hypothesis that male and female executives and executives that sort into different firm sizes have the same coefficient of risk aversion. Our estimate of the risk-aversion parameter (for all groups) is 0.534 with a standard error of 0.152, for compensation measured in millions of 2006 US\$. For example, a executive with risk-aversion parameter of 0.534 would be willing to pay \$255,199 to avoid a gamble that has an equal probability of losing or winning one million dollars. This is similar to results in Gayle and Miller (2009b), who found a risk-aversion parameter of 0.501 using data on 37 firms for the period 1944–1978 and 0.519 using data on 151 firms for the period 1993–2004. Our estimate of risk aversion is generally lower than that found in laboratory experiments and field studies (Holt and Laury, 2002, 2005; Harrison, Johnson, McInnes, and Rutström, 2005; Harrison, List, and Towe, 2007; Andersen, Harrison, Lau, Rutström, 2008; Dohmen, Falk, Huffman, and Sunde, 2010). This is plausible given that we are studying executives who, it is reasonable to assume, are more risk loving than the general population. This actually makes our estimate of the risk premium more plausible: Applying a lower risk-aversion parameter than is appropriate for the general population, we still find that the risk premium is the major component explaining the variation of compensation over rank and firm.

Table 3 displays our estimates of $\Delta_{jkt}^r(h)$, showing that, at Ranks 4 and 5, the cost of agency, measured by $\Delta_{jkt}^r(h)$, is small and insignificant in small firms, but it adjusts to \$1.5 million, \$3.3 million and \$1 million for Ranks 3, 2, and 1. Roughly 82% of the compensation of a CEO (Rank 2), versus 72% for Rank 1, 76% for Rank 3, 65% for Rank 4, and 69% for Rank 5, is due to the risk premium. The service sector pays a higher risk premium than the other two, a factor which helps close the gap between the considerably higher levels of average compensation paid in that sector and those reported in Table 3.

The risk premium increases significantly with firm size. On average an executive in a small firm receives \$1.6 million in risk premium (56% of expected compensation), \$2.8 million in a medium-size firm (85% of expected compensation), and \$4.8 million in a large firm (90% of expected compensation). These results are a further demonstration that the positive relationship between expected compensation and firm size is fully accounted for by the positive relationship between the size of the risk premium paid to executives and the size of their employer firms.

In our framework, expected compensation is the executive’s marginal product: Thus, executives with a Ph.D., who receive an average expected compensation of \$3.0 million, are more productive than those with an MBA, \$2.7 million, and without either, \$2.8 million. An executive with a Ph.D. receives a higher risk premium, \$2.3 million, than one with an MBA, \$2.1 million, but an executive with an MBA has a higher fraction of expected compensation, 78%, than one with Ph.D., 76%, as risk premium. There is a \$362,000 spike in the risk premium for new executives, but it declines by \$65,000 with each extra year of tenure and age. Consequently, the lower certainty-equivalent wage offered to first-year executives is partially hidden by data on their average compensation. Given that larger firms have more executives with MBA degrees and fewer tenured executives, the above two findings both

supports these hypothesis.

work to overstate the firm-size pay premium in the raw data. The overall effect of the interaction with firm size and rank is ambiguous: For example, the effect of Rank 1 overstates the effect of firm size while the effect of Rank 5 understates it. After controlling for the effect of rank and human capital, we find a negative relationship between firm size and certainty-equivalent wage, the main cause of which is the positive relationship between firm size and the risk premium.

8.2 Agency-Cost Decomposition

The risk premium is the manifestation of the agency cost in a moral-hazard model. This agency cost is a nonlinear function of the technology, preferences, and human capital, hence, there is no additive decomposition. There are, however, two counterfactual welfare measures that can measure the relative contribution of different sources to the overall agency cost: the gross loss from shirking to shareholders, which depends on technology parameters, and the executive's net benefit from shirking, which depends on preferences and career concerns. The second welfare measure can be further decomposed into two components measuring the relative contribution of preference and career concerns.

The gross loss to shareholders is the difference between expected return to the firm from a executive working versus shirking when all other executives are working diligently and is given by

$$\Delta_{jkt}^g(h) \equiv E_t [\pi(1 - g_{jk}(\pi, h))]. \quad (64)$$

The expected gross loss measures the importance of technology parameters to the risk premium, which can be measured by the gross output loss to the firm from switching from $f_j(\pi)$, the density of abnormal returns obtained from working, to its shirking counterpart, $f_j(\pi)g_{jk}(\pi | h)$. It also measures the quality of the signal about effort: If $g_{jk}(\pi, h) = 1$ the signal, π , is uninformative about the effort of the executive and $\Delta_{jkt}^g(h) = 0$. If there exist some point in the support of the signal, say π_0 , where $g_{jk}(\pi_0, h)$ is arbitrarily large then the signal is very informative about the effort of the executive and the first best allocation is possible and $\Delta_{jkt}^g(h)$ would be arbitrarily large. The gross loss to shareholders is obtained by first estimating the likelihood ratio of working versus shirking, $g_{jk}(\pi, h)$. This likelihood ratio is identified from the slope of the compensation schedule and the volatility of compensation under different aggregate conditions captured by bond prices as shown in Equation (50). This is the same in both the basic and extended models of moral hazard and, hence, is independent of the assumption on whether human capital evolves independent of effort. Thus, we do not have to specify $\underline{H}_{jk}(h)$ in order to estimate $\Delta_{jkt}^g(h)$.

The net benefit from shirking to the executive is denoted by $\Delta_{jkt}^{\beta^*}(h)$, it is the sum of two components: The first, denoted $\Delta_{jkt}^\beta(h)$, is the compensating differential for current utility when the executive weighs shirking against working; it measures the misalignment of incentives from the executive's perspective. Therefore it is the value an executive would place on shirking if she were paid the certainty-equivalent wage. The second, denoted $\Delta_{jkt}^B(h)$, measures the difference in the conditional continuation values from working in the current period t versus shirking; it measures by how much career concerns ameliorate the agency problem. These two components are defined as

$$\Delta_{jkt}^\beta(h) \equiv \frac{b_{t(\tau)+1}}{\rho(b_{t(\tau)} - 1)} \ln \left(\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)} \right), \quad (65)$$

and

$$\Delta_{jkt}^B(h) \equiv \frac{b_{t(\tau)+1}}{\rho} \ln \left(\frac{A_{t+1}[\overline{H}_{jk}(h)]}{B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]} \right).$$

The net benefit from shirking, $\Delta_{jkt}^{\beta^*}(h)$, is identified from data on choice probabilities, compensation schedule, the abnormal return distribution, the risk-aversion parameter and the likelihood ratio (see

Equation (52)) and is therefore identified without appealing to the functional-form assumptions on career concerns (i.e., $\underline{H}_{jk}(h)$) or exclusion restrictions. However, to separately identify $\Delta_{jkt}^B(h)$ and $\Delta_{jkt}^\beta(h)$ from $\Delta_{jkt}^{\beta^*}(h)$, we must rely on the functional-form assumptions on $\underline{H}_{jk}(h)$ given in Equation (31) and the exclusion restriction that $\beta_{jkt}(h)$ is independent of the aggregate conditions in the economy, i.e., bond prices. This is the only place in our analysis of the pay differential in the executive labor market where these assumptions are needed.

8.2.1 Firm Size and Rank

Table 4 shows our estimates of $\Delta_{jkt}^g(h)$, Table 5 reports our estimates of $\Delta_{jkt}^{\beta^*}(h)$ and Table 6 reports our estimates of $\Delta_{jkt}^B(h)$.¹⁶ We summarize in Figure 4 all the estimates by firm size. Figure 4 (and Table 4) shows that small, consumer-sector firms lose 33.6% of their equity value when a Rank-5 executive shirks, but large firms lose much less, 8%. This is in contrast to the finding in Baker and Hall (2004), whose estimates imply constant loss across firm size. Intuitively, shirking executives in small firms cause significantly more damage than they would in large firms because an executive in a smaller firm has a greater marginal impact on each unit of equity than any one executive working for a large firm. It also shows that signal quality is unambiguously poorer in larger firms. Additionally, Figure 4 (and Table 5) shows that $\Delta_{jkt}^\beta(h)$ declines in firm size, by \$3.1 and \$4.5 million for medium and large firms, respectively, and differs across sectors, \$3.8 million higher in the service sector than the consumer sector, and \$2.6 lower in the primary sector. It is evident from Figure 4 (and Table 6) that $\Delta_{jkt}^B(h)$ and career concerns do not vary by firm size. The estimates of $\Delta_{jkt}^\beta(h)$ and $\Delta_{jkt}^B(h)$ imply that the risk premium is weakly decreasing in firm size therefore the finding that signal quality is unambiguously poorer in larger firms is the reason why the risk premium is increasing in firm size. There is also a positive relationship between firm size and the expected gross loss in equity from shirking. Multiplying our estimates by the average equity value gives gross equity losses of \$102 million for a small firm, \$203 million for a medium one, and \$393 million for a large one. The gross loss in equity value from shirking would be higher in large firms; therefore the agency cost is concave increasing with firm size.

Turning to rank, the most surprising result is that the gross loss (Table 4), $\Delta_{jkt}^g(h)$, monotonically declines in rank; thus, when a Rank 1 executive in a large firm shirks, only a small proportion of equity value is lost. Similarly, the extent of destruction is lower for higher lagged ranks. These findings contradict conventional wisdom that shareholders risk more from chairmen and CEOs who shirk than lower-ranked officers; our results are consistent with the view that executives closer to the firm's operations can wreak the most havoc and therefore the excess return of the firm is a better signal of their effort. The losses are greatest in the service sector, least in the primary sector.

The value an executive would place on shirking if she were paid the certainty-equivalent wage (Table 5), $\Delta_{jkt}^{\beta^*}(h)$, is about \$10 million for a 50-year-old Rank 5 executive in a small firm in the consumer sector. It is increasing in rank; while economically significant, the rank-by-rank differential is not statistically significant. The data is consistent with a hidden-information model of human capital in which there are significant career concerns at all ranks. Career concerns reduce the differential for diligent work versus shirking (Table 6), $\Delta_{jkt}^B(h)$, by between 15% and 22%; as a percentage of the gross compensating differential, it is lowest in Rank 1 and highest in Rank 3. The lower percentage in Rank 1 reflects its position at the end of the lifecycle, while the higher percentage in Rank 3 reflects the imminent possibility of promotion to CEO. There are significant career concerns at the CEO rank,

¹⁶ Subtracting the estimates in Table 6 from those in Table 5, we obtain the gross compensating differential for diligent work versus shirking under perfect monitoring. The estimates in Tables 5 are mainly of a higher order of magnitude than those in Table 6. Therefore, the qualitative patterns of the gross compensating differential for diligent versus shirking is similar to the net differential.

19% of the gross compensating differential from working versus shirking, the same as in Rank 4 and higher than in Rank 5 (17%). We conclude that in order to maintain the hard charging life demanded of executives in our population, pursuing the goal of value maximization teaches executives to sell themselves more effectively to shareholders. The role of career concerns declines with age, tenure, executive experience and experience in different firms. In summary, the poorer quality of the signal at the higher ranks (top to the CEO) is the main reason by the risk premium is increasing over ranks.

9 Conclusion

Firm size is a major source of variation in executive pay. As in other labor markets, executives in larger firms are paid more. The empirical literature supports the importance of both assignment and sorting (Gabaix and Landier, 2008) and the agency costs (Gayle and Miller, 2009b) in explaining why executive pay increases with firm size. Our equilibrium framework incorporates both sorting and agency considerations. This allows us to estimate separately the part of the compensation due to agency and the certainty equivalent wage determined by equilibrium sorting. In contrast to previous studies, we use a hierarchy (constructed in Gayle, Golan, and Miller, 2012) to account for ranks as a source of variation in pay. We find that most of the variation of pay in firm size is due to the agency problem. A more surprising result, however, is that although the expected pay is higher in large firms, for a given skill set, the certainty-equivalent wage decreases in firm size. We further decompose the certainty-equivalent wage to quantify the different sources of pay variation. We find that the lower certainty-equivalent pay is mainly due to the lower disutility associated with diligent work in larger firms. The expected pay increases with rank, and the certainty equivalent is increasing and concave in rank. However, we find that a risk premium explains most of the variation in pay across ranks.

To explain the variation in the risk premium by firm, rank and executive characteristics, we estimate the costs and benefits of shirking, to the executive and to shareholders. We find that essentially the same reason explains why the risk premium increases with firm size and rank: executive power, or her span of control, measured in our model by the expected gross loss shareholders would incur from a shirking executive, declines significantly with firm size and rank. Consequently firm excess returns, the main signal of executive labor productivity, is more closely related to the performance of operating heads below the level of CEO than to the CEO herself, and is less informative about effort in larger firms than smaller ones (where a given executive is more likely to have a pronounced effect on firm operations). Since weak signals tend to generate large risk premiums in equilibrium, higher ranked executives in larger firms tend to receive higher risk premiums.

Our finding that executives closer to operations have a greater span of control than their more highly ranked superiors also speaks to the firm's organization. Our empirical results conform more closely to a theory of internal organization that resembles multilateral contractual obligations between self-interested parties (Alchian and Demsetz, 1972; Mirrlees, 1976), rather than the hypothesis that the firm resembles a chain of command (Williamson, 1967; Calvo and Wellisz, 1980). In the equilibrium of our model higher expected executive pay is matched to higher value of marginal productivity, and empirically CEOs are paid the most: perhaps they are paid to coordinate, not boss. Other features of our estimates support this contractual interpretation: compensation falls with tenure and nonpecuniary costs rise with tenure. The increase in the risk premium with rank is not driven by the increase in the executives' net benefit from shirking: although the loss from not providing incentives increases with rank, the differences are not significant. This finding provides only weak support for conventional wisdom that shareholders risk more from chairmen and CEOs who have greater latitude to shirk than lower-ranked officers.

Finally we decompose the role of implicit incentives in ameliorating the moral-hazard problem.

While the costs and benefits of shirking are separately identified, separating the disutility of shirking from the continuation value of shirking, both of which only occur off the equilibrium path, requires either a functional form assumption on the evolution of human capital when executives shirk or an exclusion restriction, such as age invariant preferences. Using functional-form assumptions, our empirical results show that the explicit incentives increase with age because career concerns decline as executives approach retirement. But in another twist to textbook labor economics, that higher ranked workers invest less in human capital, we find that both the CEO and executives just one rank below her have the lowest hazard rates into retirement, which leads them to forego higher pay: in other words they acquire more human capital, both public and private, than the subordinates further down in the hierarchy.

Appendix: Proofs of Lemmas and Theorems

Proof of Lemma 4.1. We proceed by induction, first showing that the expression for the value function is true for age T , and then for all $t \in \{1, \dots, T-1\}$. From Proposition 1 of Margiotta and Miller (2000, 678), the value function solving the consumption-savings problem at retirement date $T+1$ is

$$V_{T+1}(h, \xi_{T+1}, a_{t+1}) \equiv -b_{\tau(T+1)} \exp\left[-(a_{T+1} + \rho\xi_{T+1})/b_{\tau(T+1)}\right].$$

Suppose a manager works in firm and rank coordinate pair (j, k) at age T for one period and then retires. After selecting job match (j, k) , she chooses consumption and next period's endowment (c_T, ξ_{T+1}) optimally to maximize

$$-\alpha_{jkT}(h) \exp(-\varepsilon_{jkT}^*) \exp(-\rho c_T) - E_T \left[v_{jkT+1} A_{T+1}(\bar{H}_{jk}(h)) b_{\tau(T+1)} \exp\left(-\frac{a_{t+1} + \rho\xi_{T+1}}{b_{\tau(T+1)}}\right) \right], \quad (66)$$

subject to

$$E_T [\lambda_{\tau(T)+1} \xi_{T+1} | l_T, d_{jkT}, h] + \lambda_{\tau(T)} c_T \leq \lambda_{\tau(T)} \xi_T + E_T [\lambda_{\tau(T)+1} w_{jkT+1} | l_T, d_{jkT}, h]. \quad (67)$$

Then, Equation (15) of Margiotta and Miller (2000, 680) gives the value function for this problem as

$$V_{jkT}(h, \xi_T, a_t) \equiv -b_{\tau(T)} \alpha_{jkT}(h)^{1/b_{\tau(T)}} \exp(-\varepsilon_{jkT}^*/b_{\tau(T)}) E_T [v_{jkT+1}]^{1-\frac{1}{b_{\tau(T)}}} \exp\left(-\frac{a_T + \rho\xi_T}{b_{\tau(T)}}\right). \quad (68)$$

Integrating over ε_T and averaging over job matches (j, k) yields

$$V_T(h, \xi_T, a_T) \equiv -b_{\tau(T)} \left[p_{0T}(h) V_{0T}(h) + \sum_{j=1}^J \sum_{k=1}^K p_{jkT}(h) V_{jkT}(h) \right] = -b_{\tau(T)} \exp\left(-\frac{a_T + \rho\xi_T}{b_{\tau(T)}}\right) A_T(h).$$

The proof is completed with an induction for all ages $t \in \{1, \dots, T-1\}$,

$$V_{jkt}(h, \xi_t, a_t, \varepsilon_{jkt}^*) \equiv -\alpha_{jkt}(h)^{\frac{1}{b_{\tau(t)}}} \exp\left(-\frac{\varepsilon_{jkt}^*}{b_{\tau(t)}}\right) E_t [v_{jkt+1} A_{t+1}(h)]^{1-\frac{1}{b_{\tau(t)}}} b_{\tau(t)} \exp\left(-\frac{a_t + \rho\xi_t}{b_{\tau(t)}}\right). \quad (69)$$

Suppose both equations are true for all ages $s \in \{t+1, \dots, T\}$. Given job selection (j, k) , Equation (69) follows directly from the solution to the consumption-savings decision at age t by substituting t for T and $v_{jkt+1} A_{t+1}[h]$ for v_{jkT+1} in Equation (68) above. Integrating over ε_t and averaging over the

JK job matches yields

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

which follows from the recursive definition of $A_t(h)$. Substituting the expression for $V_t(h, \xi_t, a_t)$ back into the expression for $V_{jkt}(h, \xi_t, a_t, \varepsilon_{jkt}^*)$ completes the induction. ■

Proof of Theorem 4.1. The manager optimizes her expected lifetime utility at age t by choosing the highest valued conditional-valuation function, given by Equation (69), of the JK job matches and retirement. The solution can be found by taking logarithms and maximizing with respect to potential job matches and retirement. Note that $[\ln b_{\tau(t)} - (a_t + \rho e_t) / b_{\tau(t)}]$ is then an additive constant in all alternatives so it drops out of the solution. Multiplying by $b_{\tau(t)}$ then completes the proof. ■

Proof of Theorem 4.2.

We first prove the theorem for the case of private human capital. Throughout this proof, we fix (j, k, t, h) and consolidate the notation by defining

$$\gamma_1 \equiv \exp\{q_{jk}[p_t(h)]\}^{1/(1-b_{\tau(t)})} \alpha_{jkt}(h)^{1/(b_{\tau(t)}-1)} A_{t+1}[\bar{H}_{jk}(h)], \quad (70)$$

$$\gamma_2 \equiv \alpha_{jkt}(h)^{1/(1-b_{\tau(t)})} A_{t+1}[\bar{H}_{jk}(h)] \quad (71)$$

and

$$\gamma_3 \equiv \beta_{jkt}(h)^{1/(b_{\tau(t)}-1)} B_{t+1}[\underline{H}_{jk}(h), \bar{H}_{jk}(h)], \quad (72)$$

where, for convenience, we have suppressed the dependence of $(\gamma_1, \gamma_2, \gamma_3)$ on (j, k, t, h) to reduce the notational clutter. Thus, the participation and incentive-compatibility constraints can be expressed in terms of the new notation as

$$\gamma_1 E_t[v_{jkt+1}] = 1$$

and

$$\gamma_2 E_t[v_{jkt+1}] \leq \gamma_3 E_t[v_{jkt+1} g_{jkt}(\pi, h)].$$

Since the expectation operator preserves linearity, both the participation constraint (14) and the incentive-compatibility constraint (25) are rendered linear in v_{jkt+1} , after multiplying both sides of the latter by $A_{t+1}[\bar{H}_{jk}(h)] E_t[v_{jkt(\tau)+1}]$. The objective function, the expected wage bill $E_t(w_{jkt(\tau)+1})$, can be expressed as a concave function of $v_{jkt(\tau)+1}$, namely $E_t(\ln v_{jkt(\tau)+1})$. Therefore, the Kuhn Tucker Theorem applies, and the Lagrangian for the problem in which the j^{th} firm elicits diligent work from the k^{th} rank can be written as

$$E_t[\ln(v_{jkt(\tau)+1})] + \eta_0 E_t[1 - v_{jkt(\tau)+1} \gamma_1] + \eta_1 E_t[v_{jkt(\tau)+1} g_{jk}(\pi, h) \gamma_3 - v_{jkt(\tau)+1} \gamma_2],$$

where, for convenience, we have also suppressed the dependence of η_0 and η_1 on (j, k, h) . The proof now follows directly from Proposition 3 of Margiotta and Miller (2000, 713–714). ■

Proof of Theorem 4.3.

The result in this Theorem is a special case of the results in Theorem 5.3. See the proof of Theorem 5.3 below. ■

Proof of Theorem 5.1. The proof of this Theorem follows from Lemma 4.1 and Theorem 4.1 by extending the choice set to effort levels as well, and substituting $B_t(h, h')$ for $A_t(h)$ in their proofs. ■

Proof of Theorem 5.2. Follows directly from the proof of Theorem 4.2. ■

Proof of Theorem 5.3.

Appealing to the optimization problems in Theorems 4.1 and 5.1, define, for each (h, π) , the probability vector $(p_{0t}^e(h), \dots, p_{Jt}^e(h))$ and the human-capital functions $A_t^e(h)$ and $B_t^e(h, h')$ by successively substituting the compensation function,

$$w_{jkt(\tau)+1}^e(\pi, h) \equiv F_{jkt(\tau)}(h) + \frac{b_{t(\tau)+1}}{\rho} \left\{ r_{jkt(\tau)+1}^e(\pi, h) - E \left[r_{jkt(\tau)+1}^e(\pi, h) \right] \right\}, \quad (73)$$

for $w_{jkt(\tau)+1}^e(\pi, h)$ into the respective recursions, where $r_{jkt+1}^e(\pi, h)$ is defined using Equations (39) and (40). By construction, $w_{jkt(\tau)+1}^e(\pi, h)$ does not depend on future returns to the firm and therefore satisfies the no-commitment property. By inspection, $E_t[w_{jkt(\tau)+1}^e(\pi, h)|h] = F_{jkt(\tau)}(h)$. To establish that the strategies and beliefs together constitute a sequential equilibrium, as defined in Kreps and Wilson (1982), we now prove the strategies below are sequentially rational for the beliefs ascribed to firms, and that those beliefs are consistent. Recall that a manger at time t who shirks k periods loses human capital according to Equations (29) and (31) relative to a manager with the same history who works diligently in all periods. An executive at age t who has worked diligently in all period has human capital of h_t^* , and an executive with the same job history and characteristics, but who has shirked for $k \geq 1$ periods, has human capital of $h_t^{(k)}$: the human capital of a manger with t years of experience who shirked k times. Assumptions (i) and (ii) in Section 5.2 imply that, for t and for any k ,

$$B(h' = \underline{F}, h_t^{(k)}) \geq B(h' = h_t^*, h_t^{(k)}). \quad (74)$$

In other words, having the option of choosing contracts for managers with the same histories who never shirked weakly dominates the option of revealing that the manager has shirked initially. While there may be other equilibria with the same implication to the data, this assumption makes the analysis off the equilibrium path tractable and simple. In particular, for tractability purposes, it rules out the possibility of contracts that partially reveal executive types off the equilibrium path in some histories as the off-equilibrium-path analysis.

Firms' Strategies and Beliefs

1. Observing an executive with a history of contracts of the form described in Equation (73) who makes an offer $w_{jkt(\tau)+1}^e(\pi, h^*)$, shareholders believe the manger never shirked $h' = h_t^*$.
2. Observing a history with contract offers different from Equation 73, the beliefs are that the manager shirked initially and is tainted. Given ex-ante beliefs (h') , observing a current offer with a contract different from Equation (73), firms update their beliefs to assign an upper bound to her human capital of \underline{F} .
3. At the offer stage, if the manager's history does not include past and current deviation from the prescribed contract $w_{jkt(\tau)+1}^e(\pi, h^*)$, the firm accepts; any other contract offer is rejected.¹⁷

Managers' Strategies:

1. All managers offer $w_{jkt(\tau)+1}^e(\pi, h^*)$ regardless of their history.
2. They choose jobs, offers and effort level solving the problem described in Theorem 5.1.

¹⁷ It is possible to construct other sequential equilibria in which beliefs are not automatically that the manger is tainted for certain deviations from the equilibrium offers without changing any of the equilibrium implications affecting our empirical analysis. We chose the simplest equilibrium.

Sequential Rationality—Managers:

1. From the recursive definition of $w_{jkt(\tau)+1}^e(\pi, h)$, $A_t^e(h)$ and $B_t^e(h, h')$, it follows from Theorem 5.2 that the manager's job-match choices are sequentially rational when $h_t = h'$. Any higher offer is rejected, Any lower or non-incentive-compatible offer reduces utility.
2. Given that offers deviating from $w_{jkt(\tau)+1}^e(\pi, h')$ are rejected, there is nothing to be gained from managers departing from the strategies prescribed for them.

Sequential Rationality—Firms:

1. For a manager with history of offers $w_{jkt}^e(\pi, h^*) \forall t$, firms believe $h = h^*$ and hence will break even by accepting the contract.
2. If the manager offers a different contract, given the beliefs that the manager is tainted, rejecting the offer is optimal.

Consistency of Beliefs:

To demonstrate these beliefs are consistent, consider all histories where no offer has been rejected.

- (i) We suppose with probability $1/i$ a firm accepts a contract not of the form $w_{jkt}^e(\pi, h')$ and with probability $1/i$ a firm rejects a contract of the form $w_{jkt}^e(\pi, h')$.
- (ii) With probability $1/i$, a manager who has not shirked before deviates from diligent work.
- (iii) With probability $1/i$, a manager who has shirked before deviates from her prescribed strategy of diligent work.
- (iv) Managers deviate from their optimal job-match choice to one of the other choices with probability $1/i$, giving each of the other choices equal weight.
- (v) At any period $t > 1$, managers who are tainted (shirked in Period 1) demand contracts of the form $w_{jkt}(\pi, h') \neq w_{jkt}^e(\pi, h')$ with probability $1/i$. Note that in period $t = 1$ there are no histories, thus there is no private information.
- (vi) An untainted manager (worked in period 1) who deviates and shirks $k < t$ periods, deviates with probability $(1/i)^3$.

The support of the distribution of non- $w_{jkt}^e(\pi, h')$ contracts covers the entire space of such contracts. This perturbation from the conjectured equilibrium strategy is completely mixed, so the Bayes rule applies for computing the probabilities of nodes within any given information set. In particular, at any period $t > 1$, the probability of a firm being confronted with a non- $w_{jkt}^e(\pi, h')$ contract from a manager who is tainted is $(1/i)$. So, when a firm is confronted with a non- $w_{jkt}^e(\pi, h')$ contract, the probability that the manager is not tainted is less or equal to

$$\frac{(1 - 1/i)(1/i)^3}{(1 - 1/i)(1/i)^3 + (1/i)^2} = \frac{(1 - 1/i)}{(1 - 1/i) + i}. \quad (75)$$

In the limit of $i \rightarrow \infty$, this probability converges to zero, thus the firm's beliefs are consistent. ■

Proof of Theorem 6.1. There are two steps to the proof. First, for any finite positive $\hat{\rho}$, and any probability distribution function $\hat{G}(\varepsilon)$ with the same support as $\tilde{G}(\varepsilon)$, we define another parameterization, $\hat{\theta} \in \Theta$. To complete the proof, we show that the model defined by $\hat{\theta}$ generates the same data

as $\tilde{\theta}$, and is therefore observationally equivalent. Given the compensation process generated by $\tilde{\theta}$, and our construction in the first step, the conditional-choice probabilities of $\tilde{\theta}$ replicate those of θ . Thus, the only remaining task is to show that the compensation schedule generated by $\hat{\theta}$ reproduces the schedule generated by $\tilde{\theta}$. This only leaves us to prove that the contracts are the same, a second step that follows directly from the analysis of the pure moral-hazard model in Gayle and Miller (2013). Here, we prove the first step. For any finite positive $\hat{\rho}$, let $\hat{v}_{jkt+1} \equiv \exp[-\hat{\rho}w_{jkt}(\pi)/b]$ and define

$$\hat{g}_{jkt}(\pi, h) = \frac{\exp(\hat{\rho}\bar{w}_{jkt}/b) - \hat{v}_{jkt+1}^{-1}}{\exp(\hat{\rho}\bar{w}_{jkt}/b) - E_t[\hat{v}_{jkt+1}(\pi)^{-1}]} \quad (76)$$

For any probability distribution function $\hat{G}(\varepsilon)$ with the same support as $\tilde{G}(\varepsilon)$, let

$$\hat{E}_t[\exp(\varepsilon_{jkt}/b)] \equiv p_{jkt}(h_t)^{-1} \int d_{jkt} \exp(\varepsilon_{jkt}/b) d\hat{G}(\varepsilon)$$

denote the conditional expectation of ε_{jkt}/b given the choices observed in the population but integrated with respect to $\hat{G}(\varepsilon)$ rather than $\tilde{G}(\varepsilon)$. Appealing to Proposition 1 of Hotz and Miller (1993), there exists a mapping $\hat{q}(p)$ implied by $\hat{G}(\varepsilon)$ for any conditional-valuation function. Starting with $\hat{A}_t(h) = 1$ for all $t \geq R$, and given $\hat{G}(\varepsilon)$, recursively define $\hat{\alpha}_{jkt}(h)$ and $\hat{A}_t(h)$ to rationalize the choice probabilities generated by θ^* by repeatedly appealing to Equation (11) and setting

$$\hat{\alpha}_{jkt}(h_t) = \exp[\hat{q}_{jk}(p_t(h_t))] \hat{A}_{t+1} [\bar{H}_{jk}(h)]^{1-b} E_t[\hat{v}_{jkt+1}(\pi)]^{1-b}. \quad (77)$$

Finally, $\hat{\beta}_{jkt}^*(h)$ is defined as

$$\hat{\beta}_{jkt}^*(h) = \exp[\hat{q}_{jk}(p_t(h))] \hat{A}_{t+1} [\bar{H}_{jk}(h)]^{1-b} E_t[\hat{v}_{jkt+1}(\pi) \hat{g}_{jkt}(\pi, h)]^{1-b}. \quad (78)$$

In this manner, we construct another element in the parameter space, $\hat{\theta} \in \Theta$ defined by

$$\hat{\theta} \equiv \left(\hat{\alpha}_{jkt}(h), \hat{\beta}_{jkt}^*(h), \hat{\rho}, \tilde{f}(\pi), \hat{g}_{jkt}(\pi, h), \hat{G}(\varepsilon) \right).$$

The second step now follows from applying Theorem 2.1 of Gayle and Miller (2013). ■

References

- [1] Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *The American Economic Review*, 62(December), 777–795.
- [2] Altuğ, S., & Miller, R. A. (1998). The effect of work experience on female wages and labour supply. *The Review of Economic Studies*, 65(1), 45–85.
- [3] Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3), 583–618.
- [4] Antle, R., & Smith, A. (1985). Measuring executive compensation: Methods and an application. *Journal of Accounting Research*, 23(1), 296–325.
- [5] Antle, R., & Smith, A. (1986). An empirical investigation of the relative performance evaluation of corporate executives. *Journal of Accounting Research*, 1–39.

- [6] Antonovics, K., & Golan, L. (2012). Experimentation and job choice. *Journal of Labor Economics*, 30(2), 333–366.
- [7] Aron, D. J. (1988). Ability, moral hazard, firm size, and diversification. *Rand Journal of Economics*, 19(1), 72–87.
- [8] Baker, G. P., & Hall, B. J. (2004). CEO incentives and firm size. *Journal of Labor Economics*, 22(4), 767–798.
- [9] Bayer, P., Khan, S., & Timmins, C. (2011). Nonparametric identification and estimation in a Roy model with common nonpecuniary returns. *Journal of Business & Economic Statistics*, 29(2), 201–215.
- [10] Becker, G. S. (1964). *Human Capital*. New York: National Bureau for Economic Research.
- [11] Bellman, R. (1957). *Dynamic Programming*. Princeton, NJ: Princeton University Press.
- [12] Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *The Journal of Political Economy*, 75, 352–365.
- [13] Brown, C., & Medoff, J. (1989). The employer size–wage effect. *Journal of Political Economy*, 97(5), 1027–1059.
- [14] Cahuc, P., Postel-Vinay, F., & Robin, J. M. (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica*, 74(2), 323–364.
- [15] Calvo, G. A., & Wellisz, S. (1980). Technology, entrepreneurs, and firm size. *The Quarterly Journal of Economics*, 95(4), 663–677.
- [16] Chevalier, J., & Ellison, G. (1999). Career concerns of mutual fund managers. *The Quarterly Journal of Economics*, 114(2), 389–432.
- [17] Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium* Monograph 17. New Haven, CT: Cowles Foundation for Research in Economics, Yale University Press.
- [18] Dewatripont, M., Jewitt, I., & Tirole, J. (1999). The economics of career concerns, Part II: Application to missions and accountability of government agencies. *The Review of Economic Studies*, 66(1), 199–217.
- [19] D’Haultfoeulle, X., & Maurel, A. (2013). Inference on an extended Roy model, with an application to schooling decisions in France. *Journal of Econometrics*, 174(2), 95–106.
- [20] Doeringer, P. B., & Piore, M. J. (1971). *Internal Labor Markets and Manpower Analysis*. Armonk, NY: ME Sharpe.
- [21] Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review*, 100(3), 1238–1260.
- [22] Felli, L., & Harris, C. (1996). Learning, wage dynamics, and firm-specific human capital. *Journal of Political Economy*, 104(4), 838.
- [23] Fox, J. T. (2009). Firm-size wage gaps, job responsibility, and hierarchical matching. *Journal of Labor Economics*, 27(1), 83–126.

- [24] Fudenberg, D., Holmström, B., & Milgrom, P. (1990). Short-term contracts and long-term agency relationships. *Journal of Economic Theory*, 50, 1–31.
- [25] Gabaix, X., & Landier, A. (2008). Why has CEO pay increased so much? *The Quarterly Journal of Economics*, 123(1), 49–100.
- [26] Garrett, D. F., & Pavan, A. (2012). Managerial turnover in a changing world. *Journal of Political Economy*, 120(5), 879–925.
- [27] Gayle, G.-L., & Golan, L. (2012). Estimating a dynamic adverse-selection model: Labour-force experience and the changing gender earnings gap 1968–1997. *The Review of Economic Studies*, 79(1), 227–267.
- [28] Gayle, G. L., Golan, L., & Miller, R. A. (2012). Gender differences in executive compensation and job mobility. *Journal of Labor Economics*, 30(4), 829–872.
- [29] Gayle, G.-L., & Miller, R. A. (2009a). Insider information and performance pay. *CESifo Economic Studies*, 55(3–4), 515–541.
- [30] Gayle, G.-L., & Miller, R. A. (2009b). Has moral hazard become a more important factor in managerial compensation? *American Economic Review*, 99(5), 1740–1769.
- [31] Gayle, G.-L., & Miller, R. A. (2013). Identifying and testing models of managerial compensation. Working Paper, Washington University in St. Louis.
- [32] Gibbons, R., & Murphy, K. J. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *The Journal of Political Economy*, 100(3), 468–505.
- [33] Grossman, S., & Hart, O. (1983). An analysis of the principal–agent problem. *Econometrica*, 51, 7–46.
- [34] Hall, B. J., & Liebman, J. B. (1998). Are CEOs really paid like bureaucrats? *Quarterly Journal of Economics*, 113, 653–680.
- [35] Harris, M., & Holmström, B. (1982). A theory of wage dynamics. *The Review of Economic Studies*, 49(3), 315–333.
- [36] Harrison, G. W., Johnson, E., McInnes, M. M., & Rutström, E. E. (2005). Risk aversion and incentive effects: Comment. *American Economic Review*, 95(3), 897–901.
- [37] Harrison, G. W., List, J. A., & Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, 75(2), 433–458.
- [38] Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1), 169–182
- [39] Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644–1655.
- [40] Holt, C. A., & Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, 95(3), 902–904.
- [41] Hotz, V. J., & Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3), 497–529.

- [42] Jovanovic, B. (1979). Firm-specific capital and turnover. *The Journal of Political Economy*, 87, 1246–1260.
- [43] Katz, L. F., & Summers, L. H. (1989). Industry rents: Evidence and implications. *Brookings Papers on Economic Activity: Microeconomics*, 1989, 209–290.
- [44] Kreps, D. M., & Wilson, R. (1982). Reputation and imperfect information. *Journal of Economic Theory*, 27(2), 253–279.
- [45] Lee, D., & Wolpin, K. I. (2006). Intersectoral labor mobility and the growth of the service sector. *Econometrica*, 74(1), 1–46.
- [46] Li, A. (2014). Test contract. Working paper. Washington University in St. Louis.
- [47] Lucas, R. E., Jr. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 9, 508–523.
- [48] Magnac, T., & Thesmar, D. (2002). Identifying dynamic discrete decision processes. *Econometrica*, 70(2), 801–816.
- [49] Malcomson, J. M., & Spinnewyn, F. (1988). The multiperiod principal–agent problem. *Review of Economic Studies*, 55, 391–407.
- [50] Margiotta, M., & Miller, R. A. (2000). Managerial compensation and the cost of moral hazard. *International Economic Review*, 41, 669–719.
- [51] McNulty, P. J. (1984). *The Origins and Development of Labor Economics*. Cambridge, MA: MIT Press.
- [52] Miller, R. A. (1984). Job matching and occupational choice. *The Journal of Political Economy*, 92(6), 1086–1120.
- [53] Mincer, J. A. (1974). Schooling, experience, and earnings. NBER Books.
- [54] Mirrlees, J. A. (1975). The theory of moral hazard and unobservable behavior, Part 1. Working Paper. Oxford University.
- [55] Mirrlees, J. A. (1976). The optimal structure of incentives and authority within an organization. *The Bell Journal of Economics*, 7(1), 105–131.
- [56] Myerson, R. (1982). Optimal coordination mechanisms in generalized principal–agent problem. *Journal of Mathematical Economics*, 10, 67–81.
- [57] Oi, W. Y., & Idson, T. L. (1999). Firm size and wages. *Handbook of Labor Economics*, 3, 2165–2214.
- [58] Perrigne, I., & Vuong, Q. (2011). Nonparametric identification of a contract model with adverse selection and moral hazard. *Econometrica*, 79(5), 1499–1539.
- [59] Postel-Vinay, F., & Robin, J. M. (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, 70(6), 2295–2350.
- [60] Raff, D. M., & Summers, L. H. (1987). Did Henry Ford pay efficiency wages? *Journal of Labor Economics*, 5(4), S57–S86.

- [61] Rey, P., & Salanie, B. (1990). Long-term, short-term and renegotiation: On the value of commitment in contracting. *Econometrica*, 58(3), 597–619.
- [62] Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, New Series, 3(2), 135–146.
- [63] Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55.
- [64] Rosen, S. (1983). The economics of superstars: Reply. *The American Economic Review*, 73, 460–462.
- [65] Rubinstein, M. (1981). A Discrete-Time Synthesis of Financial Theory. *Research in Finance*, 3, 53-102.
- [66] Sanders, C. (2013). Skill uncertainty, skill accumulation and occupational choice. Working Paper. Washington University.
- [67] Sannikov, Y. (2014). Moral hazard and long-run incentives. Working Paper. Princeton University.
- [68] Stole, L. A., & Zwiebel, J. (1996a). Intra-firm bargaining under non-binding contracts. *The Review of Economic Studies*, 63(3), 375–410.
- [69] Stole, L. A., & Zwiebel, J. (1996b). Organizational design and technology choice under intrafirm bargaining. *The American Economic Review*, 86(1), 195–222.
- [70] Taber, C., & Vejlin, R. (2010). Estimation of a Roy/search/compensating differential model of the labor market. Working paper, University of Wisconsin–Madison.
- [71] Terviö, M. (2008). The difference that CEOs make: An assignment model approach. *The American Economic Review*, 98(3), 642–668.
- [72] Thomas, J., & Worrall, T. (1988). Self-enforcing wage contracts. *The Review of Economic Studies*, 55(4), 541–554.
- [73] Williamson, O. E. (1967). Hierarchical control and optimum firm size. *The Journal of Political Economy*, 75, 123–138.
- [74] Winter-Ebmer, R., & Zweimüller, J. (1999). Firm-size wage differentials in Switzerland: Evidence from job-changers. *The American Economic Review*, 89, 89–93.

TABLE 1: EXECUTIVE CHARACTERISTICS BY RANK

Variable	Rank				
	1	2	3	4	5
Exit	0.245	0.090	0.116	0.149	0.154
Turnover	0.027	0.032	0.028	0.021	0.016
Age	57.798 (8.220)	55.000 (7.433)	51.768 (7.363)	51.184 (8.140)	50.817 (8.804)
Female	0.019	0.015	0.025	0.053	0.06
Tenure	15.784 (11.708)	14.412 (10.672)	13.388 (10.062)	13.383 (9.754)	13.221 (9.541)
Exec. Exp.	20.331 (11.113)	18.643 (9.638)	15.738 (9.555)	15.664 (9.901)	15.871 (10.124)
NBE	0.706 (1.186)	0.689 (1.118)	0.702 (1.174)	0.959 (1.341)	1.159 (1.427)
NAE	0.887 (1.357)	0.909 (1.374)	0.764 (1.310)	0.799 (1.323)	0.841 (1.350)
Execdir	0.720	0.929	0.675	0.177	0.069
Interlocked	0.071	0.068	0.026	0.009	0.003
No College	0.232	0.212	0.236	0.178	0.144
Bachelors	0.768	0.788	0.764	0.822	0.856
MBA	0.246	0.255	0.232	0.229	0.196
MS/MA	0.155	0.172	0.168	0.212	0.214
PhD	0.151	0.150	0.135	0.183	0.257
Salary	615 (366)	719 (412)	559 (318)	397 (197)	304 (176)
Compensation	2,945 (26,035)	4,794 (26,701)	3,717 (19,009)	1,844 (11,644)	1,269 (9,438)
Observations	4,358	20,983	5,620	28,271	15,972

Sources: The data are for top managers from Standard & Poor's ExecuComp database for 1991 through 2006 matched with background data from the Marquis *Who's Who* database. *Note:* Standard deviation in parentheses; Compensation and Salary are measured in thousands of 2006 US\$; Tenure and Executive Experience (Exec. Exp.) are measured in years; NBE (NAE) is the number of times the executive changed firms before (after) becoming one of the ranks in our sample. Execdir is an indicator of whether the executive is a member of the board of directors.

TABLE 2: COMPENSATION AND MOBILITY

Compensation			Promotion					Turnover	Retirement	
π	π^2	Level	R1	R2	R3	R4	R5			
Panel A: Executive Position										
Constant	21,601 (3,859)	-9,114 (1,914)	-4,359 (2,716)							
R4	1,529 (926)	-242 (444)	103 (463)	-20.3 (20.5)	-76.2 (9.8)	-67.8 (10.4)	63.4 (4.01)	-536 (6.51)	—	
R3	2,627 (1,407)	-164 (605)	1,267 (662)	-88.1 (24.1)	-72.4 (12.8)	114 (4.43)	-404 (14.6)	-754 (19.7)	94.7 (17.6)	
R2	6,007 (1,394)	-789 (699)	3,456 (683)	-118 (25.1)	67.7 (3.90)	-393 (20.0)	-551 (18.2)	-901 (22.4)	213 (12.6)	
R1	9,839 (1,690)	-454 (987)	1,055 (797)	111 (10.1)	-290 (34.2)	-345 (33.6)	-585 (33.8)	-939 (39.0)	86.1 (23.2)	
Execdir	7,695 (570)	-848 (304)	845 (251)	-22.8 (13.2)	123 (4.03)	15.6 (5.32)	-70.4 (3.39)	-105 (6.50)	-102 (9.19)	
									—	
									—	
									-8.9 (4.28)	
									55.26 (3.23)	
									-64.72 (3.50)	
Panel B: Firm Type										
Service	3,149 (419)	88 (222)	777 (198)	—	—	—	—	—	—	
Primary	-3,609 (473)	1,537 (267)	-633 (198)	—	—	—	—	—	-18.68 (7.11)	
Medium Firm	4,079 (437)	-253 (201)	937 (214)	—	—	—	—	—	—	
Large Firm	12,703 (405)	-2,224 (212)	3,697 (190)	—	—	—	—	—	—	
Large Board	2,683 (358)	-1,203 (176)	280 (163)	—	—	—	—	—	-25.7 (5.71)	
									—	
Panel C: Human Capital and Individual Heterogeneity										
Rank 1 Lagged	12,085 (1,769)	-3,054 (987)	544 (822)						—	
Rank 2 Lagged	14,640 (1,342)	-2,875 (625)	660 (658)						-132 (18.6)	
Rank 3 Lagged	4,849 (1,389)	-1,100 (586)	597 (653)						-42.75 (19.1)	
Exec. Exp.	191 (26)	-42 (14)	1.61 (25)	-9.66 (14.5)	-32.8 (8.93)	1.01 (11.79)	14.04 (7.00)	34.76 (11.33)	82.8 (13.4)	24.44 (5.69)
Exec. Exp. sq.	—	—	—	6.09 (6.41)	8.22 (4.17)	-3.08 (5.77)	-3.79 (3.49)	-10.1 (5.49)	-20.6 (6.62)	-11.38 (2.65)
Tenure	-23 (25)	22 (14)	-40 (20)	-10.43 (9.48)	-23.03 (6.16)	-10.56 (8.15)	10.7 (4.61)	25.8 (7.99)	-302 (9.02)	24.66 (3.97)
Tenure sq.	—	—	—	4.80 (4.47)	7.28 (2.94)	4.19 (4.11)	-2.79 (2.30)	-11.85 (4.07)	88.1 (4.3)	-7.26 (1.97)
External	-12,396 (996)	2,155 (478)	-1,026 (1,255)	—	—	—	—	—	—	—
R2×External	—	—	3,840 (1,459)	—	—	—	—	—	—	—
R3×External	—	—	5,289 (1,975)	—	—	—	—	—	—	—
NBE	—	—	—	-8.99 (2.19)	-0.91 (1.57)	-6.86 (1.82)	3.11 (0.82)	8.11 (1.26)	-11.1 (2.14)	5.54 (0.62)
NAE	-484 (174)	-58 (93)	215 (80)	-1.23 (2.28)	1.35 (1.49)	-2.43 (2.00)	-0.34 (1.15)	-0.25 (1.82)	-13.7 (1.94)	4.49 (0.78)
Age	17 (23)	15 (10)	281 (85)	-9.011 (188)	0.024 (124)	174 (158)	-459 (86.1)	-847 (128)	1,948 (239)	-527.74 (58.58)
Age sq.	—	—	-3.05 (0.80)	136 (88.7)	-5.20 (60.3)	-111 (80.3)	236 (44.4)	434 (65.8)	-992 (122)	312.89 (28.81)
Female	—	—	—	—	—	—	—	—	—	17.42 (4.16)
R2×Female	—	—	2,668 (1,295)	—	—	—	—	—	-0.51 (0.24)	—
MBA	—	—	—	—	—	—	—	—	—	-84.46 (20.1)
Interlocked	6,403 (995)	-1,496 (471)	-299 (464)	—	—	—	—	—	-93.0 (28.6)	-93.0 (28.6)

Note: Standard error in parenthesis; Tenure and Experience (Exec. Exp.) measured in years; NBE (NAE) = number of firms worked in before (after) becoming a top executive. The elasticities are calculated using logit regressions.

TABLE 3: RISK PREMIUM FROM AGENCY

Variable	Constant	Age-50	Tenure	Exec. Exp.	NBE	NAE	Female	No College	MBA	MS	PhD
Constant	0.499 (0.736)	-0.046 (0.005)	-0.019 (0.004)	-0.012 (0.002)	0.032 (0.011)	0.190 (0.005)	-0.268 (0.195)	-0.178 (0.026)	0.035 (0.017)	-0.059 (0.029)	0.128 (0.017)
Rank 1	0.569 (0.125)					0.000 (0.000)	-0.660 (0.069)	0.001 (0.000)	0.000 (0.000)	0.001 (0.000)	0.032 (0.013)
Rank 2	2.836 (0.125)					-0.001 (0.000)	2.338 (0.069)	0.000 (0.000)	0.001 (0.000)	-0.001 (0.000)	0.033 (0.013)
Rank 3	1.032 (0.125)					-0.002 (0.000)	-1.120 (0.069)	0.003 (0.000)	0.001 (0.000)	0.000 (0.000)	0.032 (0.013)
Rank 4	-0.016 (0.125)					0.000 (0.000)	-0.003 (0.000)	0.002 (0.000)	0.001 (0.000)	0.000 (0.000)	0.032 (0.013)
Industrial Sector											
Primary	-0.037 (0.096)	-0.001 (0.004)	-0.001 (0.003)	0.000 (0.002)	0.012 (0.010)	0.011 (0.004)	0.142 (0.061)	0.025 (0.023)	-0.014 (0.015)	0.058 (0.026)	-0.017 (0.012)
Service	0.379 (0.098)	-0.049 (0.004)	-0.003 (0.003)	0.010 (0.002)	0.035 (0.010)	-0.061 (0.004)	-0.595 (0.062)	0.325 (0.024)	-0.166 (0.015)	0.355 (0.026)	0.096 (0.012)
Firm Size											
Medium	1.032 (0.098)	0.016 (0.004)	0.003 (0.003)	0.004 (0.002)	-0.033 (0.010)	0.007 (0.004)	0.513 (0.062)	-0.042 (0.024)	0.094 (0.015)	-0.118 (0.026)	-0.014 (0.012)
Large	3.350 (0.097)	0.030 (0.004)	0.004 (0.003)	0.001 (0.002)	-0.064 (0.010)	0.002 (0.004)	0.495 (0.061)	-0.312 (0.024)	0.126 (0.015)	-0.291 (0.026)	0.010 (0.012)
Turnover											
New Employer	0.362 (0.080)	0.008 (0.003)	-0.003 (0.003)	-0.003 (0.001)	0.012 (0.008)	0.025 (0.004)	0.258 (0.051)	0.053 (0.020)	-0.014 (0.013)	0.053 (0.022)	-0.046 (0.010)

Note: Compensation is measured in millions of 2006 US\$; Standard error in parentheses; Tenure and Executive Experience (Exec. Exp.) are measured in years; NBE (NAE) is the number of times the executive changed firms before (after) becoming one of the ranks in our sample.

TABLE 4: GROSS LOSS TO SHAREHOLDERS FROM NOT PROVIDING EXECUTIVE INCENTIVES

	$E(x(1 - g(x)))$	New Employer	Female	Individual	Characteristics
Constant	33.5963 (0.0367)	6.8678 (0.0036)	1.7380 (0.0263)	Exec. Exp.	-0.1339 (0.0006)
Rank 1	-8.0575 (0.0056)	1.0166 (0.0395)	-1.5638 (0.0358)	Exec. Exp. sq.	0.0001 (0.0001)
Rank 2	-4.2791 (0.0057)	2.8547 (0.0412)	-1.7018 (0.0359)	Tenure	0.0012 (0.0005)
Rank 3	-1.9994 (0.0057)	3.3221 (0.0440)	-1.5730 (0.0361)	Tenure sq.	-0.0001 (0.0001)
Rank 4	-0.9403 (0.0058)	2.8096 (0.0455)	-1.3255 (0.0362)	No College	-0.2616 (0.0050)
Rank 1 Lagged	-6.6667 (0.0096)			MBA	0.0026 (0.0045)
Rank 2 Lagged	-8.1900 (0.0067)			MS	-0.4054 (0.0047)
Rank 3 Lagged	-3.5289 (0.0080)			PhD	0.7338 (0.0049)
Rank 4 Lagged	-0.4527 (0.0049)			NAE	0.4477 (0.0018)
Industrial Sector				NBE	0.5651 (0.0015)
Primary	-3.7273 (0.0042)			Age-50	-0.0411 (0.0005)
Service	9.3501 (0.0043)			Age-50 sq.	0.0005 (0.0001)
Firm Size					
Medium	-12.9481 (0.0044)		0.0093 (0.0244)		
Large	-25.4104 (0.0044)		0.0139 (0.0221)		
Bond price	0.9026 (0.0021)				

Note: Gross loss to shareholders measured as a percentage of equity value; Standard error in parentheses; Tenure and Executive Experience (Exec. Exp.) are measured in years; NBE (NAE) is the number of times the executive changed firms before (after) becoming one of the ranks in our sample.

TABLE 5: THE NET COMPENSATING DIFFERENTIALS TO EXECUTIVES FROM WORKING VERSUS SHIRKING

Variable	Constant	Age-50	Age-50 sq.	Tenure	Exec. Exp.	NBE	NAE	Female	No College	MBA	MS	PhD
Constant	9.952 (0.888)	0.053 (0.019)	-0.001 (0.001)	0.110 (0.027)	0.015 (0.000)	-0.067 (0.066)	0.141 (0.031)	1.437 (0.530)	-0.518 (0.097)	0.250 (0.089)	-0.469 (0.101)	0.069 (0.079)
Rank 1	1.029 (0.798)					-0.004 (0.002)	-0.004 (0.002)	-0.378 (0.480)	-0.014 (0.005)	0.004 (0.004)	0.002 (0.004)	0.061 (0.060)
Rank 2	0.759 (0.798)					0.000 (0.002)	0.000 (0.002)	-1.082 (0.481)	-0.001 (0.005)	-0.004 (0.004)	0.016 (0.005)	0.046 (0.060)
Rank 3	0.307 (0.798)					0.006 (0.002)	0.005 (0.002)	-1.716 (0.481)	-0.027 (0.005)	-0.009 (0.004)	0.010 (0.005)	0.056 (0.060)
Rank 4	0.039 (0.798)					-0.001 (0.002)	-0.003 (0.002)	-0.120 (0.008)	-0.014 (0.005)	-0.004 (0.004)	0.008 (0.005)	0.058 (0.060)
Industrial Sector												
Primary	-2.599 (0.605)	-0.032 (0.016)	0.001 (0.001)	-0.040 (0.023)		-0.005 (0.055)	-0.080 (0.026)	-0.612 (0.419)	0.264 (0.079)	-0.164 (0.074)	0.188 (0.082)	-0.050 (0.054)
Service	3.799 (0.628)	0.060 (0.017)	-0.001 (0.001)	0.080 (0.024)		-0.050 (0.057)	0.074 (0.027)	0.788 (0.427)	-0.434 (0.082)	0.122 (0.076)	-0.562 (0.085)	0.030 (0.055)
Firm Size												
Medium	-3.105 (0.628)	-0.073 (0.017)	0.002 (0.001)	-0.079 (0.024)		0.125 (0.057)	-0.061 (0.027)	-1.041 (0.427)	0.530 (0.082)	-0.211 (0.076)	0.619 (0.085)	0.054 (0.055)
Large	-4.500 (0.621)	-0.096 (0.016)	0.002 (0.001)	-0.111 (0.024)		0.153 (0.056)	-0.105 (0.027)	-1.207 (0.425)	0.653 (0.081)	-0.306 (0.075)	0.645 (0.084)	0.038 (0.055)
Turnover												
New Employer	-4.755 (0.514)	0.051 (0.013)	-0.001 (0.001)	-0.052 (0.019)		-0.187 (0.048)	-0.189 (0.023)	-2.485 (0.355)	0.040 (0.071)	0.076 (0.066)	0.007 (0.073)	-0.259 (0.049)

Note: Compensation is measured in millions of 2006 US\$; Standard error in parentheses; Tenure and Executive Experience (Exec. Exp.) are measured in years; NBE (NAE) is the number of times the executive changed firms before (after) becoming one of the ranks in our sample.

TABLE 6: CAREER CONCERN AMELIORATION OF AGENCY PROBLEM

Variable	Constant	Age-50	Age-50 sq.	Tenure	Exec. Exp.	NBE	NAE	Female	No College	MBA	MS	PhD
Constant	-1.547 (0.003)	0.006 (0.001)	0.001 (0.001)	0.009 (0.001)	0.015 (0.001)	0.059 (0.001)	0.050 (0.001)	0.154 (0.004)	-0.080 (0.003)	-0.106 (0.002)	0.079 (0.002)	-0.044 (0.002)
Rank 1	0.013 (0.006)					-0.004 (0.002)	-0.004 (0.002)	0.061 (0.008)	-0.014 (0.005)	0.004 (0.004)	0.002 (0.004)	0.002 (0.005)
Rank 2	-0.490 (0.006)					0.000 (0.002)	0.000 (0.002)	-0.198 (0.009)	-0.001 (0.005)	-0.004 (0.004)	0.016 (0.005)	-0.012 (0.005)
Rank 3	-0.671 (0.006)					0.006 (0.002)	0.005 (0.002)	0.182 (0.009)	-0.027 (0.005)	-0.009 (0.004)	0.010 (0.005)	-0.002 (0.005)
Rank 4	-0.242 (0.006)					-0.001 (0.002)	-0.003 (0.002)	-0.120 (0.008)	-0.014 (0.005)	-0.004 (0.004)	0.008 (0.005)	0.000 (0.005)
Turnover												
New Employer	-0.101 (0.006)					-0.017 (0.002)	-0.019 (0.002)	-0.150 (0.008)	0.019 (0.005)	0.008 (0.005)	0.018 (0.005)	-0.002 (0.005)

Note: Compensation is measured in millions of 2006 US\$; Standard error in parentheses; Tenure and Executive Experience (Exec. Exp.) are measured in years; NBE (NAE) is the number of times the executive changed firms before (after) becoming one of the ranks in our sample.

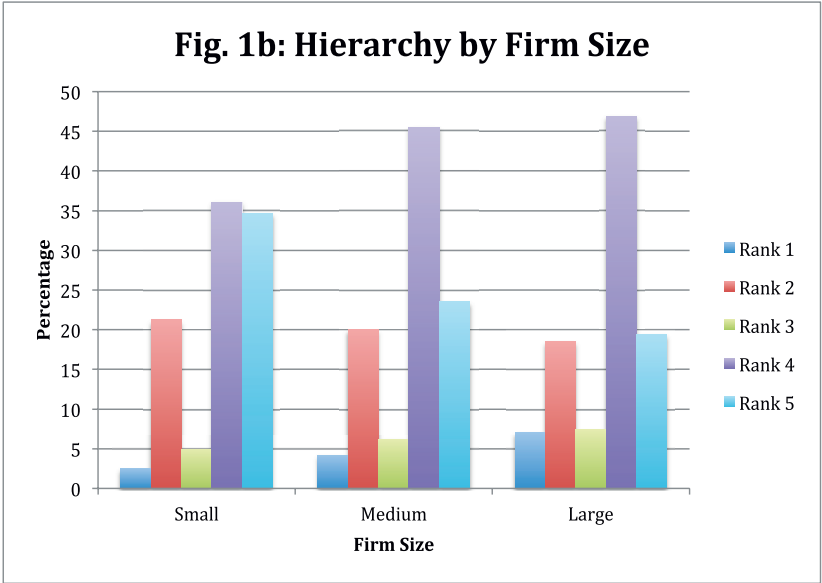
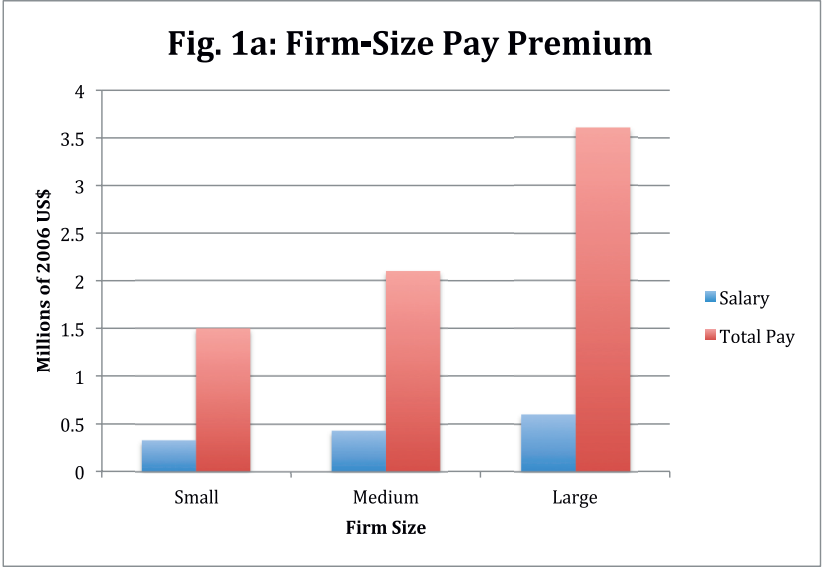


Figure 1: Pay and Hierarchy by Firm Size

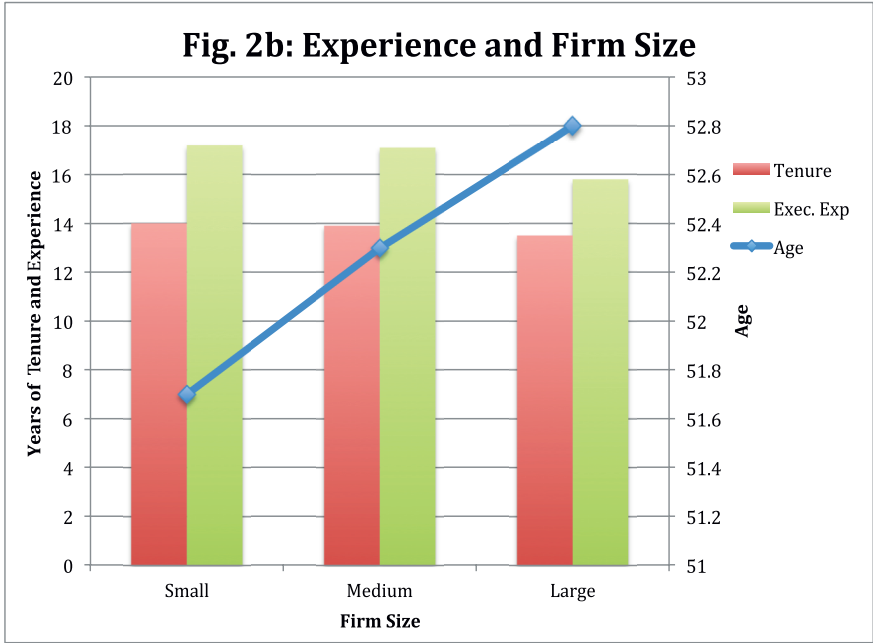
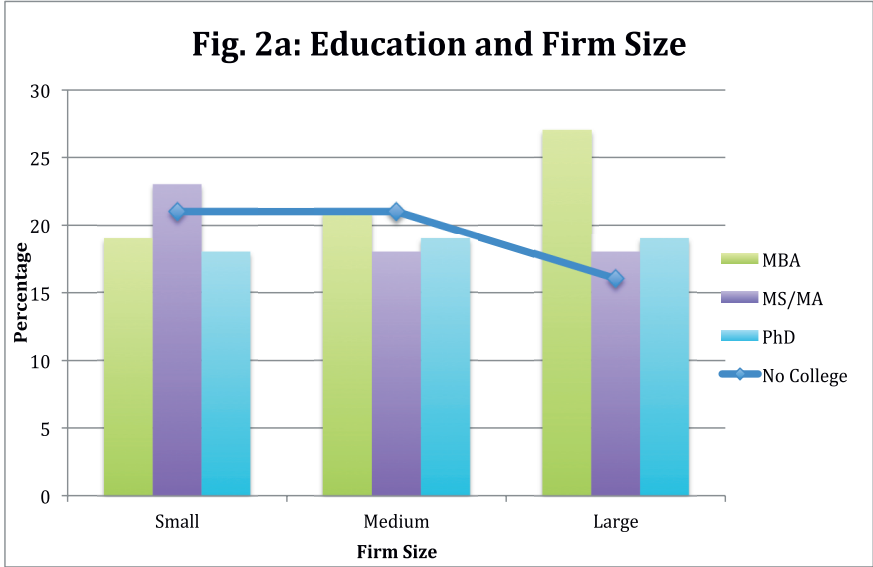
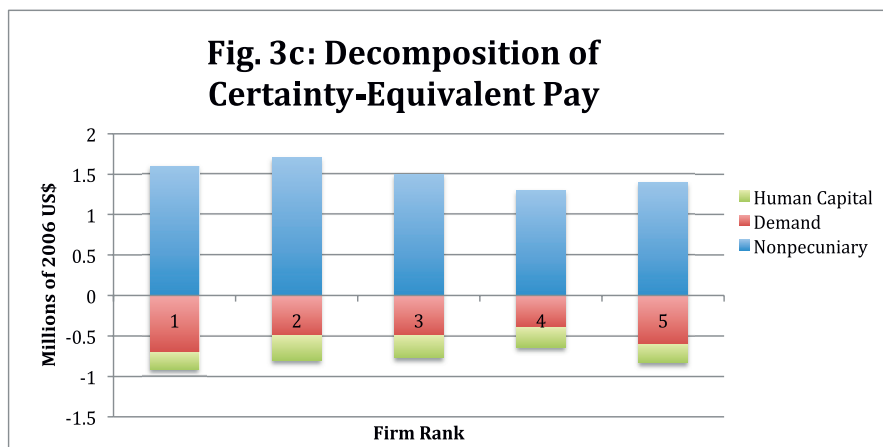
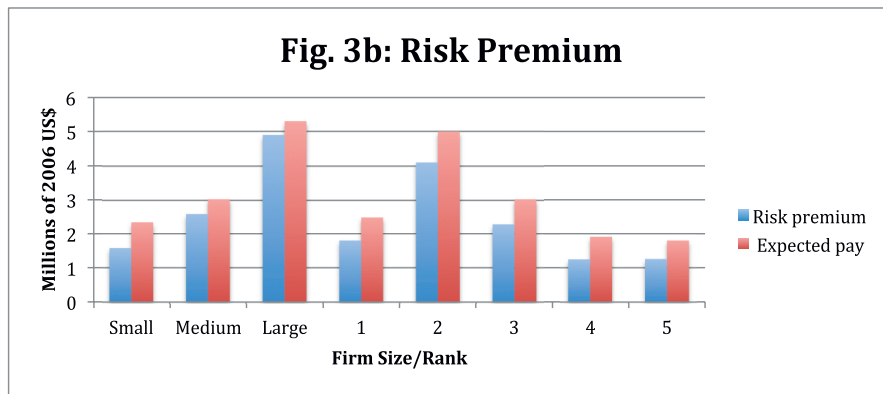
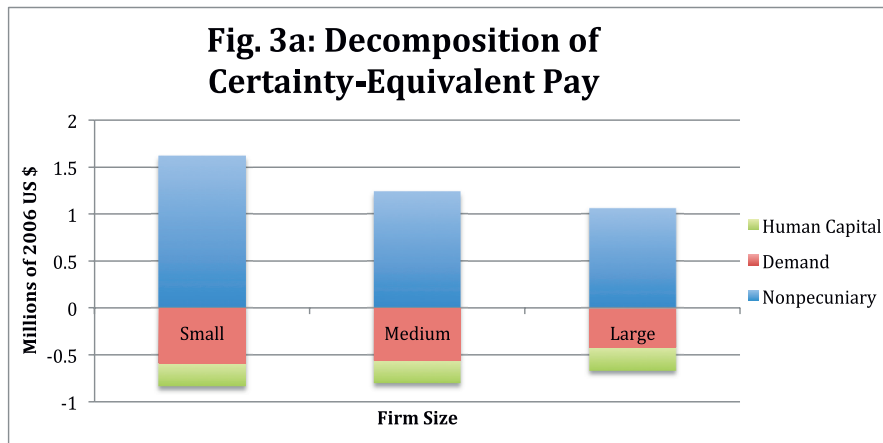
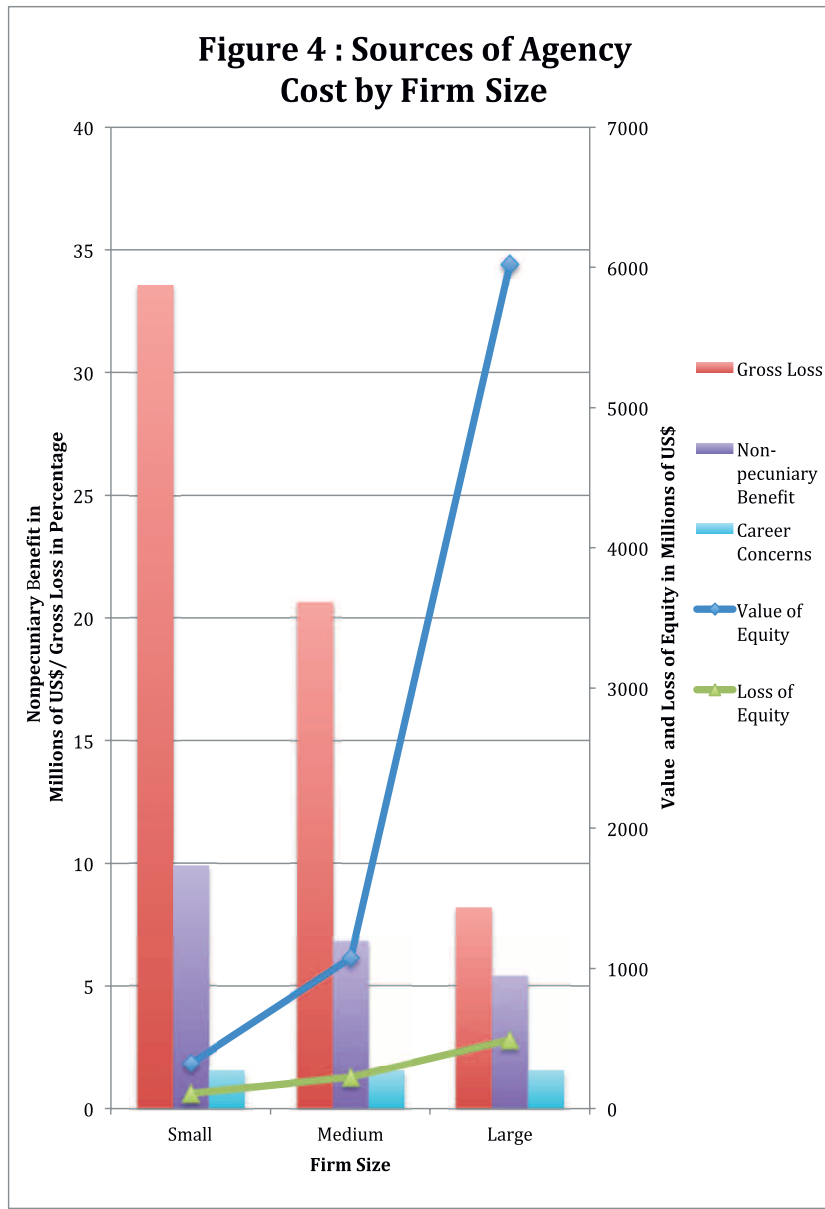


Figure 2: Education and Experience by Firm Size



Note: The certainty equivalent is the sum of human capital, demand and nonpecuniary compensating differentials.

Figure 3: Rank and Firm-Size Pay Decomposition.



Note: Gross Loss is the percentage of the firm value lost if an executive shirks instead of working diligently. Loss of Equity is the firm value lost if an executive shirks instead of working diligently. Nonpecuniary Benefit is the value to an executive of shirking relative to working diligently. Career Concerns measures by how much career concerns ameliorate the agency problem.

Figure 4: Agency Cost Decomposition