ECONOMETRICA

PROMOTION, TURNOVER, AND COMPENSATION IN THE EXECUTIVE LABOR MARKET

GEORGE-LEVI GAYLE
Washington University in St. Louis, St. Louis, MO 63130, U.S.A.

LIMOR GOLAN
Washington University in St. Louis, St. Louis, MO 63130, U.S.A.

ROBERT A. MILLER
Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

PROMOTION, TURNOVER, AND COMPENSATION IN THE EXECUTIVE LABOR MARKET

BY GEORGE-LEVI GAYLE, LIMOR GOLAN, AND ROBERT A. MILLER[1]

This paper develops a generalized Roy model with human capital accumulation, moral hazard, and career concerns. We identify and estimate the model with a large panel that matches data on publicly listed firms to information on their executives. The structural estimates obtained are used to decompose the firm-size pay gap. We find that although total compensation and incentive pay increase with firm size, certainty-equivalent pay decreases with firm size. In larger firms, and for more highly ranked executives, weaker signal quality about effort results in higher risk premiums. This risk premium accounts for roughly 80 percent of the firm-size gap in total compensation. Larger firms are also willing to pay more than smaller ones to attract executives. Finally, the estimated coefficients on human capital accumulation from formal education and experience gained from different firms are individually significant, but their collective effect on firm-size pay differentials nets out.

KEYWORDS: Agency cost, asymmetric information, career concern, compensating differential, executive compensation, firm-size pay differential, identification, moral hazard, sequential equilibrium, structural estimation.

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 demand for entrepreneurial talent. However, none 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.[2] 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. In the model, executives make sequential job and effort choices, taking into account the compensation, nonpecuniary benefits from working, and the 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 is based on two equations that hold in the sequential equilibrium we analyze. The first equation applies to a manager who is indifferent between accepting 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, up to a factor of proportionality, the slope of the contract identifies the likelihood ratio of abnormal returns for different effort choices.

---

[2]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.

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. Finally, the identification of 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.

The three extensions we undertake to the Roy (1951) model (nonpecuniary utility, agency, and human capital) are motivated by empirical regularities found in our data, a matched sample of over 30,000 executives and 2,500 firms spanning 14 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 life-cycles; 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.

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 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.

Our structural estimates show 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-sized 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 confirms 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, with larger firms having more supervisory positions and accountability more difficult, which leads 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 executive rank, contradicting conventional wisdom. Since those lower in the ranks and closer to operations can most affect excess returns 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 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 translate to a higher risk premium explain most of the differences in total compensation across ranks, a finding broadly consistent with the monitoring paradigm of McNulty (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 and 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 analyzes 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 discusses agency costs and firm size. Section 10 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 the Supplemental Material (Gayle, Golan, and Miller (2015b)).

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 private 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. To achieve identification and hence interpret the empirical results, we take a more parsimonious approach, extending Margiotta and Miller (2000) and Gayle and Miller (2015) to account for job choice and human-capital accumulation.

Only a handful of theoretical analyses study dynamic contracting with moral hazard in nonstationary environments where current actions have consequences over a long horizon (Garrett and Pavan (2012), Li (2014), Sannikov (2014), Gayle, Golan, and Miller (2015a)). Our companion paper (Gayle, Golan, and Miller (2015a)) borrows the equilibrium framework developed here to analyze the compensation of interlocked executives and inside directors. The model closest to ours is Garrett and Pavan's (2012). In both models, match quality between a firm and its managers changes stochastically over time and shocks to managerial productivity are anticipated at the time of contracting but are 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). Diversity of experience has no value in the standard general human-capital framework. 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 authors have studied identification and inference in the generalized Roy model (Bayer, Khan, and Timmins (2011), D'Haultfœuille 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 the work of Gayle and Miller (2015), who showed that the static and repeated moral hazard models are set identified only. This paper extends their 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)), whereas others, like ours, use matched firm-manager data, which allow the incorporation of firm and worker heterogeneity (Postel-Vinay and Robin (2002), Cahuc, Postel-Vinay, and Robin (2006), Taber and Vejlin (2010)). Of these, only Gayle and Golan (2012) motivated turnover and wages with information asymmetries between workers and firms, but they did 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 receive 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 are 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 percent 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 Securities and Exchange Commission (SEC). Data on the 2,818 firms for the ExecuComp database were supplemented by the COMPUSTAT North America database and monthly stock price data from the Center for Research in Security Prices 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 Marquis Who's Who, which contains biographies of about 350,000 executives. The matched data provide 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.[3] Rank 1 includes 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 is the CEO of the company. Rank 3 includes the COO, the CFO, and the chairman of the board of the company if that person holds an executive position in the company other than CEO. Other high-level corporate executives and heads of subsidiaries or regional chiefs comprise Rank 4, while Rank 5 is reserved for lower-level executives. Thus, CEOs are not in Rank 1. Since this hierarchy is based on transitions, the ranking reflects lifecycle considerations, not power or control. The ranking corroborates the institutional use of the term, 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 classify firms into three industrial sectors: primary, consumer, and service. Firms are also partitioned by size—large, medium, and small—based on the value of their assets and number of employees over the sample period. A firm is defined 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 of employees are below the median for its sector over the sample. All other firms are medium sized. We further classify firms by the number of "insiders" on their board relative to the industrial norm. A company is defined 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, reflecting our focus on executive compensation, firms are classified from the perspective of their executives: new if this is the first year the executive is working in the firm and old if the executive has worked in the firm for more than one year. This variable allows us to capture the effects of executive turnover. Summarizing, there are 36 firm types, differentiated by size, industrial sector, importance of insiders on the board, and whether the executive in question has just joined the firm.

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 executives holding firm options and changes in wealth from holding firm stock relative to a well-diversified market portfolio.[4] Hence, the change in wealth from holding their firms' stock is the

---

[3]The method for constructing the hierarchy, and a detailed description of the titles in each rank, is in Gayle, Golan, and Miller (2012).

[4]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

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 of several dimensions of labor market experience, some demographic background variables, and whether the executive is interlocked.[5] Variables we construct on labor market experience include years of tenure with the firm, years worked as top executive, number of firms in which an executive worked before becoming an executive, and the number of firms in which an executive worked after becoming an executive. We also observe educational qualifications (including MBA, MSc, PhD), gender, and age. Finally, since the price of console bonds plays a role in consumption smoothing in our model, we construct a bond price series from the Federal Reserve Economic Database (FRED). Section B of the Supplemental Material contains a full description of the construction and a data summary.

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. Documenting these basic empirical regularities is worthwhile because the previous literature on pay and firm size focused on other labor markets, and investigators studying at the executive labor market (Gabaix and Landier (2008), Gayle and Miller (2009b)) did not have data on hierarchy, education, work experience, and mobility.

Figure 1(a) 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. Figure 1(b) shows 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 might suggest that large firms have a more serious monitoring problem than

---

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, 2009b) for other papers using this measure of total executive compensation.

[5]An executive is classified as interlocked if at least one of the following is true: (i) The executive serves on the board committee that makes her compensation decisions. (ii) 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. (iii) 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.
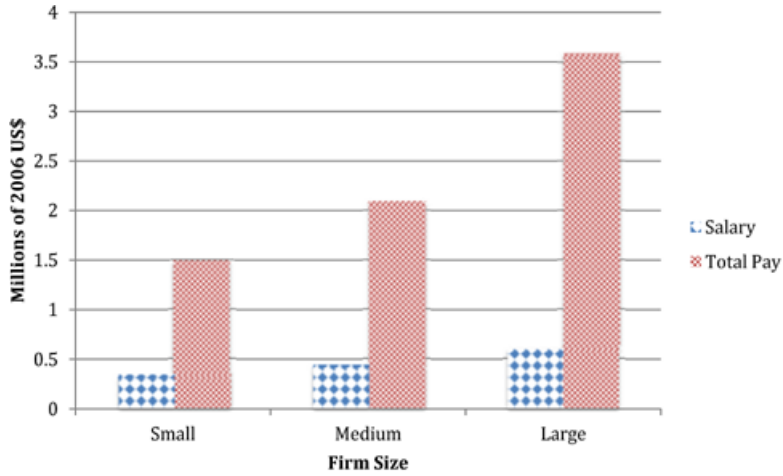
(a)
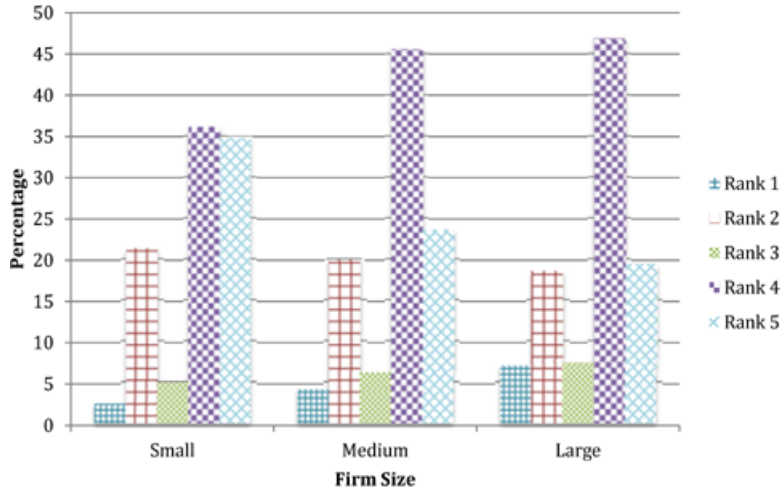
(b)

FIGURE 1.—Pay and hierarchy by firm size. (a) Firm size pay premium, (b) hierarchy by firm size.

small firms; this hypothesis has been proposed in 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 2(a) shows that executives in large firms have more formal education than executives in small firms. Among executives with formal education, there are also differences by firm size. While executives with a PhD are equally distributed across firm size, large firms have a higher concentration of execu-
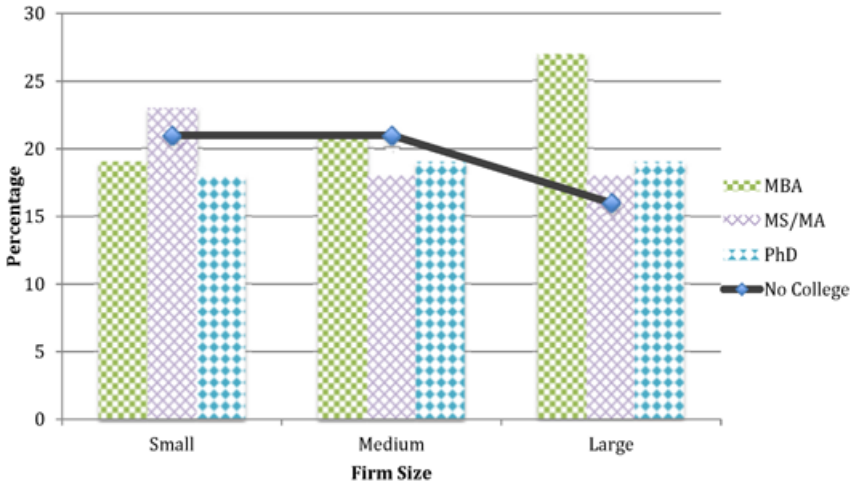
(a)
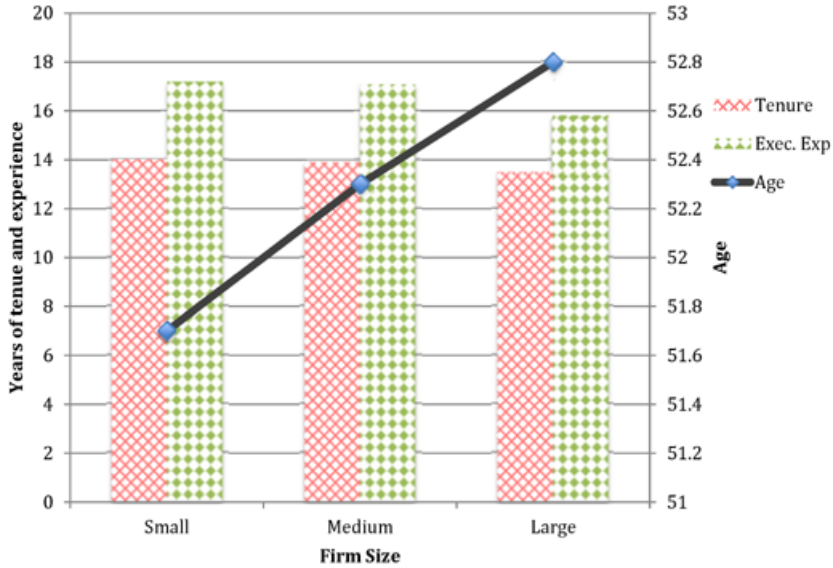
(b)

FIGURE 2.—Education and experience by firm size. (a) Education and firm size, (b) experience and firm size.

tives with an MBA but a lower concentration of nonbusiness master's degrees. This might suggest that large firms have a higher demand for talent. However, Figure 2(b) 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. Together, Figures 2(a) and 2(b) follow Mincer's (1974) arguments about

the value of schooling: Executives in large firms have less job experience and are older because they acquired more formal education. Our data are from a truncated sample of upper management executives 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 among firm size, hierarchy, and human capital, Table I 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 doctorate are most likely to be found in Ranks 4 and 5. Thus, Rank 5 executives are the most educated by every measure except MBA, while Rank 2 executives are more likely to have an MBA than an executive in any other rank. Salary, total compensation, and the likelihood of being a board member 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 I and Figure 1(a) account for interactions between firm size and hierarchy (Figure 1(b)), education (Figure 2(a)), and experience (Figure 2(b)). Table II shows the effects of using conditioning information in four regressions: on compensation and three indicators of job mobility. First is a second-order polynomial compensation regression that decomposes compensation in terms of its fixed and variable components. The first three columns of Table II report the coefficients and their estimated standard errors from this one regression on rank (panel A), firm type (panel B), and human capital plus individual heterogeneity (panel C). 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 II shows the empirical regularities in the firm-size pay premium are robust to controlling for these interactions. Panel B of Table II shows three empirical regularities with regard to compensation and firm type: (i) Larger firms compensate executives with higher fixed pay, as is customary in labor markets, and, on average, higher incentive pay as well. (ii) Firms with a larger number of insiders on their board of directors have higher incentive pay but the same fixed pay. (iii) The service sector pays the highest fixed pay and offers the highest incentive pay, while the primary sector pays the lowest fixed pay and offers the lowest incentive pay. Does (i) imply that certainty-equivalent pay is higher in large firms than small firms? Answering this question requires knowledge of the risk parameter of executives, which we obtained from an identified behavioral model that we assume generates the data.

TABLE I
EXECUTIVE CHARACTERISTICS BY RANK[a]

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
Bachelor's degree	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

[a]Standard deviations are listed 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) entering one of the ranks in our sample. Execdir is an indicator of whether the executive is a member of the board of directors. *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.

TABLE II

COMPENSATION AND MOBILITY[a]

	Compensation			Promotion					Turnover	Retirement
	π	π^2	Level	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5		
				Panel A: Executive Position						
Constant	21,601	−9,114	−4,359							
	(3,859)	(1,914)	(2,716)							
Rank 4	1,529	−242	103	−20.3	−76.2	−67.8	63.4	−536	–	–
	(926)	(444)	(463)	(20.5)	(9.8)	(10.4)	(4.01)	(6.51)		
Rank 3	2,627	−164	1,267	−88.1	−72.4	114	−404	−754	94.7	–
	(1,407)	(605)	(662)	(24.1)	(12.8)	(4.43)	(14.6)	(19.7)	(17.6)	
Rank 2	6,007	−789	3,456	−118	67.7	−393	−551	−901	213	−8.9
	(1,394)	(699)	(683)	(25.1)	(3.90)	(20.0)	(18.2)	(22.4)	(12.6)	(4.28)
Rank 1	9,839	−454	1,055	111	−290	−345	−585	−939	86.1	55.26
	(1,690)	(987)	(797)	(10.1)	(34.2)	(33.6)	(33.8)	(39.0)	(23.2)	(3.23)
Execdir	7,695	−848	845	−22.8	123	15.6	−70.4	−105	−102	−64.72
	(570)	(304)	(251)	(13.2)	(4.03)	(5.32)	(3.39)	(6.50)	(9.19)	(3.50)
				Panel B: Firm Type						
Service	3,149	88	777	–	–	–	–	–	–	–
	(419)	(222)	(198)							
Primary	−3,609	1,537	−633	–	–	–	–	–	−18.68	–
	(473)	(267)	(198)						(7.11)	
Medium sized firm	4,079	−253	937	–	–	–	–	–	–	–
	(437)	(201)	(214)							
Large firm	12,703	−2,224	3,697	–	–	–	–	–	–	–
	(405)	(212)	(190)							
Large board	2,683	−1,203	280	–	–	–	–	–	−25.7	–
	(358)	(176)	(163)						(5.71)	

(*Continues*)

TABLE II—*Continued*

	Compensation			Promotion					Turnover	Retirement
	π	π^2	Level	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5		
Panel C: Human Capital and Individual Heterogeneity										
Rank 1 Lagged	12,085	−3,054	544						–	–
	(1,769)	(987)	(822)							
Rank 2 Lagged	14,640	−2,875	660						−132	–
	(1,342)	(625)	(658)						(18.6)	
Rank 3 Lagged	4,849	−1,100	597						−42.75	–
	(1,389)	(586)	(653)						(19.1)	
Exec. exp.	191	−42	1.61	−9.66	−32.8	1.01	14.04	34.76	82.8	24.44
	(26)	(14)	(25)	(14.5)	(8.93)	(11.79)	(7.00)	(11.33)	(13.4)	(5.69)
Exec. exp. squared	–	–	–	6.09	8.22	−3.08	−3.79	−10.1	−20.6	−11.38
				(6.41)	(4.17)	(5.77)	(3.49)	(5.49)	(6.62)	(2.65)
Tenure	−23	22	−40	−10.43	−23.03	−10.56	10.7	25.8	−302	24.66
	(25)	(14)	(20)	(9.48)	(6.16)	(8.15)	(4.61)	(7.99)	(9.02)	(3.97)
Tenure squared	–	–	–	4.80	7.28	4.19	−2.79	−11.85	88.1	−7.26
				(4.47)	(2.94)	(4.11)	(2.30)	(4.07)	(4.3)	(1.97)
External	−12,396	2,155	−1,026	–	–	–	–	–	–	–
	(996)	(478)	(1,255)							
Rank 2 × External	–	–	3,840	–	–	–	–	–	–	–
			(1,459)							
Rank 3 × External	–	–	5,289	–	–	–	–	–	–	–
			(1,975)							

TABLE II—*Continued*

	Compensation			Promotion					Turnover	Retirement
	π	π^2	Level	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5		
NBE	–	–	–	−8.99	−0.91	−6.86	3.11	8.11	−11.1	5.54
				(2.19)	(1.57)	(1.82)	(0.82)	(1.26)	(2.14)	(0.62)
NAE	−484	−58	215	−1.23	1.35	−2.43	−0.34	−0.25	−13.7	4.49
	(174)	(93)	(80)	(2.28)	(1.49)	(2.00)	(1.15)	(1.82)	(1.94)	(0.78)
Age	17	15	281	−9.01	1,024	174	−459	−847	1,948	−527.74
	(23)	(10)	(85)	(188)	(124)	(158)	(86.1)	(128)	(239)	(58.58)
Age squared	–	–	−3.05	136	−5.20	−111	236	434	−992	312.89
			(0.80)	(88.7)	(60.3)	(80.3)	(44.4)	(65.8)	(122)	(28.81)
Female	–	–	–	–	–	–	–	–	–	17.42
										(4.16)
Rank 2 × Female	–	–	2,668	–	–	–	–	–	−0.51	–
			(1,295)						(0.24)	
MBA	–	–	–	–	–	–	–	–	–	−84.46
										(20.1)
Interlocked	6,403	−1,496	−299	–	–	–	–	–	−93.0	−93.0
	(995)	(471)	(464)						(28.6)	(28.6)

[a]Standard errors are listed in parentheses; tenure and experience (Exec. exp.) are measured in years; NBE (NAE) indicates the number of firms worked in before (after) becoming a top executive. The elasticities are calculated using logit regressions.

Panel C of Table II demonstrates three empirical regularities with regard 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 II 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. However, exit is convex increasing in age; older executives are more likely to be found in the highest paid ranks and are paid a premium for any rank they hold. In addition, they have substantially more incentive pay. This is more consistent with 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 of 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 evaluating 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 executive compensation in different firms, as well as its evolution with age, tenure, and experience. For pedagogical reasons, we defer our analysis of career concerns to the next section. In the basic model, shareholders are expected value maximizers 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, propose cost-minimizing contracts to shareholders, and then select their

effort levels. This process determines the returns to shareholders and executive compensation. Free entry by firms ensures that 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 $j \in \{1, \ldots, J\}$, with $j = 0$ representing retirement. There are K positions within each firm j, indexed by $k \in \{1, \ldots, K\}$ 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 $t \in \{0, 1, \ldots\}$ 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 proposes an employment contract and, after securing the agreement of shareholders, signs the contract that determines her compensation. 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

$$(4.1) \qquad d_{0t} + \sum_{j=1}^{J} \sum_{k=1}^{K} d_{jkt} = 1.$$

Summarizing, $d_t \equiv (d_{0t}, d_{11t}, \ldots, d_{JKt})$ 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 $l_t = 0$ indicates the executive shirks at age t and $l_t = 1$ indicates 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 the executive but not shareholders, and *work* 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-additive 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 $\varepsilon_t \equiv (\varepsilon_{0t}, \varepsilon_{11t}, \ldots, \varepsilon_{JKt})$, 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, as well as the firm and rank. When the executive works (setting $l_t = 1$), her nonpecuniary utility is $\alpha_{jkt}(h)$; when she shirks, it is $\beta_{jkt}(h)$. The executive's lifetime utility can thus be summarized as

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

where δ denotes the subjective discount factor and ρ is the constant absolute risk-aversion parameter. The systematic component of nonpecuniary benefits from retiring is normalized to 1. We assume there is more disutility from working than from shirking, so $\alpha_{jkt}(h) > \beta_{jkt}(h)$. The difference between $\beta_{jkt}(h)$ and $\alpha_{jkt}(h)$ captures the divergence between the shareholder and executive goals.

This formulation of the utility function captures differences in nonpecuniary costs across ranks and firms. It accounts 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 in analyses involving dynamics and uncertainty because of its tractability. Under CARA, the log of indirect utility is linear in outside wealth and additively separable in taste shocks and shifters. Consequently, outside wealth is eliminated when comparing different options. This 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 $h_{2t} = (h_{211t}, \ldots, h_{2JKt})$. Thus, the vector that captures all human capital is $h_t = (h_1, h_{2t})$. We also define a transition function, $\overline{H}_{jk}(h_{2t})$, to capture the evolution of human capital; we assume the function is deterministic and that human capital follows the law of motion:

$$(4.3) \qquad h_{2t+1} = \sum_{j=1}^{J} \sum_{k=1}^{K} d_{jkt} \overline{H}_{jk}(h_{2t}).$$

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 (i) captures by how much experience in the jth firm at rank k increases productivity in each firm and rank and (ii) allows for increments to differ by firm and rank.

Example. To illustrate the process of human-capital accumulation and help motivate the empirical application that follows, in this example we define an executive-firm match by a coordinate pair (j_1, j_2) and a triplet $h_{2t} \equiv (h_{2t}^{(1)}, h_{2t}^{(2)}, h_{2t}^{(3)})$. Thus, $j_1 \in \{0, 1\}$ is an indicator of whether the firm is new to the executive or not, where $j_1 = 1$ indicates the executive worked for this firm last period, and $j_1 = 0$ indicates she did not; $j_2 \in \{1, 2, \ldots, J_2\}$ are indicators of the firm's size and industrial sector. With regard to h_{2t}, let $h_{2t}^{(1)}$ denote executive tenure in the current firm, $h_{2t}^{(2)}$ the number of years of executive experience, and $h_{2t}^{(3)}$ the number of different firms in which the executive has worked since becoming an executive. The transition function for human capital is specified as

$$(4.4) \qquad \overline{H}_{jk}(h_{2t}) = h_{2t} + \Delta_{jk},$$

where $\Delta_{jk} \equiv (\Delta_{jk}^{(1)}, \Delta_{jk}^{(2)}, \Delta_{jk}^{(3)})$. If the executive does not retire but chooses a new firm, then $\Delta_{jk} = (-h_{2t}^{(1)}, 1, 1)$. This means she would lose all her firm-specific capital, gain an additional year of executive experience, and increase the number of firms in which she worked. On the other hand, if she remains with her current firm, then $\Delta_{jk} = (1, 1, 0)$.

While $h_{2t}^{(1)}$ and $h_{2t}^{(2)}$ are often used as variables in the learning-by-doing literature, $h_{2t}^{(3)}$ is more novel, capturing the idea that managers may acquire skills from working in different organizations. It allows for the possibility that

younger executives might 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 experimentation literature on human capital (Miller (1984), Antonovics and Golan (2012), Sanders (2013)) except that, in that body of work, learning about unknown skills takes place, whereas in this framework, the upper levels of the managerial hierarchy value certain types of experience that can most efficiently be acquired by working in multiple firms.

4.5. *Firm Technology*

Firm Production and Value. In this subsection alone, it is necessary to identify the executive pool explicitly because firms may employ more than one executive in the same position. To distinguish between lifecycle effects and aggregate technological shocks, we also track workers' ages over calendar time. We now suppose there are $N_{j\tau}$ executives who sort themselves into positions at the jth firm in period τ. Denote by $t(\tau, n)$ the age of the nth 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 jth firm's output in τ if she chooses the kth 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 let $\pi_{j,\tau+1}$ denote the (net) excess return to the jth firm. Let $\mathcal{E}_{j\tau}$ 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) $\mathcal{E}_{j\tau} \pi_{\tau+1}$ is the return on equity attributable to aggregate productivity shocks.

(iii) $\mathcal{E}_{j\tau} \pi_{j,\tau+1}$ is the excess return to the firm, $\pi_{j,\tau+1}$, whose probability distribution depends on the effort of all the executives.

The first component of the output, the summed expression involving $F_{jk}(h_{t(\tau,n)})$, is additively separable in the productivity of each executive n. 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, $h_{t(\tau,n)}$. This 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.[6] 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

---

[6]Here, we are abstracting from other costs faced by the firm, such as the wage bill for the nonexecutive workforce, by implicitly accounting for them in this term.

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 motion[7]

$$(4.5) \qquad \mathcal{E}_{j,\tau+1} \equiv \sum_{n=1}^{N} \sum_{k=1}^{K} d_{jkt(\tau,n)} \big[F_{jk}(h_{t(\tau,n)}) - w_{jk,\tau+1}^{(n)} \big] + \mathcal{E}_{j\tau}(\pi_{\tau+1} + \pi_{j,\tau+1}).$$

We show that, in equilibrium, the expected compensation to an executive fully offsets his expected contribution to output. Setting the summed expression to zero and rearranging equation (4.5) to make $\pi_{j,\tau+1}$ the subject then yields the standard definition of excess returns in the asset-pricing literature, $\mathcal{E}_{j,\tau+1}/\mathcal{E}_{j\tau} - \pi_{\tau+1}$.

Distribution of Abnormal Returns. Executive effort affects the firm only through the probability distribution determining $\pi_{j,\tau+1}$. We analyze an equilibrium where every executive works, in which case the value of $\pi_{j,\tau+1}$ is drawn from a probability density function denoted by $f_j(\pi)$. Consistent with standard asset-pricing theory, we normalize the expected value of abnormal returns in equilibrium from everyone working to zero.

If everyone except the kth ranked executive works, conditional on any level of human capital h_t, the value of $\pi_{j,\tau+1}$ is on $f_j(\pi)g_{jk}(\pi|h_t)$. Thus, the impact on production from an executive shirking is captured by $g_{jk}(\pi|h_t)$, the likelihood ratio for the density when the executive with h_t in position k shirks while all other executives work, and the density when all executives work diligently. Since equilibrium compensation depends on $\pi_{j,\tau+1}$, the kth ranked executive realizes that if he was the only one to shirk, his expected compensation would depend on $f_j(\pi)g_{jk}(\pi|h_t)$, and this consideration ultimately explains why $f_j(\pi)g_{jk}(\pi|h_t)$ helps shape equilibrium compensation.

Let $f_{0j}(\pi)$ denote the probability density function for π when the combination of who works and who shirks is chosen to maximize its expected value subject to the constraint that at least two executives shirk. The precise functional form of $f_{0j}(\pi)$ is immaterial in an equilibrium where everyone works, because $f_{0j}(\pi)$ generates only π if two or more executives deviate from their equilibrium action.[8]

---

[7]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.

[8]Margiotta and Miller (2000) assumed the distribution of π is the same when two or more executives shirk. In their specialization, $f_{0j}(\pi)$ is a primitive—namely, the common probability density for all possible work/shirk combinations of the firm's executives when at least two shirk. In our framework, we could develop notation for the density functions of all those possible work/shirk combinations and state $f_{0j}(\pi)$ in terms of those primitives. However, this would be a sterile exercise because not even $f_{0j}(\pi)$ is identified in an equilibrium where everyone works, let alone the functions from which it is derived.

In our model, a necessary condition for an equilibrium to exist where everyone works is that expected abnormal returns are maximized by everyone working. Formally, we assume

$$(4.6) \qquad 0 = \int \pi f_j(\pi)\, d\pi > \max\left\{ \int \pi f_j(\pi) g_{jk}(\pi|h_t)\, d\pi, \int \pi f_{0j}(\pi)\, d\pi \right\}.$$

The potential for conflict between executive and shareholder goals arises in this model from the preferences of executives to shirk rather than work; that is, $\alpha_{jkt}(h_t) > \beta_{jkt}(h_t)$, whereas the inequalities in (4.6) show production is greater when all executives work.

The Span of Control. The likelihood ratio $g_{jk}(\pi|h_t)$ measures the degree to which executive effort can affect a firm's returns, so we interpret it as a measure of their span of control. Since $g_{jk}(\pi|h_t)$ depends on rank within the firm, there is scope for testing whether this measure of the span of control declines with rank, a hypothesis advanced by Williamson (1967).[9] Similarly, $g_{jk}(\pi|h_t)$ depends on firm size, so we can test whether the span of control increases with firm size and, in conjunction with the other estimated parameters, calculate whether agency costs increase in firm size, a rationale Lucas (1978) postulated for diminishing returns to scale in firm size. Note that $g_{jk}(\pi|h_t)$ depends on both firm and executive characteristics, but not on the number of executives in a firm, nor their human capital. Relaxing this assumption would endogenize the optimal number of executives in each firm—and the configuration of human capital within the management team—a challenge for future research.

Effort is unobserved in our model, but $\pi_{j,\tau+1}$ is a signal of effort. In this respect, $g_{jk}(\pi|h_t)$ measures the quality of the signal. For example, if $g_{jk}(\pi', h) = 1$ for some π', then the signal is uninformative about effort. If there existed some π'' in the support of $f_j(\pi)$ such that $g_{jk}(\pi'', h)$ was arbitrarily large, then the signal would be so informative that a first-best allocation could be achieved by heavily penalizing all executives if π'' occurs, and paying a constant wage otherwise. Since executives are not paid constant wages, we assume $g_{jk}(\pi|h_t)$ is bounded. We also impose the regularity condition

$$(4.7) \qquad \lim_{\pi \to \infty} g_{jk}(\pi|h_t) = 0.$$

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

---

[9]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, $g_{jk}(\pi \mid h)$ varies across the ranks of the hierarchy. Therefore, we can test whether a manager shirking causes a larger distortions in higher ranks. In contrast, Mirrlees (1975) offered an alternative view of a firm as a decentralized contractual organization.

clearly weaker than the common monotonicity assumption requiring $g_{jk}(\pi|h_t)$ to decline in π.

4.6. *Capital Markets, Information, and Timing*

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 $\tau \in \{0, 1, 2, \ldots\}$, 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.

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. Thus, at the beginning of each calendar 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 $\mathcal{E}_{j\tau}$ and excess returns $\pi_{j\tau}$ of all J firms, while every executive also observes her own outside wealth $\xi_\tau^{(n)}$, her idiosyncratic taste shocks $\varepsilon_\tau^{(n)}$, and in addition recalls her own effort history $\{l_s^{(n)}\}_{s=0}^{t-1}$ as well. The sequence of all future consumption prices (i.e., $\{E_{\tau+i}[\lambda_s]\}_{i=0}^{\infty}$ for all $s > \tau + i$) is public and symmetric information.

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, she decides which firm to be employed in and at what rank and effort level. She approaches the firm and makes an ultimatum offer that the shareholders can only accept or reject. If the offer is rejected, the executive retires and there is no additional hiring by the firm.

4.7. *Intertemporal Consumption and Employment Choices*

As a step towards deriving the optimal contract, we first derive the value function for an executive who is constrained to work each period she is employed. The separability of preferences, the constant absolute risk-aversion assumption, and the existence of complete markets for consumption goods imply that the value function for the executive's dynamic optimization problem multiplicatively factors into two pieces, an indirect utility function for wealth and an index that represents the value of human capital.

Valuing Human Capital. The value of human capital depends on how much it will be used and how much to discount the future. Accordingly, 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 calendar period-τ prices:

$$(4.8) \qquad b_\tau \equiv E_\tau \left[\sum_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau} \right].$$

We assume throughout that executives can accurately forecast bond prices, but not necessarily other aggregate prices, let alone individual returns on stocks. Let $p_{jkt}(h)$ denote the probability of choosing (j, k) at age t conditional on h, and similarly, denote by $p_{0t}(h)$ the retirement probability.[10] Also let ε_{jkt}^* denote 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 works as

$$(4.9) \qquad A_t(h) = p_{0t}(h) E\left[e^{-\varepsilon_{0t}^*/b_\tau} \right]$$

$$+ \sum_{j=1}^{J} \sum_{k=1}^{K} p_{jkt}(h) \alpha_{jkt}(h)^{1/b_\tau} E\left[e^{-\varepsilon_{jkt}^*/b_\tau} \right]$$

$$\times \left\{ A_{t+1}\left[\overline{H}_{jk}(h) \right] E_t[v_{jk,t+1}] \right\}^{1-1/b_\tau},$$

where $v_{jk,t+1}$ is a util measure of compensation from working, annuitized for consumption smoothing purposes, defined as

$$(4.10) \qquad v_{jk,t+1} \equiv \exp\left(-\rho w_{jk,t+1}(h_t, \pi)/b_{\tau+1} \right).$$

The index $A_t(h)$ is an average of expected outcomes weighted by the conditional-choice probabilities 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. Increasing expected

---

[10]From (4.14), it is clear that $p_{jkt}(h)$ also depends on the bond price b_τ, but we suppress this dependence throughout the text to prevent the equations from becoming unwieldy.

compensation reduces $E_t[v_{jk,t+1}]$ and $A_t(h)$. Similarly, $A_t(h)$ is monotonically increasing in $\alpha_{jkt}(h)$, the utility-weighted nonpecuniary losses of job characteristics.

Example. Equation (4.9) shows that, in general, all future probabilities are used to weight the career paths that might be taken. However, the independence of irrelevant alternatives assumption dramatically simplifies the formula for $A_t(h)$, allowing us to focus on the probability of retirement alone. For example, suppose ε_t is Type I extreme value, and let $\Gamma(\cdot)$ denote the complete gamma function. We prove in the Supplemental Material that

$$(4.11) \qquad A_t(h) = p_{0t}(h)^{1/b_\tau} \Gamma\left(\frac{b_\tau + 1}{b_\tau}\right).$$

Thus, at the beginning of period t, before observing the vector of disturbances ε_t but conditioning on all the state variables plus the bond price, the higher the probability of retirement, the lower the value of human capital.

Ex ante Value Function. The consumer choice problem is standard, and it is simplified by the fact that very few securities are required to characterize the optimal financial portfolio for CARA utility functions.[11] In particular, 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:

$$(4.12) \qquad a_\tau \equiv E_\tau\left[\sum_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau}(\ln \lambda_s - s \ln \delta)\right].$$

Lemma 4.1 shows that the value function is a function of the security's price a_τ, the bond price b_τ, wealth denoted by ξ_t, and human capital h as it affects the index $A_t(h, b_\tau)$.

LEMMA 4.1: *Let $U_t(h, \xi_t, a_\tau, b_\tau)$ denote the maximized discounted sum of expected utility from age $t < R$ onward given $(h, \xi_t, a_\tau, b_\tau)$. In other words, $U_t(h, \xi_t, a_\tau, b_\tau)$ is the value function for a t-year-old 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 the constraint of working every period before retirement, when the financial securities are priced at (a_τ, b_τ) in calendar year τ. Then,*

$$(4.13) \qquad U_t(h, \xi_t, a_\tau, b_\tau) = -\lambda_\tau b_\tau \exp\left(-\frac{a_\tau + \rho \xi_t}{b_\tau}\right) A_t(h).$$

---

[11]The idea of complete markets for consumption commodities dates back to Debreu (1959, Chapter 7). Our derivation follows Margiotta and Miller (2000), who exploited an aggregation result due to Rubinstein (1981).

The term $-\lambda_\tau b_\tau \exp[-(a_\tau + \rho \xi_t)/b_\tau]$ is the value function for a retiree. Thus, equation (4.13) 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 function, equation (4.2), to solve for their employment choices.

THEOREM 4.2: *If $t \leq R$ and $l_s = 1$ for all $s \in \{t, \ldots, R\}$, then job choices d_t are picked to maximize*

$$(4.14) \quad d_{0t}\varepsilon_{0t} + \sum_{j=1}^{J}\sum_{k=1}^{K} d_{jkt}\{\varepsilon_{jkt} - \ln \alpha_{jkt}(h) - (b_\tau - 1)\ln A_{t+1}[\overline{H}_{jk}(h)]$$
$$- (b_\tau - 1)\ln E_t[v_{jk,t+1}]\}.$$

This formulation generalizes Roy's (1951) model to a dynamic framework that encompasses several models of labor market sorting. Motivated by pecuniary and nonpecuniary benefits, plus human-capital considerations, executives sort themselves into jobs within the same firm and across different firms. Compensation for current work is given as $v_{jk,t+1}$, and similarly to hedonic price models with compensating differentials, the nonpecuniary benefits from the job are given by $\alpha_{jkt}(h)$ and ε_{jkt}. The dynamics are based on $\overline{H}_{jk}(h)$, the effect of taking (j, k) on current human capital h that is valued by $A_{t+1}[\overline{H}_{jk}(h), b_{\tau+1}]$, an index of human capital that accounts for the pecuniary and nonpecuniary benefits (and thus expected) that comes from making optimal choices about future jobs and ranks. It provides a framework for analyzing the trade-off between the different types of attractions that alternative careers, which might offer different starting conditions, earnings growth, measured in terms of both pecuniary and nonpecuniary benefits that accrue over the executive's career.

Lifecycle Job Choices in Equilibrium. 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. Empirically, these choice probabilities map into the model's parameters and, therefore, play an important role in estimation. The vector of choice probability functions, $p_t(h) \equiv (p_{11t}(h), \ldots, p_{JKt}(h))$, that the executive uses to compute $A_t(h)$ in equation (4.9) are precisely the probability functions that characterize her choices when solving the optimization function described by (4.14). We appeal to Proposition 1 of Hotz and Miller (1993): A mapping exists, $q(p) \equiv (q_{11}(p), \ldots, q_{JK}(p))$, from the simplex to R^{JK} such that

$$(4.15) \quad q_{jk}[p_t(h)] = \ln \alpha_{jkt}(h) + (b_\tau - 1)\ln A_{t+1}[\overline{H}_{jk}(h)]$$
$$+ (b_\tau - 1)\ln E_t[v_{jk,t+1}].$$

Given h, the solution to the optimization problem in equation (4.14) depends only on the vector of differences $(\varepsilon_{11t} - \varepsilon_{0t}, \ldots, \varepsilon_{JKt} - \varepsilon_{0t})$ 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 (4.14), collecting terms involving d_{jkt}, and noting that the additive constant, ε_{0t}, has no effect on the optimal choices. Substituting equation (4.15) into (4.14), we see that if position (j, k) is the optimal employment choice, then $\varepsilon_{jkt} - \varepsilon_{0t} > q_{jk}[p_t(h)]$ and

$$(4.16) \quad (j, k) = \underset{(j', k')}{\arg\max} \{ \varepsilon_{j'k't} - q_{j'k'}[p_t(h)] \}.$$

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

$$(4.17) \quad (\varepsilon_{11t} - \varepsilon_{0t}, \ldots, \varepsilon_{JKt} - \varepsilon_{0t}) \equiv q[p_t(h)] \equiv (q_{11t}, \ldots, q_{JKt}).$$

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

Example. Inserting equation (4.4) into the $A_{t+1}[\overline{H}_{jk}(h)]$ function implied by (4.11), and noting that $q_{jk}[p_t(h)]$ is the familiar log-odds ratio when ε_t is Type I extreme value, (4.15) specializes to a log-linear equilibrium sorting function in the log-odds ratio between any two job-match options, including retirement. For example, if j and j' have the same rank, we obtain

$$(4.18) \quad \ln\left(\frac{p_{jkt}(h)}{p_{j'kt}(h)} \right) = -\ln\left(\frac{\alpha_{jkt}(h)}{\alpha_{j'kt}(h)} \right) - \frac{b_\tau - 1}{b_{\tau+1}} \ln\left(\frac{p_{0,t+1}(h + \Delta_{jk})}{p_{0,t+1}(h + \Delta_{j'k})} \right)$$
$$- (b_\tau - 1) \ln\left(\frac{E_t[v_{jk,t+1}]}{E_t[v_{j'k,t+1}]} \right).$$

Equation (4.18) highlights the trade-off between the four dimensions: nonpecuniary benefit, $\alpha_{jkt}(h)$; human capital accumulation, Δ_{jk}; expected utility from compensation, $E_t[v_{jk,t+1}]$; and implicitly, the privately observed idiosyncratic component to nonpecuniary utility, ε_t. Consider the three scenarios below.

(i) Suppose $\Delta_{jk} = \Delta_{j'k}$, but $\alpha_{jkt}(h) > \alpha_{j'kt}(h)$, while $w_{jk,t+1}(h_t, \pi) > w_{j'k,t+1}(h_t, \pi)$ for all (h_t, π). This implies $E_t[v_{jk,t+1}] < E_t[v_{j'k,t+1}]$. A higher value of $\alpha_{jkt}(h)$ relative to $\alpha_{j'kt}(h)$ decreases the probability of choosing firm j relative to firm j'. On the other hand, a higher $w_{jk,t+1}(h_t, \pi)$ relative to $w_{j'k,t+1}(h_t, \pi)$ lowers $E_t[v_{jk,t+1}]$ relative to $E_t[v_{j'k,t+1}]$, and this increases the probability of choosing firm j relative to firm j'. This scenario highlights the trade-off between nonpecuniary and pecuniary benefits embedded in the model.

(ii) Now say $w_{jk,t+1}(h, \pi) = w_{j'k,t+1}(h, \pi)$, but $\alpha_{jkt}(h) > \alpha_{j'kt}(h)$, and $\Delta_{j'k} \neq \Delta_{jk}$. From (4.11), j has a lower investment value than j' if and only if $p_{0,t+1}(h + \Delta_{jk}) > p_{0,t+1}(h + \Delta_{j'k})$. In that case, the additional nonpecuniary disutility from choosing j over j' is accentuated by its lower investment value, implying the choice probability for choosing j is lower than for j'.

(iii) If j dominates j' on nonpecuniary benefits, $\alpha_{jkt}(h) < \alpha_{j'kt}(h)$, compensation, $w_{jk,t+1}(h, \pi) > w_{j'k,t+1}(h, \pi)$, and investment value, $p_{0,t+1}(h + \Delta_{jk}) < p_{0,t+1}(h + \Delta_{j'k})$, then from (4.18), j is clearly more likely to be chosen: $p_{jkt}(h) > p_{j'kt}(h)$. Nevertheless, $p_{j'kt}(h) > 0$, because of the fourth factor ε_t. A fraction of executives drawing a sufficiently high $\varepsilon_{j'kt} - \varepsilon_{jkt}$ differential choose j' over j, notwithstanding its lower investment value, poorer compensation, and lower publicly observed systematic nonpecuniary benefits.

4.8. *Cost-Minimizing Contracts*

Equilibrium contracts that stimulate executives to work minimize the expected cost of attaining the equilibrium conditional-choice probabilities subject to an incentive compatibility condition deterring shirking. Proving this assertion is done by simple contradiction argument: Rather than demand an inefficient contract, the executive could have extracted more rent by offering an efficient contract that made the firm just as profitable. In our model, the cost-minimizing contract is the sum of a fixed component, called certainty-equivalent pay, plus a variable component, whose expectation is the risk premium. We derive the certainty equivalent and the incentive compatibility constraint that give rise to the variable component before presenting a theorem that establishes the cost-minimizing contract.

Certainty-Equivalent Pay. Equation (4.14) shows that, given her effort choice, the executive is indifferent between all compensation plans with the same value of $\ln E_t[v_{jk,t+1}|h]$. It immediately follows from (4.10) and (4.14) that certainty-equivalent pay, denoted by $w_{jk,t+1}^A(h)$, is the fixed amount solving

$$\ln E_t[v_{jk,t+1}|h] = \ln E_t\big[\exp\big(-\rho w_{jk,t+1}^A(h)/b_{\tau+1}\big)\big] = -\rho w_{jk,t+1}^A(h)/b_{\tau+1}.$$

Substituting $-\rho w_{jk,t+1}^A(h)/b_{\tau+1}$ for $\ln E_t[v_{jk,t+1}|h]$ in equation (4.15) yields an expression for certainty-equivalent pay in terms of the equilibrium choice probabilities $p_t(h)$ over rank and firm (j, k), the nonpecuniary benefits of the positions of each, $\alpha_{jkt}(h)$, along with their investment values $A_{t+1}[\overline{H}_{jk}(h)]$:

$$(4.19) \quad w_{jk,t+1}^A(h) = \rho^{-1}b_{\tau+1}\big\{(b_\tau - 1)^{-1}\ln\alpha_{jkt}(h) + \ln A_{t+1}[\overline{H}_{jk}(h)]$$
$$- (b_\tau - 1)^{-1}q_{jk}[p_t(h)]\big\}.$$

Example. In the specialization, equation (4.19) reduces to

$$(4.20) \quad w_{jk,t+1}^A(h) = \frac{b_{\tau+1}}{\rho} \left\{ \frac{\ln \alpha_{jkt}(h)}{b_\tau - 1} + \frac{\ln\left(p_{0,t+1}(h + \Delta_{jk})\Gamma\left[1 + \frac{1}{b_{\tau+1}}\right]\right)}{b_{\tau+1}} \right.$$

$$\left. + \frac{1}{b_\tau - 1} \ln\left(\frac{p_{jkt}(h)}{p_{0t}(h)}\right)\right\}.$$

The example illustrates the three channels through which differentials in mean compensation arise for an executive with a given set of characteristics h. First, jobs differ in $\alpha_{jkt}(h)$, the imputed nonpecuniary cost of working. Second, they differ in $p_{0,t+1}(h + \Delta_{jk})$, the value of human capital provided by experience. 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. Third, jobs have different risk premiums, as determined by the likelihood ratio and the relative disutility of working versus shirking.

Incentive Compatibility. If effort could be freely monitored and demand existed for executives giving effort, it follows from equations (4.10) and (4.15) that a cohort of executives aged t and all with human capital h, confronted with job opportunities across K ranks in J firms offering $w_{jk,t+1}^A(h)$, would sort into the jobs following the probability distribution $p_t(h)$. However, 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 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. To construct and verify an equilibrium in which everybody works, it suffices to consider what happens when just one executive deviates from the equilibrium by shirking and all the others work.

Shareholders would reject a contract that does not give the executive offering it sufficient incentive to work. Thus, the contract must yield higher expected utility to the executive from working rather than shirking. In the basic model, shirking 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 from which to choose. With reference to (4.14), the incentive compatibility constraint for the basic model is thus

$$(4.21) \quad \alpha_{jkt}(h)^{1/(b_\tau - 1)} E_t[v_{jk,t+1}] \leq \beta_{jkt}(h)^{1/(b_\tau - 1)} E_t[v_{jk,t+1} g_{jk}(\pi|h)].$$

Optimization. The goal of the executive is to minimize the risk premium, deadweight loss from the perspective of both shareholders and the executive,

subject to satisfying (4.21). To solve for the optimal contract, we define

$$(4.22) \quad r^A_{jk,t+1}(h, \pi)$$
$$\equiv \rho^{-1} b_{\tau(t+1)} \ln\left[1 - \eta^A(h)\left\{g_{jk}(\pi|h) - \left[\alpha_{jkt}(h)/\beta_{jkt}(h)\right]^{1/(b_\tau - 1)}\right\}\right],$$

where $\eta^A(h)$ is the unique positive root to

$$(4.23) \quad \int \left[\eta^{-1} + \left[\alpha_{jkt}(h)/\beta_{jkt}(h)\right]^{1/(b_\tau - 1)} - g_{jk}(\pi|h)\right]^{-1} f_j(\pi) \, d\pi = 1.$$

It is evident from (4.22) 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 $r^A_{jk,t+1}(h, \pi)$. Contracting to pay less in states that are relatively more likely to occur when there is shirking encourages the executive to work. Since $g_{jk}(\pi|h) \to 0$ as $\pi \to \infty$, it follows that $r^A_{jk,t+1}(h, \pi)$ has a finite upper bound of

$$(4.24) \quad \bar{r}_{jk,t+1}(h) \equiv \rho^{-1} b_\tau \ln\left[1 + \eta^A(h, b_\tau)\left[\alpha_{jkt}(h)/\beta_{jkt}(h)\right]^{1/(b_\tau - 1)}\right].$$

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 4.3 states that the optimal contract is the sum of certainty-equivalent pay defined by (4.19) and the variable component defined by (4.22).

THEOREM 4.3: *The cost-minimizing one-period contract that attracts a executive of age t with experience h to select the kth position in the jth firm with probability $p_t(h)$ and work is*

$$(4.25) \quad w_{jkt,+1}(h, \pi) = w^A_{jk,t+1}(h) + r^A_{jk,t+1}(h, \pi).$$

The optimal long-term contract can be implemented by a sequence of the one-period contracts defined in (4.25). 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.[12]

4.9. *Equilibrium*

We complete the characterization of equilibrium with a market-clearing condition. The executive supply, the choice probabilities of the different rank-firm

[12]See Malcomson and Spinnewyn (1988), Fudenberg, Holmström, and Milgrom (1990), and Rey and Salanie (1990).

combinations and retirement, is characterized by equation (4.15) relating the compensation ($E_t[v_{jk,t+1}|h]$) and the choice probabilities. Theorem 4.3 characterizes the cost-minimizing contract that satisfies the market-participation constraint and the incentive-compatibility constraint. The market-participation constraint relates the certainty-equivalent pay 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 pin down the demand for executives by assuming firms have free entry into the executive market, implying zero expected profit in equilibrium from hiring an executive:

$$(4.26) \qquad E_t\big[w_{jk,t+1}(h, \pi)|h\big] = F_{jk}(h).$$

The market-clearing condition, equation (4.26), states that each executive earns her expected marginal productivity in each period, conditional on accepting a contract in which the executive works. The firm makes zero expected profit, as the expected return on the net equity value is zero, conditional on all executives working. Since the contract is incentive compatible, all executives work. Solving backward to the negotiation stage, given that all other executives work, the executive extracts all the rents conditional on working. An executive cannot extract additional rents resulting from the distortion she causes by shirking instead of working because threatening to shirk is not credible, given the incentive-compatibility constraint. Therefore, the firm rejects any offer higher than the expected productivity, because the value from not filling the position is zero.

As in Rosen (1974), the market-clearing condition is achieved without frictions in hiring or finding jobs. There is no scarcity of positions, no costs associated with a vacant position, and output is additively separable across executives. Equilibrium compensation is increasing in the nonpecuniary costs and risk; it also depends on the dynamic component, the continuation value of human capital. It adjusts to attract the marginal executive with a taste shock that makes her indifferent between choosing the job and retirement, the assumption that rent sharing determines both the retirement probability and the demand threshold. Thus, equilibrium compensation determines the fraction of executives for every h assigned to each position (j, k). Ex post, a single firm may hire zero, one, or more executives at a given rank, optimal in our model because aside from compensation, no additional hiring costs are incurred when more than one executive accepts a given position.[13]

---

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

The value of accepting a job is firm specific because (i) there are firm-specific skills and (ii) executives have independently distributed taste shocks that 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. (See 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 approach to job matches is similar to that of Jovanovic (1979) and Miller (1984).

Equation (4.26) is part of an equilibrium outcome for a noncooperative game developed in the next section that generalizes the basic model developed here. In the extension, we assume executives can make ultimatum offers, and this assumption leads firms to make zero expected profits in the sequential equilibrium we analyze. Other mechanisms of surplus sharing in which the executives and shareholders share the surplus may be more realistic. Cahuc, Postel-Vinay, and Robin (2006) showed that labor market competition allows skilled workers to extract more surplus than unskilled workers. Our data sample the very top level of the U.S. executive market, where talent is scarce. It is reasonable to presume that these executives extract more rents than workers in lower echelons. Our modeling choice of the ultimatum game simplifies the empirical implementation. But in contrast to previous work estimating different rent-sharing mechanisms, we model internal promotions and ranks, differentiating between employers and jobs, and we characterize assignment within firms. Allowing for heterogeneity in executive bargaining power leads to questions about the optimal size of management and the equilibrium configuration of employment (Stole and Zwiebel (1996a, 1996b)), important issues left for future research.

Example. To recursively compute the equilibrium for the example, we specialize (4.14) by substituting equation (4.11) for $A_{t+1}[\overline{H}_{jk}(h)]$ and equation (4.4) for $\overline{H}_{jk}(h)$. From (4.25) and (4.26), certainty-equivalent pay is given by $F_{jk}(h) - E_t[r^A_{jk,t+1}(h, \pi)]$, where the formula for $r^A_{jk,t+1}(h, \pi)$ is given by (4.22). The value of each job choice in (4.14) can now be expressed as the sum of the disturbance ε_{jkt} and a deterministic component $W^A_{jkt}(h)$ defined as

$$
(4.27) \quad W^A_{jkt}(h) \equiv -\ln \alpha_{jkt}(h) - \frac{(b_\tau - 1)}{b_{\tau+1}} \ln p_{0,t+1}(h + \Delta_{jk})
$$

$$
- (b_\tau - 1) \ln \Gamma \left(1 + \frac{1}{b_{\tau+1}} \right)
$$

$$
+ \frac{\rho(b_\tau - 1)}{b_{\tau+1}} \left[F_{jk}(h) - E\left(r^A_{jk,t+1}(h, \pi) \right) \right].
$$

Since the distribution of ε_t is Type I extreme value, the equilibrium choice probabilities exhibit the usual logit form:

$$(4.28) \quad p_{0t}(h) = \frac{1}{1 + \sum_{j=1}^{J} \sum_{k=1}^{K} \exp[W_{jkt}^{A}(h)]},$$

$$p_{jkt}(h) = \frac{\exp[W_{jkt}^{A}(h)]}{1 + \sum_{j=1}^{J} \sum_{k=1}^{K} \exp[W_{jkt}^{A}(h)]}, \quad \text{if} \quad j = 1, \ldots, J.$$

Equilibrium is computed in the following five steps:

(i) Solve for $\eta^A(h)$ using equation (4.23).

(ii) Appealing to (4.25), compute $w_{jkt+1}(h, \pi)$ as the sum of certainty-equivalent pay given by equation (4.20) and the variable component of compensation given by equation (4.22).

(iii) Noting that $p_{0,T-1}(h)$ is a function of the primitives of the model and $\eta(h)$ calculated in Step (i), set $t = T - 1$ and compute $W_{jk,T-1}^{A}(h)$ and $p_{0,T-1}(h)$ from (4.27) and (4.28) for each executive, noting that $p_{0T}(h) = 1$.

(iv) Form $W_{jk,T-2}^{A}(h)$ using the primitives of the model, $p_{0,T-1}(h)$ from Step (ii) and $\eta^A(h, b_{\tau(T-2)})$ calculated in Step (iii).

(v) Using the values of $p_{0,t+1}(h)$ derived in the previous iteration, recursively iterate on Steps (iii) and (iv) for $T - 3, \ldots, t$.

5. EXTENDED MODEL

This section extends the basic model to account for career concerns within a signalling game that has a pure strategy sequential equilibrium which we analyze. In the basic model, the evolution of human capital depends on successive job matches but not on hidden effort; in this extension, it depends on both. Human capital in the extension is private information, unobserved by shareholders.[14] Thus, current hidden effort choices affect future unobserved productivity, giving rise to implicit incentives to work for career concerns about future employment choices, promotions, and pay.[15] Because the extension nests the

---

[14]Suppose, on the contrary, that investment in human capital is observed by shareholders and is a mapping of effort. Then the moral hazard problem disappears, and compensation could be based on investment in human capital. Consequently, executives are paid a constant wage as a function of h.

[15]Other ways to introduce career concerns into the basic model include having symmetric learning about executive productivity, or allowing for differential utility benefits from shirking across executives that is private information. Our formulation captures career concerns in an economically meaningful way while preserving empirical tractability.

basic model, much of the notation developed in the previous section is common to both.

5.1. *Human-Capital Accumulation and Effort*

As previously, human capital is multidimensional, and the dichotomy between h_1 and h_{2t} remains the same as in the basic model. We assume that if an executive belonging to the jth firm in rank k works, her human capital follows the same transition function as in the basic model, $\overline{H}_{jk}(h_{2t})$. If she shirks, however, her human capital evolves according to another transition function, $\underline{H}_{jk}(h_{2t})$. Therefore, the law of motion of human capital becomes

$$(5.1) \qquad h_{2t+1} = \sum_{j=1}^{J}\sum_{k=1}^{K} d_{jkt}\big[l_t\overline{H}_{jk}(h_{2t}) + (1-l_t)\underline{H}_{jk}(h_{2t})\big].$$

Thus, if $l_t = 1$, human capital evolves according to equation (4.3), as in the basic model. Also note that 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 (5.1), demonstrating that equation (5.1) nests equation (4.3).

Example. We retain equation (4.4) and specify a similar equation for $\underline{H}_{jk}(h_{2t})$, namely,

$$(5.2) \qquad \underline{H}_{jk}(h_{2t}) = h_{2t} + \underline{\Delta}_{jk},$$

where $\underline{\Delta}_{jk} \equiv (\underline{\Delta}_{jk}^{(1)}, \underline{\Delta}_{jk}^{(2)}, \underline{\Delta}_{jk}^{(3)})$. We assume that if the executive shirks, she does not gain an additional year of experience, cannot increase the number of firms for which she has worked by changing jobs, and does not add to her specific capital if she remains with her current firm. As before, she loses all her specific capital if she begins working for a new firm whether she works or shirks. Symbolically, $\underline{\Delta}_{jk}^{(2)} = \underline{\Delta}_{jk}^{(3)} = 0$ and $\underline{\Delta}_{jk}^{(1)} = (d_{0t} - 1)h_{2t}^{(1)}$.

5.2. *Firm Technology and Effort*

Firm production technology is the same in both models. However, in the extension, past effort affects current human capital h. Hence, individual marginal product, $F_{jk}(h)$, and the likelihood ratio $g_{jk}(\pi|h)$, are partly determined by h through past effort, which is unobserved. To simplify the notation and the equilibrium characterization, we make a further assumption that if $l_1 = 0$, then $F_{jk}(h_t) \equiv \underline{F}$ for all h_t.[16] This initial condition places an upper bound on output, ensuring that firms do not benefit from employing executives who shirked in their first period.

---

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

5.3. *Capital Markets, Timing, and Information*

Both models have the same financial markets for contingent consumption and the same timing assumptions. However, the information structure of the extension is more complicated, because human capital, determined by hidden effort, is now private information to the executive. Let $h'_t = (h'_1, h'_{2t})$ denote shareholder beliefs about an executive's human capital—in short, her reputation—which is distinct from her actual human capital, $h_t = (h_1, h_{2t})$. The contract is based on her reputation h'_t, not actual human capital h_t. However, if she shirks, actual human capital h_t counts, and firm returns are drawn from $g_{jk}(\pi|h_t)f_j(\pi)$, not $g_{jk}(\pi|h'_t)f_j(\pi)$.

We assume shirking by just one executive is disguised; that is, the support of excess firm returns does not depend on the level of human capital, nor on whether every executive 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. Since $F_{jk}(h_t)$ cannot be separately observed and human capital is the executive's private information, $F_{jk}(h_t)$ is private information.

Nor can an executive who shirks later be identified from her subsequent job choices. In 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. Consequently, shareholders believe that h'_t follows the law of motion $h'_{t+1} = \overline{H}_{jk}(h'_t)$ when all contracts require working in equilibrium. In truth, if an executive deviates and shirks at age t, her next-period human capital is $h_{t+1} = \underline{H}_{jk}(h_t)$. Finally, to simplify the analysis of histories off the equilibrium path, we assume that all firms observe all accepted and rejected contracts, and the full employment histories, of all executives.

5.4. *Employment and Effort Choices*

The intertemporal consumption choices of an executive remain unchanged from the basic model, but the conditions describing her employment must be embellished. In order to characterize the executive's optimal labor supply and effort choices on and off the equilibrium path, we formulate the value of job matches to the executive when $h'_t \neq h_t$.

In the model extension, the executive's choice probabilities over positions are denoted by $p_{jkt}(h, h')$ because they depend on both her true human capital and her reputation. Compensation, however, is based on reputation alone, so in place of $v_{jk,t+1}$, the risk-adjusted utility from compensation is now defined as

$$(5.3) \qquad v'_{jk,t+1} \equiv \exp\bigl(-\rho w_{jk,t+1}\bigl(h'_t, \pi\bigr)/b_{\tau+1}\bigr).$$

Analogously to the definition of $A_t(h)$ given in (4.9), we define the recursion as

$$(5.4) \qquad B_t(h, h') = p_{0t}(h, h')E_t\big[e^{-\varepsilon^*_{0t}/b_\tau}\big]$$

$$+ \sum_{j=1}^{J}\sum_{k=1}^{K} p_{jkt}(h, h')E_t\big[e^{-\varepsilon^*_{jkt}/b_\tau}\big]V_{jkt}(h, h', b_\tau),$$

where

$$(5.5) \qquad V_{jkt}(h, h')$$

$$\equiv \min\Big\{\alpha_{jkt}(h)^{1/b_\tau}\big(B_{t+1}\big[\overline{H}_{jk}(h), \overline{H}_{jk}(h')\big]E_t\big[v'_{jk,t+1}\big]\big)^{1-1/b_\tau},$$

$$\beta_{jkt}(h)^{1/b_\tau}\big(B_{t+1}\big[\underline{H}_{jk}(h), \overline{H}_{jk}(h')\big]E_t\big[v'_{jk,t+1}g_{jk}(\pi|h)\big]\big)^{1-1/b_\tau}\Big\}.$$

The difference between $A_t(h)$ and $B_t(h, h')$ stems from the minimization used to define $V_{jkt}(h, h', b_\tau)$, the conditional valuation function of match (j, k) for an executive with demographics (t, h) and reputation h'. The first element of the minimization operator in equation (5.5) is the executive's 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, in contradiction to our assumptions about private information, the executive is monitored, and is always prevented from shirking, then h reduces to h' and $B_t(h, h')$ simplifies to $A_t(h)$. The second element on the right-hand side of (5.5) is a valuation function for a similarly placed executive who shirks. 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, and her reputation subsequently diverges further from her true human capital.

Theorem 5.1 extends the job-match problem from equation (4.14) 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 maximize*

$$(5.6) \qquad \varepsilon_{0t}d_{0t} + \sum_{j=1}^{J}\sum_{k=1}^{K} d_{jkt}\big[\varepsilon_{jkt} - \ln V_{jkt}(h, h')\big].$$

Example. Denote the choice probability of retirement for an executive with demographics (t, h) and reputation h' by $p_{0t}(h, h')$. If ε_{jkt} is independent and

identically distributed as a Type I extreme value with location and scale parameters $(0, 1)$, then it is well known that

$$p_{0t}(h, h') = \left[1 + \sum_{j=1}^{J} \sum_{k=1}^{K} V_{jkt}(h, h')^{-b_\tau} \right]^{-1}.$$

Moreover, we show in the Supplemental Material that $B_t(h, h')$ simplifies to

$$(5.7) \qquad B_t(h, h') = \Gamma\left(\frac{b_\tau + 1}{b_\tau} \right) p_{0t}(h, h')^{1/b_\tau}.$$

Comparing equation (5.7) with equation (4.18), the only difference between $B_t(h, h')$ and $A_t(h)$ is that $B_t(h, h')$ depends on $p_{0t}(h, h')$ rather than $p_{0t}(h)$, reflecting the role of reputation in the extension.

5.5. *The Optimal Contract*

In the sequential equilibrium we analyze, executives always work along the equilibrium path, in which case $h = h'$. After an executive has strayed off the equilibrium path by shirking, then $h \neq h'$. We show that in equilibrium, shareholders punish executives who confess to shirking by rejecting all contracts that cannot occur on the equilibrium path. Therefore, the equilibrium response of an executive with true capital h that does not measure up to her reputation h', is to pretend her true human capital is h', by demanding an equilibrium contract as if she had not shirked. Only optimal contracts occurring on the equilibrium path are left to be derived.

The incentive compatibility constraint for inducing work is impounded within the definition of $V_{jkt}(h, h', b_\tau)$ given in equation (5.5). When $h = h'$, the executive works if the compensation schedule satisfies

$$(5.8) \qquad \alpha_{jkt}(h)^{1/(b_\tau-1)} E_t[v_{jk,t+1}] B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)]$$

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

Whenever $B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)] < B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$, career concerns ameliorate the agency problem through two channels.[17] First, equation (5.8) is satisfied with constant compensation, or a fixed wage, if and only if

$$(5.9) \qquad \ln \alpha_{jkt}(h) + (b_\tau - 1) \ln B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)]$$

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

---

[17] The inequality $B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)] < B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$ is almost tautological in models in which human capital is accumulated through effort on the job, because it simply states that working increases the value of human capital more than shirking does.

This inequality shows that when the investment value of human capital is large enough relative to the disutility from working versus shirking, the incentive compatibility constraint does not bind, obviating the need to tie remuneration to the firm's abnormal returns and pay a risk premium.

Second, when compensation is variable because (5.9) fails to hold, Theorem 5.2 shows the cost-minimizing compensation schedule decomposes into a fixed and a variable component (analogous to Theorem 4.3 for the basic model), but now the latter is defined by

$$(5.10) \quad r^B_{jk,t+1}(h, \pi) \equiv \frac{b_{\tau+1}}{\rho} \ln\left[1 - \eta^B(h)\left\{g_{jk}(\pi|h)\right.\right.$$
$$\left.\left. - \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{1/(b_\tau - 1)} \frac{B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)]}{B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]}\right\}\right],$$

where $\eta^B(h)$ is the unique positive root in η to

$$(5.11) \quad \int\left[\eta^{-1} + \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{1/(b_\tau - 1)}\left[\frac{B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)]}{B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]}\right] - g_{jk}(\pi|h)\right]^{-1}$$
$$\times f_j(\pi)\, d\pi = 1.$$

For $B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)] < B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]$, we see from (5.10) and (5.11) that career concerns, captured in the ratio of human-capital values, directly offset misaligned incentives stemming from the inequality that $\beta_{jkt}(h) < \alpha_{jkt}(h)$, thus reducing the risk premium. Both factors provide implicit incentives that substitute explicit incentives provided by incentive contracts. Since implicit incentives are larger when executives are young, explicit incentives increase as executives get closer to retirement age.

THEOREM 5.2: *Define $w^B_{jk,t+1}(h)$ by replacing $A_{t+1}[\overline{H}_{jk}(h)]$ with $B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)]$ and $p_t(h)$ with $p_t(h, h)$ in (4.19). If $h' = h$, then the cost-minimizing, one-period contract attracting an executive of age t with experience h to select the kth position in the jth firm with probability $p_t(h, h)$ and work $l_t = 1$ is*

$$(5.12) \quad w'_{jk,t+1}(h, \pi) = w^B_{jk,t+1}(h) + r^B_{jk,t+1}(h, \pi).$$

Theorem 5.2 shows there are two essential differences in the compensation schedule between the basic and extended models. They arise from the variable component in compensation and the investment value of job matches.[18]

---

[18]The optimal long-term contract cannot be implemented as a sequence of short-term contracts in the extended model. In the extended model, shirking in the current period affects the

Example. The formula for B_{t+1} derived from (5.7) simplifies (5.8) to

$$(5.13) \quad b_{\tau+1}^{-1}(b_\tau - 1)\big[\ln p_{0t+1}(h + \Delta_{jk}, h + \Delta_{jk}) - \ln p_{0t+1}(h + \underline{\Delta}_{jk}, h + \Delta_{jk})\big]$$
$$\leq \ln \beta_{jkt}(h) - \ln \alpha_{jkt}(h).$$

If this inequality is met, then $\eta^B(h) = 0$ and $r_{jk,t+1}^B(h, \pi) = 0$, so the optimal contract eliciting effort offers compensation independent of π, reducing $w_{jk,t+1}'(h, \pi)$ to certainty-equivalent pay. Thus, $w_{jk,t+1}'(h, \pi) = w_{jk,t+1}^B(h)$, where

$$(5.14) \quad w_{jk,t+1}^B(h)$$

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

$$+ \frac{1}{b_\tau - 1}\ln\left(\frac{p_{jkt}(h, h)}{p_{0t}(h, h)}\right)\Bigg].$$

In contrast to the basic model of the previous section, a constant wage can be the optimal way to induce an executive to pursue shareholder interests even though everyone knows that she would never be detected by taking an action that yields greater nonpecuniary utility.

5.6. *Equilibrium*

Given the support for the probability distributions of output and taste shocks, all outcomes and job-match choices are consistent with the belief that no executive has ever shirked. Thus, job matches and output realizations cannot serve as signals. However, executives could conceivably signal their level of human capital through the contracts they offer. Theorem 5.3 embeds the market-clearing condition (5.15) within a noncooperative game. Since it contains only one subgame, comprising the entire game, we adopt the sequential equilibrium refinement to characterize behavior.

---

firm's current and future returns, directly through F_{jk}, and also indirectly because incentive compatibility might not be achieved in the future because $h \neq h'$ in all future periods. A long-term contract punishing executives for poor past firm performance has a deterrent effect as executives contemplate their future compensation, to be used in conjunction with immediate punishment, thus endowing shareholders with additional financial tools to induce incentive compatibility. We interpret the optimal one-period contract derived in this paper in the extended model as an economically meaningful departure from the null hypothesis of ignoring career concerns entirely.

THEOREM 5.3: *A sequential equilibrium with one-period contracts exists with expected compensation equal to the marginal productivity of an executive who has never shirked*:

$$(5.15) \qquad E_t\big[w_{jk,t+1}(h', \pi)|h\big] = F_{jk}(h').$$

Executives with characteristics (t, h, h') *solve the discrete choice problem* (5.6), *and offer the cost-minimizing contract specified in equation* (5.12) *for executives with characteristics* (t, h', h'). *Shareholders accept such offers, but would reject any other offer because they believe executives making such offers have shirked and are now incapable of making profits for the firm. Along the equilibrium path, the conditional-choice probabilities solving* (5.6) *satisfy the market-clearing condition* (5.15) *and executives never shirk, so* $h' = h$.

A detailed description of strategies and beliefs on and off the equilibrium path is relegated to the Appendix with the proof. We construct a sequential equilibrium in which executives sequentially expropriate all the rent that can be extracted from one-period contracts. Along the equilibrium path, executives work every period, so $h = h'$ for all t. If the executive 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 executive who is off the equilibrium path. Having deviated from the equilibrium path by shirking once, it may be optimal for a executive to shirk at some future time, as equation (5.4) indicates. One possibility not accommodated by the construction of $B_t(h, h')$ is an executive 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 $w_{jk,t+1}(h', \pi)$, such as $w_{jk,t+1}(h, \pi)$? In the equilibrium we construct, shareholders interpret any deviation from $w_{jk,t+1}(h', \pi)$ as proof the executive has shirked initially and is therefore a liability to the firm because their marginal product is bounded above by $\underline{F}$. This assumption effectively truncates behavior off the equilibrium path because, given the shareholders' beliefs, it is a best response of an executive who has shirked to demand $w_{jk,t+1}(h', \pi)$ and follow the continuation path implied by $B_t(h, h')$.[19]

Because the incentive-compatibility condition is cheaper to enforce in the extended model, the total amount of surplus available for division between employee and employer is higher when shareholders employ executives with career concerns due to a lower risk premium, compared with employing executives whose human capital would evolve independently of their effort. Certainty-equivalent pay also differs across the two models because the investment value of job matches is different.

---

[19]We can make other assumptions and construct off-equilibrium-path behavior in which no manager truthfully reveals her type and where 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.

To illustrate, suppose career concerns are exogenously introduced into just one type of job match—say, because the technology of capital accumulation changes from (4.4) to (5.1). Some executives who previously spurned this job match in favor of another option, such as retirement, now take it, because the reduced risk premium absorbs a smaller portion of the marginal product, more than offsetting the negative effects of the idiosyncratic shock that previously deterred them. Thus, the primary effect of this change in technology is to increase the equilibrating value of the nonpecuniary loss for the marginal executive, thereby increasing the number of executives taking that particular job match rather than pursuing other options. There is, however, a secondary effect. With higher certainty-equivalent pay coming from the introduction of career concerns in this particular job match, which is chosen with strictly positive probability at some future date, the investment value of all job matches increases. In equilibrium, executives extract all the surplus, so a higher investment value ratchets up the nonpecuniary loss an executive is willing to incur when the new technology is introduced, reducing the probability of retirement.

Example. All the differences between the basic model and the extension trace back to reputation h', the choice probabilities for retirement morphing from $p_{0t}(h)$ into $p_{0t}(h, h')$, and the value of human capital from $\Gamma[(b_\tau + 1)/b_\tau]p_{0t}(h)^{1/b_\tau}$ into $\Gamma[(b_\tau + 1)/b_\tau]p_{0t}(h, h')^{1/b_\tau}$. The equilibrium contract in the extension is specified in terms of h', regardless of whether the executive is on the equilibrium path, but the variability in (5.16) ultimately arises from π, which is generated by h, not h'. Similarly, the fixed component of pay, $F_{jk}(h')$, depends on h', but nonpecuniary benefits depend on h.

The variable component of pay in the basic model, defined by (4.22) and (4.23), is computed without recourse to recursion. However, when there are career concerns, this component and the associated multiplier are computed recursively because they depend on future retirement probabilities, denoted by $p_{0,t+1}(h' + \Delta_{jk}, h' + \Delta_{jk})$ and $p_{0,t+1}(h' + \underline{\Delta}_{jk}, h' + \Delta_{jk})$. Thus (5.10) simplifies to

$$(5.16) \quad r^B_{jk,t+1}(h, h', \pi)$$

$$= \frac{b_{\tau+1}}{\rho} \ln\left[1 - \eta^B(h')\left\{g_{jk}(\pi|h)\right.\right.$$

$$\left.\left. - \left[\frac{\alpha_{jkt}(h')}{\beta_{jkt}(h')}\right]^{1/(b_\tau - 1)}\left[\frac{p_{0,t+1}(h' + \Delta_{jk}, h' + \Delta_{jk})}{p_{0,t+1}(h' + \underline{\Delta}_{jk}, h' + \Delta_{jk})}\right]^{1/(b_{\tau+1})}\right\}\right],$$

where $\eta^B(h')$, defined by (5.11), solves

$$(5.17) \quad \int\left\{\eta^{-1} + \left[\frac{\alpha_{jkt}(h')}{\beta_{jkt}(h')}\right]^{1/(b_\tau - 1)}\left[\frac{p_{0,t+1}(h' + \Delta_{jk}, h' + \Delta_{jk})}{p_{0,t+1}(h' + \underline{\Delta}_{jk}, h' + \Delta_{jk})}\right]^{1/(b_{\tau+1})}\right.$$

$$\left. - g_{jk}(\pi|h')\right\}^{-1} f_j(\pi)\, d\pi = 1.$$

We modify the analog to $W_{jkt}(h, b_\tau)$ defined for the basic model in (4.27) to account for situations in which $h \neq h'$:

$$(5.18) \quad W_{jkt}^B(h, h') = -\ln \alpha_{jkt}(h) - \frac{(b_\tau - 1)}{b_{\tau+1}} \left[\ln p_{0,t+1}(h + \Delta_{jk}, h' + \Delta_{jk}) \right.$$

$$+ b_{\tau+1} \ln \Gamma \left(1 + \frac{1}{b_{\tau+1}} \right)$$

$$\left. + \rho \{ F_{jk}(h') - E[r_{jk,t+1}^B(h, h', \pi)] \} \right].$$

To interpret (5.18), note that nonpecuniary losses from (j, k), captured by $\alpha_{jkt}(h)$, depend on actual human capital, the investment value of the job depends on the upper term in the square brackets, while certainty-equivalent pay, $F_{jk}(h') - E[r_{jk,t+1}^B(h, h', \pi)]$, depends on h' because the contract is based on reputation, and also h, because the draw from the firm's excess return distribution, and hence the risk premium, depends on the executive's true human capital. Substituting $W_{jkt}^B(h, h')$ for $W_{jkt}^A(h)$ in equation (4.28) yields the equilibrium conditional-choice probabilities for the extension.

The recursion to derive the equilibrium choice probabilities exploits the fact that in the period immediately preceding retirement, the investment value of a job is zero and there are no career concerns. Thus, in period $t = T - 1$, both models have the same solution:

(i) For each executive, set $t = T - 1$. Follow Steps (i) and (ii) in the equilibrium construction of the basic model, solving for $\eta^A(h)$ with equation (4.23) and then computing $w_{jk,T-1}^A(h, \pi) = w_{jk,T-1}^B(h, \pi)$, the sum of (4.20) and (4.22).

(ii) Compute $W_{jk,T-1}^B(h, h')$ and $p_{0,T-1}(h, h')$, a function of the primitives and $\eta(h)$ calculated in Step (i).

(iii) For each executive, set $t = T - 2$. Solve for $\eta^B(h)$ using equation (5.17) and use it to compute $r_{jk,T-2}^B(h, h', \pi)$ defined in equation (5.16) with $p_{0,T-1}(h, h')$ calculated in Step (ii).

(iv) Compute $W_{jk,T-2}^B(h, h')$ and $p_{0,T-2}(h, h')$ using the primitives of the model, $p_{0,T-1}(h, h')$ from Step (ii), and $r_{jk,T-2}^B(h, h', \pi)$ calculated in Step (iii).

(v) Recursively repeat Steps (iii) and (iv) for $T - 3, \ldots, t$.

6. IDENTIFICATION

Our data consist of matched panel data on firms and their executives in different time periods, consisting of job-match choices d_{jkt} over the firms j and ranks k, compensation w_{jkt} indexed by age t, executive demographic information and employment histories h_{it}, excess firm returns $\pi_{j\tau}$ indexed by calendar time τ, and bond prices b_τ, again indexed by calendar time. The basic

model is characterized by its preference and technology parameters. The preference parameters include the coefficient of risk aversion ρ, the disutility from working $\alpha_{jkt}(h)$, the disutility from shirking $\beta_{jkt}(h)$, and an idiosyncratic taste shock associated with each job match $G(\varepsilon_t)$. The technology parameters are the marginal product of work $F_{jk}(h)$, the probability density function of excess returns when every executive works $f_j(\pi)$, and the likelihood ratio $g_{jk}(\pi|h)$ that essentially defines the density $f_j(\pi)g_{jk}(\pi|h)$ when everybody except for one executive in rank k at firm j works, and the human-capital transition function $\overline{H}_{jk}(h)$. The extended model has one additional parameter, $\underline{H}_{jk}(h)$, the human-capital transition associated with shirking.

There are potentially two situations to investigate, depending on whether or not it is optimal to pay executives a constant wage. The latter arises when career concerns are so pronounced that the incentive-compatibility constraint is not binding, meaning (5.9) is satisfied, or when the cost-minimizing risk premium is so high relative to the net losses from shirking that executives are optimally paid to shirk. All the executives in our data receive compensation awards that depend on excess firm returns, leading us to focus on the former situation, when it is optimal for executives to work because the incentive-compatibility constraint is met with equality in equilibrium.

We assume the data are generated by an equilibrium in which every executive works. Thus, $F_{jk}(h)$ is identified from the conditional expectation of $w_{jk,t+1}$ on d_{jkt}, h_t, and t using the rent extraction condition equation (5.15); $f_j(\pi)$ is identified from observations on $\pi_{j\tau}$; while $\overline{H}_{jk}(h)$ is identified from the empirical distribution of h_{t+1} at $t+1$ conditional on d_{jkt} and h_{it} at t. Since Magnac and Thesmar (2002) have shown that the distribution of unobserved idiosyncractic shocks is not identified nonparametrically in dynamic discrete choice models, we assume $G(\varepsilon_t)$ is known. This leaves only ρ, $\alpha_{jkt}(h)$, $\beta_{jkt}(h)$, $g_{jk}(\pi|h)$, and $\underline{H}_{jk}(h)$ to identify. To explain our identification and estimation strategy, we notate $\beta_{jkt}(h)$, the shirking parameter, by $\beta_{jkt}^A(h) \equiv \beta_{jkt}(h)$ when the data are generated by the basic model, and by $\beta_{jkt}^B(h) \equiv \beta_{jkt}(h)$ when the data are generated by the extension.

Our approach to identification mimics the one we used to explain the model. First, we analyze identification of ρ and $\alpha_{jkt}(h)$, the preference parameters that generate the job-match choices observed in the data, when (i) the equilibrium choice is to work each period and (ii) the compensation schedule inducing the effort choice is given. This part of the model establishes identification in dynamic Roy models where there is human-capital accumulation. Then we turn to the agency issues arising from moral hazard in the basic model, showing how $\beta_{jkt}^A(h)$ and $g_{jk}(\pi|h)$ are identified. The last part deals with career concerns developed in the extension; we show ρ, $\alpha_{jkt}(h)$ and $g_{jk}(\pi|h)$ are identified in exactly the same manner as in the basic model and then focus on the remaining primitives, $\underline{H}_{jk}(h)$ and $\beta_{jkt}^B(h)$.

6.1. *Sorting Over Job Matches*

The conditional-choice probability (CCP) vector, $p_t(h)$, is identified by the conditional expectation of $d_{ijk\tau}$, on $(h_{i\tau}, t_{i\tau}, b_\tau)$. Hotz and Miller (1993, p. 501) showed that if $G(\varepsilon_t)$ is known, there exist known mappings $\varphi_{jk}[p_t(h), b_\tau] \equiv E[\exp(-\varepsilon^*_{jkt}/b_\tau)]$ that can be written as a known function of $p_t(h)$. Exponentiating equation (4.15) and then raising it to the power of $1/b_\tau$ yields[20]

$$(6.1) \qquad \alpha_{jkt}(h)^{1/b_\tau} \big\{ E_t[v_{jk,t+1}] A_{t+1}\big[\overline{H}_{jk}(h), b_{\tau+1}\big] \big\}^{1-1/b_\tau} = \exp\big[q_{jk}\big(p_t[h]\big)/b_\tau\big].$$

Now, substituting $\varphi_{jk}[p_t(h), b_\tau]$ for $E[\exp(-\varepsilon^*_{jkt}/b_\tau)]$ and the right-hand side of equation (6.1) for the left in equation (4.9) yields a representation for $A_t(h, b_\tau)$ in terms of the CCP vector $p_t(h)$:

$$(6.2) \qquad A_t(h, b_\tau) = p_{0t}(h)\varphi_0\big[p_t(h), b_\tau\big]$$
$$+ \sum_{j=1}^{J}\sum_{k=1}^{K} p_{jkt}(h)\varphi_{jk}\big[p_t(h), b_\tau\big]\exp\big[q_{jk}\big[p_t(h)\big]/b_\tau\big]$$
$$\equiv \mathcal{A}_t\big[p_t(h), b_\tau\big].$$

Substituting $\mathcal{A}_t[p_t(h), b_\tau]$ into (6.1), upon rearrangement we obtain

$$(6.3) \qquad \alpha_{jkt}(h_t) = \frac{\exp\big(q_{jk}\big[p_t(h)\big]\big)}{\mathcal{A}_{t+1}\big(p_{t+1}\big[\overline{H}_{jk}(h)\big], b_{\tau+1}\big)^{b_{\tau}-1}} E\big[e^{-\rho w_{jk,t+1}(h,\pi)/b_{\tau+1}}|h_t, j\big]^{1-b_\tau}.$$

Equation (6.3) is an equilibrium sorting condition characterized by $q_{jk}[p_t(h)]$ that, with reference to (4.15), accounts for certainty-equivalent pay, the value of human capital $\mathcal{A}_{t+1}(p_{t+1}[\overline{H}_{jk}(h)], b_{\tau+1})$, a shrinkage factor that raises the value of job matches, and a market-clearing condition captured by $q_{jk}[p_t(h)]$ that equilibrates the idiosyncratic individual taste disturbances.

The compensation schedules offered by different ranks and firms can be interpreted as choices over lotteries with different nonpecuniary characteristics. Thus, (6.3) can be used to identify both $\alpha_{jkt}(h_t)$ and ρ when exclusion restrictions exist that limit the dependence of the taste parameters on variables the help determine the contract. Define $z_{jkt}(h, b_\tau, b_{\tau+1})$ as

$$(6.4) \qquad z_{jkt}(h, b_\tau, b_{\tau+1})$$
$$\equiv \exp\big\{q_{jk}\big[p_t(h)\big]/(b_\tau - 1)\big\}/\mathcal{A}_{t+1}\big(p_{t+1}\big[\overline{H}_{jk}(h), b_{\tau+1}\big]\big);$$

---

[20]Henceforth, the dependence of $A_t(h)$ and $B_t(h, h')$ on b_τ is made explicit. In identification and estimation, b_τ plays a critical role; for example, in Gayle and Miller (2009b), the exclusion restriction on b_τ is one of the main sources of identification.

since $q_{jk}[p_t(h)]$ and $\mathcal{A}_{t+1}(p_{t+1}[\overline{H}_{jk}(h)], b_{\tau+1})$ are identified from (4.15) and (6.2) from the CCP vector $p_t(h)$ assuming $G(\varepsilon_t)$ is known, so is $z_{jkt}(h, b_\tau, b_{\tau+1})$. Identification of ρ and $\alpha_{jkt}(h)$ then follow 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 $\alpha_{jkt}(h)$. Note that 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 that affects her choices. This can be ascertained by checking for variation in the CCP vector. Hence, they qualify as valid instruments if they do not affect preferences as well. For example, human capital is a candidate for an exclusion restriction.[21] Similarly, b_τ is a valid instrument if, as we later 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 $f(\pi)$. Substituting $z_{jkt}(h, b_\tau, b_{\tau+1})$ into (6.3), rearranging to make $z_{jkt}(h, b_\tau, b_{\tau+1})$ the subject of the equation, and taking expectations conditional on x yields

$$(6.5) \qquad E\big[z_{jkt}(h, b_\tau, b_{\tau+1})|x\big]$$
$$= E\left[\alpha_{jkt}(h)^{1/(b_\tau-1)} \exp\left(\frac{-\rho w_{jk,t+1}(\pi, h)}{b_{\tau+1}}\right) \frac{f_j(\pi)}{f(\pi)}\Big|x\right].$$

Thus, ρ and $\alpha_{jkt}(h)$ are identified from the conditional expectations function (6.5), thus establishing identification of the basic model up to $G(\varepsilon_t)$.

6.2. *Moral Hazard in the Basic Model*

From the data, the equilibrium compensation schedule, $w_{jk,t+1}(h_t, \pi)$, is identified by the conditional expectation of individual observations of compensation on $(d_{jkt}, \pi_{j\tau}, h_t, t, b_\tau)$.[22] The finite-upper-bound property of $r_{jk,t+1}(h, \pi)$ in equation (4.24) and the optimal compensation schedule in equation (5.12) imply that compensation is bounded and the executive's maximum compensation is

$$(6.6) \qquad \lim_{\pi \to \infty} w_{jk,t+1}(h, \pi) = w_{jk,t+1}^A(h) + \overline{r}_{jk,t+1}(h) \equiv \overline{w}_{jk,t+1}(h).$$

Thus, $\overline{w}_{jk,t+1}(h_t)$ is identified by the maximum of $w_{jk,t+1}$ conditional on $(d_{jkt}, h_t, t, b_\tau)$.

Theorem 6.1 adapts Theorem 2.1 of Gayle and Miller (2015) to incorporate human-capital accumulation and sorting in our model. It demonstrates

---

[21] In this paper, we assume that (i) ρ is independent of an executive's human capital and (ii) the nonpecuniary cost of switching firms or ranks does not depend on some dimension of human-capital accumulation. In estimation, we use previous ranks as an instrument.

[22] In this way, we allow for observations on compensation to be measured with independent error.

that, in equilibrium, $g_{jk}(\pi|h)$ is a mapping of the identified functions $p_t(h)$, $w_{jk,t+1}(h, \pi)$, $\overline{w}_{jk,t+1}(h)$, and ρ. Intuitively, (6.7) shows $g_{jk}(\pi|h)$ is identified from the curvature of $w_{jkt+1}(h, \pi)$.

THEOREM 6.1: *In equilibrium,*

$$(6.7) \qquad g_{jk}(\pi|h) = \frac{e^{\rho \overline{w}_{jk,t+1}(h)/b_{\tau+1}} - e^{\rho w_{jk,t+1}(h,\pi)/b_{\tau+1}}}{e^{\rho \overline{w}_{jk,t+1}(h)/b_{\tau+1}} - E\left[e^{\rho w_{jk,t+1}(h,\pi)/b_{\tau+1}}|h, j\right]}.$$

Having identified the working preference parameter $\alpha_{jkt}(h)$ from (6.3) and the likelihood ratio $g_{jk}(\pi|h)$ from (6.7), the shirking preference parameter $\beta_{jkt}^A(h)$ is now identified from the incentive-compatibility constraint (4.21), which holds with equality when compensation varies with π:

$$(6.8) \qquad \beta_{jkt}^A(h) = \frac{\exp(q_{jk}[p_t(h)])}{\mathcal{A}_{t+1}(p_{t+1}[\overline{H}_{jk}(h)], b_{\tau+1})^{b_{\tau}-1}}$$

$$\times E\left[e^{\rho w_{jk,t+1}(h,\pi)/b_{\tau+1}} g_{jk}(\pi|h_t)|h, j\right]^{1-b_\tau}.$$

6.3. *Career Concerns in the Extended Model*

It is instructive to highlight the similarities and differences between the basic and extended models in identification by defining, for the extended model, a virtual shirking parameter as

$$(6.9) \qquad \beta_{jkt}^*(h) \equiv \left(1 - 1\{\text{private}\}\right)\beta_{jkt}^A(h)$$

$$+ 1\{\text{private}\}\beta_{jkt}^B(h)\left\{\frac{B_{t+1}\left[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_\tau\right]}{B_{t+1}\left[\overline{H}_{jk}(h), \overline{H}_{jk}(h), b_\tau\right]}\right\}^{(b_\tau-1)},$$

where $1\{\text{private}\}$ denotes an indicator function taking a value of 1 if human capital is private (as in the extension), and zero if not (the basic model). Substituting $\beta_{jkt}^*(h)$ for $\beta_{jkt}^A(h)$ and $B_t(h, h)$ for $A_t(h)$ throughout the previous subsection demonstrates, by exactly the same logic, that ρ, $g_{jk}(\pi|h)$, $\alpha_{jkt}(h)$, and $\beta_{jkt}^*(h)$ are identified up to $G(\varepsilon_t)$ in the extended model. We exploit this fact in estimation.

The only remaining question is whether $\underline{H}_{jk}(h)$ and $\beta_{jkt}^B(h)$ can be identified from the virtual shirking parameter $\beta_{jkt}^*(h)$, which is itself identified but is clearly not a primitive. Imagine the data are generated by the extended model and substitute the virtual parameter $\beta_{jkt}^*(h)$ defined in equation (6.9), the incentive-compatibility constraint for the extended model, into equation (5.8). This gives the incentive-compatibility constraint for the basic model, (4.21), with $\beta_{jkt}^*(h)$ replacing $\beta_{jkt}^A(h)$. Neither (4.14) nor (4.15) depend on $\beta_{jkt}^A(h)$ or the information structure because the executive works in the equilibrium of

both models. Replacing $\beta_{jkt}^*(h)$ with $\beta_{jkt}^A(h)$ and $B_t(h, h, b_\tau)$ for $A_t(h, b_\tau)$, it is straightforward to check that, with these changes, the solution for the extension solves the optimal contract problem for the basic model, which is given by equations (4.22), (4.23), and (4.25). This argument suggests that data generated by models with private information inducing career concerns are difficult to distinguish from data generated by models that do not have career concerns. Specifically, $\beta_{jkt}^*(h)$ indexes observationally equivalent models that differ only in their specification of $\underline{H}_{jk}(h)$ and $\beta_{jkt}(h)$.[23]

THEOREM 6.2: *Let Θ denote the class of models under consideration, consisting of elements*

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

Suppose $b_\tau = b$ for all τ and $(w_{ijk}, d_{ijk}, \pi_j, h_i, t_i)$ is generated by $\widetilde{\theta}$. For every $\widehat{\rho} > 0$ and all proper probability distribution functions $\widehat{G}(\varepsilon)$ defined on the same support as $\widetilde{G}(\varepsilon)$, there exists a unique $\widehat{\theta}$ solving equations (5.6), (5.12), (5.15), (6.3), (6.7), and (6.9) that is observationally equivalent to $\widetilde{\theta}$.

Imposing exclusion restrictions on preferences or the technology of human capital accumulation does, however, distinguish the basic model from the extension. To illustrate, consider the following three restrictions:

(i) Suppose $\beta_{jkt}^B(h)$ does not depend on the executive's age, meaning $\beta_{jkt}^B(h) = \beta_{jk}^B(h)$ for all t, and there is a maximum retirement age T. Recalling that, at age T, there is no investment value from human capital or career concerns, then

$$B_T\big[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau(T)}\big] = B_T\big[\overline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau(T)}\big] = 1.$$

In this case, the shirking parameter is identified from (6.9) and after substituting $B_t(h, h, b_\tau)$ for $A_t(h, b_\tau)$ equation (6.8) as

$$\begin{aligned}
\beta_{jk}^B(h) &= \beta_{jk,T-1}^*(h) \\
&= \exp\big(q_{jk}\big[p_{T-1}(h)\big]\big)E\big[e^{\rho w_{jkT}(h,\pi)/b_{\tau(T)}}g_{jk}(\pi|h)|h, j, k\big]^{1-b_\tau}.
\end{aligned}$$

Intuitively, the basic and extended models have exactly the same predictions if the executive is of age $T - 1$ and has not shirked before, so the distinction between $\beta_{jk}^A(h)$ and $\beta_{jk}^B(h)$ is moot. Having identified $\beta_{jk}^B(h)$, the continuation

---

[23]We formally state this result for the case where bond prices are constant over time. However, a more general result holds when b_τ varies over time, providing the parameters are also permitted to vary with calendar time.

value associated with shirking the first time is also identified from (6.9) for all $t \leq T - 2$ as

$$B_{t+1}\big[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}\big]$$
$$= \big[\beta_{jkt}^*(h)/\beta_{jk}^B(h)\big]^{1/(b_\tau - 1)} B_{t+1}\big[\overline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}\big].$$

In this way, the importance of career concerns at younger ages can be compared by showing how the identified continuation value of shirking for the first time varies over the lifecycle. Note that the basic model does have empirical content against the extension that nests it: Under the null hypothesis of no career concerns, $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}] = B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}]$.

(ii) Similarly, suppose $\beta_{jkt}(h)$ is independent of aggregate shocks in the economy, more specifically, bond prices b_τ. In this case, given (j, k, t, h) and two bond prices $b_{\tau'} \neq b_{\tau''}$, equation (6.9) yields two equations in three unknowns—namely, $\beta_{jkt}^B(h)$, $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau'}]$, and $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau''}]$. Relative to the normalization $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau'}] = 1$, the other two parameters are identified.

(iii) If $\underline{H}_{jk}(h)$ is known, then $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}]$ can be numerically calculated in recursive fashion starting from $t = T$ using equation (5.4). The parameter $\beta_{jkt}^B(h)$ now follows from (6.9).

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 $G(\varepsilon)$ in this manner are evident from the formulas for $A_t(h, b_\tau)$ and $B_t(h, h', b_\tau)$ in equation (5.7) and the expression for $q_{jk}[p_t(h)]$ in equation (4.18). On and off the equilibrium path, the human-capital transition functions are deterministic; see equations (4.4) and (5.2) for $\overline{H}_{jk}(h)$ and $\underline{H}_{jk}(h)$, respectively. We use a four-step procedure, which directly follows the approach of our identification strategy, to estimate and test our models:

(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 $\alpha_{jkt}(h)$ from sample moments formed from population moments implied by (6.5), 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 (6.7) and (6.8) to estimate $g_{jk}(\pi|h)$ and $\beta_{jkt}^A(h)$ 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 $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}]$ recursively, assuming that $\beta_{jkt}(h)$ is independent of b_τ and that $\underline{H}_{jk}(h)$ is known, and test the implied overidentifying restrictions.

An alternative estimation strategy is to exploit the equilibrium computation algorithm outlined at the end of Section 5.6. It involves computing a nested fixed-point algorithm to calculate $\eta(h, b_\tau)$ and $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}]$ 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, but in practical applications is also 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 $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}]$ are imposed only when we estimate the effects of career concerns.

Step 1. The state space for the dynamic system is the Cartesian product of the executive's age, t, and personal background, $h_t \in \{1, \ldots, H\}$, at the beginning of each period, as well as a vector that includes her employer firm during the last period, $j_{t-1} \in \{1, \ldots, 36\}$, management rank last period, $k_{t-1} \in \{0, 1, \ldots, 5\}$, 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's rule: Given background h, the (joint) probability, $p_{jkt}(h_t)$, is the product of the probability of choosing the jth firm conditional on choosing the kth 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 the conditional expectation function (6.5):

$$(7.1) \quad E\big[z_{jkt}(h, b_\tau, b_{\tau+1})x\big]$$
$$= E\bigg[\alpha_{jkt}(h)^{1/(b_\tau-1)} \exp\bigg(\frac{-\rho w_{jk,t+1}(\pi, h)}{b_{\tau+1}}\bigg)\frac{f_j(\pi)}{f(\pi)}x\bigg].$$

Upon substituting (4.15) and (5.7) into (6.4), $z_{jkt}(h)$ simplifies to

$$(7.2) \quad z_{jkt}(h) \equiv \Gamma\bigg(\frac{b_{\tau+1}+1}{b_{\tau+1}}\bigg)^{-1} p_{0,t+1}\big(\overline{H}_{jk}(h)\big)^{-1/b_{\tau+1}}\bigg[\frac{p_{0t}(h)}{p_{jkt}(h)}\bigg]^{1/(b_\tau-1)}.$$

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 (7.2). Sample

analogs for the CCP vector, the compensation schedule, and conditional and unconditional densities of the abnormal return from Step 1 are substituted into equation (7.1). Consistent estimates of ρ and $\alpha_{jkt}(h)$ are then obtained from the approximate sample moments along with (consistent 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, PhD, 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 assumption of absolute risk aversion 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 (7.1) 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 nonrejected restrictions on the results and re-estimate 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. This reduces the number of parameters to $(16 \times 36 + 5 \times 16 + 1) = 657$. We obtain similar results from both the restricted and unrestricted versions; hence, only the restricted version is reported.

Step 3. We form $\widehat{w}(h, \pi)$, 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, $E_t[\exp(-\widehat{\rho}\widehat{w}(h, \pi)/b_{\tau+1}]$, by integration using the nonparametrically estimated density of π for a given j, from Step 1, and compute $\overline{w}_{jk,t+1}(h)$ using the maximum $\widehat{w}(h, \pi)$ for each value of (j, k, t, h). Finally, our estimate of $g_{jk}(\pi|h)$ is obtained by substituting our estimates of $\overline{w}_{jk,t+1}(h)$, ρ and $E_t[v_{jk,t+1}(\rho, \pi)]$ into equation (6.7). The sample analog of the CCP vector, $\widehat{w}(h, \pi)$, and the estimates of $g_{jk}(\pi|h)$ are now substituted into a sample average of equation (6.8) to obtain an estimate for $\beta_{jkt}^{A}(h)$, which is $\beta_{jkt}^{*}(h)$ in the extended model

Step 4. In (6.8), we replace $\beta_{jkt}^{A}(h)$ with $\beta_{jkt}^{*}(h)$, and substitute (5.7), the formula for $B_t(h, h', b_{\tau})$ under Type I extreme values, for $A_t(h, b_{\tau})$ to obtain an expression for $\beta_{jkt}^{*}(h)$ in the extended model. The resulting expression replaces $\beta_{jkt}^{*}(h)$ in equation (6.9), and using the log-odds form of $q_{jk}[p_t(h)]$, we

rearrange the equation to obtain

$$(7.3) \qquad \beta_{jkt}^B(h) \equiv \frac{p_{0t}(h, h)}{p_{jkt}(h, h)} B_{t+1}\big[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}\big]^{1-b_\tau}$$

$$\times \left\{ \frac{E_t[v_{jk,t+1}] - \overline{v}_{jk,t+1}^{-1}}{1 - \overline{v}_{jk,t+1} E_t\big[v_{jk,t+1}^{-1}\big]} \right\}^{1-b_\tau}$$

for all (j, k, t, h). Estimates of $\beta_{jkt}^B(h)$ and $B_t(h, h', b_\tau)$ are obtained recursively. Noting that $B_{T+1}(h, h', b_{\tau+1}) \equiv 1$ and substituting our estimated risk-aversion parameter and conditional-choice probabilities into equation (7.3) yields $\beta_{jkT}^B(h)$. Substituting $\beta_{jkT}^B(h)$ into equation (5.5) yields $V_{jkT}(h, h', b_\tau)$ and hence $B_T(h, h', b_\tau)$, using equation (5.7). More generally, given $B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h), b_{\tau+1}]$, $\beta_{jkt}^B(h)$ is obtained from equation (7.3); hence, estimates of $V_{jkt}(h, h', b_\tau)$ and $B_t(h, h', b_\tau)$ are produced from equations (5.5) and (5.7), respectively.

8. PAY DIFFERENTIALS IN THE EXECUTIVE LABOR MARKET

This section presents our estimates of different components comprising the sources of pay differentials across ranks and firms in the executive labor market. Expected compensation is the sum of certainty-equivalent pay and a risk premium. First, we report on the estimated risk premium and the coefficient of risk aversion, from which it is derived. Then we decompose certainty-equivalent pay into three additive components, arising from compensating variation in utility, due to permanent and idiosyncratic job and executive characteristics, plus the investment value of the job in developing human capital. All the results in this section, plus the results on the span of control discussed in the next section, can be interpreted within the context of either model. As foreshadowed in our analysis of identification, only the results on career concerns explicitly draw on the extension. Following the precedent set in the previous sections, we revert to the notation for the extended model only when we discuss our results on career concerns.

8.1. *The Risk Premium*

The risk premium is a compensating differential to risk-averse executives for bearing risk in the form of firm-denominated securities. In our model, it is measured by the difference between expected compensation and certainty-equivalent pay defined in equation (4.19). From (4.26), expected compensation is the expected value of the executive's marginal product:

$$(8.1) \qquad \Delta_{jkt}^r(h) \equiv E_t\big[r_{jk,t+1}(h, \pi)\big] = F_{jk}(h) - w_{jk,t+1}^A(h).$$

In our model, $\Delta_{jkt}^r(h)$ measures the cost of agency. Note that since the executive works in equilibrium, even in the extended model $\Delta_{jkt}^r(h)$ does not directly

depend on $\underline{H}_{jk}(h)$, $\beta_{jkt}(h)$, or $B_{t+1}(\cdot)$, terms that characterize what occurs to human capital, utility, and the continuation value if the executive shirks. Thus, $\Delta^r_{jkt}(h)$ is computed the same way in both models.

Variation in Compensation Across Firm Size and Rank. Figure 3 presents the components of expected pay decomposition by firm size and rank, evaluated at
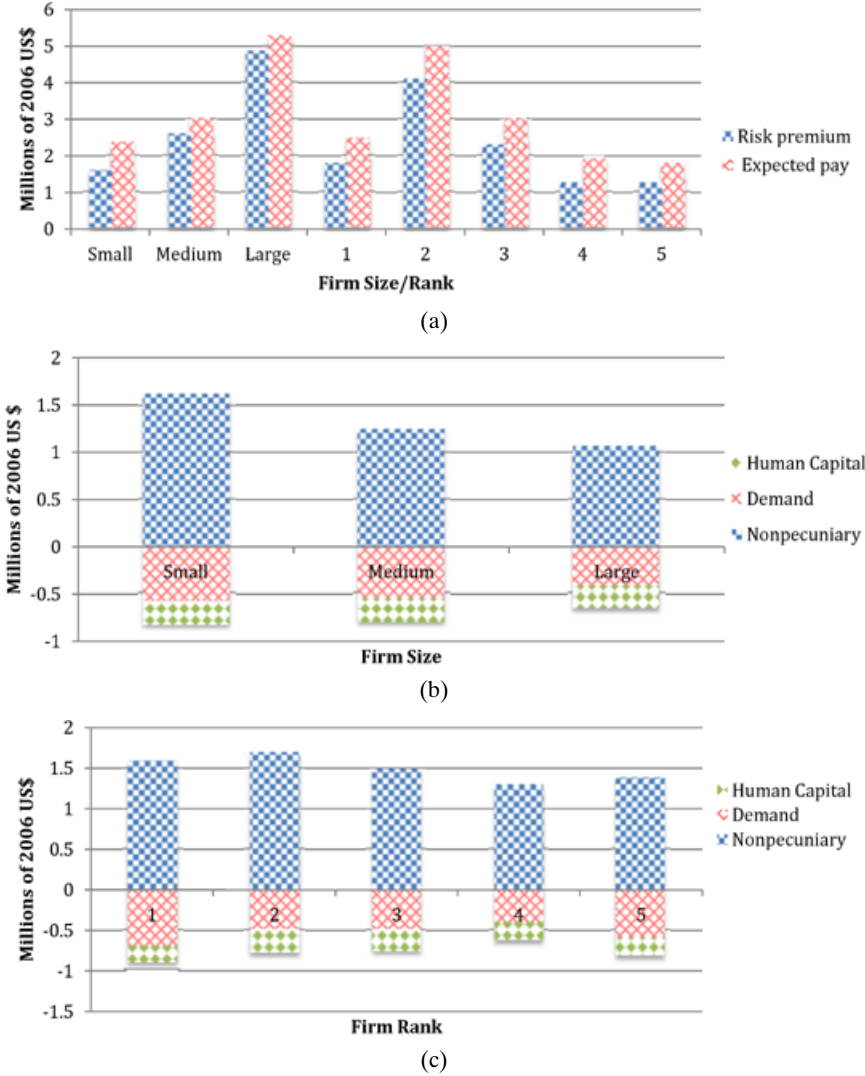

FIGURE 3.—Rank and firm-size pay decomposition. (a) Risk premium, (b) decomposition of certainty-equivalent pay, (c) decomposition of certainty-equivalent pay. *Note*: The certainty equivalent is the sum of human capital, demand, and nonpecuniary compensating differentials.

the median bond price for the sample, and averaged over the other characteristics. The risk premium accounts for most of the variation in pay across ranks and firms of different sizes. Figure 3(a) shows that expected compensation is greater in large firms and in higher ranks (up to Rank 2) because the risk premium has the same pattern. Indeed, the magnitude and differentials in the risk premium dominate expected pay so much that the difference between them, certainty-equivalent pay, falls with firm size.

With regard to firm size, on average an executive in a small firm receives $1.6 million in risk premium (56 percent of expected compensation), $2.8 million in a medium-size firm (85 percent of expected compensation), and $4.8 million in a large firm (90 percent of expected compensation).

Table III reports more detail on our estimates of $\Delta_{jkt}^r(h)$. At Ranks 4 and 5, $\Delta_{jkt}^r(h)$ 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, respectively. Roughly 82 percent of the compensation of a Rank 2 executive, versus 72 percent for Rank 1, 76 percent for Rank 3, 65 percent for Rank 4, and 69 percent for Rank 5, is due to the risk premium. The service sector pays a higher risk premium than the other two, a factor that helps close the gap between the considerably higher levels of average compensation paid in that sector and those reported in Table III.

Coefficient of Risk Aversion. The coefficient of risk aversion plays a vital role in estimating the risk premium. Not only does that parameter directly affect the optimal contract in the theory, but it also plays a critical role in identification and in estimation. For these reasons, we examined the robustness of the estimates of the coefficient of risk aversion and compared them with previous published estimates. We initially specified the risk-aversion parameter as a function of gender and firm size, but at the 1 percent level could not reject the null hypothesis that male and female executives and executives sorting into firms of different size and sector 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, an 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 1 million dollars.

Gayle and Miller (2009b) 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 those 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, and Rutström (2008), Dohmen, Falk, Huffman, and Sunde (2010)). This discrepancy is plausible because those with greater risk tolerance are more likely to accept jobs that entail greater risk, and executive compensation is much more volatile than wages in most occupations.

TABLE III

TABLE III

RISK PREMIUM FROM AGENCY[a]

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

[a]Compensation is measured in millions of 2006 US$; standard errors are listed 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) entering one of the ranks in our sample.

8.2. *Certainty-Equivalent Pay*

From equation (4.19), certainty-equivalent pay factors into three additive components:

$$(8.2) \qquad w^*_{jkt(\tau)+1}(h) = \Delta^{\alpha}_{jkt}(h) + \Delta^{A}_{jkt}(h) - \Delta^{q}_{jkt}(h),$$

where $\Delta^{\alpha}_{jkt}(h)$ is a compensating differential due to the nonpecuniary utility gain or loss incurred by working in (j, k) relative to the outside option, $\Delta^{A}_{jkt}(h)$ is the investment value of (j, k) from accumulating human capital, and $\Delta^{q}_{jkt}(h)$ is a compensating differential that induces selection on the unobserved idiosyncratic preference shocks:

$$\Delta^{\alpha}_{jkt}(h) \equiv \left[\rho(b_{\tau} - 1)\right]^{-1} b_{\tau+1} \ln \alpha_{jkt}(h),$$

$$\Delta^{A}_{jkt}(h) \equiv \rho^{-1} b_{\tau+1} \ln A_{t+1}\left[\overline{H}_{jk}(h), b_{\tau+1}\right],$$

$$\Delta^{q}_{jkt}(h) \equiv \left[\rho(b_{\tau} - 1)\right]^{-1} b_{\tau+1} q_{jk}\left[p_t(h, h)\right].$$

Note that both $\Delta^{\alpha}_{jkt}(h)$ and $\Delta^{q}_{jkt}(h)$ have static analogs; $q_{jk}[p_t(h, h)]$ 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 at market-clearing pay. Inframarginal executives making the same (j, k) choice, who have higher values of $\varepsilon_{jkt} - \varepsilon_{0t}$ but are otherwise identical to the marginal executive, garner producer surplus in equilibrium. Following the literature, we call $q_{jk}[p_t(h, b_{\tau})]$ the demand effect. The only structural parameters needed to estimate the certainty equivalent and its decomposition that 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.

Firm Size and Rank. Our discussion of Figure 3(a) foreshadowed the most striking result of Figure 3(b): Certainty-equivalent pay decreases with firm size. To interpret the histograms in Figures 3(b) and 3(c), the human capital and demand pieces below zero reduce certainty-equivalent pay and, therefore, should be subtracted from the nonpecuniary pieces above zero to obtain total certainty-equivalent pay. Thus, average certainty-equivalent pay 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 greater demand effect; higher compensating differentials are paid to attract 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.

In addition to the negative relationship between firm size and nonpecuniary benefit from working, the distribution of ranks across firm size, as demon-

strated in Figure 1(b), contributes to the difference between the average compensation and the certainty equivalent by firm size. Figure 3(c) shows that certainty-equivalent pay 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. Thus, Rank 3 executives have higher certainty-equivalent pay, $730,000, than Rank 1 executives, but Rank 1 executives have slightly higher certainty-equivalent pay than Rank 4 executives, $660,000. This ordering follows that of the average total compensation by executive rank reported in Table I, which ranges from $1,269,000 (for Rank 5) to $4,794,000 (for Rank 2), and the compression of certainty-equivalent pay mirrors the outsized role of the risk premium.

The demand effect is lowest for Rank 5, highest for Rank 4, and then declines through to Rank 1. Similarly, Table S-VIII in the Supplemental Material shows that, compared to small firms, medium sized firms pay an additional $32,000 and large firms an extra $170,000 to attract executives that experience greater idiosyncratic disutility from employment. There is a trade-off between higher fixed pay to executives with career concerns and higher risk premiums to those who lack them. In equilibrium, large firms offer greater certainty-equivalent pay to attract the type of executives who can be induced to work for less variable pay with a lower compensating risk premium, rather than paying an even higher risk premium to attract those types of executives who require more variable pay to meet the incentive-compatibility constraint.

Investment Value. The lifecycle theory of human capital predicts that as executives age, human-capital investment becomes less important. In support of the theory, Table I 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 3(c) also shows that executives give up more compensation for human-capital investment as they progress through the ranks until they reach Rank 1, where the trend falls off. In our model, the investment value of human capital is inversely related to the probability of exit. This pattern is reflected in the exit probability, which from Table II is lowest in Rank 2, 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, in which everybody retires at the same age, overpredict human-capital investment in the lower ranks and underpredict investment in higher ranks. As a fraction of the 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. This new finding on human-capital investment pairs with another: Even late in the career cycle, variety in job experience adds to human capital, and the value of human capital is higher in large firms. Our findings suggest that in the top

ranks of executive management, general human capital might increase from gaining management experience in different environments.[24]

The Role of Education, Tenure, and Rank Composition. We also investigated several other factors that might help explain the pay differential in executive compensation between small and large firms. In our framework, expected compensation is the executive's marginal product. Consequently, we interpret executives with a PhD, who receive an average expected compensation of $3.0 million, as being more productive on average than those with an MBA, $2.7 million, and those without either, $2.8 million. An executive with a PhD receives a higher risk premium, $2.3 million, than one with an MBA, $2.1 million, but an executive with a PhD receives a smaller fraction of expected compensation in the form of a risk premium than an executive with an MBA, 76 percent versus 78 percent. 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. Because larger firms have more executives with MBA degrees and fewer tenured executives, both findings overstate the firm-size pay premium in the raw data. Finally, 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. In summary, these other factors, albeit significant, are too small in magnitude to rationalize the pay differential in executive compensation between firms of different sizes.

9. AGENCY COSTS AND FIRM SIZE

Our empirical findings show the risk premium largely explains why mean executive compensation in large firms is higher than in small firms, in the process revealing the surprising result that certainty-equivalent pay is higher in small firms than in large ones. The premium, however, is not itself a primitive of the model, but given an equilibrium, a mapping from the technology and preference parameters. In this section, we investigate why the risk premium is so much higher in large firms, by turning to the agency issues that produce it. We report on two measures of the relative contribution of different sources to the overall agency cost to show how they vary with firm size. The first is the gross loss to shareholders from an executive unilaterally deviating from the equilibrium by shirking; we interpret this measure as her span of control. The second is the benefit an executive would extract from shirking for one period, leaving aside its effect on compensation she receives that period.

---

[24]Table S-IX in the Supplemental Material, which shows that the value of human capital increases with turnover by roughly $13,000, supports this hypothesis.

9.1. *Span of Control*

First used in Margiotta and Miller (2000) and Gayle and Miller (2009a, 2009b, 2015), the difference in the expected abnormal return π to firm j, generated by every executive working, versus everyone except a single rank k executive working, is given by

$$(9.1) \qquad \Delta^g_{jkt}(h) \equiv E_t\big[\pi\big(1 - g_{jk}(\pi|h)\big)\big].$$

It is measured by the gross output loss to the firm from switching from $f_j(\pi)$, the density of abnormal returns obtained from everyone working, to its shirking counterpart, $f_j(\pi)g_{jk}(\pi|h)$. To obtain the gross loss to shareholders, we first estimate the likelihood ratio of working versus shirking, $g_{jk}(\pi, h)$, a part of the production technology defined in Section 4.5.

The likelihood ratio $g_{jk}(\pi|h)$ is not just a technology parameter. It also shapes the signal π used by shareholders to enforce incentive compatibility (equation (4.21)) through the variable component of equilibrium compensation (equation (4.25)) and, hence, can be identified from the curvature with respect to abnormal returns (equation (6.7)). Theorem 6.1 establishes that identifying $g_{jk}(\pi|h)$ and $\Delta^g_{jkt}(h)$ does not depend on aggregate conditions, bond prices b_τ, or the aggregate return π_τ. Our estimates of $g_{jk}(\pi|h)$ and $\Delta^g_{jkt}(h)$ are robust to the specification of $\underline{H}_{jk}(h)$; they are not affected by whether or not human capital evolves with effort.

Figure 4 shows (with Table IV providing greater detail) that small consumer-sector firms lose 33.6 percent of their equity value when a Rank 5 executive shirks, but large firms lose much less, 8 percent. This contrasts with a finding by 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. 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 sized 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 from Table IV is that $\Delta^g_{jkt}(h)$ monotonically declines in rank. 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 overturn the 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 ex-
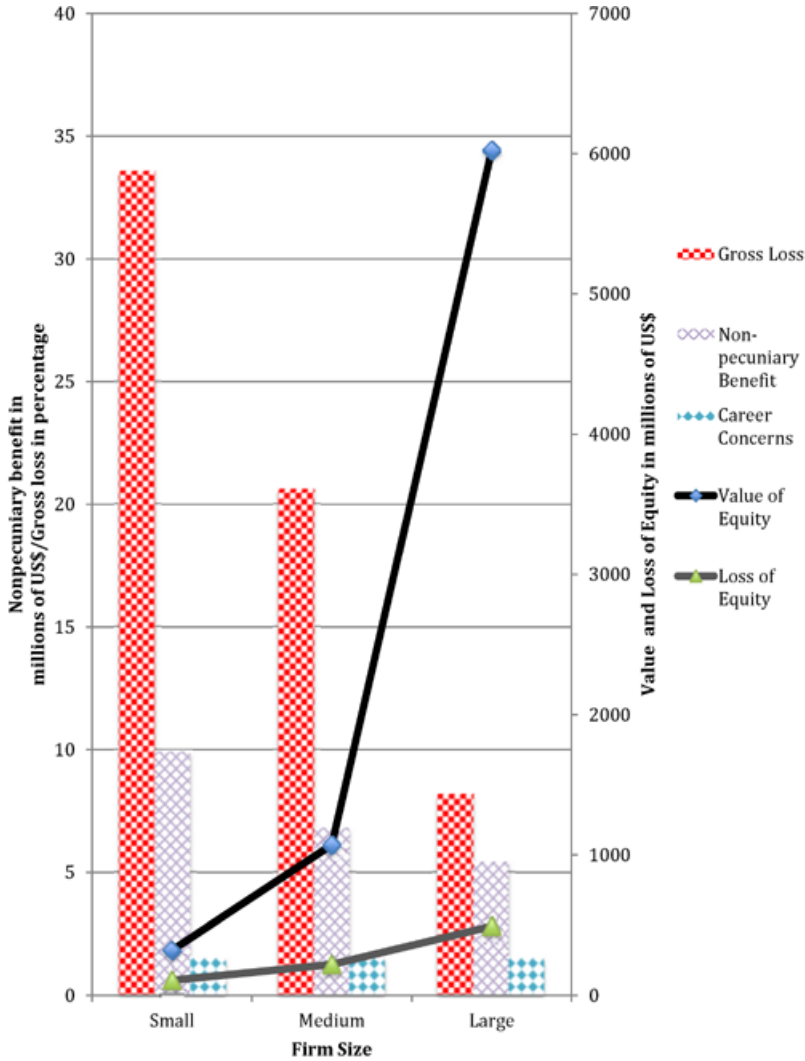
FIGURE 4.—Agency cost decomposition. Sources of agency cost by firm size. *Note*: Gross loss is the percentage of the firm value lost if an executive shirks instead of working. Loss of equity is the firm value lost if an executive shirks instead of working. Nonpecuniary benefit is the value to an executive of shirking relative to working. Career concerns measures the extent to which career concerns ameliorate the agency problem.

cess return of the firm is a better signal of their effort. The losses are greatest in the service sector and least in the primary sector.

Interpreted as a measure of signal quality, the flatter the $g_{jk}(\pi|h)$, the less information it conveys. Figures 5(a) and 5(b) show that the strength of the

TABLE IV
Gross Loss to Shareholders From Not Providing Executive Incentives[a]

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

[a]Gross loss to shareholders measured as a percentage of equity value; standard errors are listed 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) entering one of the ranks in our sample.

signal weakens with both firm size and executive rank. Thus, a chairman of a large firm receives more variable compensation and is consequently paid a larger risk premium, because she has less control and thus transmits a weaker signal of her effort than a lower-ranked executive employed in a small firm who is closer to operations and therefore more directly accountable, and hence transmits a stronger signal about effort, thus reducing the cost of achieving incentive compatibility. For this reason, the chairman of a large firm receives a larger risk premium than a lower-ranked executive in a small firm.
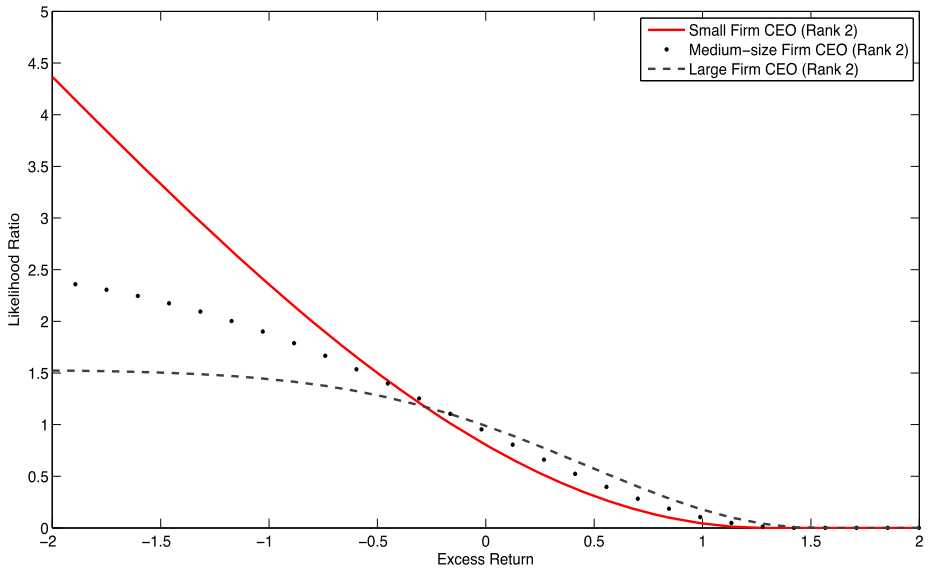
(a)
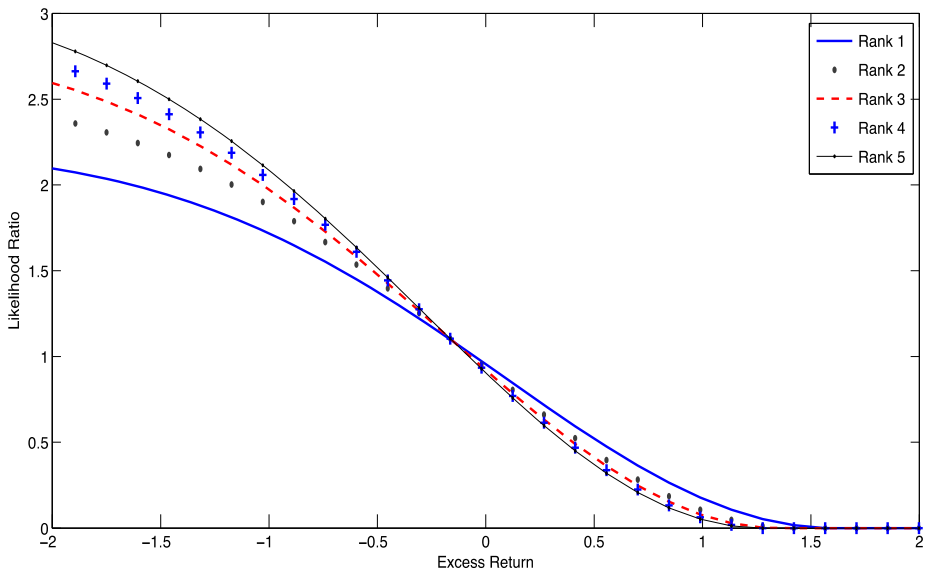
(b)

FIGURE 5.—Likelihood ratio. (a)Likelihood ratio by firm size for a CEO, (b) Likelihood ratio by rank 1 for a medium size firm. *Note*: Likelihood ratios are calculated at the average of the sample for the appropriate groups.

9.2. *Career Concerns*

Aside from issues related to current compensation, executives weigh two other factors when making the effort decision in the extended model. The first, denoted by $\Delta_{jkt}^{\beta}(h)$, is the compensating differential for current utility when the executive weighs shirking against working, the value an executive places on shirking over working:

$$\Delta_{jkt}^{\beta}(h) \equiv \left[\rho(b_\tau - 1)\right]^{-1} b_{\tau+1} \left[\ln \alpha_{jkt}(h) - \ln \beta_{jkt}(h)\right].$$

As such, $\Delta_{jkt}^{\beta}(h)$ measures the misalignment of preferences from the executive's perspective and also applies to models of moral hazard without career concerns, including our basic model in Section 4.[25] Substituting $\beta_{jkt}(h)$ for $\alpha_{jkt}(h)$ in (4.19), the certainty-equivalent pay equation, highlights the monetized value of shirking over working. The second, $\Delta_{jkt}^{B}(h)$, measures the difference in the continuation value from working in the current-period t versus shirking. It measures by how much career concerns ameliorate the agency problem:

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

This effect is best seen in (5.8), the incentive compatibility constraint for the extended model (which holds with equality in equilibrium). Hence, the net benefit to the executive from shirking can be expressed as a sum:

$$\Delta_{jkt}^{\beta^*}(h) \equiv \Delta_{jkt}^{\beta}(h) + \Delta_{jkt}^{B}(h).$$

Our estimates for $\Delta_{jkt}^{\beta^*}(h)$ apply equally to both models; from equation (6.9) the estimates for $\Delta_{jkt}^{\beta^*}(h)$ are valid for $\Delta_{jkt}^{\beta}(h)$ obtained for the basic model. More generally, all the preceding empirical results and those for $\Delta_{jkt}^{\beta^*}(h)$ apply without qualification to both the basic model and its extension. The net benefit from shirking, $\Delta_{jkt}^{\beta^*}(h)$, is identified from data on choice probabilities, the compensation schedule, the abnormal return distribution, the risk-aversion parameter, and the likelihood ratio. (Replacing $\beta_{jkt}^{A}(h_t)$ with $\Delta_{jkt}^{\beta^*}(h)$ equation (6.8) refers.) Therefore, $\Delta_{jkt}^{\beta^*}(h)$ is identified without appealing to the functional form assumptions about $\underline{H}_{jk}(h)$, which are the basis for career concerns in our model, or exclusion restrictions. However, the observational equivalence between the basic and extended models established in Theorem 6.2 implies

---

[25]See Margiotta and Miller (2000) and Gayle and Miller (2009a, 2009b, 2015) for estimates of $\Delta_{jkt}^{\beta}(h)$ in models of executive compensation where career concerns are absent.

that $\Delta_{jkt}^{B}(h)$ and $\Delta_{jkt}^{\beta}(h)$ cannot be separately identified from $\Delta_{jkt}^{\beta*}(h)$ without functional form assumptions on $\underline{H}_{jk}(h)$, such as those given in equation (5.2), along with an exclusion restriction that $\beta_{jkt}(h)$ is independent of bond prices.

Table V reports our estimates of $\Delta_{jkt}^{\beta*}(h)$ and Table VI reports our estimates of $\Delta_{jkt}^{B}(h)$.[26] Figure 4 and Table V show that $\Delta_{jkt}^{\beta}(h)$ declines with firm size, by $3.1 and $4.5 million for medium sized and large firms, respectively, and differs across sectors, $3.8 million higher in the service sector than the consumer sector, and $2.6 million lower in the primary sector. It is also evident from Figure 4 and Table VI that career concerns, $\Delta_{jkt}^{B}(h)$, do not vary with firm size.

From Table V, $\Delta_{jkt}^{\beta*}(h)$, the net benefit from shirking, is about $10 million for a 50-year-old Rank 5 executive in a small firm in the consumer sector and increasing in rank. While economically significant, the differential across ranks is not statistically significant. Our estimates of $\Delta_{jkt}^{B}(h)$ in Table VI show that there are significant career concerns at all ranks, reducing the differential for working versus shirking by between 15 and 22 percent. 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, with 19 percent of the gross compensating differential from working versus shirking, equal to Rank 4, and higher than the 17 percent for Rank 5. Career concerns decline with age, tenure, executive experience, and experience gained from working in different firms. In equilibrium, variable pay is used more extensively as a tool to motivate these older, more experienced executives. Because the resulting risk premium is a deadweight loss stemming from a second-best solution, as Figure 3(c) shows, demand for such executives is less than for their younger, less experienced, lower-ranked counterparts, who are willing to work, as opposed to shirk, for less variable pay.

10. 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 separately estimate the share of the compensation due to agency and the certainty-equivalent wage

---

[26]By subtracting the estimates in Table VI from those in Table V, we obtain the gross compensating differential for working versus shirking under perfect monitoring. The estimates in Table V are mainly of a higher order of magnitude than those in Table VI. Therefore, the qualitative patterns of the gross compensating differential for diligence versus shirking are similar to the net differential.

TABLE V

THE NET COMPENSATING DIFFERENTIALS TO EXECUTIVES FROM WORKING VERSUS SHIRKING[a]

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

[a]Compensation is measured in millions of 2006 US$; standard errors are listed 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) entering one of the ranks in our sample.

TABLE VI

CAREER CONCERN AMELIORATION OF AGENCY PROBLEM[a]

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

[a]Compensation is measured in millions of 2006 US$; standard errors are listed 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) entering one of the ranks in our sample.

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 the 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 forgo higher pay. In other words, they acquire more human capital, both public and private, than the subordinates farther 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 calendar time τ, and then for all $t \in \{1, \ldots, T - 1\}$. From Proposition 1 of Margiotta and Miller (2000, p. 678), the value function solving the consumption savings problem at retirement date $T + 1$, for all h in our model, in present value terms, is

$$U_{T+1}(h, \xi_{T+1}, a_{\tau+1}, b_{\tau+1})$$
$$\equiv -\lambda_{\tau+1} b_{\tau+1} \exp[-(a_{\tau+1} + \rho \xi_{T+1})/b_{\tau+1}].$$

Suppose an executive works in firm and rank coordinate pair (j, k) at age T for one period and then is forced into mandatory retirement. After selecting job match (j, k), she chooses consumption and the next-period's endowment (c_T, ξ_{T+1}) optimally to maximize

$$(A.1) \quad -\alpha_{jkT}(h)\delta^T \exp(-\rho c_T - \varepsilon_{jkT})$$
$$- E_T\left[v_{jk,T+1} \lambda_{\tau+1} b_{\tau+1} \exp\left(-\frac{a_{\tau+1} + \rho \xi_{T+1}}{b_{\tau+1}}\right)\right]$$

subject to

$$(A.2) \quad E_T[\lambda_{\tau+1} \xi_{T+1} | l_T, d_{jkT}, h] + \lambda_\tau c_T \le \lambda_\tau \xi_T.$$

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

$$(A.3) \quad U_{jkT}(h, \xi_T, a_\tau, b_\tau, \varepsilon_{jkT})$$
$$\equiv -\lambda_\tau b_\tau \alpha_{jkT}(h)^{1/b_\tau} e^{-\varepsilon_{jkT}/b_\tau} E_T[v_{jk,T+1}]^{1-1/b_\tau} \exp\left(-\frac{a_T + \rho \xi_T}{b_\tau}\right).$$

Define ε_{jkT}^* as the value of ε_{jkT} when (j,k) is selected at T. Integrating over ε_{jkT} and averaging over job matches (j,k) yields

$$
\begin{aligned}
&U_T(h, \xi_T, a_\tau, b_\tau) \\
&\equiv -\lambda_\tau b_\tau \Bigg[p_{0T}(h) U_{0T}\big(h, \xi_T, a_\tau, b_\tau, \varepsilon_{0T}^*\big) \\
&\qquad + \sum_{j=1}^{J} \sum_{k=1}^{K} p_{jkT}(h) E\big[U_{jkT}(h, \xi_T, a_\tau, b_\tau, \varepsilon_{jkT}^*)|h, \xi_T, a_\tau, b_\tau\big] \Bigg] \\
&= -\lambda_\tau b_\tau \exp\left(-\frac{a_\tau + \rho \xi_T}{b_\tau}\right) A_T(h),
\end{aligned}
$$

where in the second line we make use of the recursive definition of $A_t(h)$ given in (4.9) and the fact that $A_{T+1}[\overline{H}_{jk}(h)] = 1$. The proof is completed with an induction by showing that (4.13) is true for all ages $t \in \{1, \ldots, T-1\}$. Suppose both equations are true for all ages $s \in \{t+1, \ldots, T\}$. Appealing to Bellman's (1957) principle, the executive's problem at age t is to maximize

$$
\begin{aligned}
&-\alpha_{jkt}(h)\delta^t \exp(-\rho c_t - \varepsilon_{jkt}) \\
&\quad - E_t\left[A_{t+1}\big[\overline{H}_{jk}(h)\big] v_{jk,t+1} \lambda_{\tau+1} b_{\tau+1} \exp\left(-\frac{a_{\tau+1} + \rho \xi_{\tau+1}}{b_{\tau+1}}\right) \right].
\end{aligned}
$$

Given job selection (j,k), equation (A.4) below follows directly from the definition of $A_{t+1}[\overline{H}_{jk}(h)]$ and the solution to the consumption savings decision at age t by substituting t for T and $v_{jk,t+1} A_{t+1}[\overline{H}_{jk}(h)]$ for $v_{jk,T+1}$ in equation (A.3):

$$
\begin{aligned}
\text{(A.4)} \quad &U_{jkt}(h, \xi_t, a_\tau, b_\tau, \varepsilon_{jkt}) \\
&\equiv -\lambda_\tau b_\tau \exp\left(-\frac{a_\tau + \rho \xi_t}{b_\tau}\right) \alpha_{jkt}(h)^{1/b_\tau} \\
&\quad \times \exp\left(-\frac{\varepsilon_{jkt}}{b_\tau}\right) E_t\big[v_{jk,t+1} A_{t+1}\big[\overline{H}_{jk}(h)\big]\big]^{(b_\tau - 1)/b_\tau}.
\end{aligned}
$$

Integrating over ε_t and averaging over the $JK+1$ job matches and appealing to the recursive definition of $A_t(h)$ then yields (4.13), as required. *Q.E.D.*

PROOF OF THEOREM 4.2: The executive optimizes her expected lifetime utility at age t by choosing the highest-valued conditional valuation function, given by equation (A.4), 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 $[-\ln \lambda_\tau - \ln b_\tau - (a_\tau + \rho \xi_t)/b_\tau]$ is then an additive constant in all alternatives, so it drops out of the solution. Multiplying by b_τ then completes the proof. *Q.E.D.*

PROOF OF THEOREM 4.3: Define $\gamma_2 \equiv \alpha_{jkt}(h)^{1/(1-b_\tau)}$, $\gamma_3 \equiv \beta_{jkt}(h)^{1/(b_\tau-1)}$, and

$$(A.5) \qquad \gamma_1 \equiv \exp\{q_{jk}[p_t(h)]\}^{1/(1-b_\tau)} \alpha_{jkt}(h)^{1/(b_\tau-1)} A_{t+1}[\overline{H}_{jk}(h)],$$

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 constraint can be expressed in terms of the new notation as $\gamma_1 E_t[v_{jkt+1}] = 1$, while the incentive compatibility condition is

$$(A.6) \qquad \gamma_2 E_t[v_{jk,t+1}] \le \gamma_3 E_t[v_{jk,t+1} g_{jkt}(\pi, h)].$$

Since the expectation operator preserves linearity, both the participation constraint (4.15) and the incentive-compatibility constraint (2330) are rendered linear in $v_{jk,t+1}$ after multiplying both sides of the latter by $A_{t+1}[\overline{H}_{jk}(h)] \times E_t[v_{jk,t+1}]$. The objective function, the expected wage bill $E_t(w_{jk,t+1})$, can be expressed as a concave function of $v_{jk,t+1}$, namely, $E_t(\ln v_{jk,t+1})$. Therefore, the Kuhn–Tucker theorem applies and the Lagrangian for the problem in which the jth firm elicits diligent work from the kth rank can be written as

$$(A.7) \qquad E_t[\ln(v_{jk,t+1})] + \eta_0 E_t[1 - v_{jk,t+1}\gamma_1]$$
$$+ \eta_1 E_t[v_{jk,t+1} g_{jk}(\pi, h)\gamma_3 - v_{jk,t+1}\gamma_2],$$

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, pp. 713–714). *Q.E.D.*

PROOF OF THEOREM 5.1: The proof of this theorem follows from Lemma 4.1 and Theorem 4.2 by extending the choice set to effort levels as well, and substituting $B_t(h, h')$ for $A_t(h)$ in their proofs. *Q.E.D.*

PROOF OF THEOREM 5.2: Setting

$$(A.8) \qquad \gamma_1 \equiv \exp\{q_{jk}[p_t(h)]\}^{1/(1-b_\tau)} \alpha_{jkt}(h)^{1/(b_\tau-1)} B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)],$$

$$(A.9) \qquad \gamma_2 \equiv \alpha_{jkt}(h)^{1/(1-b_\tau)} B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)],$$

and

$$(A.10) \qquad \gamma_3 \equiv \beta_{jkt}(h)^{1/(b_\tau-1)} B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)],$$

the proof now follows directly from the proof of Theorem 4.3. *Q.E.D.*

PROOF OF THEOREM 5.3: In this game, each executive with characteristics (t, h, h') makes a contract offer for a rank j at a single firm k of her choice, which we denote by $w_{jk,t+1}^o(\pi, h')$. If the shareholders accept the offer, the executive chooses her effort, is compensated at the beginning of the next period according to the provisions in the contract, and updates her state variables according to the transitions defined in the text. If shareholders reject the contract, the executive retires.[27] Finally, we assume (i) the executive is employed for at most T periods for some $T < \infty$ and (ii) that the optimal contract involves working every period. The proof proceeds by setting up some notation that defines a compensation function, and then applying the definition of sequential equilibrium given in Kreps and Wilson (1982), to show that the strategies of executives and shareholders are sequentially rational and that the beliefs of shareholders are consistent.

Compensation Function. Appealing to the optimization problem in Theorem 5.1, define, for each (h, π), the probability vector $(p_{0t}^e(h), \dots, p_{Jt}^e(h))$ and the human-capital function $B_t^e(h, h', b_\tau)$ by successively substituting the compensation function:

$$(A.11) \quad w_{jk,t+1}^e(\pi, h) \equiv F_{jkt}(h) + r_{jk,t+1}^e(\pi, h) - E\big[r_{jk,t+1}^e(\pi, h)\big],$$

for $w_{jk,t+1}(\pi, h)$ into the respective recursions, where $r_{jk,t+1}^e(\pi, h)$ is defined using equations (5.10) and (5.11). By inspection, $E_t[w_{jk,t+1}^e(\pi, h)|h] = F_{jkt}(h)$.

Prescribed Strategy of Executives. They choose jobs, offers, and effort level solving the problem described in Theorem 5.1. All executives offer $w_{jk,t+1}^e(\pi, h')$ regardless of their history.

Shareholder Beliefs and Prescribed Strategy. If $w_{jk,t+1}(\pi, h') = w_{jk,t+1}^e(\pi, h)$, shareholders believe the executive never shirked and $h' = h$. Alternatively, if $w_{jk,t+1}(\pi, h') \neq w_{jk,t+1}^e(\pi, h)$, shareholders believe that the executive previously shirked on $t = 0$ and is now tainted. In that case, shareholders update their beliefs to assign an upper bound to her human capital of $\underline{F}$. Thus, shareholders accept the contract if $w_{jk,t+1}(\pi, h') = w_{jk,t+1}^e(\pi, h)$, but reject it otherwise.

Sequential Rationality. From the recursive definition of $w_{jk,t\tau+1}^e(\pi, h)$ and $B_t^e(h, h')$, it follows from Theorem 5.2 that $w_{jk,t\tau+1}^e(\pi, h)$ is the most lucrative

---

[27]One can show that if the bargaining game is expanded to give executives an opportunity to make multiple offers, either sequentially to the same firm or simultaneously to different firms, then a sequential equilibrium exists with the same distribution of outcomes. In this more general bargaining framework, the game tree would still state (i) what executives do if all the offers they are permitted to make are rejected and (ii) what information shareholders have about previous rejected offers. To deal with the added complications of multiple offers, beliefs and strategies must be defined for many more contingencies that occur off the equilibrium path. These include beliefs of shareholders about executives whose previous offers have been rejected and beliefs of executives whose equilibrium offers have been rejected by shareholders.

contract in which the executive's job-match choices are sequentially rational. Since every contract that would be accepted is less lucrative, the other offer is rejected, compelling her to retire, which yields less utility if it was not the optimal choice when she made her initial selection. After agreeing on a contract, it is sequentially rational for the executive to work rather than shirk because of the incentive-compatibility constraint. Therefore, the strategy of the executive is sequentially rational. Since $w_{jk,t+1}(\pi, h') = w^e_{jk,t+1}(\pi, h)$, shareholders believe with probability 1 that $h = h'$ and the executive will work if the contract is accepted. Consequently, the shareholders believe they will make zero profits from accepting the contract, so it is a best response to accept the offer.

A Completely Mixed Strategy. To demonstrate these beliefs are consistent, consider the following perturbation from the conjectured equilibrium strategy. With probability i^{-1}, shareholders accept a contract offer of $w_{jk,t+1}(\pi, h') \neq w^e_{jk,t+1}(\pi, h)$, and with probability i^{-1}, they reject a contract offer of $w_{jk,t+1}(\pi, h') = w^e_{jk,t+1}(\pi, h)$. With probability i^{-1}, an executive deviates from her prescribed effort strategy. Thus, i^{-1} is the probability that the executive deviates by shirking in the first period and becomes tainted. Executives deviate from their optimal job-match choice to one of the other choices with probability i^{-1}, giving each of the other choices equal weight. At any period $t > 1$, let i^{-3} be the probability that an untainted executive makes a contract offer of $w_{jk,t+1}(\pi, h') \neq w^e_{jk,t+1}(\pi, h)$, and let i^{-1} be the probability that a tainted executive offers $w_{jk,t+1}(\pi, h') \neq w^e_{jk,t+1}(\pi, h)$. Thus, $(1 - \frac{1}{i})^{t-1}$ is the probability that a tainted executive does not make any contract demands off the equilibrium path in the $t - 1$ periods following the first.

Consistency of Beliefs. We let $\Psi\{t|l_1, i\}$ denote the probability of an executive holding employment beyond $t - 1$ conditional on (i) shareholders following their prescribed strategy and (ii) the executive following the prescribed contract offers after putting effort of l_1 initially. Note that $\Psi\{t|l_0, i\}$ depends on i because the executive may deviate off the work prescription by shirking. It is, however, straightforward to show that $0 < \Psi\{t|l_0, i\} < 1$ and $\Psi\{t|l_0, i\} \to \Psi\{t|l_0\} \in (0, 1)$ as $i \to \infty$. Denote by $\Pr\{\text{tainted}|\overline{w} \text{ at } t\}$ the probability that an executive demanding $w_{jk,t+1}(\pi, h') \neq w^e_{jk,t+1}(\pi, h)$ in period t is tainted. Appealing to Bayes's rule,

$$\Pr\{\text{tainted}|\overline{w} \text{ at } t\}$$

$$= \frac{\frac{1}{i}\Psi\{t|0, i\}\left(1 - \frac{1}{i}\right)^{t-1}\frac{1}{i}}{\left(1 - \frac{1}{i}\right)\Psi\{t|1, i\}\left(1 - \frac{1}{i^3}\right)^{t-1}\frac{1}{i^3} + \frac{1}{i}\Psi\{t|0, i\}\left(1 - \frac{1}{i}\right)^{t-1}\frac{1}{i}}$$

$$= \left\{\left(\frac{i-1}{i^2}\right)\left[\frac{\Psi\{t|1, i\}}{\Psi\{t|0, i\}}\right]\left(\frac{i^3 - 1}{i^3 - i^2}\right)^{t-1} + 1\right\}^{-1}.$$

In the limit of $i \to \infty$, this probability converges to 1 because $\Psi\{t|1, i\}/\Psi\{t|0, i\}$ converges to a positive constant and $(i^3 - 1)/(i^3 - i^2)$ to 1, thus demonstrating that the firm's beliefs are consistent. *Q.E.D.*

PROOF OF THEOREM 6.1: Following Gayle and Miller (2015), and using the notation defined in the proof of Theorem 4.3, take the expectation of the first-order conditions of (A.7), and also take the limit as $\pi \to \infty$ to obtain the two equations:

$$(A.12) \quad E_t[v_{jk,t+1}^{-1}] = \gamma_1 - \eta_1\gamma_3 + \eta_1\gamma_2,$$

$$(A.13) \quad \overline{v}_{jk,t+1}^{-1} = \gamma_1 + \eta_1\gamma_2.$$

Differencing (A.12) and (A.13) gives

$$(A.14) \quad \overline{v}_{jk,t+1}^{-1} - E_t[v_{jk,t+1}^{-1}] = \eta_1\gamma_3.$$

Subtracting the first-order conditions of (A.7) from (A.13) gives

$$(A.15) \quad \overline{v}_{jk,t+1}^{-1} - v_{jk,t+1}^{-1} = \eta_1 g_{jkt}(\pi, h)\gamma_3.$$

The proof of the theorem is completed by taking the quotient of (A.14) and (A.15), which yields (6.7). *Q.E.D.*

PROOF OF THEOREM 6.2: There are two steps to the proof. First, for any finite positive $\widehat{\rho}$ and any probability distribution function $\widehat{G}(\varepsilon)$ with the same support as $\widetilde{G}(\varepsilon)$, we define another parameterization, $\widehat{\theta} \in \Theta$. We show that the model defined by $\widehat{\theta}$ generates the same data as $\widetilde{\theta}$ and is therefore observationally equivalent. Given the compensation process generated by $\widetilde{\theta}$, and our construction in the first step, the conditional-choice probabilities of $\widehat{\theta}$ replicate those of $\widetilde{\theta}$. The second step is to prove that the compensation schedule generated by $\widehat{\theta}$ reproduces the schedule generated by $\widetilde{\theta}$—in other words, that the contracts are the same.

To prove the first step, let $\widehat{v}_{jkt+1} \equiv \exp[-\widehat{\rho}w_{jkt}(\pi)/b]$ for any finite positive $\widehat{\rho}$, and define

$$(A.16) \quad \widehat{g}_{jkt}(\pi, h) = \frac{\exp(\widehat{\rho}\overline{w}_{jkt}/b_\tau) - \widehat{v}_{jkt+1}^{-1}}{\exp(\widehat{\rho}\overline{w}_{jkt}/b_\tau) - E_t[\widehat{v}_{jkt+1}(\pi)^{-1}]}.$$

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

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

denote the conditional expectation of ε_{jkt}/b_τ given the choices observed in the population but integrated with respect to $\widehat{G}(\varepsilon)$ rather than $\widetilde{G}(\varepsilon)$. Appealing to Proposition 1 of Hotz and Miller (1993), there exists a mapping $\widehat{q}(p)$ implied by $\widehat{G}(\varepsilon)$ for any conditional valuation function. Starting with $\widehat{A}_t(h) = 1$ for all $t \geq R$, and given $\widehat{G}(\varepsilon)$, recursively define $\widehat{\alpha}_{jkt}(h)$ and $\widehat{A}_t(h)$ to rationalize the choice probabilities generated by $\widehat{\theta}$ by repeatedly appealing to equation (4.9) and setting

$$(A.17) \quad \widehat{\alpha}_{jkt}(h_t) = \exp\big[\widehat{q}_{jk}\big(p_t(h_t)\big)\big]\widehat{A}_{t+1}\big[\overline{H}_{jk}(h), b_{\tau+1}\big]^{1-b_\tau} E_t\big[\widehat{v}_{jk,t+1}(\pi)\big]^{1-b_\tau}.$$

Finally, $\widehat{\beta}_{jkt}(h)$ is defined as

$$(A.18) \quad \widehat{\beta}_{jkt}(h) = \exp\big[\widehat{q}_{jk}\big(p_t(h)\big)\big]\widehat{A}_{t+1}\big[\overline{H}_{jk}(h), b_\tau\big]^{1-b_\tau}$$
$$\times E_t\big[\widehat{v}_{jk,t+1}(\pi_t)\widehat{g}_{jkt}(\pi, h)\big]^{1-b_\tau}.$$

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

$$\widehat{\theta} \equiv \big(\widehat{\alpha}_{jkt}(h), \widehat{\beta}_{jkt}(h), \widehat{\rho}, \widetilde{f}(\pi), \widehat{g}_{jkt}(\pi, h), \widehat{G}(\varepsilon)\big).$$

The second step follows directly from applying Theorem 2.1 of Gayle and Miller (2015).   *Q.E.D.*

REFERENCES

ALCHIAN, A. A., AND H. DEMSETZ (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62, 777–795. [2359]

ALTUG, S., AND R. A. MILLER (1998): "The Effect of Work Experience on Female Wages and Labour Supply," *Review of Economic Studies*, 65, 45–85. [2299]

ANDERSEN, S., G. W. HARRISON, M. I. LAU, AND E. E. RUTSTROM (2008): "Eliciting Risk and Time Preferences," *Econometrica*, 76, 583–618. [2346]

ANTLE, R., AND A. SMITH (1985): "Measuring Executive Compensation: Methods and an Application," *Journal of Accounting Research*, 23, 296–325. [2301]

——— (1986): "An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives," *Journal of Accounting Research*, 24, 1–39. [2301]

ANTONOVICS, K., AND L. GOLAN (2012): "Experimentation and Job Choice," *Journal of Labor Economics*, 30, 333–366. [2298,2313]

ARON, D. J. (1988): "Ability, Moral Hazard, Firm Size, and Diversification," *Rand Journal of Economics*, 19, 72–87. [2296]

BAKER, G. P., AND B. J. HALL (2004): "CEO Incentives and Firm Size," *Journal of Labor Economics*, 22, 767–798. [2343,2351]

BAYER, P., S. KHAN, AND C. TIMMINS (2011): "Nonparametric Identification and Estimation in a Roy Model With Common Nonpecuniary Returns," *Journal of Business & Economic Statistics*, 29, 201–215. [2298]

BECKER, G. S. (1964): *Human Capital*. New York: National Bureau for Economic Research. [2298,2325]

BELLMAN, R. (1957): *Dynamic Programming*. Princeton, NJ: Princeton University Press. [2361]

BEN-PORATH, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75, 352–365. [2298]

BROWN, C., AND J. MEDOFF (1989): "The Employer Size-Wage Effect," *Journal of Political Economy*, 97, 1027–1059. [2299]

CAHUC, P., F. POSTEL-VINAY, AND J. M. ROBIN (2006): "Wage Bargaining With on-the-Job Search: Theory and Evidence," *Econometrica*, 74, 323–364. [2299,2325]

CALVO, G. A., AND S. WELLISZ (1980): "Technology, Entrepreneurs, and Firm Size," *Quarterly Journal of Economics*, 95, 663–677. [2359]

CHEVALIER, J., AND G. ELLISON (1999): "Career Concerns of Mutual Fund Managers," *Quarterly Journal of Economics*, 114, 389–432. [2298]

DEBREU, G. (1959): *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Cowles Foundation for Research in Economics Monograph, Vol. 17. New Haven, CT: Yale University Press. [2318]

DEWATRIPONT, M., I. JEWITT, AND J. TIROLE (1999): "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies," *Review of Economic Studies*, 66, 199–217. [2298]

D'HAULTFOEUILLE, X., AND A. MAUREL (2013): "Inference on an Extended Roy Model, With an Application to Schooling Decisions in France," *Journal of Econometrics*, 174, 95–106. [2298]

DOERINGER, P. B., AND M. J. PIORE (1971): *Internal Labor Markets and Manpower Analysis*. Armonk, NY: ME Sharpe. [2296]

DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, 100, 1238–1260. [2346]

FELLI, L., AND C. HARRIS (1996): "Learning, Wage Dynamics, and Firm-Specific Human Capital," *Journal of Political Economy*, 104, 838–868. [2325]

FOX, J. T. (2009): "Firm-Size Wage Gaps, Job Responsibility, and Hierarchical Matching," *Journal of Labor Economics*, 27, 83–126. [2295]

FUDENBERG, D., B. HOLMSTRÖM, AND P. MILGROM (1990): "Short-Term Contracts and Long-Term Agency Relationships," *Journal of Economic Theory*, 50, 1–31. [2323]

GABAIX, X., AND A. LANDIER (2008): "Why Has CEO Pay Increased so Much?" *Quarterly Journal of Economics*, 123, 49–100. [2293,2294,2297,2301,2356]

GARRETT, D. F., AND A. PAVAN (2012): "Managerial Turnover in a Changing World," *Journal of Political Economy*, 120, 879–925. [2298]

GAYLE, G.-L., AND L. GOLAN (2012): "Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968–1997," *Review of Economic Studies*, 79, 227–267. [2299]

GAYLE, G.-L., AND R. A. MILLER (2009a): "Insider Information and Performance Pay," *CESifo Economic Studies*, 55, 515–541. [2301,2351,2355]

——— (2009b): "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *American Economic Review*, 99, 1740–1769. [2293,2294,2296,2301,2337,2346,2351,2355,2356]

——— (2015): "Identifying and Testing Models of Managerial Compensation," *Review of Economic Studies*, 82, 1074–1118. [2298,2299,2338,2351,2355,2365,2366]

GAYLE, G.-L., L. GOLAN, AND R. A. MILLER (2012): "Gender Differences in Executive Compensation and Job Mobility," *Journal of Labor Economics*, 30, 829–872. [2300,2359]

GAYLE, G.-L., L. GOLAN, AND R. A. MILLER (2015a): "Interlocked Executives and Insider Board Members: An Empirical Analysis," Working Paper, Carnegie Mellon University. [2298]

——— (2015b): "Supplement to 'Promotion, Turnover, and Compensation in the Executive Labor Market'," *Econometrica Supplemental Material*, 83, http://dx.doi.org/10.3982/ECTA11020. [2297]

GIBBONS, R., AND K. J. MURPHY (1992): "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, 100, 468–505. [2298]

HALL, B. J., AND J. B. LIEBMAN (1998): "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics*, 113, 653–680. [2301]

HARRIS, M., AND B. HOLMSTROM (1982): "A Theory of Wage Dynamics," *Review of Economic Studies*, 49 (3), 315–333. [2325]

HARRISON, G. W., E. JOHNSON, M. M. MCINNES, AND E. E. RUTSTRÖM (2005): "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95, 897–901. [2346]

HARRISON, G. W., J. A. LIST, AND C. TOWE (2007): "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion," *Econometrica*, 75, 433–458. [2346]

HOLMSTROM, B. (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66, 169–182. [2298]

HOLT, C. A., AND S. K. LAURY (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655. [2346]

——— (2005): "Risk Aversion and Incentive Effects: New Data Without Order Effects," *American Economic Review*, 95, 902–904. [2346]

HOTZ, V. J., AND R. A. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529. [2319,2337,2366]

JOVANOVIC, B. (1979): "Firm-Specific Capital and Turnover," *Journal of Political Economy*, 87, 1246–1260. [2325]

KATZ, L. F., AND L. H. SUMMERS (1989): "Industry Rents: Evidence and Implications," *Brookings Papers on Economic Activity: Microeconomics*, 1989, 209–290. [2296]

KREPS, D. M., AND R. WILSON (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894. [2363]

LEE, D., AND K. I. WOLPIN (2006): "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica*, 74, 1–46. [2299]

LI, A. (2014): "Test Contract," Working Paper, Washington University in St. Louis. [2298]

LUCAS, R. E., JR. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9, 508–523. [2296,2297,2315]

MAGNAC, T., AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816. [2336]

MALCOMSON, J. M., AND F. SPINNEWYN (1988): "The Multiperiod Principal–Agent Problem," *Review of Economic Studies*, 55, 391–407. [2323]

MARGIOTTA, M., AND R. A. MILLER (2000): "Managerial Compensation and the Cost of Moral Hazard," *International Economic Review*, 41, 669–719. [2298,2301,2314,2316,2318,2351,2355, 2360,2362]

MCNULTY, P. J. (1984): *The Origins and Development of Labor Economics*. Cambridge, MA: MIT Press. [2297]

MILLER, R. A. (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086–1120. [2298,2313,2325]

MINCER, J. A. (1974): *Schooling, Experience, and Earnings*. Boston, MA: NBER Books. [2303]

MIRRLEES, J. A. (1975): "The Theory of Moral Hazard and Unobservable Behavior, Part 1," Working Paper, Oxford University. [2315]

——— (1976): "The Optimal Structure of Incentives and Authority Within an Organization," *Bell Journal of Economics*, 7, 105–131. [2359]

OI, W. Y., AND T. L. IDSON (1999): "Firm Size and Wages," in *Handbook of Labor Economics*, Vol. 3. Amsterdam: North-Holland, 2165–2214. [2293,2299]

POSTEL-VINAY, F., AND J. M. ROBIN (2002): "Equilibrium Wage Dispersion With Worker and Employer Heterogeneity," *Econometrica*, 70, 2295–2350. [2299]

RAFF, D. M., AND L. H. SUMMERS (1987): "Did Henry Ford Pay Efficiency Wages?" *Journal of Labor Economics*, 5, S57–S86. [2296]

REY, P., AND B. SALANIE (1990): "Long-Term, Short-Term and Renegotiation: On the Value of Commitment in Contracting," *Econometrica*, 58, 597–619. [2323]

ROSEN, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34–55. [2324]

——— (1983): "The Economics of Superstars: Reply," *American Economic Review*, 73, 460–462. [2297]

ROY, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers, New Series*, 3, 135–146. [2294,2295,2298,2319]

RUBINSTEIN, M. (1981): "A Discrete-Time Synthesis of Financial Theory," *Research in Finance*, 3, 53–102. [2318]

SANDERS, C. (2013): "Skill Uncertainty, Skill Accumulation and Occupational Choice," Working Paper, Washington University in St. Louis. [2298,2313]

SANNIKOV, Y. (2014): "Moral Hazard and Long-Run Incentives," Working Paper, Princeton University. [2298]

STOLE, L. A., AND J. ZWIEBEL (1996a): "Intra-Firm Bargaining Under Non-Binding Contracts," *Review of Economic Studies*, 63, 375–410. [2325]

——— (1996b): "Organizational Design and Technology Choice Under Intrafirm Bargaining," *American Economic Review*, 86, 195–222. [2325]

TABER, C., AND R. VEJLIN (2010): "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," Working Paper, University of Wisconsin–Madison. [2299]

TERVIÖ, M. (2008): "The Difference That CEOs Make: An Assignment Model Approach," *American Economic Review*, 98, 642–668. [2293,2294,2297]

THOMAS, J., AND T. WORRALL (1988): "Self-Enforcing Wage Contracts," *Review of Economic Studies*, 55, 541–554. [2325]

WILLIAMSON, O. E. (1967): "Hierarchical Control and Optimum Firm Size," *Journal of Political Economy*, 75, 123–138. [2315,2359]

WINTER-EBMER, R., AND J. ZWEIMÜLLER (1999): "Firm-Size Wage Differentials in Switzerland: Evidence From Job-Changers," *American Economic Review*, 89, 89–93. [2299]

Dept. of Economics, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, U.S.A.; glgayle@me.com,

Dept. of Economics, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, U.S.A.; lim.golan@gmail.com,

and

Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, U.S.A.; ramiller@andrew.cmu.edu.