Structural Estimation of Executive Compensation

Suggested Citation: Robert A. Miller and Yizhen Xie (2025), "Structural Estimation of Executive Compensation", : Vol. XX, No. XX, pp 1–XX. DOI: 10.1561/XXXXXXXXX.

Robert A. Miller
Carnegie Mellon University
ramiller@cmu.edu

Yizhen Xie
Carnegie Mellon University
yizhenxi@andrew.cmu.edu

now
the essence of knowledge
Boston — Delft

Contents

Structural Estimation of Executive Compensation

Robert A. Miller[1] and Yizhen Xie[2]

[1] *Tepper School of Business, Carnegie Mellon University; ramiller@cmu.edu*

[2] *Tepper School of Business, Carnegie Mellon University; yizhenxi@andrew.cmu.edu*

ABSTRACT

Structural estimation of executive compensation combines cross-sectional and longitudinal data relating firm performance to the wages, grants and wealth holdings of managers to quantify principal-agent models characterized by asymmetric information. The estimated models are used to measure the importance of information asymmetries, such as the degree of conflict between executives and the firms they manage, the role of human capital in mitigating conflicting interests, and the social welfare loss

Robert A. Miller and Yizhen Xie (2025), "Structural Estimation of Executive Compensation", : Vol. XX, No. XX, pp 1–XX. DOI: 10.1561/XXXXXXXXX.

from moral hazard. Following a brief guide to this survey and a short review of related literature, we begin by describing the data used to estimate these models, explain the theory behind a simple static model of moral hazard, and provide a first approach to estimating them, before analyzing identification in more depth. The latter sections then show how the simplest models of moral hazard can be extended to account for other sources of hidden information and dynamic considerations that arise from the life cycle aspirations of managers.

1

Introduction

This monograph examines recent advances in the structural estimation of agency costs arising from information asymmetry between shareholders and executive management. Multiple factors contribute to the information differential between executives and resource owners. Executives possess superior knowledge of firm performance, operational capabilities, and financial arrangements through their direct engagement with operational activities. Furthermore, executives have better knowledge of their managerial capabilities, which significantly influence organizational outcomes while remaining largely unobservable to shareholders. Additionally, executives must allocate their efforts between shareholder objectives and various competing priorities, including stakeholder relationships and professional advancement considerations, aspects which remain predominantly opaque to shareholders. While partial alignment exists between shareholder and executive incentives, fundamental divergences persist. This misalignment of interests constitutes a significant agency problem that has garnered extensive scholarly attention, particularly regarding executive compensation and firm performance.

Structural estimation methodology employs empirical observations of managerial compensation and firm performance to derive inferences about managerial preferences, incentive effects on firm performance, and costs imposed by information asymmetry. This approach requires the construction of models that incorporate behavioral assumptions about managers and shareholders while maintaining computational tractability for numerical estimation. The fundamental premise of optimal contracting posits that shareholders, through explicit or implicit mechanisms, design contracts that maximize profit by aligning managerial incentives with firm performance. Managers, being utility-maximizing agents, will not consistently act in alignment with shareholder interests absent appropriate incentive structures. Although explicit incentive contracts constitute a primary mechanism for aligning interests, Gibbons and Murphy, 1992, Holmström, 1999, and Dewatripont *et al.*, 1999, demonstrate that career concerns can function as implicit incentives for executives, potentially reducing the need for explicit contractual incentives.

The study of managerial compensation and executive-shareholder relationships represents a fundamental domain of inquiry within social sciences, particularly in accounting, economics, and finance. The substantial economic value embodied in non-owner-managed firms and the pivotal role of executives in determining organizational outcomes underscore the significance of understanding incentive mechanisms.

Corporate governance policy exercises considerable influence over executive compensation practices and organizational oversight. Regulatory frameworks, exemplified by the Sarbanes-Oxley Act and the Dodd-Frank Act, have been implemented to enhance transparency, mitigate excessive risk-taking behavior, and address principal-agent conflicts between executives and shareholders.

The implications of research on managerial compensation extend beyond individual firms to encompass broader economic and policy considerations. For organizations, empirical insights derived from structural estimation facilitate the optimization of compensation structures to enhance performance, minimize agency costs, and attract superior managerial talent. For policymakers, rigorous empirical analysis informs the development and refinement of regulatory frameworks governing executive compensation, corporate disclosure requirements, and governance standards.

The methodological contributions of structural estimation manifest themselves in three distinct dimensions. First, it enables the decomposition and quantification of various components of the model. In the context of executive compensation with asymmetric information, these components encompass: (i) the moral hazard cost to the firm, quantified as the risk premium paid to executive management for bearing uncertainty in firm-performance-linked compensation; (ii) the incremental risk premium attributable to information asymmetry between shareholders and executives regarding firm conditions; (iii) the dynamic effects of asymmetric information regarding executive capability, human capital acquisition, and effort allocation on both short-term and long-term firm performance.

Second, the structural framework provides a systematic methodology for evaluating the impact of regulatory interventions on the principal-agent relationship, as exemplified by the Sarbanes-Oxley Act and its associated regulatory modifications for publicly traded companies.

Third, structural estimation enables the analysis of counterfactual scenarios, such as proposed modifications to tax policy affecting high-income earners, capital gains treatment, and provisions governing the temporal distribution of compensation.

The inferences and predictions of structural estimation are inherently constrained by both the empirical data utilized in estimation procedures and the underlying assumptions incorporated into model construction. Consequently, structural estimation's primary contribution lies in providing qualitative and quantitative predictions within an internally consistent framework that captures essential aspects of a broader, albeit incomplete, perspective. At minimum, structural estimation imposes methodological discipline by explicitly delineating the theoretical connections between empirical predictions and observed data through a well-specified model.

1 Outline

This monograph synthesizes contributions from Margiotta and Miller, 2000, Gayle and Miller, 2009a, Gayle and Miller, 2009b, Gayle *et al.*, 2012, Gayle and Miller, 2015, Gayle *et al.*, 2015, and Gayle *et al.*, 2022, on structural estimation of executive compensation. The analysis proceeds as follows.

Section 2 presents a comprehensive examination of the primary data sources employed in structural estimation of executive contracting, delineating the construction of key variables, and examining the empirical patterns that motivate structural estimation. The empirical framework encompasses samples of firms and executives, incorporating total compensation (salary, bonuses, stock and option grants) and executives' holdings of firm-denominated securities, firms' financial returns, accounting earnings, capital structure (debt-equity ratios), and demographic characteristics of executives (age, education, tenure, career history).

Section 3 establishes a foundational static model of pure moral hazard between a single principal and one agent, examining its implications for optimal contracting. This framework serves as the analytical benchmark for investigating identification and estimation methodologies, while providing the theoretical foundation for subsequent model extensions.

The initial extensions distilled in Section 4 address two fundamental characteristics of executive compensation: its dynamic nature and the presence of multiple agents. Executives typically maintain long-term employment relationships with firms and have access to various mechanisms for wealth accumulation over their careers. Moreover, firms implement incentive structures not only for the Chief Executive Officer but for multiple senior executives. Consequently, Section 4 extends the static framework to incorporate dynamic environments and multiple agents. Section 5 develops parametric estimators and presents empirical findings derived from the data described in Section 2 for this dynamic multi-agent model. Two significant empirical findings emerge regarding agency costs: while firms would incur substantial equity losses by disregarding agency considerations, the cost of incentivizing managers to mitigate agency concerns is relatively modest, suggesting that aligning managerial and shareholder interests is relatively cheap and effective.

The empirical analysis presented in Section 5, which employs Maximum Likelihood (ML) and Generalized Methods of Moments (GMM) estimators, does not fully address the identification challenges inherent in this nonlinear framework. Section 6 examines these methodological issues by expanding the conceptual framework from point identification to set identification and providing precise characterization of identifiable parameters within the model. This analysis demonstrates how this enhanced identification framework can be applied to key empirical objects, including agency cost decomposition and counterfactual analysis. Section 7 introduces a semi-parametric estimation approach and presents empirical findings suggesting that while moral hazard constitutes a central mechanism in executive contracting, it does not fully characterize the information asymmetry between shareholders and executives.

Beyond moral hazard considerations, executives possess private information that may be utilized for their financial gains, as Gayle and Miller, 2009b, document. Consequently, Section 8 expands the moral hazard model to incorporate hidden information. In this extended framework, managers privately observe the firm's production state and may strategically misreport this information in accounting statements to optimize their compensation. Section 8.2 provides empirical support for this hybrid model of moral hazard and hidden information, demonstrating non-empty overlapping sets of estimated risk-aversion parameters across various industries and firm sizes.

Another dynamic consideration is human capital accumulation—the potential for earnings growth through job transitions and rank advancement. Section 9 introduces human capital accumulation into the framework. This extension generalizes Roy's 1951 job-sorting model to incorporate both human capital accumulation and moral hazard. A subsequent extension considers the case where effort choices influence human capital accumulation, thereby introducing career concerns as executives internalize the impact of current performance on future compensation and career opportunities. This theoretical development builds on Gibbons and Murphy's 1992 demonstration that optimal contracts must account for both explicit incentives and implicit career concerns. Section 9.2 elaborates on this concept, incorporating human capital accumulation as an additional incentive mechanism. Section 10 presents empirical evidence from these human capital models, revealing that while human capital investment significantly influences compensation levels, the predominant source of pay variation stems from increasing firm size and the associated risk premium.

2 Literature

This monograph relates to a growing literature on structural estimation of executive compensation[1]. The theoretical literature presents three non-exclusive views of executive compensation patterns: shareholder value maximization through optimal contracting, rent extraction by executives[2], and institutional regulation. Edmans *et al.*, 2017, survey these theoretical perspectives, demonstrating that no single view fully explains the observed patterns in executive compensation. Frydman and Jenter, 2010, review the empirical patterns in CEO compensation since the 1930s and also conclude that no single approach is fully consistent with the empirical evidence.

Moral hazard has been a primary explanation for the correlation between managerial compensation and firm performance. While Jensen and Murphy, 1990, find that CEO pay-performance sensitivity appears suboptimally low, Hall and Liebman, 1998, demonstrate that equity-based compensation makes CEO wealth

[1]Bertomeu *et al.*, 2023, provide a primer on structural estimation in accounting research. Strebulaev and Whited, 2012, review the literature on structural estimation in corporate finance, including a brief review of its applications in executive compensation.

[2]For example, Sannikov, 2008, derives optimal principal-agent contracts with moral hazard in continuous time, while DeMarzo and Sannikov, 2017, extend this framework to account for asymmetric learning about firm profitability, showing how optimal contracts control information rents through distorted termination decisions.

highly sensitive to performance. Margiotta and Miller, 2000, show through structural estimation that observed incentive levels sufficiently deter managers from shirking. Gayle and Miller, 2009a, attribute the increasing trend in managerial compensation to the growth in firm size and the associated risk premium.

Private information and executive discretion in accounting reports are another source of agency frictions that affect executive contracts, as the evidence in Gayle and Miller, 2009b, suggests. Gayle and Miller, 2015, establish identification for moral hazard models and provide evidence that supports a hybrid model with both moral hazard and private information. Armstrong *et al.*, 2010b, find that firms with higher levels of stock-based CEO compensation are less subject to accounting manipulation. Zakolyukina, 2018, estimates modest expected penalties for earnings manipulation, while Bertomeu *et al.*, 2020, show that such misstatements can substantially reduce firm value.

Another strand of literature studies the sorting and turnover of executive-firm matching. The theoretical foundations for size-based matching originate from Lucas, 1978, who demonstrates how managerial talent determines firm size distribution. Gabaix and Landier, 2008, and Terviö, 2008, develop assignment models showing that small differences in CEO talent can generate large pay differentials when matched to firms of different sizes. Gabaix and Landier, 2008, extend these insights to explain the increase in CEO compensation through an equilibrium matching model. Taylor, 2010, attributes infrequent forced executive turnover to high perceived firing costs by firms, while Pan, 2017, estimates complementarities between executive and firm attributes in matching outcomes.

Regulations, tax policy, and accounting rules also affect executive compensation. Murphy, 2013, provides a comprehensive review of institutional influences on executive compensation. Empirical evidence from Frydman and Molloy, 2011, Rose and Wolfram, 2002, and Goolsbee, 2000, demonstrates that tax policy changes significantly influence both the level and the structure of executive compensation. Changes in accounting standards also shape compensation practices. For example, Carter *et al.*, 2007, and Hayes *et al.*, 2012, show that firms adjust the mixture of stock options and restricted stocks in response to evolving accounting treatments for options, while the total compensation remains unchanged. Armstrong *et al.*, 2010a, document improved information quality following the adoption of International Financial Reporting Standards (IFRS) in Europe. Gayle *et al.*, 2022, demonstrate that the Sarbanes-Oxley Act reduces shareholder-CEO conflicts while increasing agency costs.

Finally, human capital and learning introduce dynamic considerations that affect executive compensation. Human capital can also be private information of the manager, as supported by Taylor, 2013, who documents asymmetric pay responses to news about CEO ability. Gayle *et al.*, 2015, generalize the job-sorting model in Roy, 1951, to incorporate moral hazard and human capital accumulation, while accounting for the sorting and assignment of executives to firms. They extend the model further to account for career concerns, that is, how current performance affects future career trajectories, which builds on Gibbons and Murphy, 1992, Dewatripont *et al.*, 1999, and Holmström, 1999. Harris and Holmstrom, 1982, characterize optimal long-term labor contracts under incomplete information. Fee *et al.*, 2013, provide evidence that managerial human capital significantly affects firm policies, and Pan *et al.*, 2015, estimate the relative importance of uncertainty about CEO ability versus the firm's fundamental cash flow uncertainty in contributing to stock return volatility. Miller and Nguyen, 2024, analyze optimal long-term contracts incorporating learning about match quality.

2

The Data

This Section describes the data sources for the variables used to estimate structural econometric models of managerial compensation, how the variables are constructed, and the patterns or stylized facts that describe their covariation. The two main variables for estimating structural econometric models of managerial compensation are abnormal financial returns to shareholders and total executive compensation. The first is a measure of executive productivity; the second measures their pay. Unlike most empirical studies on compensation in labor economics, the data connects what a worker is paid, with what he or she is paid for. Here, abnormal returns are measured as total dividends plus capital gains as a proportion of firm equity relative to some norm, such as financial returns on the stock market index.

The argument for this measure of executive productivity is that movement in the aggregate index represents factors in the economy over which managers within any one firm have no control. In this context, the adjective "total" means that compensation should measure the change in wealth that occurs because of the executive's position; it is reasonable to assume that if the executive were not employed by his firm, he would hold the stock market index rather than his firm's financial securities, so that his financial portfolio is not exposed to the firm's abnormal returns. The product of equity held in his firm and changes in the abnormal returns are therefore one component of his compensation.

Executive duties are tailored to the type of firm they work for. Conditional on their effort level, the nature of the work affects the amenity value, or the nonpecuniary benefit from being employed. For this reason, background variables typically include characteristics of the firm, such as industry, firm size by asset value, employment, and debt-to-equity ratio. For a given industry, larger firms are typically more cumbersome to manage than smaller ones; this factor also partly determines the number of executives who play some role in firm management. Similarly, a larger labor force in the firm could bring more human resource issues to address. A higher debt-to-equity ratio increases the risk of insolvency issues; a lower ratio might be associated with dispersed ownership and weaker governance, or a more concentrated but less diversified ownership that keeps tighter control of the executive suite. These influences are difficult to assess without a detailed model-based explanation, but in the absence of such a mechanism, it is reasonable to include these factors determining firm heterogeneity that might affect the productivity of managers and their nonpecuniary benefits on the job.

Including background variables on the executive in empirical studies of managerial compensation is also useful because their duties are sometimes partly tailored to skills and reputations acquired with experience. Researchers in this area do not have access to the detail boards have when considering candidates for management in their own firms. Variables like tenure with the firm, executive experience, age, and formal education, are poor but useful substitutes.

Accounting statements and reports diminish the knowledge gap between executive management and shareholders in two fundamental ways. They give executive management a voice with which to communicate unverifiable information at their discretion to shareholders; through audits, they require management to disclose information about the firm that can be checked for which they are legally accountable. Penalties for malfeasance are used to enforce truthful disclosures that managers are legally obligated; designing compensation that depends on accounting earnings statements to incentivize truth-telling behavior is a way of ensuring that managers report truthfully.

1 Data sources

Pioneering work, Masson, 1971, and Antle and Smith, 1985, set the stage for how data would be collected and used in this field. Margiotta and Miller, 2000, adopt the executive compensation data from those earlier studies. Typically, the data sets for these empirical investigations might comprise a sample of firms and their executives that include: total compensation (including salary and bonus plus stock and option grants) and inside wealth (stocks and options in the firm); the firm's financial returns (including dividends and capital gains), accounting earnings and capital structure (such as debt-to-equity ratio); other characteristics of the firm (identifying its industry, firm employment, assets); and demographic details describing the executive (such as age, education, tenure with the firm, total executive experience, employment spells with the firm in various executive positions).

The data for the empirical study are assembled in Gayle and Miller, 2009a, Gayle and Miller, 2015, and Gayle *et al.*, 2015. The three sources of data involve executive compensation, firm performance, and executive background. The main data source is the S&P Compustat ExecuComp database, which tracks board members and annual compensation for each executive in publicly traded firms that constitute the S&P 1500 index, covering large-cap, mid-cap and small-cap US firms. Data on these firms are supplemented by the S&P Compustat North America database and monthly stock prices from the Center for Research in Securities Prices (CRSP) database. In addition, Gayle *et al.*, 2015, collect background histories for a subsample of 16,300 executives by matching the 30,614 executives from the COMPUSTAT database with records in *Who's Who*. The matching process used full names, years of birth, and gender to link the data to biographies of approximately 350,000 executives. The resulting matched data set provides unprecedented access to detailed firm characteristics, including accounting and financial data, as well as comprehensive information about the executives. This includes the main components of their compensation—including pensions, salaries, bonuses, stock options, and stock grants and holdings—as well as their demographic details, including age, gender, and education. In addition, it offers a detailed account of their career paths, tracing annual transitions across positions and firms.

2 Variables

For an overview of the data set, this Section presents the construction of the total compensation variable, abnormal returns and accounting returns, executive background, and a job hierarchy within executive ranks.

Compensation

Total compensation includes the sum of salary and bonus, the value of restricted stocks and options granted, the value of retirement and long-term compensation schemes—all itemized in ExecuComp, plus changes in wealth from holding firm options and changes in wealth from holding firm stock relative to a well-diversified market portfolio.[1]

Changes in wealth from holding firm stock and options reflect the opportunity cost a manager incurs from restrictions in selling these securities. Antle and Smith, 1985, argue persuasively for including the returns from insider wealth in the measure of executive compensation. Managers recognize that part of the return from their firm-denominated securities should be attributed to aggregate market performance, but the rest is firm-specific risk, which is undiversifiable. To offset this exposure,

[1] See Antle and Smith, 1985, Antle and Smith, 1986, and Hall and Liebman, 1998, for other papers in addition to the main papers referenced that use this measure of total executive compensation.

they reduce their holdings of other stocks when constructing their overall portfolio of real and financial assets. Hence, the change in wealth from holding their firms' stock is the value of the stock at the beginning of the period multiplied by the abnormal return relative to the market return. Options are valued using the Black-Scholes formula. This construct overlooks the difference between granting and vesting financial securities, in effect assuming that all granted securities will be vested.

Abnormal returns and bond prices

As the sample of firms is all publicly traded, they use abnormal returns from the firm's stocks as the primary measure of firm performance, which is distinct from the firm's accounting returns based on accounting reports. Bertomeu *et al.*, 2022, compare the information content in firm's market price and the accounting reports and find market prices to be more informative in executive compensation. The abnormal return of a firm's stock is defined as the difference between the return on the firm's stock and the return on the market portfolio. Gayle and Miller, 2009a, use alternative specifications of abnormal return benchmarking against industry and sector factors, but that does not affect the results significantly. That implies firms may be able to hedge their risk exposure to industry and sectoral factors.

Managers optimally smooth consumption over time, as is standard in the asset pricing literature (Debreu, 1959, Chap 7). This amounts to annuitizing compensation across periods. To facilitate the calculation of annuitized compensation, Gayle and Miller, 2009a, detail the method to build a price series for annuity bonds from the Federal Reserve's Economic Research Database. The annuity bond price b_t is the price of a synthesized bond that pays one consumption unit each year starting from t, evaluated based on the treasury yields.

Table 2.1 displays the longitudinal characteristics of the data.[2] As the benchmark, stock market returns during this period varied widely, from a yield of 45% in one year to a loss of 14% in another. However, the variation in returns among individual firms far exceeds that of the overall market. The abnormal return is measured as the firm's stock returns net of the stock market index return. A commonly used measure of accounting performance is the difference between the change in assets and the changes in liabilities plus dividends, called comprehensive income. Normalizing comprehensive income, Gayle and Miller, 2015, define the accounting return π_{nt} for the firm n in period t

$$\pi_{nt} = \frac{Asset_{nt} - Debt_{nt} + Dividend_{nt}}{Asset_{n,t-1} - Debt_{n,t-1}}. \tag{2.1}$$

In particular, average accounting returns closely track abnormal

[2]The number of observations is consistent across years, except for 2005, when only firms with financial years ending before December are included.

returns, showing a strong positive correlation as they rise and fall almost in tandem.

Table 2.1
Summary of Firm Characteristics and Compensation by Year

Year	Bond	Assets	Employees	$\frac{\text{Debt}}{\text{Equity}}$	Abnormal Return	Accounting Return	Compensation	N
1993	15.90	8896	18.02	2.83	1.19	1.18	1854	1574
		(26,269)	(46.15)	(7.24)	(0.45)	(0.51)	(12,412)	
1994	13.72	7770	16.18	2.87	0.97	1.07	2714	1876
		(25,284)	(43.41)	(5.04)	(0.29)	(2.52)	(10,909)	
1995	14.00	8187	16.43	3.45	1.26	1.18	1781	1867
		(28,650)	(44.41)	(33.40)	(0.47)	(0.64)	(13,252)	
1996	13.79	8357	17.31	2.41	1.16	1.17	3257	1926
		(29,029)	(45.92)	(17.20)	(0.38)	(0.87)	(14,824)	
1997	13.67	8770	17.94	2.76	1.30	1.22	4691	1997
		(31,797)	(47.96)	(41.40)	(0.48)	(3.06)	(17,791)	
1998	15.00	9486	17.67	3.91	1.05	1.20	2726	2012
		(40,145)	(45.91)	(71.30)	(0.53)	(1.11)	(18,530)	
1999	13.97	10,303	18.34	2.84	1.14	1.31	1652	1970
		(43,087)	(45.75)	(11.57)	(0.76)	(8.27)	(21,631)	
2000	13.18	10,484	19.59	2.64	1.14	1.18	4624	1865
		(45,936)	(54.08)	(8.31)	(0.68)	(1.50)	(21,641)	
2001	14.16	12,015	20.10	2.69	1.08	1.17	3314	1851
		(52,064)	(56.50)	(14.90)	(0.54)	(1.86)	(18,842)	
2002	14.32	12,115	19.47	4.69	0.86	1.00	3165	1877
		(57,166)	(54.51)	(105.00)	(0.42)	(2.43)	(16,077)	
2003	14.87	13,869	19.15	2.51	1.45	1.53	3151	1814
		(66,331)	(52.85)	(35.20)	(0.64)	(16.10)	(18,830)	
2004	14.17	14,429	21.05	2.77	1.16	1.11	4069	1687
		(70,812)	(64.83)	(9.39)	(0.37)	(1.38)	(17,195)	
2005	13.89	20,925	22.19	2.63	1.07	1.16	4397	751
		(89,832)	(52.34)	(12.27)	(0.36)	(1.63)	(19,992)	

Note: Excerpt from Gayle and Miller, 2015, Table 2. Standard deviations are in parentheses. Assets are in millions of 2000 US$, Employees are in thousands.

Classifying firms

The sample firms are classified into three industrial sectors, primary, consumer, and service. Firms are further grouped by size—large, medium, and small—based on the value of their assets and the number of employees relative to the median in their sector. They also categorize firms according to the number of insiders on the board relative to the median number of insiders in the industry. Finally, they label a firm as new or old from the perspective of an executive based on whether this is the first year for him to work in this firm, indicating executive turnover. In total, there are 36 types of firms based on firm size, industrial sector, size of insider board, and whether the firm is a new firm for the executive.

Table 2.2 provides a summary of the cross-section of firms in this data set. The three sectors investigated are broadly representative of all sectors and align with the trends identified in previous studies using other data sets. Firms in the service sector have the highest average size when measured by total assets, while those in the consumer sector are the smallest by this metric. A similar sectoral difference exists in the debt-equity ratio, as the sector with the largest firms by assets also exhibits the highest leverage. However, this ranking is reversed when firm size is measured by the number of employees. Consequently, they use both total assets and employment as their measures of firm size and include the debt-equity ratio as a factor that influences the distribution of abnormal returns, which in turn impacts managerial compensation, summarized in a triplet (A, W, D), where each component can be S (small) or L (large).

Table 2.2
Summary of Firm Characteristics by Sector

Variable / Sector	Primary	Consumer	Service
Observations	8980	6762	11,144
Assets (millions of 2000 US$)	6322	5277	17,776
	(27,773)	(22,124)	(67,133)
Market Value (millions of 2000 US$)	6480	7811	11,664
	(25,160)	(21,975)	(35,002)
Employees (thousands)	15.8	32.23	11.9
	(40.8)	(78.75)	(26.59)
Debt-equity ratio	2.07	1.94	4.56
	(40.9)	(26.21)	(50.63)
Accounting return	1.15	1.13	1.28
	(4.54)	(1.68)	(7.26)

Note: Excerpt from Gayle and Miller, 2015, Table 1. Standard deviations are in parentheses.

Accounting earnings

To capture the idea that the CEO can exercise discretion in reporting accounting returns to shareholders, they label the reported state of the firm as good or bad, when the firm's accounting return is higher or lower than the average for all firms within the same sector, size, and capital structure categories in the same period.

Table 2.3 displays the number of observations in each sector and size category, and the probability that the accounting report is good. For the most part, the probability of being in the bad state is higher than 0.5, implying that the reported state is skewed towards the bad state.[3]

[3]There are exceptions, such as $(A, W, D) = (\mathrm{S}, \mathrm{S}, \mathrm{L})$ in the primary and consumer sectors.

Table 2.3
Estimated Probability Distribution of Accounting Reports

Firm type	Primary			Consumer			Service			
(A, W, D)	N	Good	Bad	N	Good	Bad	N	Good	Bad	Total
(S,S,S)	2598	0.0917	0.1975	2023	0.1227	0.1764	3483	0.1249	0.1877	8103
(S,L,S)	319	0.0141	0.0214	268	0.0121	0.0275	210	0.0040	0.0149	797
(S,L,L)	469	0.0257	0.0266	418	0.0229	0.0389	1210	0.0337	0.0749	2097
(S,S,L)	1326	0.0763	0.0713	961	0.0725	0.0696	952	0.0434	0.0421	3239
(L,S,S)	541	0.0272	0.0331	498	0.0308	0.0427	760	0.0248	0.0434	1799
(L,L,S)	1105	0.0635	0.0595	734	0.0593	0.0493	927	0.0164	0.0668	2766
(L,L,L)	2398	0.1118	0.1552	1686	0.0879	0.1614	3056	0.0865	0.1878	7140
(L,S,L)	224	0.0127	0.0123	175	0.0145	0.0114	546	0.0262	0.0227	945
Total	8980	0.423	0.577	6762	0.423	0.577	11,144	0.360	0.640	26,886

Note: Excerpt from Gayle and Miller, 2015, Table 3. A report is classified as 'Good' if the firm's accounting is higher the expected value of accounting return—the yearly sample average for a firm type and sector—and 'Bad' otherwise. N is the number of observations. Firm type is measured by the triplicate (A, W, D), where A is assets, W is the number of workers, and D is the debt–equity ratio with each corresponding to whether that element is above (L) or below (S) its industry median.

Executive background

For each executive, the individual characteristics contain work experience, education (such as MBA, M.Sc., Ph.D., etc.), gender, age, and whether the executive is interlocked by being on the board or connected to the board.[4] Work experience is measured in several dimensions, including years of tenure in the firm, years worked as a top executive, and the number of firms in which the executive worked before and after becoming an executive.

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

Promotion hierarchy

Following the methodology outlined in Gayle *et al.*, 2012, managerial roles are tiered into a hierarchy of five ranks. They define a career hierarchy as a rational (complete and transitive) ordering over a set of job titles on the basis of transitions and independent of compensation. The idea is that transitions are more likely to occur from a lower rank to a higher rank; and that if rank A is higher than B, and B higher than C, then rank A is higher than C. Specifically, let J denote an unordered finite collection of job titles, denoted $j \in \{1, \ldots, J\}$. Denote the probability of switching from the j th to the k th job by p_{jk}. If $p_{kj} \geq p_{jk}$, they write $j \succeq k$, i.e. rank j is higher than rank k. The property of transitivity is imposed such that if $j \succeq k$ and $k \succeq j'$, then $j \succeq j'$. If $j \succeq k$ and $k \succeq j$, then $j \sim k$, in which case they say that there is indifference between the the j th job and the k th job. If $j \succeq k$ but $j \nsim k$, then $j \succ k$, in which case they say that the j th job ranks higher than the k th. Thus, indifference occurs if $p_{jk} = p_{kj}$ or if, for example, $p_{kj} > p_{jk}$ (implying $j \succeq k$) but there exists a j' such that $p_{j'k} \geq p_{kj'}$ and $p_{jj'} \geq p_{j'j}$ (which implies $k \succeq j$ by transitivity). An ordered rank is assigned to each of the distinct indifference sets, with rank 1 being the highest rank in the hierarchy.

Since there is only a finite set of jobs, the algorithm above ensures that the ranking is complete, that is, every pair of job titles in the set can be compared: for any two job titles j and k, either $j \succeq k$, or $k \succeq j$, or $j \sim k$. Another desirable property is that the algorithm gives the maximum number of ranks. If the condition for indifference is relaxed, for example, for some $p > 0$, let $-p \leq p_{kj} - p_{jk} \leq p$ be a sufficient condition for indifference between j and k, this leads to a coarser partition of the hierarchy.

The following is a rough description of the job titles in each rank.

Rank	**Description**
Rank 1	Chairman of the board of the company or chairman of a subsidiary who does not have any other executive positions in the firm.
Rank 2	CEO of the company.
Rank 3	COO, CFO, and chairman of the board of the company who holds some other executive position in the company other than CEO.
Rank 4	Other high-level corporate executives and heads of subsidiaries or regional chiefs.
Rank 5	Other lower-level executives.

Note that CEOs are not in *Rank 1* but instead in *Rank 2*: Since this hierarchy is based on transitions, this could reflect a lifecycle consideration more than control. However, it aligns with the institutional usage of the term Rank, which emphasizes the supervisory roles of managers over their subordinates—for example, the chairman of the board of directors (in *Rank 1*) monitors the CEO of the firm (in *Rank 2*). They set aside a position called *Execdir* in *Rank 3* for executives who serve on the board of directors of their own company. This is because directors are of special interest to corporate governance and their presence on the board brings their span of control into question, as is highlighted by the literature on the firm-size pay premium. More details on the construction of the hierarchy are provided in Gayle *et al.*, 2012.

3 Empirical patterns

Components of compensation

Table 2.4 provides a decomposition of total compensation for three industrial sectors: aerospace, chemicals, and electronics. The primary components of total compensation include salary and bonus, the value of restricted options granted, the value of restricted stock granted, changes in wealth from holdings of firm options, and changes in wealth from holdings of firm stock. Taken together, for CEO ranks, the first three components constitute approximately 80% of total compensation. In contrast, changes in wealth from holding firm options and stocks, while contributing less to the level of compensation, account for a significant portion of its variability. This pattern is much more pronounced for CEOs than for non-CEO ranks.

Salary and bonus increased nearly four times from the 1960s to the 2000s. Meanwhile, total compensation has grown by a magnitude higher. A major driver of this shift is the increase in options granted to managers. In the three sectors, the average

Table 2.4
Components of Compensation across Rank and Sample

	Rank	(1) 1944-1978 Three Sectors	(2) 1993-2002 Three Sectors	(3) 1993-2002 All Sectors
Total compensation	All	528	4,121	2,319
		(1,243)	(19,283)	(12,121)
	CEO	729	6,109	5,320
		(1,472)	(24,250)	(19,369)
	Non-CEO	400	2,256	1,562
		(1,026)	(12,729)	(9,303)
Salary and bonus	All	219	838	667
		(114)	(1,066)	(905)
	CEO	261	1,037	1,127
		(115)	(1,365)	(1,282)
	Non-CEO	179	640	552
		(97)	(576)	(738)
Value of options granted	All	79	2,401	903
		(338)	(13,225)	(3,753)
	CEO	111	3,402	1,782
		(439)	(18,172)	(7,169)
	Non-CEO	51	1,401	681
		(198)	(4,237)	(2,106)
Value of restricted stock granted	All	11	187	152
		(95)	(1,633)	(936)
	CEO	8	242	298
		(72)	(2,021)	(1,464)
	Non-CEO	13	133	115
		(112)	(1,118)	(743)
Change in wealth from options held	All	5	785	281
		(134)	(14,636)	(8,710)
	CEO	7	1,667	1,474
		(167)	(17,078)	(13,567)
	Non-CEO	3	-76	-18
		(94)	(11,706)	(6,939)
Change in wealth from stock held	All	-3	-40	125
		(439)	(5,681)	(4,350)
	CEO	0.434	-14	264
		(479)	(6,712)	(6,791)
	Non-CEO	-7	-64	90
		(398)	(4,496)	(3,473)

Note: Excerpt from Gayle and Miller, 2009a, Table 3 and Table 4. In thousands of US$ (2000). Standard deviations in parentheses. The source of the 1944-1978 sample is Antle and Smith, 1985. The sample covers three industrial sectors including aerospace, chemicals, and electronics. The 1993-2002 sample is based on S&P Execucomp, covering all firms in S&P 1500. To facilitate comparison with the old sample, Column (2) uses the new sample restricted to the firms in the three sectors in the Old sample based on the Global Industry Classification Standard (GICS). Column (3) is based on the unrestricted new sample.

value of the options granted has increased more than 30 times, making up more than half of the total compensation. A similar pattern arises across all sectors. In both samples in the 2000s, options grants represent the largest component of managerial pay. Stock grants, a relatively minor component, have further declined in importance—the ratio of the value of options granted to stock granted increased from 7:1 to 14:1 from the old sample to the new three-sector sample.

Holding financial securities in CEOs' own firms brings substantial uncertainty. Among all compensation components, changes in wealth from holding firm options exhibit the greatest dispersion. This contrasts to cash, bonuses, and grants, which are less directly related to firm performance and are partially explained by industry dynamics, firm size, and macroeconomic conditions such as GDP. Changes in wealth from holding firm stock also contribute significantly to overall volatility, with standard deviations exceeding those of other components. Over time, the standard deviations for changes in stock and option wealth have increased more than 100 times, underpinning the growing variability in managerial compensation.

Pay-performance sensitivity

Pay-performance sensitivity was estimated using ordinary least-squares regressions for all executives in the sample covering

1944 to 1978 from Antle and Smith, 1985, using various definitions of executive income. The results of these regressions are summarized in Table 2.5. Almost all estimated coefficients are statistically significant. Following the approach outlined by Jensen and Murphy, 1990, the sensitivity of executive compensation to changes in shareholder wealth is evaluated by predicting the change in compensation associated with a $1,000 increase in shareholder wealth, using the estimated linear models.

Assuming the change in salary and bonus as permanent, as in Column 4, after accounting for lagged effects, the effect of a $1,000 increase in shareholder wealth raises the executive compensation by $11.03 - $1.19 = $9.84. Furthermore, the return on stocks held contributes $4.00, while the return on options held adds another $2.73. Adding these components, a $1,000 increase in shareholder wealth due to favorable abnormal returns raises the most comprehensive measure of executive compensation by $9.84 + $4.00 + $2.73 = $16.57. This result is significantly higher than Jensen and Murphy's 1990 estimate of $3.25, yet it aligns with their broader conclusion: the sensitivity of CEO pay to performance remains relatively low in the 1960s. In general, shareholders are exposed to far greater financial risk than executives managing their firms.

Table 2.5
OLS Estimates of Pay-Performance Sensitivity Coefficients

	1 Change in Pretax Salary and Bonus	2 Change in Pretax Salary and Bonus	3 Change in Total Pay	4 Change in Sum of Total Pay and PV of Salary and Bonus	5 Return on Stock Held	6 Return on Options Held
	All	All	All	All	All	All
Intercept	3.577	3.639	−5.296	9.536	−3.700	6.423
	(0.4251)	(0.4297)	(22.050)	(18.069)	(8.020)	(2.439)
Change in shareholders' wealth from abnormal returns	0.00011	0.00007	0.00947	0.01103	0.00400	0.00253
	(0.00002)	(0.00003)	(0.00167)	(0.0014)	(0.0004)	(0.00018)
Lagged change in shareholders' wealth from abnormal returns		−0.000058	−0.00720	−0.00119		0.00020
		(0.00002)	(0.0011)	(0.00099)		(0.00012)

Note: Excerpt from Margiotta and Miller, 2000, Table 6. Dependent Variables in Thousands of 1967 US$. Standard Errors in Parentheses. The variable change in shareholders' wealth from abnormal returns was constructed by multiplying abnormal returns by common equity. The variable total pay is after-tax compensation minus return on stock held and return on options held. The variable PV (salary and bonus) is the present value of an annuity equal to the change in pretax salary and bonus starting the following year and ending the year in which the executive turns 65, computed using the interest rates during the period.

Hidden Information

If the state of production is private information of the manager, it poses an incentive for managers to provide misleading accounting reports. Table 2.6 provides a cross-sectional summary of abnormal returns and compensation conditional on the reported state, by sector and firm type. With no exceptions, the average abnormal return and average compensation are both higher when a good state is reported.

Insider holdings may perform better if managers can leverage insider information to adjust their insider holdings for their personal gain. If so, it is reasonable to hypothesize that the information embedded in future returns could help explain the portfolio choices managers have made in advance. For a risk-averse manager that maximizes expected utility, insider trading is optimal if and only if the manager has insider information about the abnormal return in the next period. The results presented in Table 2.7 support this hypothesis. The coefficient of the lead abnormal return, 2.30, is positive and statistically significant, as predicted for a risk-averse expected utility maximizer. This suggests that managers do possess insider information about their firm's state of production.

Table 2.6
Returns and Compensation by Firm Type and Sector

Firm type	Primary		Consumer		Service	
(A,W,D)	Good	Bad	Good	Bad	Good	Bad
			Abnormal Return			
(S,S,S)	0.11	-0.10	0.10	-0.15	0.28	-0.05
	(0.56)	(0.43)	(0.57)	(0.47)	(0.94)	(0.70)
(S,L,S)	-0.03	-0.14	-0.01	-0.07	0.02	-0.10
	(0.40)	(0.34)	(0.31)	(0.35)	(0.43)	(0.46)
(S,L,L)	-0.03	-0.11	0.07	-0.12	0.10	-0.05
	(0.36)	(0.37)	(0.38)	(0.36)	(0.41)	(0.34)
(S,S,L)	0.07	-0.11	0.05	-0.11	0.20	-0.11
	(0.52)	(0.44)	(0.62)	(0.55)	(0.75)	(0.82)
(L,S,S)	-0.01	-0.10	0.01	-0.12	0.08	-0.09
	(0.34)	(0.39)	(0.41)	(0.40)	(0.71)	(0.52)
(L,L,S)	-0.07	-0.11	-0.06	-0.12	0.15	-0.05
	(0.29)	(0.33)	(0.32)	(0.34)	(0.61)	(0.47)
(L,L,L)	-0.03	-0.13	-0.01	-0.16	0.02	-0.06
	(0.27)	(0.30)	(0.41)	(0.38)	(0.32)	(0.37)
(L,S,L)	0.02	-0.13	0.07	-0.24	0.13	-0.12
	(0.30)	(0.40)	(0.47)	(0.49)	(0.85)	(0.59)
			Compensation			
(S,S,S)	3889	670	3397	−1501	6063	1701
	(14,651)	(10,779)	(19,178)	(15,235)	(20,034)	(17,316)
(S,L,S)	4384	2339	4922	−486	8015	−1183
	(9381)	(14,243)	(30,677)	(23,882)	(24,615)	(25,740)
(S,L,L)	3742	521	9194	821	7096	2274
	(11,903)	(15,710)	(19,898)	(11,820)	(14,740)	(14,363)
(S,S,L)	2522	721	3977	908	4154	−150
	(9855)	(8851)	(14,844)	(11,504)	(16,068)	(14,255)
(L,S,S)	3079	−850	4235	−510	3386	1629
	(20,381)	(15,773)	(20,107)	(16,940)	(18,844)	(19,287)
(L,L,S)	4154	2422	4727	−429	8035	5496
	(13,375)	(16,220)	(20,989)	(21,784)	(24,244)	(26,472)
(L,L,L)	5781	2200	6897	2775	9846	5595
	(12,807)	(12,208)	(19,288)	(19,118)	(24,075)	(19,936)
(L,S,L)	4396	−3729	4742	−2442	5647	1718
	(14,831)	(18,890)	(19,288)	(14,448)	(20,347)	(17,612)

Note: Excerpt from Gayle and Miller, 2015, Table 4. A report is classified as 'Good' if the firm's accounting is higher the expected value of accounting return—the yearly sample average for a firm type and sector—and 'Bad' otherwise. Abnormal return is firm's stock returns less the return on the market portfolio. Compensation in thousands of 2000 US\$. Firm type is measured by the triplicate (A, W, D), where A is assets, W in number of workers, and D is debt–equity ratio with each corresponding to whether that element is above (L) or below (S) its industry average. Standard deviations are in parentheses.

Table 2.7
Return Predictability of Managers' Stock Holdings Changes

	Yearly Change in Number of Shares Managers Hold in Their Firms
Ratio of Salary and Bonus to Total Compensation	-0.77
	(2.13)
Lead Abnormal Return	2.30
	(1.11)
Constant	80.34
	(50.21)
R^2	0.12
Observations	67,769

Note: Excerpt from Gayle and Miller, 2009b, Table 3.

Regulatory effects

In 2002, the US government enacted the Sarbanes–Oxley (SOX) Act as a response to corporate governance failures in many prominent companies. The SOX legislation brought greater accountability to financial statements, more rigorous enforcement of property rights in governance, and higher penalties for fraud, discouraging managers from breaking the law. An empirical analysis of S&P 1500 firms shows that SOX reduced the conflict of interest between shareholders and their CEOs.

Table 2.8 presents the pre- and post-SOX summary of firm characteristics and compensation by type, with the compensation further tabulated by the accounting reports. The main sample covers the pre-SOX era from 1993 to 2001 and the post-SOX era as the years 2004 and 2005—omitting the two years 2002 and 2003, a transition phase of the legislation. All monetary values are adjusted to US dollars (2006). The top panel shows that firms in the service sector have more assets and higher leverage on average than the other two sectors. All three sectors experienced significant growth. The debt-to-equity ratio rose in the primary sector while declining in the service sector.

The bottom panel provides a comparison of total compensation, across sectors, and from the pre- to post-SOX era, conditional on the reported state. The average compensation is higher in large firms, although this difference is fully explained by the risk premium (Gayle and Miller, 2015). Compensation is also higher in the service sector than in the other two sectors in most types and accounting states. Compensation is consistently lower in the bad state[5] for all types. Following the implementation of SOX, average compensation increased significantly in the primary sector but remained unchanged in the other two sectors. The variance in compensation within each type either declined or remained unchanged[6] from the pre- to post-SOX eras.

The structural estimation of the effects of SOX is deferred to Section 8.3. The gross financial loss to shareholders if the CEO is not incentivized is reduced; the benefits to CEOs from pursuing their own goals on the job within the firm if their

[5] In the bad state, compensation is negative for several firm types, reflecting substantial losses incurred by CEOs when the value of the stocks and options they hold in their firms declines sharply.

[6] There is a single exception—highly leveraged large firms in the service sector during the good state.

Table 2.8
Firm Characteristics and Compensation

	Primary Sector			Consumer Sector			Service Sector		
	Pre	Post	t-/F-stat	Pre	Post	t-/F-stat	Pre	Post	t-/F-stat
Total assets	4460	6971	6.7	3460	5091	4.0	13,678	20,311	3.3
	(7436)	(11,549)	(0.4)	(8242)	(11,358)	(0.5)	(42,619)	(78,265)	(0.3)
Debt-to-equity	1.792	2.092	3.5	1.559	1.498	−0.9	3.651	2.798	−6.9
	(1.428)	(2.646)	(0.3)	(1.589)	(1.779)	(0.8)	(5.273)	(4.231)	(1.6)
Accounting return	1.110	1.135	3.2	1.117	1.070	−4.7	1.180	1.190	−11.7
	(0.233)	(0.231)	(1.0)	(0.283)	(0.258)	(1.2)	(0.342)	(0.258)	(1.8)

		Total Compensation								
		Overall			Bad			Good		
		Pre	Post	t-/F-stat	Pre	Post	t-/F-stat	Pre	Post	t-/F-stat
Primary	(S, S)	1837	4704	3.6	387	2268	2.1	3593	8600	3.5
		(12,126)	(13,797)	(0.8)	(8132)	(12,283)	(0.4)	(15,480)	(15,174)	(1.0)
	(S, L)	1099	4675	4.7	−131	3173	4.5	2818	6927	2.4
		(7970)	(8906)	(0.8)	(7520)	(5679)	(1.8)	(8267)	(11,970)	(0.5)
	(L, S)	4300	10,074	4.4	3225	8525	3.3	5772	11,924	8.7
		(11,983)	(16,369)	(0.5)	(9827)	(15,074)	(0.4)	(14,305)	(17,717)	(0.7)
	(L, L)	4204	8362	5.2	3147	7106	4.3	5595	10,115	3.2
		(11,571)	(14,816)	(0.6)	(10,662)	(12,921)	(0.7)	(12,540)	(17,003)	(0.5)
Consumer Goods	(S, S)	1425	2151	0.6	−1874	−2903	−0.9	6056	7875	0.9
		(21,403)	(18,474)	(1.3)	(15,927)	(12,315)	(1.7)	(26,629)	(22,279)	(1.4)
	(S, L)	1702	2355	0.4	−691	−1516	−0.4	4754	5120	0.1
		(14,362)	(13,605)	(1.1)	(11,263)	(9682)	(1.4)	(17,079)	(15,316)	(1.2)
	(L, S)	6074	6793	0.2	1355	1399	0.0	12,012	10,903	−0.2
		(32,090)	(31,953)	(1.0)	(23,206)	(24,290)	(0.9)	(39,848)	(36,345)	(1.2)
	(L, L)	7297	9015	1.0	3460	4590	0.6	12,710	13,283	0.2
		(26,565)	(28,798)	(0.9)	(21,908)	(22,864)	(0.9)	(31,228)	(33,068)	(0.9)
Service	(S, S)	3757	2149	−1.9	590	−953	−1.8	7766	5700	−1.3
		(21,304)	(18,728)	(1.3)	(15,543)	(14,180)	(1.2)	(26,350)	(22,352)	(1.4)
	(S, L)	3311	4318	0.6	2041	1725	−0.2	4706	7579	0.9
		(16,672)	(19,656)	(0.7)	(13,242)	(12,581)	(1.1)	(19,696)	(25,703)	(0.6)
	(L, S)	11,231	7135	−1.6	6065	205	−2.1	18,877	14,515	−1.0
		(38,738)	(32,155)	(1.5)	(31,249)	(25,182)	(1.5)	(46,727)	(36,918)	(1.6)
	(L, L)	9438	9185	−0.2	6383	6221	−0.1	14,114	14,365	0.1
		(26,040)	(24,173)	(1.2)	(22,689)	(19,051)	(1.4)	(29,875)	(30,536)	(1.0)

Note: Excerpt from Gayle *et al.*, 2022, Table 1. In the columns "Pre" and "Post" indicating the pre- and post- SOX eras, standard deviation is listed in parentheses below the corresponding mean. The columns "t-/F-stat" report the statistics of a two-sided t-test on equal mean with critical value equal to 1.96 at the 5% confidence level, and the one-sided F-test on equal variance with critical value equal to 1. Firm type is measured by the coordinate pair (A, C) where A is assets and C is the debt-to-equity ratio with each corresponding to whether that element is above (L) or below (S) its industry median. Accounting return is classified as "Good (Bad)" if it is greater (less) than the industry average. Assets (Compensation) is measured in millions (thousands) of 2006 U.S. dollars.

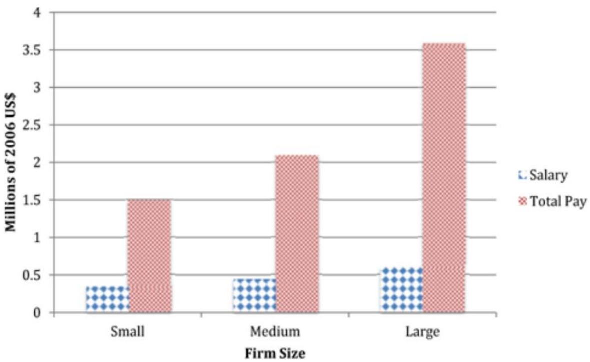
compensation is fixed becomes less diffuse across diverse firm types. This suggests that strengthening regulations to impose harsher penalties for misreporting can help mitigate CEOs' misaligned incentive.

Lifecycle patterns

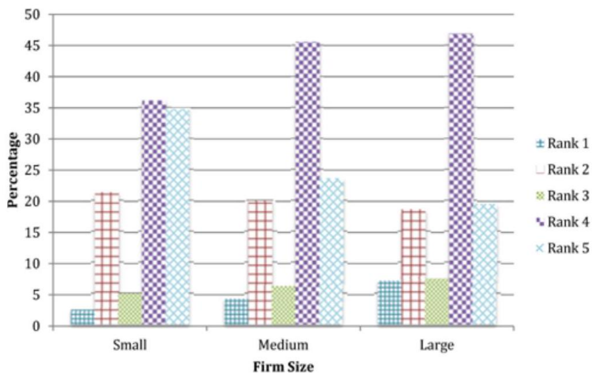
Firm-size differences arise not only in compensation but also in hierarchy, education, and experience. Figure 2.1(a) shows that both the total compensation and the base salary increase with the size of the firm. However, total compensation grows at a much higher rate than salary. For example, the average 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 does compensation increase with firm size, so does incentive pay. Figure 2.1(b) shows that hierarchy also varies with firm size. For example, large firms are more likely than small firms to separate the roles of CEO (*Rank* 2) from chairman of the board (*Rank* 1). This might indicate that large firms face more significant monitoring challenges than small firms, a hypothesis that has been suggested in the literature as a reason for the firm-size pay premium. Also, *Rank* 5 is more common in a small firm than in a large firm, while the reverse is true for *Rank* 4.

Figure 2.1
Pay and Hierarchy by Firm Size

(a) Firm Size Pay Premium



(b) Hierarchy by Firm Size

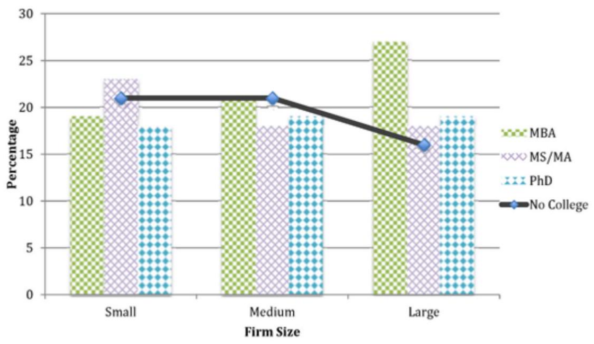


Note: Excerpt from Gayle *et al.*, 2015, Figure 1. This figure contains two bar charts illustrating the pay and hierarchy by firm size. The top bar chart, titled "Firm size pay premium," compares salaries and total pay across small, medium, and large firms. The bars show that both salary and total pay increase with firm size, with large firms offering significantly higher total pay. The bottom bar chart, titled "Hierarchy by firm size," represents the distribution of employees across different hierarchical ranks (Rank 1 to Rank 5) within small, medium, and large firms. The stacked bars indicate that higher-ranking employees (e.g., Rank 1) are more prevalent in larger firms.

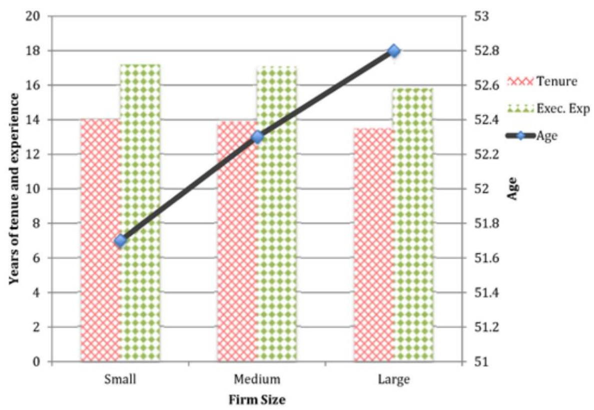
Figure 2.2(a) shows that executives in large firms tend to have more formal education (college and above) than executives in small firms. Among executives with formal education, there are also differences in education by firm size. Although the distribution of executives with a PhD is similar across firm size, large firms exhibit a higher concentration of executives with an MBA but a lower concentration of non-business master's degrees. This might suggest that large firms have a higher demand for managerial expertise. However, Figure 2.2(b) reveals a trade-off, as both tenure and years of executive experience decrease with firm size. On the flip side, age increases with firm size. Together, Figures 2.2(a) and 2.2(b) are consistent with the value of schooling (Mincer, 1974): Executives in large firms have fewer years of managerial experience, but more age because they acquired more formal education. Our data concern upper-management executives in publicly held companies, which limits the extent to which they can infer about the incubation and selection into executive positions. However, conditional on their human capital upon entering management, they can infer the value of human capital acquired through on-the-job experience by investigating the transitions within the executive hierarchy and their subsequent career trajectories.

Figure 2.2
Education and Experience by Firm Size

(a) Education and Firm Size



(b) Experience and Firm Size



Note: Excerpt from Gayle *et al.*, 2015, Figure 2. The figure contains two bar charts illustrating the relationship between education, experience, and firm size. The top bar chart, titled "Education and firm size," presents the percentage distribution of employees with different educational backgrounds (MBA, MS/MA, PhD, and No College) across small, medium, and large firms. The bars indicate variations in educational attainment across firm sizes, with larger firms having more employees with MBA and fewer without a college degree, and small firms having more employees with PhD. The bottom bar chart, titled "Experience and firm size," shows the years of tenure and experience of employees in firms of different sizes. The bar segments represent tenure and executive experience, while a black line traces the average age of employees. The chart indicates that executive age increases with firm size while tenure and average executive experience declines slightly with firm size.

Patterns across rank

Table 2.9 breaks down 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. The 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 MBA than an executive in any other rank. Salary, total compensation, and the likelihood of being a board member increase with rank advancing, peak at *Rank* 2, and then decline at *Rank* 1.

To account for the interactions between firm size and hierarchy (Figure 2.1), education (Figure 2.2), and experience (Figure 2.2), Gayle *et al.*, 2015, estimate the effects using conditioning information in regressions in Table 2.10 on four dependent variables: compensation and three indicators of job mobility. The first regression breaks down compensation into fixed and variable components (second-order polynomial terms of the firm's performance) as regressors in the first three columns. The ex-

Table 2.9
Executive Characteristics by Rank

	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	0.060
Tenure	15.784 (11.708)	14.412 (10.672)	13.388 (10.062)	13.383 (9.754)	13.221 (9.541)
Executive experience	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	2945 (26,035)	4794 (26,701)	3717 (19,009)	1844 (11,644)	1269 (9438)
Observations	4358	20,983	5620	28,271	15,972

Note: Excerpt from Gayle *et al.*, 2015, Table 1. 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.

planatory variables include rank (panel A), firm type (panel B), and human capital plus individual heterogeneity (panel C). The second regression summarizes promotion, and estimates a multivariable logit of the probability of each rank. The third and fourth regressions are logit that explains the probability of changing firms and retirement, respectively. Panel A of Table 2.10 demonstrates that the firm-size pay premium holds even after accounting for these interactions. Panel B of Table 2.10 shows that the fixed pay and incentive pay are both higher in larger firms, highest in the service sector, and lowest in the primary sector; companies with more insiders on their board of directors offer higher incentive pay, despite the same level of fixed pay.

Table 2.10
Compensation and Mobility

	Compensation			Promotion						
	π	π^2	Level	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Turnover	Retirement
Panel A: Executive Position										
Constant	21,601	−9114	−4359							
	(3859)	(1914)	(2716)							
Rank 4	1529	−242	103	−20.3	−76.2	−67.8	63.4	−536	–	–
	(926)	(444)	(463)	(10.4)	(4.01)	(6.51)	(4.43)	(14.6)		
Rank 3	2627	−164	1267	−88.1	−72.4	114	−404	−754	94.7	–
	(1407)	(605)	(662)	(24.1)	(12.8)	(19.7)	(22.4)	(18.2)	(17.6)	
Rank 2	6007	−789	3456	−118	67.7	−393	−551	−901	213	−8.9
	(1394)	(699)	(683)	(25.1)	(3.90)	(20.0)	(33.6)	(39.0)	(12.6)	(4.28)
Rank 1	9839	−454	1055	111	−290	−345	−585	−939	86.1	55.26
	(1690)	(987)	(797)	(34.2)	(8.15)	(33.8)	(33.9)	(39.0)	(23.2)	(3.23)
Execdir	7695	−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	3149	88	777	–	–	–	–	–	–	–
	(419)	(222)	(198)							
Primary	−3609	1537	−633	–	–	–	–	–	−18.68	(7.11)
	(473)	(267)	(198)							
Medium-sized firm	4170	−253	937	–	–	–	–	–	–	–
	(437)	(201)	(214)							
Large firm	12,703	−2224	3697	–	–	–	–	–	–	–
	(405)	(212)	(190)							
Large board	2683	−1203	280	–	–	–	–	–	−25.7	(5.71)
	(358)	(176)	(163)							

(Continues)

Panel C of Table 2.10 demonstrates the significance of human capital in executive compensation. The effect of tenure is highly nonlinear and varies by rank. Tenure in a given rank affects the variable pay but not the fixed pay, similarly for years of executive experience. The last seven columns of Table 2.10 show that human capital affects promotion, turnover, and exit, whereas firm size does not have such effects.

In summary, human capital, promotion, and turnover all play an important role in executive compensation, in addition to firm

Compensation and Mobility (Continued)

	Compensation			Promotion						
	π	π^2	Level	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Turnover	Retirement
Panel C: Human Capital and Individual Heterogeneity										
Rank 1 Lagged	12,085	−3054	544	−9.66	−32.8	1.01	14.04	34.76	82.8	24.44
	(1769)	(987)	(822)	(14.5)	(8.93)	(11.79)	(7.00)	(11.33)	(13.4)	(5.69)
Rank 2 Lagged	14,640	−2875	660	–	–	–	–	–	–	–
	(1342)	(625)	(658)							
Rank 3 Lagged	4849	−1100	597	–	–	–	–	–	–	–
	(1389)	(586)	(653)							
Exec. exp.	191	−42	1.61	–	–	–	–	–	–	–
	(26)	(14)	(25)							
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	2155	−1026	–	–	–	–	–	–	–
	(996)	(478)	(1255)							
Rank 2 × External	–	–	–	3840	–	–	–	–	–	–
				(1459)						
Rank 3 × External	–	–	–	5289	–	–	–	–	–	–
				(1975)						
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	1024	174	−459	−847	1948	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	–
	–	–	(1295)						(0.24)	
MBA	–	–	–	–	–	–	–	–	–	−84.46
	–	–	–	–	–	–	–	–	–	(20.1)
Interlocked	6403	−1496	−299	–	–	–	–	–	−93.0	−93.0
	(995)	(471)	(464)						(28.6)	(28.6)

Note: Excerpt from Gayle *et al.*, 2015, Table 2. Standard errors are listed in parentheses; tenure and 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. The elasticities are calculated using logit regressions.

and sector variation. Firm size and sector differences primarily influence compensation—rather than promotion, turnover, or exit, suggesting that a static model of compensating differentials might account for them.[7] However, older executives are more likely to be in higher ranks and paid more in both fixed and variable pay controlling for the rank, while aging accelerates exit. This is more compatible with a non-stationary dynamic model with career concerns in which aging executives have increasing productivity but declining willingness to remain employed with the firm. Turnover complicates the picture because newly hired executives at *Ranks* 2 and 3 receive a substantial sign-on bonus, packaged with declining compensation as their tenure increases. Similarly, newly hired executives seem to have a break-in period when their compensation is less performance-oriented than peers with a longer tenure at the given rank. Finally, the variation in rank distribution and human capital across firm sizes suggests that evaluating the determinants of the firm-size pay premium requires a comprehensive model that accounts for all these factors.

[7]A noteworthy exception is that the primary sector exhibits lower mobility, which could be due to technological considerations and specialized training.

Gender differences in compensation and promotion

While there is a large literature on gender gaps in the labor market, few studies focus on the gender gap for top executives in publicly traded firms, such as Bertrand and Hallock, 2001. Using the rich background characteristics of the executive in their matched sample, Gayle *et al.*, 2012, study the gender differences in executive compensation and promotion. They find that fewer women than men become executives, and women executives tend to earn less over the course of their careers. Additionally, women are more likely to hold junior managerial positions compared to their male counterparts and have a higher likelihood of exiting the occupation earlier. However, after controlling for executive rank and background, they reveal that female executives earn more than men on average. Despite this, women who remain in these positions are more likely to be promoted to higher ranks, notably, to CEO. In general, while women face significant barriers to reaching executive positions and higher exit rates, contributing to the observed gender gap in executive compensation, those who manage to remain in the managerial profession are more likely to rise to the top.

3

A Static Model of Pure Moral Hazard

Much of the intuition supporting the structural estimation of principal agent models is conveyed by the following simple model. At the beginning of the period, a risk-neutral *principal*, representing a firm or shareholders for example, proposes to a risk-averse *agent*, such as a CEO, a *compensation plan* that depends on revenue to the principal realized at the end of the period. The plan may be an explicit *contract* or an implicit agreement. The agent decides whether to accept or reject the principal's (implicit) offer. If he rejects the offer, he receives a fixed utility from an *outside option*. If he accepts the offer, the agent chooses between maximizing the principal's expected revenue, called *working*, and accepting employment from the principal but following the objectives he would pursue if he were paid a fixed wage, called *shirking*. The decision to accept or reject the offer is observed by the principal, but the work routine is not. After revenue is realized at the end of the period, the agent receives compensation according to the explicit contract or implicit agreement, and the remaining revenue is profit to the principal. A notion of commitment by the principal is embedded within this static model: after the agent has made his choice, there are gains from renegotiating the contract to eliminate the uncertainty the agent is exposed to under the optimal contract we derive below. We assume throughout this review that public corporations can engage in these kinds of commitments.

The assumption of a risk-neutral principal roughly corresponds to shareholders having diversified portfolios. Moral hazard arises since the agent (manager) chooses whether to work or shirk, which remains unobserved by shareholders. Assuming a risk-averse agent makes sense, since more than 75% of the variation in managerial compensation can be explained by firms' excess return (Gayle and Miller, 2009a).

Choices. We denote the agent's workplace employment decision by an indicator $l_0 \in \{0, 1\}$, where $l_0 = 1$ means the agent rejects the principal's offer and pursues an outside option. Working is defined by the indicator $l_2 \in \{0, 1\}$, where $l_2 = 1$ means the agent works, and shirking is defined by the indicator $l_1 \in \{0, 1\}$, where $l_1 = 1$ means the agent shirks. Since taking the outside option, working and shirking are mutually exclusive activities, $l_0 + l_1 + l_2 = 1$.[1] In the context of managerial compensation, working means following the objectives of the principal; shirking means following his own objectives when he is paid a fixed wage to be employed. The data available for empirical analysis largely dictate this restrictive choice set. A more flexible specification of the agent's choice set would define a vector space of activities in which the agent exercises discretion in selecting a level for each activity. However, researchers typically do not observe anything that meaningfully corresponds to these dimensions, let alone its level.

Compensation. Let x denote a random variable drawn from a probability distribution determined by the agent's work routine. After x is revealed to both the principal and the agent at the end of the period, the agent receives compensation according to the contract or the implicit agreement. To reflect its potential dependence on x, we denote compensation by $w(x)$. Key to interpreting $w(x)$ in empirical work on managerial compensation is that this function defines how the agent's wealth would change through his employment by the principal if he did not consume anything: for example, supposing the CEO holds firm specific

[1] We maintain the (l_0, l_1, l_2) notation throughout to facilitate the heterogeneity analysis in later sections.

financial securities (such as stock and options in the firm) that he would otherwise trade for a more diversified portfolio at the beginning of the period, then adjustments to his wealth at the end of the period should include the financial consequences of holding a more specialized portfolio.

Production and profits. In this simple model, we assume that x is gross revenue to the principal, that the principal's profit is revenue less compensation to the agent, $x - w(x)$, and that the principal maximizes expected profits. In the empirical analyses we discuss below, x denotes financial returns to the firm, the sum of capital gains and dividends multiplied by shareholder equity that would apply if the agent was paid nothing. Denoting by V the value of shareholder equity at the beginning of the period, and π the financial return on its assets, $x = V\pi + w(x)$.

Denote by $f(x)$ the probability density function for x conditional on the agent working, and let $f(x)g(x)$ denote the probability density function for revenue when the agent shirks. The assumption of profit maximization and the definition of working imply

$$E[x] \equiv \int xf(x)\mathrm{d}x > \int xf(x)g(x)\mathrm{d}x \equiv E[xg(x)]. \qquad (3.1)$$

The firm maximizes expected profits by inducing the agent to work, but faces moral hazard: the agent's actions cannot be directly contracted upon since they are unobserved. Instead, the firm must rely on the observed revenue x as an imperfect signal of the agent's actions, since the probability distribution of revenue differs between working, $f(x)$, and shirking, $f(x)g(x)$.

The span of control Since $f(x)$ and $f(x)g(x)$ are densities of revenue that characterize the production technology with working and with shirking, the ratio of the two densities, $g(x)$, is a likelihood ratio—$g(x)$ is nonnegative for all x and

$$E[g(x)] \equiv \int g(x)f(x)\mathrm{d}x = 1. \tag{3.2}$$

This likelihood ratio measures the degree to which executive effort can affect a firm's returns in a discretionary manner by shirking versus working, so can be interpreted as a measure of his *span of control.*

Assume there is an upper range of revenue that might be achieved from working but is extremely unlikely to occur if the agent shirks. Formally

$$\lim_{x\to\infty}[g(x)] = 0. \tag{3.3}$$

Intuitively, this assumption states that truly extraordinary performance can only be attained if the agent works.[2] We further

[2]This is an assumption of convenience that in principle can be relaxed at the expense of adding an additional parameter to be treated at the identification and estimation stages of the empirical analysis. Thus, (3.3)

impose that $g(x)$ is bounded. These two regularity conditions on $g(x)$, ensure that the distinction between the two likelihoods is informative about the agent's action, but does not allow the principal to distill shirking from working fully on set of x with positive measure.[3]

The premise of our econometric analysis is that shareholders (supported by the compensation committee) design contracts reflecting their interests, and that CEOs respond rationally, while the actions of CEOs are unobserved by the shareholders and thus uncontractible. Shareholders calculate the expected gross benefits from employing a CEO to shirk and also to work, offset those benefits with the expected CEO compensation from the respective cost-minimizing contracts, and select the action that maximizes expected firm value if it is positive. The contract is not based on whether the CEO works or shirks, which the

---

could be replaced with $g(x) \to c$ as $x \to \infty$ for some $c \geq 0$.

[3]If $g(x)$ was unbounded for all x in an open interval (x_1, x_2), then $f(x) = 0$ for those values of $x \in (x_1, x_2)$, and the principal could be assured the agent definitely shirked. In this case, the agent could be induced to work for a fixed wage by punishing the agent sufficiently harshly if $x \in (x_1, x_2)$, generating a first-best solution to the agency problem, as Mirrlees, 1976, first explained. This is inconsistent with the empirical pattern that compensation varies systematically with revenue. Hence, we study situations where the first best is not attainable, and accordingly impose that $g(x)$ is bounded to rule out the first-best solution inconsistent with the data. However, the simple model does not rule out situations in which executive management is penalized for some outcomes that are either internally administered (such as dismissal coupled with failing to vest financial securities previously granted) or externally imposed (perhaps subject to criminal prosecution), provided that the penalties are not sufficiently severe to deter the agent from shirking unless additional incentives are provided.

principal does not observe. Rather, shareholders design the contract based on contractable events, such as the performance of the firm, which are publicly observed.

In competitive executive labor markets, the principal's optimization problem is constrained by information asymmetry: the principal cannot observe the agent's action and thus cannot achieve the first-best outcome. Rather than eliminating information rents entirely, competitive labor markets limit such rents and ensure that executives receive compensation commensurate with their outside options. The additional pay moving from first-best—where actions are observable and contractible—to second-best outcomes—where the principal must design incentive-compatible contracts to induce the preferred action under asymmetric information—represents the cost of moral hazard.

1 The agent's maximization problem

In the simplest of models, the agent maximizes his expected utility exponential in compensation

$$-l_0 - l_1 \beta E\left[e^{-\gamma w(x)} g(x)\right] - l_2 \alpha E\left[e^{-\gamma w(x)}\right], \qquad (3.4)$$

where γ is the coefficient of absolute risk aversion, α is a utility parameter measuring the distaste from working, while β is a similar parameter measuring a distaste for shirking. The definitions of working and shirking imply $\alpha > \beta$. Normalizing the utility of the outside option to negative one is without loss of generality.

2 The firm's cost minimization

The firm's primary objective is to maximize firm value by designing optimal contracts that induce the agent to work diligently. This involves solving a sequential problem: first, for any given action (work or shirk), the firm determines the cost-minimizing contract; then, it selects the action that maximizes net firm value after accounting for compensation costs. Cost minimization is thus a sub-problem within the broader goal of profit maximization.

To induce the agent to shirk versus taking the outside option, it suffices to propose a contract giving the agent an expected utility of at least minus one (the utility from taking the outside option). With reference to (3.4), to meet this *participation constraint*, $w(x)$ must satisfy the inequality

$$\beta E\left[e^{-\gamma w(x)} g(x)\right] \leq 1. \tag{3.5}$$

To elicit work from the agent, the principal must offer a contract that gives the agent a higher expected utility than shirking provides. The inequality defining this condition is called the *incentive compatibility constraint*

$$\alpha E\left[e^{-\gamma w(x)}\right] \leq \beta E\left[e^{-\gamma w(x)} g(x)\right]. \tag{3.6}$$

The firm's mechanism design problem can be stated concisely as follows: To induce working, the principal chooses a compensation schedule $w(x)$ to maximize expected profits $E[x - w(x)]$ which is equivalent to minimizing expected cost $E[w(x)]$ subject to the agent's participation constraint (3.5) and incentive compatibility constraint (3.6). To induce shirking, the principal only needs to satisfy the participation constraint (3.5) while minimizing $E[w(x)g(x)]$. The solution to each cost-minimization problem, characterized in the following theorem, determines the optimal contracts for working and shirking respectively. The firm then compares the net profits under each contract to determine which action to induce.

Theorem 3.1 (Optimal Contract for Pure Moral Hazard Model, Equation (10) in Gayle and Miller, 2009a)**.** The minimal cost of employing an agent to shirk is $\gamma^{-1} \ln \beta$, which gives him just enough to be indifferent to the outside option. To minimize the cost of inducing the agent to accept employment and work diligently, the board offers the contract

$$w^o(x) \equiv \gamma^{-1} \ln \alpha + \gamma^{-1} \ln \left[1 + \lambda \left(\frac{\alpha}{\beta}\right) - \lambda g(x)\right] \tag{3.7}$$

which depends on the revenue x. The Lagrange multiplier λ is a constant given by the unique positive solution to

$$E\left[\frac{g(x)}{\alpha + \lambda[(\alpha/\beta) - g(x)]}\right] = E\left[\frac{(\alpha/\beta)}{\alpha + \lambda[(\alpha/\beta) - g(x)]}\right]. \quad (3.8)$$

This theorem is the static model version of Propositions 3 and 4, Margiotta and Miller, 2000. It is straightforward to prove that the cost-minimizing contract for employing a CEO to shirk is $\gamma^{-1} \ln \beta$. Because the data on CEO compensation shows that their pay depends on the firm's excess returns, we focus on the problem of deriving the cost-minimization problem for inducing the CEO to work. To derive (3.7) and (3.8), we define $v(x) \equiv \exp[-\gamma w(x)]$ and denote things in utility terms in our derivation. To see things in compensation terms, $w(x) = -\frac{1}{\gamma} \log v(x)$.

The participation constraint for work can be expressed as

$$\alpha E[v(x)] \leq 1 \quad (3.9)$$

where $v(x)$ is such that the agent weakly prefers working (gets less disutility) to the outside option.

Similarly, the incentive-compatibility constraint for work can be expressed as

$$\alpha E[v(x)] \leq \beta E[v(x)g(x)] \quad (3.10)$$

which means that the agent weakly prefers working to shirking.

To minimize expected compensation subject to (3.9) and (3.10), we choose $v(x)$ to maximize the Lagrangian[4]

$$E\{\log[v(x)]\} + \lambda_0 E\left[1 - \alpha v(x)\right] + \lambda_1 E\left[\beta g(x) v(x) - \alpha v(x)\right]. \tag{3.11}$$

The first-order condition is, for any x in its support,

$$v(x)^{-1} = \lambda_0 \alpha + \lambda_1 \alpha - \lambda_1 \beta g(x). \tag{3.12}$$

Multiplying both sides by $v(x)$ and taking expectations yields

$$1 = \lambda_0 \alpha E[v(x)]$$

since the complementary-slackness condition for incentive compatibility implies

$$\lambda_1 E\left[\beta g(x) v(x) - \alpha v(x)\right] = 0.$$

The complementary-slackness condition for the participation constraint is given by

$$\lambda_0 \alpha E[v(x)] - \lambda_0 = 0,$$

[4]The use of Lagrangian methods in this mechanism design problem requires certain technical conditions to ensure the validity of the first-order approach. The boundedness condition on $g(x)$ imposed earlier is sufficient to guarantee that the agent's optimization problem is well-behaved and that the first-order conditions characterize the global optimum.

and substituting 1 for $\lambda_0 \alpha E[v(x)]$ into the above proves that $\lambda_0 = 1$ and consequently

$$\alpha E[v(x)] = 1. \tag{3.13}$$

Thus, the first-order condition simplifies to

$$\begin{aligned} v(x)^{-1} &= \alpha + \lambda_1 \alpha - \lambda_1 \beta g(x) \\ &= \alpha \left[1 + \lambda \left(\alpha / \beta \right) - \lambda g(x) \right], \end{aligned} \tag{3.14}$$

where $\lambda \equiv \lambda_1 \beta / \alpha$. Substituting for $v(x) \equiv \exp[-\gamma w(x)]$ and taking logarithms then yields (3.7), the optimal work compensation equation. A contradiction argument establishes that the incentive-compatibility constraint (3.10) also holds with equality. Substituting (3.14) into the incentive-compatibility condition and imposing equality gives the solution to λ, namely (3.8).

Similarly, the optimal contract for shirking is found by setting $\lambda_1 = 0$, i.e. to remove the incentive compatibility constraint for working, and substituting β for α in (3.11) to impose the participation constraint for shirking, then solving for the first-order condition to obtain the shirking contract, $w(x) = \gamma^{-1} \log(\beta)$.

The derivation above shows the participation constraint is met with equality in both cases, pinning down the certainty-equivalent wage. There is no point in exposing the agent to uncertainty in a shirking contract by tying compensation to revenue. Hence, an agent paid to shirk is offered a fixed wage that just offsets his nonpecuniary benefits, $\gamma^{-1}\ln\beta$. The certainty equivalent of the cost-minimizing contract that induces diligent work is $\gamma^{-1}\ln\alpha$, higher than the optimal shirking contract to compensate for the lower nonpecuniary benefits because $\alpha > \beta$. Moreover, the agent is paid a positive risk premium of $E\left[w^{o}(x)\right] - \gamma^{-1}\ln\alpha$.[5] These two factors, that diligence is less enjoyable than shirking and less uncertainty in compensation is preferable, explain why compensating an agent to align his interests with those of the principal is more expensive than merely paying them enough to accept employment.

3 The firm's profit maximization

The principal's profit maximization determines which cost-minimizing contract should be offered. The profits from inducing the agent to work diligently are $x - w^{o}(x)$, while the profits from employing the agent to shirk are $xg(x) - \gamma^{-1}\log(\beta)$. Thus, work is preferred by the principal if and only if

$$\max\{0, \gamma E[xg(x)] - \ln\beta\} \leq \gamma E[x - w^{o}(x)], \tag{3.15}$$

[5]To prove $E\left[w^{o}(x)\right]$ is greater than its certainty equivalent, $\gamma^{-1}\ln\alpha$, we note that in the cost-minimizing contract inducing diligence, (3.6), is met with equality. This implies $\alpha E\left[e^{-\gamma w(x)}\right] = 1$ or $\gamma^{-1}\ln\alpha = -\gamma^{-1}\ln E\left[e^{-\gamma w(x)}\right]$. Thus, $E\left[w^{o}(x)\right]$ exceeds $\gamma^{-1}\ln\alpha$ if and only if $\exp E\left[\gamma w^{o}(x)\right]$ exceeds $E\left[e^{-\gamma w(x)}\right]$, which is true by Jensen's inequality if and only if $\gamma > 0$.

which implies that it is weakly more profitable for the principal to employ the agent to work than to employ the agent to shirk or not to employ the agent.while a shirking contract is offered if and only if

$$\max\{0, \gamma E[x - w^o(x)]\} \leq \gamma E[xg(x)] - \ln \beta. \tag{3.16}$$

which means that employing the agent to shirk is profitable and more so than employing the agent to work.

Otherwise, the principal does not employ the agent if neither shirking nor working yields positive expected profit, i.e. if and only if

$$\max\{\gamma E[x - w^o(x)], \gamma E[xg(x)] - \ln \beta\} \leq 0. \tag{3.17}$$

Empirically, the null hypothesis that CEO compensation does not vary with the firm's state or with the firm returns is rejected by the data, as Section 2 illustrates. Therefore, for the most part, we will focus on the cost-minimizing contract that incentivizes working, as given in (3.7) and (3.8), rather than shirking. In the equilibrium of the model derived above, the only way to induce the CEO to work is to offer a contract that depends on the firm's performance x.

4 Cost decomposition

The decomposition of compensation serves to measure the cost of moral hazard, which is the additional compensation that the principal needs to pay the agent to align the interests due to moral hazard, versus the first-best where there is no moral hazard. This cost is compared with the gross revenue loss to shareholders if moral hazard is ignored, which reflects how much shareholders evaluate the moral hazard problem. From the agent's point of view, the conflict of interest is measured by the differential between the certainty equivalent wage for working and for shirking. By structurally estimating these components, we can assess the magnitude of agency costs and their implications for firm value and social welfare, providing insights that are essential for policy analysis but impossible to obtain from reduced-form approaches.

The conflict of interest arises between the principal and the agent, because the latter prefers shirking, which brings less distaste than working ($\beta < \alpha$), yet the principal prefers the agent to work, because the expected revenue would be higher ($E[xg(x)] < E[x]$). To quantify this conflict of interest, we compare the amount the principal would lose from the agent shirking with how much the agent gains from shirking when he is paid a fixed wage. We also decompose the expected compensation of the agent into two additive pieces, how much he would be paid if the principal observed his actions, and the expected extra pay the agent receives—the risk premium attributable to the agency problem. The formulas defining these measures are the objects of structural estimation, and this Section explains how they arise in the context of our model.

The principal gets higher expected revenue from the agent working versus shirking. The revenue loss to the principal from the agent shirking versus working is

$$\tau_1 \equiv \int [x - xg(x)] f(x) dx. \tag{3.18}$$

This is the loss in expected revenue that the principal would incur based on the production technology if it ignores the moral hazard problem. It is also what the firm would be willing to pay for a perfect monitoring device to eliminate moral hazard.

On the agent's side, shirking generates less disutility than working. To measure the conflict of interest from the agent's view, we make the nonpecuniary benefits comparable with the pecuniary measures by converting the utility into the *consumption equivalent*—the amount of consumption units that would generate this much utility. The consumption equivalent of working in (3.4) is $-\gamma^{-1} \ln(\alpha)$. It is the reservation wage for working, rather than taking the outside option, for which the utility is normalized to one. Similarly, the consumption equivalent of shirking is $-\gamma^{-1}\ln(\beta)$.

Therefore, the compensating differential in nonpecuniary benefits between working and shirking denominated in consumption units is

$$\tau_2 \equiv -\gamma^{-1}\ln(\alpha/\beta), \tag{3.19}$$

which measures how much the agent evaluates the conflict of interest.

However, simply paying the agent a fixed reservation wage for working is insufficient to deter shirking, as the principal cannot directly monitor the agent's actions. To align the agent's incentives with the firm's objectives, the principal designs a performance-based contract, which ties the agent's compensation to the firm's performance. To compensate the risk-averse agent for the risk in his compensation, the principal pays an additional risk premium to the agent, as detailed below.

The expected total compensation to the agent to induce work is

$$\tau_3 = \int w(x)f(x)\mathrm{d}x. \tag{3.20}$$

If the agent's actions could be perfectly monitored by the shareholders, the shareholders, who are risk-neutral, would pay the risk-averse manager a fixed wage to fully insure him from any risk in the firm's performance, called the *certainty equivalent*. It is the reservation wage for working

$$\tau_4 = \gamma^{-1} \ln \alpha. \tag{3.21}$$

We interpret the certainty equivalent for being employed as CEO and working. This represents what would be paid if the CEO's actions could be perfectly monitored and contracted upon, eliminating the need for performance-based incentives.

Since shareholders cannot observe the agent's actions, they incentivize him to work through performance-based compensation. However, making the agent subject to risk is costly because the agent is risk averse. The risk premium paid to the agent to align his incentives with those of the principal is the difference between expected compensation and its certainty equivalent

$$\tau_5 \equiv \tau_3 - \tau_4. \tag{3.22}$$

To summarize, there are four measures related to the moral hazard problem: τ_1, the expected gross revenue loss shareholders incur from the manager shirking versus working; τ_2, the welfare cost to the agent to work versus shirk, defined as the compensating differential in nonpecuniary benefits between working and shirking; τ_4, the certainty equivalent wage for the agent to work if shareholders could perfectly monitor; and τ_5, the risk premium paid to the manager due to the agency problem.

The risk premium is the cost of moral hazard as it captures the additional compensation to the agent with moral hazard compared with perfect monitoring. These four measures directly address the six questions posed in Abowd and Kaplan's 1999 survey, and are estimated for the pure moral hazard model in Section 5, and for model extensions in Section 8.2 and Section 10.

4

Towards Structural Estimation

The static model is defined by the preferences of the agent (α, β, γ), plus the probability density functions for working, $f(x)$, and shirking, $f(x)g(x)$. It provides an attractive template for empirical research in executive compensation: the conflict between managers and owners, and the cost of realigning incentives, is clearly portrayed. There are, of course, other sources of asymmetric information complicating the analysis of agency; this review discusses them in Sections 8 and 9, and provides the corresponding empirical applications in Sections 8.2 and 10. Aside from these sources, many other extraneous factors affect compensation, which could bias estimates obtained from models that ignore them. Broadly speaking, heterogeneity amongst firms and executives are two obvious factors that also affect compensation. A third factor is the manager's position, and relatedly, how activities are coordinated between managers within the same firm. Finally, the time dimension (or when information is revealed) plays a role in how compensation is measured, the utility compensation confers upon managers (who do not immediately consume everything they earn), and determining the length of the contract between shareholders and their managers. We now consider how these factors can be incorporated in a parsimonious way, before locking the structural model to the data-generating process.

Then we extend the model to a multiperiod setting, for simplicity focusing on the case in which there is only one executive.[1] The latter issue opens up questions about how compensation is spent over the lifecycle, and also how the optimal contract is adapted to a multiperiod setting underlying simple structure of management.

1 Heterogeneity

In empirical applications, it is useful to distinguish between different executive positions within the firm, denoted by $j \in \{1, \ldots, J\}$, and types of firm, denoted by $k \in \{1, \ldots, K\}$. Examples of different executive positions are the Chief Executive

[1] Dealing with both issues within the one model is straightforward. See Margiotta and Miller, 2000.

Officer (CEO) and Chief Financial Officer (CFO). Firms might be classified by industry and size. Data on executive characteristics might include individual traits such as race, gender, age, education, plus employment experience, here denoted by $h \in \mathbb{H}$, a vector space and/or a set of categorical variables. A starting point for estimation is to treat (h, j, k) as strictly exogenous. Rather than allowing for a complete set of interactions, which is cumbersome and data intensive, we could estimate a parameterization of the model with functional forms denoted by $\alpha_{jk}(h)$, $\beta_{jk}(h)$, $\gamma_{jk}(h)$, $f_{jk}(x|h)$ and $g_{jk}(x|h)$.

2 Coordination between agents

An unsatisfactory feature of this approach is that it does not reconcile the roles of several agents employed by the same principal to undertake a joint venture. In the executive suite of a large corporation, several executives manage the firm, and for the most part are employed for several years in that position. To incorporate this feature, suppose there are K principals employing J agents each. For notational convenience, we subsume executive characteristics, representing the preferences of agent (j, k) by γ, a risk aversion parameter that is invariant across agents, and amenity value factors $(\alpha_{jk}, \beta_{jk})$ that are determined both by the agent's identity and by the principal's identity. In this extended set-up, the principal's revenue depends on the employment and effort choices of all the agents she approaches. Analogous to (l_0, l_1, l_2) in the single agent model, let $l_{jk} \equiv (l_{0jk}, l_{1jk}, l_{2jk})$ denote the choices of agent (j, k), and let $l_{-j,k} \equiv (l_{1k}, \ldots, l_{j-1,k}, l_{j+1,k}, \ldots, l_{Jk})$ denote the choices of all the agents approached by principal k aside from the j^{th} agent's choices. We now define $f_k(x \mid l_{1k}, \ldots, l_{Jk})$, the probability density function for x_k, conditional on the effort levels of the agents, by

$$f_k(x|l_{1k}, \ldots, l_{Jk}) = \begin{cases} f_k(x) & \text{if } \sum_{j=1}^{J} l_{2jk} = J \\ f_k(x)\, g_{jk}(x) & \text{if } \sum_{j=1}^{J} l_{2jk} = J-1,\ l_{1jk} = 1. \end{cases}$$

We are only considering the two cases where all managers work diligently, or all but one manager work diligently. This is motivated by the empirical observation that all managers' compensation are correlated with the abnormal returns of the firm, so the contract simplifies to incentivizing each manager j to work, as a Nash response to $-j$ (all managers other than j) working diligently. We do not model the case where more than one manager shirks, because the shirking of any one manager is not observed empirically and thus off-equilibrium behavior.

For principal k, to solve for a cost minimizing contract given that agents choose the actions $l \equiv (l_{1k}, \cdots, l_{Jk})$, the principal chooses the compensation schedules to minimize the total expected compensation

$$\int \sum_{j=1}^{J} w_{jk}(x)\mathrm{d}x$$

subject to the participation constraints for all agents j employed at k that

$$\beta_{jk} \int e^{-\gamma w_{jk}(x)} f_k\left(x\right) g_{jk}\left(x\right) \mathrm{d}x \leq 1,$$

and the additional incentive compatibility constraint for each agent j working

$$\beta_{jk} \int e^{-\gamma w_{jk}(x)} f_k\left(x\right) g_j\left(x\right) \mathrm{d}x \leq \alpha_{jk} \int e^{-\gamma w_{jk}(x)} f_k\left(x\right) \mathrm{d}x.$$

In this way, the optimal contract embeds a Bayesian Nash equilibrium in which, for each agent, given their own compensation schedule, and the actions of the other agents, the best response of each is to follow the equilibrium actions. Having computed the cost minimizing contract for each combination of agents, the principal then selects the profit maximizing staff and effort assignment. Appealing to the revelation principle (Myerson, 1982), the participation and incentive compatibility constraints ensure that the agents follow the principals' employment and effort assignment.

Consistent with the simplest of asset-pricing models, we normalize the expected value of abnormal returns in equilibrium from everyone working to zero. 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

$$0 = \int x f_k(x) \mathrm{d}x > \int x f_k(x) g_{jk}(x) \, \mathrm{d}x, \tag{4.1}$$

where $f_k(x)$ is the distribution of returns where everyone works, $g_{jk}(x)$ is the likelihood ratio if all managers work except manager j. 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, $\alpha_{jk} > \beta_{jk}$; whereas the inequalities for all j in (4.1) show production is greater when all executives work.

This model applies to situations where several executives jointly manage the company. We assume they play a Bayesian Nash equilibrium in a noncooperative game. Shareholders then set an optimal contract subject to those conditions.

Li, 2024, tests two multi-agent moral hazard models, which distinguish between a team perspective and an individual perspective. The results support the team perspective under which multiple managers deviate simultaneously, rejecting the individual perspective under which each manager can unilaterally shirk.

3 A dynamic model

To simplify the exposition, we have thus far assumed a static model where managers consume their compensation immediately, and the principal designs a one-period contract to maximize expected revenue net of compensation. However, in reality, both shareholders and managers face dynamic problems. Managers typically accumulate wealth during their high-earning years while smoothing consumption, and shareholders, following the financial economics literature, are forward-looking and diversified against idiosyncratic shocks affecting firm value.

Adapting the static model to a dynamic setting opens up several questions: (i) instead of instantaneous gratification, the agent receives utility over several periods, typically extending beyond his employment with the principal; (ii) if the agent is employed for more than one period, to what extent should contracts be written over the term of his employment?

The extension builds on the dynamic pure-moral-hazard models analyzed in Malcomson and Spinnewyn, 1988, Fudenberg *et al.*, 1990, and Rey and Salanié, 1990. This body of theoretical work has been applied in empirical studies of executive compensation, such as Margiotta and Miller, 2000, Gayle and Miller, 2009a, Gayle and Miller, 2009b, Gayle and Miller, 2015, and Edmans *et al.*, 2012.

Instead of (3.4), suppose the agent's preferences are captured by a utility function that is additively separable over periods and multiplicatively separable between consumption and work activity within periods, with lifetime utility as

$$-\sum_{t=0}^{\infty} \delta^t \left[l_{0t} + \beta l_{1t} + \alpha l_{2t} \right] \exp\left(-\gamma c_t\right) \tag{4.2}$$

where δ is the constant subjective discount factor, γ is the constant absolute level of risk aversion (CARA), and α and β are the respective utility parameters that measures the distaste from working and shirking, $\alpha > \beta$. The CARA assumption simplifies the analysis because it implies that the preferences of a CEO towards his risk exposure are independent of his wealth. This implies that the optimal contract, derived below, is independent of outside wealth, personal assets with returns that are independent of his own firm's idiosyncratic abnormal return. Relaxing CARA would be straightforward if the data on each CEO included holdings of all his assets (financial and real), not just those related to his firm.

While the actions of the agent are not contractible in this model, because they are not fully observed by the principal, we assume there are no other impediments to trade. Thus, shareholders are well-diversified, and the CEO has access to a well-functioning market to smooth consumption streams with wealth they have accumulated. Formally, we assume a complete set of markets for all publicly disclosed events exists, and attribute all deviations from the law of one price to the information asymmetries of pure moral hazard and private information.[2]

Extending the static framework to dynamic settings barely affects the notation of revenues or profits. Revenue incurred in period t from actions in the previous period is denoted by x_t, drawn from a probability distribution with *pdf* $f(x)$ when the agent works and $f(x)g(x)$ when he shirks, where $f(x)$ and $g(x)$ have the properties discussed in the static model. The definition for the measure of the conflict of interest by the firm, (3.18), remains unchanged, τ_1, the gross revenue loss from shirking in a period. τ_3, the expected compensation, also remains as defined in (3.20). However, three other measures, τ_2 (the compensating differential in nonpecuniary benefits between working and shirking), τ_4 (the certainty-equivalent wage for working under perfect monitoring), and τ_5 (the risk premium shareholders pay due to moral hazard), adjust because the timing of the nonpe-

[2]The assumption of complete markets, and tests of the assumption, have been used when applying structural econometric models to panel data of consumption and labor supply—Altug and Miller, 1990, Altug and Miller, 1998; housing size and labor supply—Miller and Sieg, 1997; first home purchase, fertility and labor supply—Khorunzhina and Miller, 2022; as well as managerial compensation—Margiotta and Miller, 2000, Gayle and Miller, 2009b, Gayle and Miller, 2015, and Gayle *et al.*, 2015. A common alternative assumption in structural labor econometrics is that all current income is immediately consumed. The alternative is inappropriate for this context, because the CEOs of these large companies are at this point in their life primarily saving for their retirement and have access to many financial vehicles for saving.

cuniary benefits to the agent and his pay is nonsynchronous: nonpecuniary utility accrues in the current period but pay is received next period.

To make these two sources of benefit comparable, we convert the utility gain into its consumption equivalent in the next period, i.e. the amount of consumption units in the next period that would deliver the same utility. Consider an annuity bond that pays one consumption unit each period from the current period onward. Let b denote the current price of the annuity bond, b' the bond price next period and i the current one-period interest rate. Noting that $b - 1 = (1+i)^{-1} b'$, we define the certainty equivalent differential gain, denominated in next period's consumption units, to a CEO from shirking instead of working in the current period as

$$\tau_2 \equiv -(1+i)\,\gamma^{-1}\ln(\alpha/\beta) = -b'\left[(b-1)\,\gamma\right]^{-1}\ln(\alpha/\beta). \quad (4.3)$$

If the actions of the CEO were perfectly monitored by shareholders, they would pay him a fixed wage, denoted by τ_4. It is easy to prove that in terms of next period's consumption units

$$\tau_4 \equiv (1+i)\,\gamma^{-1}\ln(\alpha) = b'\left[(b-1)\,\gamma\right]^{-1}\ln\alpha. \quad (4.4)$$

We now define the agency cost as

$$\tau_5 \equiv \tau_3 - \tau_4. \quad (4.5)$$

The agency cost represents the difference between the expected total compensation under moral hazard and the fixed wage under perfect monitoring. It is the expectation of the additional amount that the firm pays the CEO because shareholders cannot monitor his activities. In the optimal contract derived below, τ_5 is also the risk premium paid to the CEO with certainty equivalent τ_4.

There are no gains from a long-term arrangements between shareholders and the CEO in this framework, because the distribution of the firm's financial returns is independent of his actions taken more than one period ago, and his private information is only useful for forecasting returns one period ahead. The benefits from a long-term contract arise if, for example, the hidden actions of the CEO were only revealed some years after they were taken. Consequently, the optimal long-term contract between shareholders and the CEO in this model decentralizes to a sequence of short-term one-period contracts (Malcomson and Spinnewyn, 1988, Fudenberg *et al.*, 1990, Rey and Salanié, 1990, and Gayle and Miller, 2015).

When comparing two employment activities, such as shirking and working, the manager weighs the nonpecuniary benefits of the activity in the current period against the probability distribution defining its pecuniary benefits next period. To render these two sources of benefit comparable, we convert the utility gain as its consumption equivalent in the next period. Similar to the calculation of τ_2 in (4.3), the total value of working diligently in the current period (relative to the outside option) and being paid $w(x)$ in the next period, denominated in units of next period's consumption, is therefore

$$-b' \left[(b-1)\,\gamma\right]^{-1} \ln(\alpha) + w(x). \tag{4.6}$$

When the manager evaluates the compensation, he optimally smooths his consumption over the remaining periods of his life, as is standard in the asset pricing literature (Debreu, 1959). This amounts to valuing additional wealth by a factor that scales up his current utility function by an annuitized amount, a property that derives directly from the exponential utility (or CARA) assumption. Compensation depends on contractible events, including the bond price and subsequent excess returns in the next period. Then the annuitized value[3] of $w\,(x)$ starting with (the payment of) a constant flow of consumption units starting from the next period is $w\,(x)\,/b'$, which yields an indirect utility of

$$v(x) \equiv \exp\left[-\gamma w\,(x)\,/b'\right]. \tag{4.7}$$

[3]We annuitize the wage because the manager will smooth consumption over all remaining periods. We focus on annuity bonds rather than fixed-term bonds because we cannot observe when executives pass away, though alternative specifications using fixed-term bonds could be considered if the lengths of executives' lifecycle were observable.

This representation of the manager's indirect utility considerably simplifies the shareholders' contracting problem: appealing to (4.6) and (4.7) when considering the manager's responses to the contract terms, we only need to compare expressions like $\alpha^{1/(b-1)}v(x)$ (the annuitized indirect utility from the monetary measure in (4.6)), which evaluates the indirect utility from the next period's compensation, against the disutility from working (or shirking, substituting β for α) from the current period, both annuitized into a constant flow of utility from next period onward. This representation inherently accounts for the risk aversion of the manager when dealing with risk in $w(x)$, allowing us to assess the expected utility by taking the expectation based on the probability distribution $f(x)$ (or $f(x)g(x)$).

Appealing to the revelation principle (Myerson, 1982), the optimal contract is solved by a direct mechanism: in the current period t shareholders choose $w(x)$ incurred in the next period for each potential x to minimize the expected cost of managerial compensation subject to the constraints that the manager prefers shirking to the outside option and that he prefers working to shirking.

Participation To induce a candidate for CEO to accept employment with the firm and work diligently, his annuitized expected utility from working must exceed the utility obtained from taking the outside option, which is normalized to one. The candidate would only accept the position if

$$\int \alpha^{1/(b-1)} v(x) f(x) \mathrm{d}x \leq 1. \tag{4.8}$$

Incentive compatibility Given his decision to be employed, the incentive compatibility constraint induces the CEO to prefer working rather than shirking. Shirking yields an immediate nonpecuniary benefit, because $\beta^{1/(b-1)} < \alpha^{1/(b-1)}$, yet the expectation of the (annuitized consumption) value of compensation $v(x)$ is taken with respect to the $g(x)f(x)$ density rather than $f(x)$. Comparing the annuitized expected utility from working to that from shirking, the incentive compatibility constraint is

$$\int \alpha^{1/(b-1)} v(x) f(x) \mathrm{d}x \leq \int \beta^{1/(b-1)} v(x) g(x) f(x) \mathrm{d}x. \tag{4.9}$$

Optimization Minimizing expected compensation is equivalent to choosing $v(x)$ that maximizes

$$\int \ln [v(x)] f(x) \mathrm{d}x. \tag{4.10}$$

Appealing to the Kuhn-Tucker theorem, there is a unique positive solution. The Lagrangian and the first-order conditions for the dynamic model are derived in the appendix of Margiotta and Miller, 2000, Proposition 3 and 4.

5

Structural Estimates: A First Look

This Section reviews parametric approaches to estimate the static model, and explains how to generalize the model to handle the extensions discussed in the previous Section. The latter parts report on the costs and welfare measures for moral hazards, and how they have changed.

1 Parameterizing the static model

Parameterizing the model amounts to specifying the distribution functions for revenue conditional on effort. The empirical work reviewed below assumes the distribution functions for working and for shirking are both left truncated normal with the same support, differing only in their parent means, μ_2 and μ_1

respectively, with $\mu_2 > \mu_1$

$$f(x) = \left[\Phi\left(\frac{\mu_2 - \psi}{\sigma}\right)\sigma\sqrt{2\pi}\right]^{-1} \exp\left[-\frac{1}{2}\left(\frac{x-\psi}{\sigma}\right)^2\right] \tag{5.1}$$

$$g(x) = \frac{\Phi\left[\left(\mu_2 - \psi\right)/\sigma\right]}{\Phi\left[\left(\mu_1 - \psi\right)/\sigma\right]} \exp\left[\left(\frac{\mu_2^2 - \mu_1^2}{2\sigma^2} - \frac{(\mu_2 - \mu_1)}{\sigma^2}x\right)\right]. \tag{5.2}$$

Three reasons motivate this particular parameterization. First, for the three industries that Margiotta and Miller, 2000, investigate,[1] there is evidence that the *pdf* for abnormal returns conditional on working is nonmonotone, single peaked and skewed to the right. Second, total compensation increases with financial returns, which occurs in this model if and only if $g(x)$ is monotone decreasing in x, guaranteed by this parameterization because

$$\frac{\partial g(x)}{\partial x} = \frac{(\mu_1 - \mu_2)}{\sigma^2}g(x) < 0.$$

Third, the lower truncation point ψ may be loosely interpreted as a bankruptcy state or liquidation condition. That the parent variance σ is constant can be relaxed, by replacing σ with σ_2 in (5.1) and defining

$$f(x)g(x) = \left[\Phi\left(\frac{\mu_1 - \psi}{\sigma_1}\right)\sigma\sqrt{2\pi}\right]^{-1} \exp\left[-\frac{1}{2}\left(\frac{x-\psi}{\sigma_1}\right)^2\right].$$

The benefit is that the compensation equation is modeled more flexibly, while the computational cost comes from complicating

[1] The evidence we present in Table 2.4 in Section 2 shows that the three sectors they investigated are quite representative of all sectors and exhibit trends that have been found in previous studies using other data.

the estimation of $g(x)$ with an additional parameter.

2 Estimation

Parameter estimates for $f(x)$ can be obtained with a nonlinear Limited Information Maximum Likelihood (LIML) estimator

$$(\mu_2, \sigma, \psi^*) = \arg\min \frac{1}{N}\sum_{n=1}^{N}\left\{\ln\sigma + \ln\Phi\left(\tfrac{\mu_2-\psi^*}{\sigma}\right) + \frac{1}{2}\left(\tfrac{x-\mu_2}{\sigma}\right)^2\right\}.$$

Shirking is off the equilibrium path, as a flat wage is never observed, hence the data does not yield direct estimates of the outcomes from shirking without imposing structure on the model. The conditions characterizing the cost minimizing contract provide equations for estimating (α, β, γ), the taste parameters, plus μ_1, the remaining parameter for the shirking distribution. The sample analog to the participation constraint is

$$\frac{1}{N}\sum_{n=1}^{N}\left\{\alpha^{-1} - \exp\left[-\gamma w(x_n)\right]\right\} = 0.$$

Similarly the sample analog to the incentive compatibility constraint is

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\left\{\alpha - \beta\frac{\Phi\left(\frac{\mu_2-\psi}{\sigma}\right)}{\Phi\left(\frac{\mu_1-\psi}{\sigma}\right)}\exp\left[\left(\frac{\mu_2^2-\mu_1^2}{2\sigma^2} - \frac{(\mu_2-\mu_1)}{\sigma^2}x_n\right)\right]\right\}}{\exp\left[\gamma w(x_n)\right]} = 0.$$

The compensation equation provides the only other equation

$$w^o(x) \equiv \gamma^{-1}\ln\alpha + \gamma^{-1}\ln\left[1 + \lambda\left(\frac{\alpha}{\beta}\right) - \lambda g(x)\right].$$

Overall, the model fits the data reasonably well. The parameter estimates in Table 5.1 align with the theoretical model's assumptions. The coefficients on the explanatory variables are estimated in a linear regression. Here we report the estimates for the electronic industry; the results for the other industries are detailed in Gayle and Miller, 2009a. Panel A shows that there have been changes in the probability distributions of abnormal returns, with higher dispersion and more negative lower truncation point. The average variance of firms' diligent-returns is consistent with the unconditional standard deviation in the data, suggesting that our parametrization of the truncated normal is a good fit. Panel B provides estimates of the shirking distribution. The last line of panel B shows a notable loss in abnormal returns if the manager shirks, which is a loss of 2.46 percent and had not fallen over time. Panel C shows the estimates of managers' preferences. The estimates of α show that the market for managers has become more differentiated and the relative premium for managing a larger versus smaller firm, whether measured in terms of employees or assets, has increased, a result that corroborates the work of Gabaix and Landier, 2008, Murphy and Zabojnik, 2004, and Murphy and Zabojnik, 2006. On the other hand, the estimates of α/β show that, if anything, the conflict between a firm with a given set of characteristics and its executives has declined. The managerial preferences for risk, γ, have remained stable in an economic sense.

Panel D reports three measures related to moral hazard. The first measure, τ_1, represents the gross revenue loss to shareholders if they failed to design compensation contracts that align managers' actions with shareholder interests. The magnitude of this loss has increased dramatically by 16 times, from approximately \$100 million to more than \$1.6 billion.

On the other hand, the increase in the second measure, τ_2 (defined as the nonpecuniary benefits of shirking to managers), is modest. The estimates show that risk-averse managers receive only a four-fold increase in expected compensation to accept fluctuations in their wealth due to the volatility of firm returns relative to the market portfolio.

The third measure, τ_5, represents the cost of moral hazard, which is measured by the additional compensation needed to motivate managers to work diligently. This measure shows that the cost of motivating the manager to work diligently is modest compared to the substantial benefits of aligning managers' actions with shareholder interests through the compensation schedule.

The welfare cost to the CEO for working versus shirking (τ_2) has increased drastically. On the other hand, the rate of increase in the average certainty equivalent wage is about the same as the growth rate in national income per capita. Therefore,

Table 5.1
Parameter Estimates, Nonpecuniary Benefit and Welfare Cost

Parameter	**Explanatory Variable**	**Old**	**New**
Panel A: Parameter estimates of the diligent-returns distribution			
σ^2	Constant	-2.07	-7.12
		(0.33)	(0.46)
	Asset to equity ratio	-1.12	8.93
		(139)	(13.5)
	Number of employees	0.36	0.88
		(0.28)	(0.22)
	GDP	-16.60	5.44
		(2.13)	(0.47)
ψ		-0.61	-1.60
		(0.15)	(0.48)
$\sqrt{\text{Var}\left(x_{nt} \mid l_{2nt}=1\right)}$		22.09	37.90
		(5.52)	(5.98)
Panel B: Parameter estimates of the shirking-returns distribution			
μ_1	Constant	-0.00	-0.01
		(0.00)	(0.00)
	Asset to equity ratio	-0.03	-0.03
		(0.00)	(0.00)
	Number of employees	-0.02	-0.01
		(0.00)	(0.00)
	GDP	-0.06	-0.02
		(0.00)	(0.00)
$E\left(x_{nt} \mid l_{1nt}=1\right)$		-2.14	-2.46
		(1.56)	(0.23)
Panel C: Nonpecuniary benefits from diligence relative to the outside option, and to shirking			
α	Constant	0.99	3.91
		(0.05)	(0.00)
	Assets	-0.48	2.31
		(0.10)	(0.31)
	Employees	1.08	2.82
		(0.10)	(0.12)
α/β	Constant	3.23	1.85
		(0.32)	(0.02)
	Assets	0.11	2.72
		(0.02)	(0.02)
	Employees	5.06	1.90
		(0.25)	(0.07)
γ		0.52	0.50
		(0.00)	(0.00)
Panel D: Cost of moral hazard, nonpecuniary benefits of shirking, and welfare cost			
τ_1		99,910	1,613,960
		(894,490)	(4,204,250)
τ_2		747	3048
		(432)	(387)
τ_5		278	4873
		(1257)	(17,285)

Note: Excerpt from Gayle and Miller, 2009a, Tables 5,6,7,8,9,10. τ_1, τ_2, τ_5 are in thousands of US$ (2000). Standard deviations in parentheses.

the outsized increase of the expected executive compensation relative to the growth of national income per capita is attributed to the rising cost of moral hazard.

They further attribute the increase in the cost of moral hazard primarily to the increase in firm size. With the assumption that expected abnormal returns when the manager works diligently are normalized to zero, adapting the definition to account for firm size ν, τ_1 from (3.18) becomes

$$\tau_1 = -\nu \int x f(x) g(x) dx. \tag{5.3}$$

The dominant role of firm size in explaining the large increase in the cost of ignoring moral hazard is evident from the decomposition of the change in estimates during two regimes (Equation (22) from Gayle and Miller, 2009a)

$$\begin{aligned} -\Delta(\tau_1) = (\Delta\nu) \int x g(x) f(x)\, dx + \nu \int x \left[\Delta g(x)\right] f(x)\, dx \\ + \nu \int x g(x) \left[\Delta f(x)\right]\, dx. \end{aligned} \tag{5.4}$$

The measures of moral hazard depend on the preferences of managers, what shareholders observe about their behavior, the distribution of abnormal returns accruing to firms, and the characteristics of the firms they manage. The exogenously changing distribution of firm size is the primary driver of the steep increase in the cost of moral hazard and managerial compensation.

Over the years, an average firm size has increased multi-fold, in terms of increased sales and assets, and reduced labor count. Their model does not explain the distribution of firm size, but does distinguish between the direct effect of firm size through the market-clearing certainty-equivalent wage, and its indirect effect through moral hazard and its welfare costs.

Their empirical results support the view that the cost of moral hazard have increased due to larger firms being more vulnerable to corporate governance issues, where managers have greater opportunities to act against shareholders' interest, a widely recognized intuition that traces back to the early work of Berle and Means, 1932, on corporations. Their model explains about half of the observed variation in managerial compensation in both datasets they studied. This leads us to conclude that the sharp rise in the cost of moral hazard, driven by the exogenous shifting composition of firms, is the most significant factor explaining the increasing level and variation in managerial compensation.

6

Identification

This Section establishes identification of the static model without imposing the parametric form of $f(x)$ and $g(x)$ in the previous section. The parameters of the model are characterized by $f(x)$ and $g(x)$, which together define the probability density functions of revenue, (α, β), the preference parameters for diligent work and shirking (relative to the normalized utility from taking the outside option), and the risk-aversion parameter, γ. The estimation method described in the previous Section is based on the presumption that there is only one set of parameters that could have generated the data. It is worth questioning this presumption. Although two parameter vectors might have the same data-generating process, they typically yield different predictions when perturbed by a counterfactual, i.e. what happens when the model parameters change.

Since one of the appealing features of structural estimation is to predict counterfactuals, lack of point identification, where the model restrictions do not uniquely determine the parameter, might seem a serious drawback. However, lack of point identification does not doom structural modeling. Set identification refers to the case where a set of different parameter vectors that leads to the same data-generating process—the data could be generated by any of the parameter vectors in this set. Rather than making a prediction based on a point, the prediction for a counterfactual becomes a set. The principal agent model provides a useful paradigm for illustrating this general principle.

Without loss of generality, we assume the data comprise repeated independent cross- sectional draws of revenue and compensation, (x_n, w_n), for a sample of N observations generated in equilibrium from identical principal–agent pairs, or a time series of (x_n, w_n) from the same agent and profit-maximizing principal in a stationary environment. However, the analysis extends to situations with observed heterogeneity in all the model parameters. Unobserved fixed effects that help determine the manager's preferences could also be identified from panel data.[1] There are three cases to investigate: When is it optimal

[1]For example suppose the panel tracks managers for T periods, and the

for managers to shirk? When is it optimal for managers to work? When is it optimal for one type of principal to induce working and another type to induce shirking?

Whether the agent works or shirks is identified from the variation in observed compensation, w_n: In the case of shirking, (3.16) holds, a fixed compensation is prescribed, so compensation does not depend on revenue. In this case, the density $f(x)g(x)$ can be identified from observations on revenue, the compensation is constant at $w^{(1)} = \gamma^{-1} \ln \beta$, but nothing more can be gleaned from the data about the structure of the model. Loosely speaking, this variation on the model is under-identified, and is indistinguishable from a model where there are no moral-hazard considerations.

When the agent works, (3.15) holds and w_n depends non-trivially on revenue, x_n. Because of the relevance of managerial compensation to firm performance, we focus on the case in which the principal seeks to overcome a moral hazard problem through the provision of performance pay. In this case, $f(x)$ is identified by the marginal distribution of x and the optimal contract $w^o(x)$ is identified by the conditional expectation of W given x

$$w^o(x) = E[W|x].$$

risk-aversion parameter is fixed over time, taking the form $\gamma_n = m(z_n) + \gamma\widetilde{\gamma}_n$, where z_n are some covariates for manager n, and $\gamma\widetilde{\gamma}_n$ is an unobserved manager-specific fixed effect. Then our identification and empirical-content results would apply.

Identification of the remaining parameters $g(x), \alpha, \beta$, and γ proceeds in the following steps. First, we show that if γ is known, then the remaining parameters are point-identified from the cost-minimization problem. This means that the set of observationally equivalent parameters can be indexed by the positive real number γ, the risk-aversion parameter. Secondly, we show that the principal's preference for working over shirking provides an additional inequality that helps delineate the values of observationally equivalent γ. Thirdly, we prove that the set of restrictions we have derived in the first two steps fully characterize the identified set.

1 Restrictions from cost minimization

To begin this analysis, we temporarily suppose that the risk parameter, γ, is known, and define the mapping $\upsilon : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^+$ as

$$\upsilon\left(x, \gamma\right) \equiv \exp\left(\gamma w(x)\right).$$

The first-order condition for the cost minimization is

$$e^{\gamma w(x)} = \lambda_0 \alpha + \lambda_1 \alpha - \lambda_1 \beta g(x),$$

i.e.

$$\upsilon\left(x, \gamma\right) = \lambda_0 \alpha + \lambda_1 \alpha - \lambda_1 \beta g(x).$$

Since $E\left[g(x)\right]=1$ taking expectations over x gives

$$E\left[\upsilon\left(x,\gamma\right)\right]=\lambda_0\alpha+\lambda_1\alpha-\lambda_1\beta.$$

Also, the regularity condition that $g(x)\to 0$ as $x\to\infty$ implies that

$$\underset{x\to\infty}{lim}\,\upsilon\left(x,\gamma\right)=\lambda_0\alpha+\lambda_1\alpha\equiv\overline{\upsilon} \tag{6.1}$$

where $\overline{\upsilon}=\exp(\gamma\overline{w})$ is identified from the maximum wage $\overline{w}$. Substituting this back in gives

$$\upsilon\left(x,\gamma\right)=\overline{\upsilon}-\left\{\overline{\upsilon}-E\left[\upsilon\left(x,\gamma\right)\right]\right\}g(x), \tag{6.2}$$

which upon rearrangement gives

$$g(x)=g(x,\gamma)=\frac{\overline{\upsilon}-\upsilon\left(x,\gamma\right)}{\overline{\upsilon}-E\left[\upsilon\left(x,\gamma\right)\right]}.$$

Hence, the shirking distribution $f(x)g(x)$ is identified when γ is known.

Turning now to the taste parameters α and β, from the participation and incentive compatibility constraints, we see that

$$\alpha(\gamma)\equiv\left\{E\left[e^{-\gamma w^o(x)}\right]\right\}^{-1},$$

and, upon substituting for $g(x)$ that

$$\beta(\gamma)\equiv\left\{E\left[e^{-\gamma w^o(x)}g(x,\gamma)\right]\right\}^{-1}.$$

Theorem 6.1 (Theorem 2.1 of Gayle and Miller, 2015). Suppose the data on x_n and w_n are generated by a parameterization of a pure moral hazard model with risk aversion γ^*. Then

$$\alpha = \alpha\left(\gamma^*\right) \equiv \left\{E\left[e^{-\gamma w^o(x)}\right]\right\}^{-1} \tag{6.3}$$

$$\beta = \beta\left(\gamma^*\right) \equiv \frac{1 - E\left[e^{\gamma w^o(x) - \gamma \bar{w}}\right]}{E\left[e^{-\gamma w^o(x)}\right] - e^{-\gamma \bar{w}}} \tag{6.4}$$

$$g(x) = g\left(x, \gamma^*\right) \equiv \frac{e^{\gamma \bar{w}} - e^{\gamma w^o(x)}}{e^{\gamma \bar{w}} - E\left[e^{\gamma w^o(x)}\right]}. \tag{6.5}$$

These mappings are derived from the compensation equation (3.7), the participation constraint (3.5), and the incentive-compatibility constraint (3.6). All three mappings inherit the basic structure of the model for any positive value of γ. That is, $g(x, \gamma)$ is a likelihood ratio, $\beta(\gamma)$ and $\alpha(\gamma)$ are positive, and $\beta(\gamma) < \alpha(\gamma)$.

Integrating (6.5) over x, we get $E[g(x, \gamma)] = 1$ for all $\gamma > 0$. Also by definition $\bar{w} \geq w$, so $e^{\gamma \bar{w}} \geq E\left[e^{\gamma w^o(x)}\right]$ and $e^{\gamma \bar{w}} \geq e^{\gamma w^o(x)}$ for all $\gamma > 0$. Therefore, $g(x, \gamma) \geq 0$ for all $\gamma > 0$. Furthermore, as $x \to \infty$, from (6.1), we see that $v(x, \gamma) \to \overline{v}$, and hence $g(x, \gamma) \to 0$, as stipulated by the regularity condition in (3.3). This proves $g(x, \gamma)$ can be interpreted as a likelihood ratio satisfying (3.3) for any $\gamma > 0$.

To provide intuition about the identification of the likelihood ratio, consider the following cases in (6.2). When $g(x) = 1$, the observed revenue realization makes working and shirking equally likely, so optimal compensation equals the expected compensation level. When $g(x) < 1$, working is more likely than shirking given the observed revenue, so optimal compensation is higher than the expected compensation level. Conversely, when $g(x) > 1$, shirking is more likely than working for the observed revenue, so optimal compensation is lower than the expected compensation level. In the limiting case where $g(x) \to 0$, working is almost certain given the observed revenue, so the optimal compensation equals the maximum compensation level. In this way, the likelihood ratio is directly identified by the shape of the compensation schedule: the principal rewards outcomes that are more likely to be generated by working, and punishes outcomes that are more likely to be generated by shirking.

Next, consider $\alpha(\gamma)$ and $\beta(\gamma)$. If γ is given, $\alpha(\gamma)$ and $\beta(\gamma)$ are identified from the certainty equivalents of the compensation with respect to the working distribution $f(x)$ and the shirking distribution $f(x)g(x)$. Clearly $\alpha(\gamma) > 0$ because $e^{-\gamma w^o(x)} > 0$. Similarly, $\beta(\gamma)$ is also positive. Rearranging the expression for the ratio of $\beta(\gamma)$ and $\alpha(\gamma)$, we obtain

$$\frac{\beta(\gamma)}{\alpha(\gamma)} = \frac{e^{\gamma \overline{w}} - E\left[e^{\gamma w^o(x)}\right]}{e^{\gamma \overline{w}} - \left\{E\left[e^{-\gamma w^o(x)}\right]\right\}^{-1}}. \tag{6.6}$$

Since the inverse function is convex, Jensen's inequality gives $E\left[e^{-\gamma w^o(x)}\right] > \frac{1}{E\left[e^{\gamma w^o(x)}\right]}$ or $\frac{1}{E\left[e^{-\gamma w^o(x)}\right]} < E\left[e^{\gamma w^o(x)}\right]$, and consequently $\beta(\gamma) < \alpha(\gamma)$ for all positive γ as stipulated by the theoretical model.

It is worth emphasizing the generality of this model framework. Until we impose the profit-maximization constraint that bounds the risk-aversion parameter, the analysis requires only that the principal minimizes compensation costs, regardless of what specific objective they ultimately maximize. This generality allows the model to accommodate various forms of misalignment in performance measures and multi-tasking environments. The framework remains robust whether "low effort" refers to shirking on a single dimension, suboptimal allocation of effort across multiple tasks, or more complex deviations from the principal's preferred actions, making the identification strategy broadly applicable across different agency relationships.

An optimal shirking contract Now consider the third case, and suppose there exists some unobserved heterogeneity in the types of principals; some of them are just as we have described above and satisfy (3.15), but the revenue generation process for the remainder is $f(x)g(x)$ regardless of whether the agent works or shirks. In equilibrium, the latter pay a fixed wage of $w^{(1)}$, and the former pay variable compensation of (3.7). There is a discontinuity in the distribution function for compensation data at $w^{(1)}$, and the size of the jump determines the fraction of principals who induce shirking. The density $f(x)$ is identified from data on revenues to principals not paying $w^{(1)}$, and $f(x)g(x)$ is identified from data on revenues to principals paying $w^{(1)}$. Taking the quotient identifies $g(x)$. This only leaves β, α, and γ to identify. In the optimal contract, the participation constraint for both types of principals holds with equality, as does the incentive-compatibility constraint for the principal who induces work. Thus (3.5) and (3.6) reduce to

$$\beta E\left[e^{-\gamma w^{o}(x)}g(x)\right] = \beta e^{-\gamma w^{(1)}} = \alpha E\left[e^{-\gamma w^{o}(x)}\right] = 1. \quad (6.7)$$

Define $\psi(\xi) \equiv E\left\{e^{-\xi\left[w^{o}(x)-w^{(1)}\right]}g(x)\right\}$. The first two equalities in (6.7) imply $\psi(\gamma) = 1$. By inspection, $\psi(0) = 1$, $\psi'(0) < 0$ and $\psi''(\xi) > 0$.[2] Thus, $\psi(\xi)$ is a convex function with a unique nonzero solution at $\psi(\gamma) = 1$. This equality identifies γ. Substituting the solution into the second two equalities of (6.7) identifies β and α. Therefore, all the parameters are identified from the cost-minimization equations alone. Indeed this variant on the model is over-identified, because $w^{o}(x)$ must satisfy (3.7) and (3.8) for each x, a very strong exclusion restriction that relates the two types of principals to each other. However, relaxing the restriction would necessitate a separate analysis of the first two cases.

[2]To prove $\psi'(0) < 0$, first note that $\psi'(0) = w^{(1)} - E\left[w^{o}(x)g(x)\right]$. Second, $w^{(1)}$ is the certainty equivalent of $w^{o}(x)$ under the probability density $f(x)g(x)$. Hence $w^{(1)} < E\left[w^{o}(x)g(x)\right]$.

2 Restrictions from profit maximization

The restrictions from cost minimization tie down all the parameters up to γ, but place no restrictions at all on γ. Imposing profit maximization, as opposed to cost minimization only, does limit the set of admissible γ. Since profit maximization implies that the expected profits from paying the agent $w^o(x)$ are higher than paying him $\gamma^{-1}\ln\beta$, it follows from (3.15) that

$$E[x] - E\left[w^o(x)\right] - E[xg(x)] + \gamma^{-1}\ln\beta \geq 0. \tag{6.8}$$

Substituting $g(x)$ and β from (6.5) and (6.3) into the LHS of (6.8) implies

$$\begin{aligned} Q_0(\gamma) \equiv E[x] - E\left[w^o(x)\right] - E\left[x\frac{e^{\gamma\overline{w}} - e^{\gamma w^o(x)}}{e^{\gamma\overline{w}} - E\left[e^{\gamma w^o(x)}\right]}\right] \\ + \gamma^{-1}\ln\left(\frac{1 - E\left[e^{\gamma w^o(x) - \gamma\overline{w}}\right]}{E\left[e^{-\gamma w^o(x)}\right] - e^{-\gamma\overline{w}}}\right) \geq 0. \end{aligned} \tag{6.9}$$

This inequality restricts the set of γ that are admissible for the model.

3 Tight and sharp bounds

When the risk aversion γ is not given, the identifying restrictions cannot pin down γ uniquely, and the model is set-identified. Set identification means that, instead of a unique point, there is a set of possible parameterizations that are observationally equivalent to the true data-generating process. A set is tight if it covers the identified set. Sharp means that every element of the set is observationally equivalent, and thus included in the identified set. Characterizing the identified set amounts to establishing that the set given by model restrictions is both tight and sharp.

Theorem 6.1 exploits the first-order conditions, participation and incentive-compatibility constraints, and an inequality derived from the optimization problem. The second-order conditions of the cost minimization problem are satisfied for all $\gamma > 0$. Are there any other restrictions? The short answer is no. We now establish that, given the underlying data-generating process, every positive γ satisfying (6.9) is admissible. Thus Γ, a Borel set of risk-aversion parameters defined as

$$\Gamma \equiv \{\gamma > 0 : Q_0(\gamma) \geq 0\}, \tag{6.10}$$

indexes all the parameterizations that are observationally equivalent to the true model.

Tight means that Γ covers the identified set. By construction Γ is tight. Sharp means that every element in Γ is admissible for the data-generating process. A subset of the restrictions from the model might define a sharp set, in which case adding other restrictions to the same model would not shrink that set. Theorem 2.2 in Gayle and Miller, 2015, establishes that Γ is also sharp. That is, for any $\gamma > 0$ satisfying (6.9), Equations

(6.5), (6.3) and (6.4) define the primitives for the principal–agent models considered here that generate a compensation schedule from the optimal contract, given in (3.7), that matches the data-generating process. The constructed α and β satisfy the inequalities $0 < \beta < \alpha$; the constructed $g(x)$ is positive with $E[g(x)] = 1$ and $\lim_{x\to\infty}[g(x)] = 0$; every draw from the data set (x_n, w_n) satisfies $w_n = w(x_n)$, where $w(x)$ is the optimal contract for the constructed model. Since Γ is both tight and sharp, the model is set-identified, with Γ being the identified set of γ, which maps to the other parameters of the model based on Theorem 6.1.

Note that if only negative values of γ can make the expression in (6.9) non-negative, then the model is rejected. Does there exist a positive γ that rationalizes this model of pure moral hazard for any joint probability distribution of (x, w)? The next lemma answers this question in the negative by showing that the model is rejected by some data-generating processes. It proves that the profit inequality embodies restrictions that have empirical content: Some joint distributions of revenue and compensation are incompatible with all positive risk-aversion parameters.

Lemma 6.2 (Corollary 2.1 of Gayle and Miller, 2015)**.** There exist joint distributions of (x, w) such that $Q_0(\gamma) < 0$ for all $\gamma > 0$.

The process of identifying the pure moral hazard model

(PMH) and determining its empirical content can be summarized in four steps. First, given the probability density of revenue conditional on working and the compensation schedule as a function of revenue, the taste or ability parameter for working is recovered. This is achieved using the participation constraint by computing the certainty equivalent of compensation. Specifically, this involves integrating exponentiated compensation, scaled by the risk-aversion parameter, over the revenue distribution when the agent works. The taste parameter is normalized with respect to the agent's outside option.

Second, the likelihood ratio of revenue densities for shirking versus working is derived. This step relies on the mapping from profits to compensation as defined by the first-order condition for cost minimization, along with a regularity condition asserting that very high profits are unlikely when the agent shirks. By multiplying the revenue density from working by this likelihood ratio, the revenue density conditional on shirking is obtained.

Third, the preference for shirking is recovered using the incentive-compatibility constraint, which holds with equality in the cost-minimizing contract. The same approach used for working is applied here: exponentiated compensation, scaled by the risk-aversion parameter, is integrated over the revenue distribution when the agent shirks.

Finally, the fourth step uses the profit-maximization condition to partially identify the risk-aversion parameter, providing empirical content to the PMH model. This is achieved by identifying the subset of positive real numbers representing risk-aversion parameters that satisfy an inequality reflecting the principal's greater profits—revenue minus expected compensation—from a variable-compensation working contract over a fixed-wage shirking contract. Notably, this subset may be empty, in which case the model is rejected. That gives the model empirical content, that is, whether the model is plausible to the data.

In the empirical literature on managerial compensation, a significant challenge in determining whether current compensation practices are efficient lies in a measurement problem. This comes from the fact that the key elements required to evaluate the efficiency of executive compensation are inherently unobservable. A critical question, therefore, is whether models can always be constructed to fit any observed empirical regularity.[3]

The identification theorems establish that the PMH model is identified, meaning it cannot explain any and all empirical regularities, thereby confirming its empirical content. Similar identification results extend to various model adaptations. Section 8 introduces the hybrid moral hazard model (HMH) from Gayle and Miller, 2015, which incorporates hidden information

[3]See Abowd and Kaplan, 1999, and Oyer *et al.*, 2011, for comprehensive reviews of this literature.

alongside moral hazard. Section 9 discusses a human capital model with moral hazard from Gayle *et al.*, 2015, which builds on the PMH framework by incorporating job turnover and career concerns.

4 Multiple bond prices

In the dynamic version of the PMH model, we could reduce the identified set of γ exploiting distinct bond prices across different states.

As before, α, β, and $g(x)$ are identified if γ is known. Set identification for γ in the dynamic PMH model is established in Gayle and Miller, 2015, using profit-maximizing conditions. Furthermore, the appendix of Gayle and Miller, 2009a, shows that, when bond prices vary across different states (dates or sectors), it changes the intertemporal trade-off between current nonpecuniary benefits and the risky pecuniary compensation in the next period, leading to different compensation schedules, which helps identify the risk aversion γ that is constant.

A natural place to begin investigating the identification of γ^* is the participation constraint for working

$$\psi_t(\gamma) \equiv E_t\left[\exp\left(-\frac{\gamma w_t(x)}{b_{t+1}}\right)\right] = \left(\frac{1}{\alpha}\right)^{1/(b_t-1)} \tag{6.11}$$

where b_t is the bond price in period t as in Section 3. We assume that $\alpha > 1$, meaning that working brings more disutility than the outside option, then the participation constraint implies a lower bound for the risk-aversion parameter γ as the expected compensation for working must be positive to compensate for the higher disutility relative to the outside option. Since the participation constraint holds for each date while α remains constant, variation in bond prices across dates r and s introduces moment conditions of the form

$$\psi_r(\gamma)^{b_r-1} = \psi_s(\gamma)^{b_s-1}. \tag{6.12}$$

Similarly, we can exploit the cross-sectional variation in risk profiles to introduce identification conditions for γ. If there are at least two positions within a firm or sectors across firms, denoted as state r and s, where executives with identical preferences (α) face different compensation schemes $w_r(x)$ and $w_s(x)$, along with different firm risks $f_r(x)$ and $f_s(x)$, γ can be identified by the intersection of the two participation constraints

$$\psi_r(\gamma) = \psi_s(\gamma). \tag{6.13}$$

Figure 6.1 illustrates that if there is a unique intersection provided $\alpha > 1$, then γ is point-identified. We pin down the shape of $\psi(\cdot)$ as follows. By its definition, $\psi_t(0) = 1$, while the assumption above implies

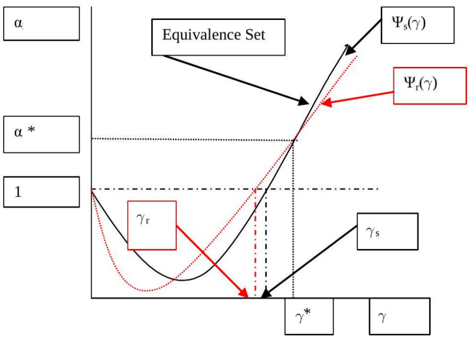
$$\psi_t'(0) = \frac{\partial}{\partial \gamma} E\left[\exp\left(-\frac{\gamma w_t(x)}{b_{t+1}}\right)\right]_{\gamma=0} = -E\left[\frac{w_t(x)}{b_{t+1}}\right] < 0.$$

Also $\psi_t(\gamma)$ is convex in γ because

$$\frac{\partial^2}{\partial \gamma^2}\left[\exp\left(-\frac{\gamma w_t(x)}{b_{t+1}}\right)\right] = \left(\frac{w_t(x)}{b_{t+1}}\right)^2 \exp\left(-\frac{\gamma w_t(x)}{b_{t+1}}\right) > 0$$

and the expectations operator preserves convexity. Assuming $\alpha > 1$, it now follows that $\psi_t(\gamma)$ crosses the unit level from below just once at say γ_t, which implies $\psi_t(\gamma) > 1$ for all $\gamma > \gamma_t$. This rules out the possibility that $\gamma^* \leq \gamma_t$. Intuitively, each participation constraint is satisfied by different combinations of γ and α satisfying $\gamma > \gamma_t$ and $\alpha = \psi_t(\gamma)^{1-b_t}$ as we see in Figure 6.1; we cannot distinguish between the alternative parametrizations of (γ, α) in the equivalence set given by a single participation constraint. Provided the participation constraints for several states, although there may still be multiple roots for γ, if there is a unique root satisfying $\gamma > \gamma_t$, then γ is point-identified.

Figure 6.1
Equivalence Set



Note: Excerpt from Gayle and Miller, 2009a, Online Appendix. This figure illustrates how to exploit variations across states to narrow the identified set. Each curve represents the equivalence set of pairs of (γ, α). If there is a unique intersection where $\alpha > 1$ between equivalence sets for different states, as the graph illustrates, then the model is point-identified.

7

A Semiparametric Approach to Estimation

This Section develops and applies a semiparametric estimation approach for the pure moral hazard model (PMH). We present a three-step estimation procedure that allows us to test the PMH model against the data. The approach is particularly valuable because it avoids imposing restrictive parametric assumptions on the distribution of returns and compensation. We demonstrate how this methodology can be used to make meaningful inferences about the model parameters.

1 A semiparametric estimator

This Section outlines a general approach to estimation, testing and inference based on our identification analysis. It then illustrates the approach within the context of executive compensation. The application tests pure moral hazard models (PMH) of executive compensation and estimates the importance of asymmetric information in the variations not rejected by the data.

Top executives in publicly traded companies are paid mostly with firm-denominated securities, but their total compensation also varies positively with accounting benchmarks. The PMH model is the standard paradigm for explaining why executives are paid in firm-denominated securities instead of a fixed wage. Yet in the simple PMH, α, β, $f(x)$, and $g(x)$ do not depend on accounting reports, which we can easily reject.

The PMH model with multiple states discussed below can simultaneously justify why compensation consists mostly of firm-denominated securities and depends on accounting benchmark if accounting reports are verifiable.

Now assume that before the agent chooses effort, there are two states $s \in \{1, 2\}$ about the production process, and the probability that state s occurs is distributed with probability $\varphi_s \in (0, 1)$ that sums up to one. If state s occurs, revenue is drawn from the probability density function $f_s(x)$ if the agent works and from $g_s(x)f_s(x)$ if the agent shirks. The state is observed by both the agent and the principal. The same identification results from Section 6 hold for each state. The estimation and testing procedure outlined in this Section assumes the data comprise independent draws of states, profits, and compensation payout, (s_n, x_n, w_n), for a sample of N observations generated in equilibrium. We propose a three-step procedure to estimate and test these models:

- Step 1: Nonparametrically estimate $w_s(x)$, $\overline{w}_s$, $f_s(x)$, $h(x)$, $\overline{h}$ and φ_s from the data (s_n, x_n, w_n).
- Step 2: Using the definitions of Γ in (6.10), estimate the confidence interval of γ. Reject the specification if the confidence interval is empty.
- Step 3: Using the formulas from Theorems 6.1, estimate the confidence interval for $g_s(x)$, β and α by replacing γ with values in the estimated confidence interval from Step 2 and $w_s(x)$, $\overline{w}_s$, $f_s(x)$, $h(x)$, $\overline{h}$ and φ_s with their estimates from Step 1.

Step 1(a): Compensation schedules The analysis focuses on cases when work is induced, (3.15), (8.15), (8.16) and (8.17) hold, and compensation, w_n, depends nontrivially on revenue x_n. Given $s \in \{1, 2\}$ and any x, we estimate $w_s^o(x)$ with the nonparametric kernel regression

$$w_s^{(N)}(x) = \left[\sum_{n=1}^{N} \boldsymbol{I}\{s_n = s\} K\left(\frac{x - x_n}{\varsigma_N}\right)\right]^{-1} \sum_{n=1}^{N} \boldsymbol{I}\{s_n = s\} K\left(\frac{x - x_n}{\varsigma_N}\right) w_n. \quad (7.1)$$

where $K\left(\frac{x - x_n}{\delta_N}\right)$ is the kernel, ς_N is the bandwidth associated with x, and $\boldsymbol{I}\{s_n = s\}$ is an indicator function that takes the value one if $s_n = s$ and zero otherwise.

Step 1(b): Probability densities We use the simple frequency estimator of φ_s defined by

$$\varphi_1^{(N)} = N^{-1} \sum_{n=1}^{N} \boldsymbol{I}\{s_n = 1\}. \tag{7.2}$$

The probability density, $f_s(x)$, is nonparametrically estimated by

$$f_s^{(N)}(x) = \frac{\sum_{n=1}^{N} \boldsymbol{I}\{s_n = 1\} K_1\left(\frac{x-x_n}{\delta_{1N}}\right)}{\delta_{1N} \sum_{n=1}^{N} \boldsymbol{I}\{s_n = s\}} \tag{7.3}$$

where $K_1\left(\frac{x-x_n}{\delta_{1N}}\right)$ is the kernel and δ_{1N} is the bandwidth associated with x.

Step 1(c): Boundary conditions For each state $s \in \{1, 2\}$, we estimate $\overline{w}_s$ with the super-consistent bound estimator

$$\overline{w}_s^{(N)} \equiv \max\left\{w_1 \boldsymbol{I}\{s_1 = s\}, \ldots, w_N \boldsymbol{I}\{s_N = s\}\right\}. \tag{7.4}$$

Let $x(q)$ correspond to the q^{th} highest value of x_n within the subset of data formed from observations for which $s_n = 1$.

Step 2: Risk parameter We obtain Borel sets that asymptotically cover Γ in the PMH model with probability greater than or equal to $1 - \delta$ for $\delta \in (0, 1)$ by inverting tests of the null hypothesis that γ is the true value for each $\gamma > 0$. Suppose $T^{(N)}(\gamma)$ is a test statistic and $c^{(N)}_{1-\delta}(\gamma)$ is the corresponding critical value for the test with significance level δ. Then the $(1-\delta)$ confidence interval is

$$CI^{(N)}[\gamma] \equiv \{\gamma > 0 : T^{(N)}(\gamma) \leq c^{(N)}_{1-\delta}(\gamma)\}. \tag{7.5}$$

Under the null hypothesis, the probability that $\Gamma \subseteq CI^{(N)}[\gamma]$ for the PMH model converges to $(1-\delta)$.

The test statistic for the PMH model is based on a sample analogue of $Q_0(\gamma)$ defined in (6.9), which we denote by $Q^{(N)}(\gamma)$. Under the null hypothesis $\min\{0, Q_0(\gamma)\}^2 = 0$. Lemma (6.2) shows that such a test has power against the null hypothesis. Our test statistic is

$$T^{(N)}_{\text{MMM}}(\gamma) = \min\left\{0, N^{1/2}Q^{(N)}(\gamma)\right\}^2. \tag{7.6}$$

There are several methods for obtaining the critical value; we follow Chernozhukov *et al.*, 2007.[1]

The asymptotic distribution of $T^{(N)}(\gamma)$ is the asymptotic distribution of its most slowly converging components. In the PMH model, there is only one boundary condition estimate,

[1]This is called the moment selection t-test (MMM). See Andrews and Soares, 2010, for a discussion of this class of critical-value functions. An alternative statistic is the quasilikelihood ratio (QLR) statistic, defined as $T^{(N)}_{\text{QLR}}(\gamma) = \inf_{t\in R_{+,\infty}}\left(N^{1/2}Q^{(N)}(\gamma) - t\right)^2$. See Pakes *et al.*, 2005, Chernozhukov *et al.*, 2007, and Romano and Shaikh, 2010 for studies using MMM. See Rosen, 2008, Andrews and Guggenberger, 2009, Andrews and Soares, 2010, and Andrews and Barwick, 2012, for studies using QLR.

$\overline{w}_s^{(N)}$; it converges to $\overline{w}_s$ at a rate faster than $N^{1/2}$ if x is observed without error and the estimator defined in (7.4) is used.

Step 3: Cost of effort and likelihood ratio Let $X^o = \{\underline{x}, \ldots, \overline{x}\}$ be a grid of points that span the support of observed profits, x_n. In the PMH model, suppose that $CI^{(N)}[\gamma]$ is nonempty. Then, from (6.3), the set of β induced by γ is

$$CI^{(N)}[\beta] = \left\{ \frac{1 - \sum_{s=1}^{2} \varphi_s^{(N)} \sum_{x^o \in X^o} \left[f_s^{(N)}(x^o)\, e^{\gamma\left(w_s^{(N)}(x^o) - \overline{w}_s^{(N)}\right)} \right]}{\sum_{s=1}^{2} \varphi_s^{(N)} \sum_{x^o \in X^o} \left[f_s^{(N)}(x^o)\, e^{-\gamma w_s^{(N)}(x^o)} \right] - e^{\gamma \max\left[\overline{w}_1^{(N)}, \overline{w}_2^{(N)}\right]}} : \forall \gamma \in CI^{(N)}[\gamma] \right\}, \tag{7.7}$$

and $CI^{(N)}[\alpha]$ is defined accordingly using (6.4). Using (6.5) the set of $g_s(x^o)$ for each x^o in X^o is, for $s = 1, 2$

$$CI^{(N)}[g_s(x^o)] = \left\{ \frac{e^{\gamma \overline{w}_s^{(N)}} - e^{\gamma w_s^{(N)}(x^o)}}{e^{\gamma \overline{w}_s^{(N)}} - \sum_{x \in X^o} \left[f_s^{(N)}(x) e^{\gamma w_s^{(N)}(x)} \right]} : \quad \forall \gamma \in CI^{(N)}[\gamma] \right\}, \tag{7.8}$$

where $\varphi_s^{(N)}$, $f_s^{(N)}(x)$, $w_s^{(N)}(x^o)$ and $\overline{w}_s^{(N)}$ are the estimates obtained from Step 1.

2 Application

We present the results applying the semi-parametric estimator for the PMH model across firm types and sectors. Table 7.1 provides the estimated confidence intervals in the least restrictive version of the PMH model. In this specification, γ is allowed to vary across firm types and sectors, but not over time or by accounting return. Conversely, α and β are allowed to vary across firm types, sectors, accounting returns, and periods. The unrestricted PMH model is not rejected at the 5% significance level. The intersection of the estimated intervals is (0.02, 0.21), which is non-empty. Consequently, we cannot reject the hypothesis that a common γ applies to all firm types within all sectors.

Table 7.1
95% Confidence Region of Risk-Aversion (Unrestricted PMH)

Firm type (A, W, D)	Primary Sector	Consumer Sector	Service Sector
(S, S, S)	(0.01, 13.40)	(0.01, 0.43)	(0.01, 1.61)
(S, L, S)	(0.01, 1.61)	(0.01, 6.61)	(0.01, 1.78)
(S, L, L)	(0.01, 2.66)	(0.01, 3.61)	(0.01, 0.24)
(S, S, L)	(0.01, 4.88)	(0.01, 16.40)	(0.01, 3.26)
(L, S, S)	(0.01, 9.90)	(0.01, 0.29)	(0.01, 0.21)
(L, L, S)	(0.02, 4.00)	(0.01, 20.10)	(0.01, 0.35)
(L, L, L)	(0.02, 2.66)	(0.01, 4.88)	(0.01, 0.43)
(L, S, L)	(0.01, 4.88)	(0.01, 0.39)	(0.01, 18.20)
Observations	7,796	5,600	8,536

Note: Excerpt from Gayle and Miller, 2015, Table 5. The subsampling procedure was performed using 100 replications of subsamples with 3000 observations each using a grid of 1000 equally spaced points on the interval [9.112E-04, 50]. Firm type is measured by the triplicate (A,W,D), where A is assets, W is number of workers, and D is debt–equity ratio with each corresponding to whether that element is above (L) or below (S) its industry average. All specifications automatically imposed cost minimization.

Table 7.2 presents the estimated identified set of γ from applying additional exclusion restrictions, which impose the independence of the taste parameters α and β on accounting returns. There is no common overlap region for γ across all 24 firm and sector types. The bottom panel of Table 7.2 presents results under the restriction that β is equal across the two

Table 7.2

95% Confidence Region of Risk-Aversion (Restricted PMH)

Firm type (A, W, D)	Primary Sector	Consumer Sector	Service Sector
Cost of shirking ($\alpha_{1t} = \alpha_{2t}$)			
(S, S, S)	(0.4, 0.5) ∪ (2.0, 13.4)	(0.01, 0.43)	(0.29, 1.61)
(S, L, S)	(0.16, 0.39)	(0.01, 0.48)	(0.06, 0.21)
(S, L, L)	(0.01, 0.43)	(0.07, 0.16)	(0.01, 0.24)
(S, S, L)	(0.17, 0.72)	(0.17, 0.59)	(2.18, 3.26)
(L, S, S)	(0.19, 9.90)	(0.09, 0.29)	(0.13, 0.21)
(L, L, S)	(0.02, 0.43)	(0.02, 0.59)	(0.06, 0.26)
(L, L, L)	(0.02, 0.29)	(0.01, 0.43)	(0.01, 0.21)
(L, S, L)	(0.08, 0.53)	(0.16, 0.24)	(3.26, 16.40)
Cost of working ($\beta_{1t} = \beta_{2t}$)			
(S, S, S)	(2.18, 8.09)	(0.26, 0.43)	(0.48, 1.61)
(S, L, S)	(0.01, 0.02)	(1.31, 1.97)	(0.01, 0.03)
(S, L, L)	(0.03, 2.18)	(2.41, 3.99)	(0.06, 0.24)
(S, S, L)	(0.01, 4.88)	(1.45, 16.40)	(2.41, 3.26)
(L, S, S)	(0.01, 0.03) ∪ (1.31, 2.95)	(0.17, 0.29)	(0.05, 0.21)
(L, L, S)	(0.02, 0.07)	(0.02, 0.03)	(0.03, 0.35)
(L, L, L)	(0.02, 0.05)	(0.02, 0.04)	(0.17, 0.43)
(L, S, L)	(0.01, 0.02)	(0.01, 0.02)	(5.40, 18.20)
Observations	7,796	5,600	8,536

Note: Excerpt from Gayle and Miller, 2015, Table 6. The subsampling procedure was performed using 100 replications of subsamples with 3000 observations each using a grid of 1000 equally spaced points on the interval [9.112E-04, 50]. Firm type is measured by the triplicate (A,W,D), where A is assets, W is number of workers, and D is debt–equity ratio with each corresponding to whether that element is above (L) or below (S) its industry average. All specifications imposed cost minimization and profit maximization. The risk-aversion parameter from the interaction across sector, size, and leverage is $\gamma \in [0.02, 0.21]$ for the pure moral hazard based on only cost minimization and profit maximization.

accounting return states. Similarly, the bottom panel shows no overlap across the three columns. Even when γ is allowed to vary across all 24 firm and sector combinations, the data rejects the hypothesis that α and β are independent of accounting returns.[2]

[2] Notably, the identified set for γ is empty in three of the firm-sector combinations when intersecting the intervals from the top and bottom panels. The three firm-sector combinations in which γ regions do not intersect in their corresponding left and right panels are (L,S,L) in the primary sector, (S,L,L) in the consumer sector and (S,L,S) in the service sector.

These limitations of the pure moral hazard model motivate us to introduce the framework in the next Section to incorporate the case of hidden information where the state of production is unobserved by the principal.

8

The Hybrid Model with Hidden Information

Motivated by the evidence in Gayle and Miller, 2009b, that executives extract abnormal returns from holding firm securities, we extend the model to account for the private information that executives exploit for their financial gains at the expense of shareholders. When the state of production is unobservable to the principal and must instead be reported by the agent, it creates another agency issue—hidden information. The manager reports the firm's financial condition through accounting statements, exercising significant discretion over the reported values. The manager can influence the firm's balance sheets in various ways, such as selecting different valuation methods for assets and liabilities or timing the write-off of nonperforming assets strategically. This flexibility provides the manager the flexibility to signal the state of the firm to shareholders. The pure moral hazard model (PMH) can be extended to situations in which the agent has hidden information about the state of production besides moral hazard—a model we refer to as hybrid moral hazard (HMH). The agent makes the employment decision, l_0, then privately observes the state $s \in \{1, 2\}$, reports the state, $r \in \{1, 2\}$, and then makes the effort choice. The probability of state s is ψ_s, which adds up to 1. Assume the regularity conditions (3.1), (3.2), (3.3) apply to both states individually. If the agent reports the second state, $r = 2$, indicating good revenue prospects, then the principal can independently verify it. This reflects legal considerations that constrain the agents not to overstate revenue prospects. In other words, the agent cannot lie that the state is good reportedly ($r = 2$) when the bad state occurred ($s = 1$). However, given a good state $s = 2$, the agent may underreport, $r = 1$, unless incentives are provided to dissuade him from doing so.

Define $h(x) \equiv \varphi_2 f_2(x)/\varphi_1 f_1(x)$ as the weighted likelihood ratio of the second state occurring relative to the first given any observed value of excess returns $x \in R$ conditional on the agent working, and assume

$$\lim_{x \to \infty}[h(x)] = \sup_{x \in R}[h(x)] \equiv \overline{h} < \infty. \tag{8.1}$$

The boundedness of $h(x)$ precludes the case where the principal could infer the state based on some realization of revenue and punish the agent severely if his reported state was inconsistent with the subsequently realized revenue. If $h(x)$ was unbounded for some value of x' (i.e. $f_2(x') > 0$ and $f_1(x') = 0$), then the principal could enforce truth telling about $s = 2$ without cost by committing to severely punish the agent if $r = 1$ is reported but x' is subsequently drawn as the revenue outcome. Thus, (8.1) rules out this possibility by bounding $h(x)$. The first equality in (8.1) captures the idea that if the agent works, then the likelihood of the second state is highest relative to that of the first state when the revenue attains its highest values. The agent's compensation, denoted by $w_r(x)$, depends on $r \in \{1, 2\}$—what he discloses about the state of production—and its subsequent performance, x, revealed to both the principal and the agent at the end of the period.

Truth telling and sincerity constraints Contracts between the principal and the agent that induce working in both states and honest reporting in state 2 must satisfy a participation constraint, two incentive-compatibility constraints (one for each state), and two additional conditions inducing the agent to truthfully reveal his private information (one for each state). Define $v_s(x) \equiv \exp[-\gamma w_s(x)]$ as the multiplicative utility value from the payoff $w_s(x)$. We rewrite the incentive-compatibility constraint conditional on the state as

$$\begin{aligned}
&\int \left[1-(\beta/\alpha)\, g_s(x)\right] v_s(x) f_s(x)\mathrm{d}x \\
&\quad\equiv E_s\left\{\left[1-(\beta/\alpha)\, g_s(x)\right] v_s(x)\right\} \\
&\quad\leq 0, \quad s\in\{1,2\},
\end{aligned} \tag{8.2}$$

and the participation constraint for working (unconditional on the state) as

$$\sum_{s=1}^{2}\varphi_s \int \left[v_s(x)\right] f_s(x)\mathrm{d}x \equiv E\left[v_s(x)\right] \leq \alpha^{-1}. \tag{8.3}$$

To induce the agent to report truthfully in each state, we append (8.2) and (8.3) with two further constraints. Conditional on the agent working, the principal can prevent the agent from lying about the second state by requiring contracts to satisfy

$$\int \left[v_2(x)-v_1(x)\right] f_2(x)\mathrm{d}x \equiv E_2\left[v_2(x)-v_1(x)\right] \leq 0. \tag{8.4}$$

An optimal contract also induces the agent not to understate and shirk in the second state, behavior we describe as sincere. The sincerity constraint that prevents the agent from lying about the second state and shirking is

$$\begin{aligned}&\int \left[v_2(x) - (\beta/\alpha)\, v_1(x) g_2(x)\right] f_2(x)\mathrm{d}x \\ &\equiv E_2\left[v_2(x) - (\beta/\alpha)\, v_1(x) g_2(x)\right] \\ &\leq 0, \end{aligned} \tag{8.5}$$

where $-\beta v_1(x)$ is the utility obtained from shirking and announcing the first state, and $f_2(x)g_2(x)$ is the probability density function associated with shirking when the second state occurs.

Several cases If the agent's choice of working or shirking can be observed, and the optimal action does not depend on the state, then the principal would pay the agent a fixed wage, $\gamma^{-1}\ln\alpha$ to work or $\gamma^{-1}\ln\beta$ to shirk, and would only demand accounting statements that certified agreements. The situation changes if the optimal action depends on the unobserved state. In this case the constraints for participation (8.3) and truth telling (8.4) apply, but not those for incentive compatibility (8.2) or sincerity. Then the principal maximizes

$$\sum\nolimits_{s=1}^{2} \int \varphi_s \log\left[v_s(x)\right] f_s(x)\mathrm{d}x \equiv E\left[\log v_s(x)\right] \tag{8.6}$$

subject to (8.3) and (8.4), a concave problem with linear constraints.

Optimal contracting in the hybrid model Since $v_s(x)$ is monotone decreasing in $w_s(x)$, deriving $w_s(x)$ to minimize expected compensation for inducing work in both states is tantamount to choosing $v_s(x)$ for each (s, x) to maximize (8.6) subject to (8.2), (8.3), (8.4), (8.5).

To induce work and truth telling in both states, the principal maximizes the Lagrangian

$$\sum_{s=1}^{2} \varphi_s \int \left\{ \log\left[v_s(x)\right] + \eta_0 \left[\alpha^{-1} - v_s(x)\right] \right\} f_s(x) \mathrm{d}x$$
$$+ \sum_{s=1}^{2} \varphi_s \eta_s \int_{\underline{x}}^{\infty} v_s(x) \left(\frac{\beta}{\alpha} g_s(x) - 1 \right) f_s(x) \mathrm{d}x$$
$$+ \varphi_2 \int \left\{ \eta_3 \left[v_1(x) - v_2(x)\right] + \eta_4 \left[\frac{\beta}{\alpha} v_1(x) g_2(x) - v_2(x) \right] \right\} f_2(x) \mathrm{d}x \quad (8.7)$$

with respect to $v_s(x)$, where η_0 through η_4 are the shadow values assigned to the linear constraints. Since each constraint is a convex set, their intersection is too. Also, $\log v$ is concave increasing in v, the expectation operator preserves concavity, so the objective function is concave in $v_s(x)$ for each x. Hence, the Kuhn Tucker theorem guarantees there is a unique positive solution to the equation system formed from the first-order conditions augmented by the complementary-slackness conditions.

The differences between the cost-minimization problems for the PMH and HMH models are evident from (8.7). In the PMH model $\eta_3 \equiv \eta_4 \equiv 0$ because the truth-telling and sincerity constraints do not figure into the formulation of the problem. The first-order conditions for this problem are

$$v_1(x)^{-1} = \eta_0 + \eta_1 \left[(\alpha/\beta) - g_1(x)\right] - \eta_3 h(x) - \eta_4 \left(\beta/\alpha\right) g_2(x) h(x)$$

$$v_2(x)^{-1} = \eta_0 + \eta_2 \left[(\alpha/\beta) - g_2(x)\right] + \eta_3 + \eta_4. \tag{8.8}$$

The following lemma is helpful for interpreting the first-order conditions.

Lemma 8.1 (Lemma 3.1 of Gayle and Miller, 2015)**.** The Lagrange multipliers satisfy

1. $\eta_0 = \alpha$
2. $\eta_3 + \eta_4 = E_2\left[v_2(x)\right]^{-1} - E\left[v_s(x)\right]^{-1}$.

From the second equality in Lemma 8.1, we infer that if, as in the PMH model, $\eta_3 = \eta_4 = 0$, then

$$E_2\left[v_2(x)\right] = E\left[v_s(x)\right] = E_1\left[v_1(x)\right].$$

In words, if neither the truth-telling nor the sincerity constraints bind, or if the state is directly observed by the principal, then the PMH model applies, and expected utility is equalized across states. Otherwise, $(\eta_3 + \eta_4)$ is strictly positive, which implies that the expected utility of the PMH model straddles the expected utility attained in the HMH model

$$E_2\left[v_2(x)\right] < E\left[v_s(x)\right] < E_1\left[v_1(x)\right].$$

When the agent has private information, he is rewarded for announcing $s = 2$ and penalized for $s = 1$; in other words, the optimal contract pays him for luck.

There are three other contracts the principal might design. All three involve the agent shirking in at least one state. The cost-minimizing contract for shirking in both states is found by setting $\eta_1 = \eta_2 = 0$, and in both states the agent is paid $\gamma^{-1} \ln(\beta_1)$. To make the agent work in the first state and shirk in the second, the principal sets $\eta_2 = 0$ in the cost minimization problem. From the second part of Lemma 8.1, the agent receives a certain utility in the second state that exceeds his expected utility in the first state because to instill incentives in the first state, the agent must be rewarded to reveal when it does not occur. Finally, when the agent chooses work in the second state and shirking in the first, at least one of the multipliers, η_3 or η_4, is strictly positive: From its first-order condition, $v_1(x)$ also depends on revenue, through $h(x)$ and possibly $g_2(x)$. Rather than load all the risk premiums into the second state, compensation in the first state optimally depends on revenue, not to induce work, but to induce truth telling and sincerity. The principal completes the optimization by comparing the profits from each of the four contract types using the solutions to the respective cost-minimization problems.

1 Identification

The primitives of the HMH model differ from those of the PMH model of the previous Section only because the states are unobserved. We assume cross-sectional data is available on (r, x, w), where $r \in \{1, 2\}$ is a report by the agent on the firm's state. In equilibrium, agents truthfully reveal the state, implying $s = r(s)$, so $r = s$ in the data generated by the HMH model. Consequently, φ_s and $f_s(x)$ are identified from observations on (r, x), and similarly $w_s(x) = w_r(x)$ is identified from observations on (r, x, w). All that remains is to identify (α, β, γ) and $g_s(x)$ for $s \in \{1, 2\}$. We follow the same procedure as in the previous section.

The most important differences between the identification of PMH in Section 6 and this one arise from the inequalities and equations that define equilibrium. These differences complicate the analysis of the identification in the HMH model. Nevertheless, the main thrust of the results derived for the easier PMH model also hold for the HMH model. First, if γ, the risk parameter is known, then the remaining parameters are nonparametrically point-identified. Second, if γ is unknown then all the parameters are only set-identified. Third, we obtain sharp and tight bounds.

To set the stage for the theorem on tightness, let $\overline{v}_s(\gamma) \equiv \sup\left[e^{-\gamma w_s}\right]$ and define the real-valued mappings

$$\hat{\alpha}(\gamma) \equiv \left[\int \sum_{s=1}^{2} \varphi_s v_s(x,\gamma) f_s(x) \mathrm{d}x\right]^{-1} \equiv E\left[v_s(x,\gamma)\right]^{-1}$$

$$\hat{\beta}(\gamma) \equiv \hat{\alpha}(\gamma) \left\{ \frac{\left[\overline{v}_2(\gamma)\right]^{-1} - E_2\left[v_2(x,\gamma)\right]^{-1}}{\left[\overline{v}_2(\gamma)\right]^{-1} - E_2\left[v_2(x,\gamma)^{-1}\right]} \right\}.$$

For any given γ, we interpret $\hat{\beta}(\gamma)$ and $\hat{\alpha}(\gamma)$ as taste parameters because one can show, by following the same arguments used to characterize $\beta(\gamma)$ and $\alpha(\gamma)$ in the PMH case, that $0 < \hat{\beta}(\gamma) < \hat{\alpha}(\gamma)$ for all $\gamma > 0$. Also, let

$$g_2(x,\gamma) \equiv \frac{\overline{v}_2(\gamma)^{-1} - v_2(x,\gamma)^{-1}}{\overline{v}_2(\gamma)^{-1} - E_2\left[v_2(x,\gamma)^{-1}\right]}.$$

As in the previous section, $g_2(x,\gamma)$ is positive with $E_2\left[g_2(x,\gamma)\right] = 1$ and therefore can be interpreted as a likelihood ratio function of x for all $\gamma > 0$. Finally, we sequentially define $g_1(x,\gamma)$ by first defining $\eta_4(\gamma)$, then $\eta_3(\gamma)$ and $\eta_1(\gamma)$ as

$$g_1(x,\gamma) \equiv \frac{\overline{v}_1(\gamma)^{-1} - v_1(x,\gamma)^{-1} + \eta_3(\gamma)\left[\overline{h} - h(x)\right] - \eta_4(\gamma) g_2(x) h(x) \frac{\hat{\beta}(\gamma)}{\hat{\alpha}_2(\gamma)}}{\eta_1(\gamma)}, \tag{8.9}$$

where

$$\eta_4(\gamma) \equiv \frac{\frac{E_1[v_1(x,\gamma)]}{E[v_s(x,\gamma)]} - 1 - E_1\left[v_1(x,\gamma)h(x)\right]\left\{E_2\left[v_2(x,\gamma)\right]^{-1} - E\left[v_s(x,\gamma)\right]^{-1}\right\}}{\frac{\hat{\beta}(\gamma)}{\hat{\alpha}_2(\gamma)} E_1\left[v_1(x,\gamma) g_2(x) h(x)\right] - E_1\left[v_1(x,\gamma)h(x)\right]} \tag{8.10}$$

$$\eta_3(\gamma) \equiv E_2\left[v_2(x,\gamma)\right]^{-1} - \eta_4(\gamma) - E\left[v_s(x,\gamma)\right]^{-1} \tag{8.11}$$

$$\eta_1(\gamma) \equiv \frac{\hat{\beta}(\gamma)}{\hat{\alpha}_2(\gamma)} \left\{\overline{v}_1(\gamma)^{-1} - E\left[v_s(x,\gamma)\right]^{-1} + \eta_3(\gamma)\overline{h}\right\}. \tag{8.12}$$

By inspection, all the mappings above can be computed as population moments given a value for the risk-aversion parameter.[1]

[1]Note that $E\left[v_s(x,\gamma)\right]$, $E_s\left[v_s(x,\gamma)^{-1}\right]$, $E_s\left[v_s(x,\gamma)\right]$, $\overline{v}_s(\gamma)$, $\overline{h}$, $h(x)$, φ_s

We do not claim that $g_1(x,\gamma)$ is a likelihood ratio for all $\gamma > 0$, nor that $\eta_i(\gamma) \geq 0$ for each $i \in \{1,3,4\}$ so we cannot necessarily interpret them as Kuhn Tucker multipliers for all $\gamma > 0$. Nevertheless, $g_1(x,\gamma^*)$ is a likelihood ratio, by Theorem 8.2 below. This theorem is the analogue to Theorem 6.1. It shows that if an HMH model with parameter γ^* generates the data, then the remaining parameters are point identified by $\hat{\beta}(\gamma^*)$, $\hat{\alpha}(\gamma^*)$ and $g_s(x,\gamma^*)$. In other words, if the risk parameter is known, the HMH model is also point-identified without making any further parametric assumptions.

Theorem 8.2 (Theorem 3.2 of Gayle and Miller, 2015)**.** Suppose the data (x_n, r_n, w_n) is generated by a parameterization of the HMH model with a positive risk-aversion parameter γ^*. Then,

$$\beta = \hat{\beta}(\gamma^*)$$

$$\alpha = \hat{\alpha}(\gamma^*)$$

$$g_1(x) = g_1(x,\gamma^*)$$

$$g_2(x) = g_2(x,\gamma^*) .$$

Additional restrictions The HMH model imposes truth-telling and sincerity constraints. Since these constraints help shape the optimal contract as a function of the parameters, they provide several restrictions on the population moments that are not

and $f_s(x)$ can be expressed as population moments of the data-generating process given γ.

imposed in the PMH model. Define $\Psi_2(\gamma)$ through $\Psi_4(\gamma)$ as

$$\begin{aligned}
\Psi_2(\gamma) &\equiv E_1\left[1\left\{g_1(x,\gamma)\right\}-1\right] \\
\Psi_3(\gamma) &\equiv E_2\left[v_1(x,\gamma)-v_2(x,\gamma)\right] \\
\Psi_4(\gamma) &\equiv E_2\left[\hat{\alpha}_1(\gamma)v_1(x,\gamma)g_2(x,\gamma)-\hat{\alpha}(\gamma)v_2(x,\gamma)\right]. \qquad (8.13)
\end{aligned}$$

The truth-telling constraint (8.4) implies $\Psi_3\left(\gamma^*\right)\geq 0$, while the sincerity constraint (8.5) implies $\Psi_4\left(\gamma^*\right)\geq 0$. The equality $\Psi_3\left(\gamma^*\right)\Psi_4\left(\gamma^*\right)=0$ guarantees that at least one of the constraints holds strictly. Since $g_1(x)$ is a likelihood ratio in the HMH model, we ensure $\hat{g}_1\left(x,\gamma^*\right)\geq 0$ with unit mass by imposing the restriction that $\Psi_2\left(\gamma^*\right)\geq 0$. Three more inequalities ensure $\eta_1\left(\gamma^*\right)$, $\eta_3\left(\gamma^*\right)$ and $\eta_4\left(\gamma^*\right)$ are positive, a necessary condition for being Kuhn Tucker multipliers. Similarly, complementary-slackness conditions for truth telling and sincerity must be satisfied, meaning $\Psi_3\left(\gamma^*\right)\eta_3\left(\gamma^*\right)=0$ and $\Psi_4\left(\gamma^*\right)\eta_4\left(\gamma^*\right)=0$.

Another exclusion restriction imposed throughout is that β does not depend on the state.[2] To characterize this exclusion restriction, define

[2]This exclusion restriction is a natural one to impose in our application, but is easy to relax. Noting (8.14) defines $\Psi_1\left(\gamma^*\right)$, and (8.18) defines the constraint set $\hat{\Gamma}$, we redefine $\hat{\Gamma}$ by omitting the equation $\Psi_1\left(\gamma^*\right)=0$ from $\hat{\Gamma}$.

$$\Psi_1(\gamma) \equiv [\overline{v}_1(\gamma)]^{-1} - \eta_3(\gamma)\overline{h} + E_1\left[v_1(x,\gamma)^{-1}\right] - \eta_1(\gamma)$$
$$- \eta_3(\gamma)\varphi_2/\varphi_1 - \eta_4(\gamma)\left(\hat{\beta}(\gamma)/\hat{\alpha}_2(\gamma)\right)E_1\left[g_2(x,\gamma)h(x)\right]. \quad (8.14)$$

Gayle and Miller, 2015, prove in the appendix that if an HMH model with parameter γ^* generates the data, this restriction can be formally stated as $\Psi_1(\gamma^*) = 0$.

Turning now to the effort level induced by the principal in the HMH model, we first remark that if shirking is demanded in both states, then compensation is determined in Lemma 3.1 for the one-state PMH model. Since this is suboptimal,

$$\Lambda_1(\gamma) \equiv E\left[x - w_s(x)\right] - E\left[xg_s(x,\gamma)\right] + \gamma^{-1}\log\left[\hat{\alpha}_1(\gamma)\right] \quad (8.15)$$

is positive at γ^*. The principal induces work in both states when the expected profits from doing so are higher than when the agent shirks in one state; this remark yields two additional restrictions on γ^* to be utilized in identification. For any $\gamma \in R^+$, we denote by $w_s^{(s')}(x,\gamma)$ the cost-minimizing compensation in state s when the agent works in state $s' \in \{1,2\}$ and shirks in the other state. Given the parameterization indexed by γ, the difference in the value to the principal of demanding work in both states versus working in the first state only is

$$\begin{aligned}\Lambda_2(\gamma) &= \varphi_1 E_1\left[w_1^{(1)}(x,\gamma) - w_1(x)\right] \\ &\quad + \varphi_2 E_2\left\{x - w_2(x) - g_2(x,\gamma)\left[x - w_2^{(1)}(x,\gamma)\right]\right\}. \end{aligned} \tag{8.16}$$

At $\gamma = \gamma^*$, this expression is positive when it is optimal to require work in both states. In a similar fashion, we define

$$\begin{aligned}\Lambda_3(\gamma) &= \varphi_1 E_1\left\{x - w_1(x) - g_1(x,\gamma)\left[x - w_1^{(2)}(x,\gamma)\right]\right\} \\ &\quad + \varphi_2 E_2\left[w_2^{(2)}(x,\gamma) - w_2(x)\right]. \end{aligned} \tag{8.17}$$

and note that $\Lambda_3(\gamma^*) \geq 0$ for the same reason.

Sharp and tight bounds Consolidating the restrictions directly applied to the HMH model, we define $\hat{\Gamma}$, a Borel set of risk-aversion parameters, as

$$\hat{\Gamma} \equiv \left\{ \gamma > 0 \;\begin{array}{l} \Lambda_i(\gamma) \geq 0 \text{ for } i \in \{1,2,3\} \\ \eta_j(\gamma) \geq 0 \text{ for } j \in \{1,3,4\} \\ \Psi_1(\gamma) = 0 \text{ and } \Psi_k(\gamma) \geq 0 \text{ for } k \in \{3,4\} \\ \Psi_3(\gamma)\Psi_4(\gamma) = \Psi_3(\gamma)\eta_3(\gamma) = \Psi_4(\gamma)\eta_4(\gamma) = 0 \end{array} \right\}. \tag{8.18}$$

By construction, $\hat{\Gamma}$ is tight because every observationally equivalent parameterization must satisfy its restrictions. The last theorem demonstrates that the tight bounds constructed for the HMH model also exclude every parameterization that cannot be rationalized by the data.

Theorem 8.3 (Theorem 3.2 of Gayle and Miller, 2015). $\hat{\Gamma}$ is sharp.

2 An empirical application

This Section provides the empirical application from Gayle and Miller, 2015, which tests the hybrid moral hazard (HMH) and pure moral hazard (PMH) models and estimates the importance of hidden information using plausible model specifications.

Their application of the principal-agent model to executive compensation focuses on the role of accounting information: whether such information is common knowledge, or nonverifiable reports by managers with private information about the state of the firm. The notion that managers have more knowledge than shareholders about the state of the firm is well embedded in the delegation. Several empirical studies find that trading by corporate insiders appears profitable, which provide reasonable grounds for believing that managers have private information they exploit to their advantage when the opportunity arises.[3]

[3]See Lorie and Niederhoffer, 1968, Jaffe, 1974, Finnerty, 1976, and Seyhun, 1986, who find that insiders tend to buy before an abnormal rise in stock prices and sell before an abnormal decline. Seyhun, 1992a, and Seyhun, 1992b, present evidence showing that insiders earn over 5 percent abnormal returns on average, and determines that insider trades predict up to 60 percent of the total variation in one-year-ahead returns. Hayes and Schaefer, 2000, present evidence that the unexplained variation in current compensation predicts the future variation in firm performance. Gayle and Miller, 2009b, construct a simple self-financing dynamic portfolio strategy based on changes in asset holdings by managers that significantly outperforms the market portfolio, realizing over 90 percent of the gains that could have been achieved with perfect foresight.

Reduced form analysis shows that executive compensation correlates positively with accounting returns. However, it does not resolve whether managers have private information because both models—the pure moral hazard model with publicly-verifiable states and the hybrid moral hazard model with nonverifiable accounting information—predict that executive compensation depends on accounting performance. In both models, accounting events have information value for shareholders that helps them design the optimal contract by informing them about the state of the firm. The distinction lies in whether accounting reports can be manipulated by managers, and hence shareholders internalize the incentives to elicit managers to report truthfully. Distinguishing between the two regimes is crucial as they lead to differing implications for the value of accounting information.

They test several versions of the PMH and HMH models and analyze the measures of moral hazard with the models unrejected by the data. There are four ways to evaluate the moral hazard in managerial compensation, as defined in Section 3 for the dynamic version of the PMH model. The expected gross output loss to the firm for switching from the distribution of abnormal returns for working to the distribution for shirking, denoted as τ_1 in (3.18), scaled by the value of the firm at the beginning of the period; the nonpecuniary benefits to the manager from shirking, denoted as τ_2 in (4.3); the fixed reservation wage for working under perfect monitoring, if the shareholders can perfectly monitor, denoted as τ_4 in (4.4); and the firm's willingness to pay to eliminate the agency problem, denoted as τ_5 in (4.5).

Both models share the prediction with the literature on this topic that a manager would be paid a fixed salary if he shirks, that the revenue x is drawn from the probability distribution conditional on him working, and that asymmetric information explains why managerial compensation varies with abnormal firm returns, an almost universal finding of a large empirical literature subject only to the caveat that all the components comprising CEO compensation be included in the definition.

The hybrid moral hazard model concerns the role of asymmetric information in reported accounting returns. The manager has discretion in reporting the accounting measures, which may not align with the private information that the manager has about the state of production. The accounting returns are measured as in (2.1). For a given firm type, let $E\left[\pi_{nt}\right]$ denote the expected accounting return of π_{nt} for firm n at the beginning of period t before the manager compiles the accounting reports. To indicate the good and bad reported states, set $r_{nt} = 2$ if the firm's accounting return is higher than the average accounting returns for firms in the given type in period t, and $r_{nt} = 1$ otherwise. In contrast, the PMH model with two states assumes that the actual states of production are observed by shareholders and thus not distinguished from the reported states.

The estimated results of the HMH model are presented below, and compared with the estimation results of variants of the PMH model reported in Section 7. Then estimates of the welfare measures are obtained for the unrejected models, including the unrestricted PMH model and the HMH model.

Specification tests As discussed in Section 6, the model is set-identified, with the identified set of γ indexing the identified set of the model, while the other parameters can be expressed as mappings of γ. Table 8.1 reports the 95% confidence interval for the γ identified set for a restricted HMH model. The restriction they impose is that β and α depend only on the sector and type of the firm, and not on the accounting return or the calendar time, and assume a common γ between all sectors and types of firms. This model cannot be rejected at the 5% significance level for any sector. Intersecting across sectors, $\gamma \in (0.37, 0.42)$ corresponds to the identified set of the HMH model.

Table 8.1
95% Confidence Region of Risk-Aversion for the HMH Model

Sector	Observations	Confidence Region
Primary	7,796	$(0.002, 0.26) \cup (0.37, 0.42)$
Consumer	5,600	$(0.002, 0.13) \cup (0.19, 0.57)$
Service	8,536	$(0.27, 0.53)$

Note: Excerpt from Gayle and Miller, 2015, Table 7. The subsampling procedure was performed using 100 replications of subsamples with 3000 observations each using a grid of 1000 equally spaced points on the interval [9.112E-04, 50]. The restrictions imposed are profit maximization and equalization of preference parameters across size, leverage, and time.

Compared to the PMH model, the HMH model offers a more parsimonious fit to the data, which does not need the degree of heterogeneity in α, β, γ in the unrestricted PMH model. As reported in Section 7, also unrejected is the unrestricted PMH model where the taste parameters α and β are most flexible—they can vary by accounting return and over time, not just by firm sector and type (the flexibility that the restricted PMH allows for). The risk-aversion γ is assumed to vary across firm sector and type in the PMH models.

Gross loss from shirking (τ_1) The top panel of Table 8.2 presents the estimated gross losses for both the pure and hybrid moral hazard models. The gross losses to firms from shirking are estimated as percentages of the firms' assets. The differences between these two model specifications are relatively minor compared to the variation observed across firm types. For example, the median minimum distance between the confidence intervals of the pure and hybrid models is 0.32% in the primary sector, 1.74% in the consumer sector, and 1.49% in the service sector. In contrast, the variation in confidence intervals across firm types is significantly larger. For the pure moral hazard model, the range is between 7.84% and 14.89%, while for the hybrid model, it spans from 17.35% to 24.57%, depending on the sector. Given that the average annual stock market return during this period was approximately 10%, the expected gross return would have been negative for more than half of the firm and sector types in both model specifications if shareholders had ignored the moral-hazard problem.

Table 8.2

Confidence Regions for the Agency Costs

Firm Type	Primary		Consumer		Service	
(A, W, D)	PMH	HMH	PMH	HMH	PMH	HMH
Gross losses to firms from shirking (as percentages of assets)						
(S,S,S)	(12.53, 13.36)	(12.22, 12.44)	(15.15, 16.35)	(14.21, 14.53)	(15.32, 17.05)	(13.80, 14.19)
(S,L,S)	(7.27, 8.09)	(6.57, 6.72)	(14.11, 15.22)	(14.74, 14.99)	(11.72, 14.11)	(1.46, 1.53)
(S,L,L)	(3.45, 4.17)	(5.97, 6.42)	(8.71, 9.86)	(19.41, 20.87)	(8.19, 9.14)	(6.97, 7.15)
(S,S,L)	(8.00, 8.58)	(14.44, 14.76)	(16.55, 17.46)	(17.45, 17.60)	(16.07, 17.32)	(23.22, 23.35)
(L,S,S)	(17.73, 18.88)	(17.35, 17.65)	(10.14, 11.72)	(21.90, 22.45)	(11.29, 13.30)	(9.60, 10.03)
(L,L,S)	(4.64, 5.52)	(4.60, 4.63)	(10.45, 11.52)	[0.00, 3.83)	(11.23, 13.79)	(9.01, 9.51)
(L,L,L)	(3.17, 3.99)	[0.00, 6.74)	(9.09, 10.47)	(7.97, 8.29)	(10.54, 12.00)	(10.77, 11.01)
(L,S,L)	(10.61, 11.66)	(10.58, 10.72)	(9.89, 10.80)	(7.04, 7.21)	(16.13, 17.65)	(22.62, 26.11)
Risk premium from agency (in millions of 2000 US$)						
(S,S,S)	(0.020, 0.201)	(0.369, 0.416)	(0.042, 0.435)	(0.807, 0.909)	(0.044, 0.451)	(0.813, 0.916)
(S,L,S)	(0.033, 0.308)	(0.526, 0.586)	(0.092, 0.939)	(1.716, 1.930)	(0.113, 1.172)	(2.140, 2.425)
(S,L,L)	(0.025, 0.240)	(0.425, 0.476)	(0.029, 0.297)	(0.692, 0.781)	(0.026, 0.274)	(0.533, 0.604)
(S,S,L)	(0.007, 0.076)	(0.141, 0.159)	(0.024, 0.247)	(0.438, 0.493)	(0.021, 0.222)	(0.446, 0.505)
(L,S,S)	(0.046, 0.477)	(0.868, 0.981)	(0.048, 0.503)	(0.947, 1.070)	(0.056, 0.581)	(1.017, 1.149)
(L,L,S)	(0.028, 0.288)	(0.511, 0.577)	(0.062, 0.629)	(1.144, 1.288)	(0.113, 1.169)	(2.036, 2.300)
(L,L,L)	(0.023, 0.234)	(0.443, 0.500)	(0.056, 0.580)	(1.046, 1.182)	(0.075, 0.767)	(1.371, 1.528)
(L,S,L)	(0.037, 0.376)	(0.848, 0.960)	(0.023, 0.233)	(0.552, 0.622)	(0.035, 0.367)	(0.703, 0.795)

Note: Excerpt from Gayle and Miller, 2015, Table 8. Firm size and leverage is measured by the triplicate of (A, W, D), where A is assets, W is number of workers ,and D is debt-equity ratio with each corresponding to whether that element is above (L) or below (S) its industry average. The numbers of observations are 7796, 5600 and 8536 in the primary, consumer, and service sectors respectively. All calculations in this table were performed using the median bond price in the data. The risk-aversion parameters used are the interaction across sector, size and leverage of the most parsimonious unrejected specification—i.e. $\gamma \in (0.02, 0.21)$ for the pure moral hazard model and $\gamma \in (0.37, 0.42)$ for the hybrid moral hazard model.

To get a rough estimate of the annual dollar losses implied by τ_1 within each sector, they averaged the bounds of the confidence intervals across firm types within sectors and multiplied these averages by the corresponding average market values provided in Table 2.1. Under the PMH model, the estimated average annual losses to firms range from \$545 million to \$601 million in the primary sector, \$918 million to \$1.00 billion in the consumer sector, and \$1.46 billion to \$1.66 billion in the service sector. Similarly, under the HMH model, the estimated losses range from \$580 million to \$648 million in the primary sector, \$1.00 billion to \$1.07 billion in the consumer sector, and \$1.42 billion to \$1.49 billion in the service sector annually.

Certainty equivalent wages The compensating differential between working and shirking, τ_2, is the difference between the manager's certainty equivalent (and reservation wage) for working and for shirking. Table 8.3 presents the estimated identified sets of these reservation wages, $w^{(1)}$ for shirking and $w^{(2)}$ for working, under both the unrestricted PMH model and the restricted HMH model. In the unrestricted PMH model, reservation wage is specified for each accounting report r, whereas in the restricted HMH model, β and α remain constant across r. Consequently, the PMH model reports twice as many regions as the HMH model.

The top panel of Table 8.3 presents the estimated identified set of $w^{(1)}$ for both models. In the PMH model, $w^{(1)}$ is consistently higher in the good state compared to the bad state. For 18 out of 24 firm and sector types, the hybrid $w^{(1)}$ falls within the range of the PMH estimates for $w^{(1)}$; in the remaining six cases, the HMH $w^{(1)}$ is below the PMH region for $w^{(1)}$ in the bad state.[4] In the PMH model, the shirking wage $w^{(1)}$ is negative[5] for over half of the firm types in the bad state, but in the good state, the manager demands positive compensation to shirk in 22 out of 24 firm and sector types. Approximately half of the $w^{(1)}$ estimates in the HMH model are positive. Since the HMH estimates generally lie between the PMH $w^{(1)}$ for the two states, the HMH estimates are typically lower in magnitude.

The bottom panel of Table 8.3 presents the identified set of $w^{(2)}$, the reservation wage for working. In the PMH model, $w^{(2)}$ is negative for 9 of 24 firm types in the bad state—in these cases, the manager would be willing to pay for the opportunity to remain employed when the firm reports poor performance relative to its industry. While this result is striking, it is consistent with Table 2.6, which shows that the same nine firm-sector types also have negative average compensation. Since the difference between expected compensation and its certainty equivalent reflects the risk premium, the former must exceed the latter for risk-averse agents.

[4]Three of these six exceptions occur in the (S,S,S) firm type, spanning all three sectors.

[5]A negative $w^{(1)}$ or $w^{(2)}$ indicates that, conditional on his actions, the manager would pay shareholders for the privilege of holding the position.

Table 8.3
Confidence regions of certainty-equivalent wages

Firm type (A, W, D)	Primary			Consumer			Service		
	PMH		HMH	PMH		HMH	PMH		HMH
	Bad	Good	All	Bad	Good	All	Bad	Good	All
Certainty-equivlanet wage for shirking									
(S,S,S)	(-0.78, -0.78)	(0.67, 0.67)	(-1.20, -1.18)	(-3.67, -3.66)	(-1.99, -1.98)	(-4.62, -4.59)	(-1.13, -1.10)	(0.87, 0.88)	(-1.37, -1.36)
(S,L,S)	(1.21, 1.24)	(2.96, 2.96)	(0.49, 0.50)	(-6.05, -6.04)	(0.70, 0.81)	(-4.27, -4.18)	(-5.53, -5.43)	(3.82, 3.87)	(0.09, 0.12)
(S,L,L)	(-0.24, -0.22)	(2.17, 2.18)	(0.70, 0.72)	(-0.72, -0.71)	(6.88, 6.91)	(1.84, 1.84)	(0.68, 0.70)	(3.55, 3.56)	(0.40, 0.41)
(S,S,L)	(-0.42, -0.42)	(1.34, 1.34)	(0.54, 0.54)	(-1.40, -1.40)	(0.69, 0.70)	(-0.27, -0.26)	(-2.21, -2.21)	(0.76, 0.78)	(-1.34, -1.34)
(L,S,S)	(-3.84, -3.84)	(-2.68, -2.65)	(-4.46, -4.43)	(-3.00, -2.98)	(1.65, 1.71)	(-1.21, -1.20)	(-1.19, -1.16)	(0.66, 0.70)	(-0.61, -0.57)
(L,L,S)	(0.72, 0.74)	(2.96, 2.98)	(1.81, 1.83)	(-4.66, -4.65)	(1.38, 1.41)	(-0.78, -0.74)	(1.09, 1.20)	(3.90, 3.94)	(1.89, 2.05)
(L,L,L)	(1.24, 1.26)	(4.63, 4.65)	(2.60, 2.62)	(0.51, 0.54)	(3.21, 3.25)	(0.78, 0.79)	(1.45, 1.48)	(6.34, 6.41)	(3.42, 3.44)
(L,S,L)	(-5.60, -5.57)	(1.07, 1.08)	(-3.07, -3.01)	(-4.28, -4.28)	(3.50, 3.51)	(0.23, 0.26)	(-1.59, -1.57)	(2.44, 2.46)	(0.31, 0.32)
Certainty-equivlanet wage for working under perfect monitoring									
(S,S,S)	(0.35, 0.48)	(3.16, 3.43)	(1.05, 1.09)	(-2.19, -1.94)	(2.40, 2.99)	(-0.77, 0.67)	(0.42, 0.81)	(4.07, 4.49)	(1.41, 1.51)
(S,L,S)	(1.78, 2.17)	(4.32, 4.39)	(2.52, 2.58)	(-1.46, -0.83)	(5.37, 6.66)	(-0.36, 0.15)	(-1.47, -0.24)	(4.98, 5.44)	(-1.38, 1.09)
(S,L,L)	(0.16, 0.37)	(2.94, 3.14)	(1.30, 1.35)	(0.68, 0.85)	(8.52, 8.95)	(3.10, 3.19)	(1.72, 1.96)	(6.14, 6.36)	(2.77, 2.84)
(S,S,L)	(0.30, 0.35)	(1.96, 2.04)	(1.08, 1.10)	(0.60, 0.77)	(2.58, 2.85)	(1.36, 1.42)	(-1.02, -0.88)	(3.20, 3.46)	(8.37, 8.95)
(L,S,S)	(-1.22, -0.97)	(2.19, 2.81)	(-0.17, 0.06)	(-1.09, -0.75)	(3.76, 4.35)	(0.39, 0.51)	(0.91, 1.46)	(2.65, 3.10)	(0.98, 1.11)
(L,L,S)	(1.67, 1.93)	(3.88, 4.12)	(2.52, 2.59)	(-0.87, -0.30)	(3.54, 4.10)	(0.89, 1.03)	(4.25, 5.41)	(5.92, 6.44)	(3.47, 3.73)
(L,L,L)	(1.84, 2.06)	(5.15, 5.32)	(2.97, 3.03)	(2.44, 2.90)	(6.11, 6.71)	(3.15, 3.28)	(4.69, 5.27)	(8.85, 9.76)	(5.22, 5.40)
(L,S,L)	(-4.22, -3.75)	(3.46, 3.67)	(-0.89, 0.78)	(-2.55, -2.34)	(4.41, 4.62)	(0.95, 1.02)	(0.52, 0.83)	(4.92, 5.27)	(2.44, 2.54)

Note: Excerpt from Gayle and Miller, 2015, Table 9. A report is classified as "Good" if the firm's accounting is higher than the expected value of accounting return—the yearly sample average for a firm type and sector—and "Bad" otherwise. Firm size and leverage is measured by the triplicate of (A, W, D), where A is assets, W is number of workers ,and D is debt-equity ratio with each corresponding to whether that element is above (L) or below (S) its industry average. The numbers of observations are 7796, 5600 and 8536 in the primary, consumer, and service sectors respectively. The shirking wage is $w_t^{(1)} = [\gamma(b_t - 1)]^{-1} b_{t+1} \log(\beta)$ and the working wage under perfect monitoring is $w_t^{(2)} = [\gamma(b_t - 1)]^{-1} b_{t+1} \log(\alpha)$. All calculations in this table were performed using the median bond price in the data. The risk-aversion parameters used are the interaction across sector and firm types of the most parsimonious unrejected specification—i.e. $\gamma \in (0.02, 0.21)$ for the pure moral hazard model and $\gamma \in (0.37, 0.42)$ for the hybrid moral hazard model.

In the HMH model, although 4 out of 24 firm types have a negative lower bound, the confidence region always includes positive values. This implies that managers under the HMH model are consistently paid a positive certainty equivalent for working. In contrast, the parameter estimates of the PMH model suggest that managers receive a negative $w^{(2)}$ in the bad state, potentially reflecting the manager's willingness to pay for the privilege of holding the position when the firm does poorly. This finding provides evidence supporting the restricted HMH model over the unrestricted PMH model.

Willingness to pay to eliminate the agency cost (τ_5) The bottom panel of Table 8.2 presents estimates of the identified set of τ_5, the risk premium shareholders pay managers to mitigate agency issues. The results show that τ_5 is consistently higher in the HMH model compared to the PMH model for every firm type, often by several hundred thousand dollars. Despite these quantitative differences, the qualitative comparisons between firm types and industry sectors are largely aligned across the two models.

For both the HMH and PMH specifications, τ_5 is generally lower in the primary sector than in the consumer sector, with one single firm type in exception: (L,S,L). Similarly, τ_5 in the consumer sector is generally lower than in the service sector.

Controlling for assets and employment, firm types with a higher debt-equity ratio tend to have a lower τ_5 compared to their counterparts with a lower debt-equity ratio in both the PMH and HMH models. This suggests that managers face greater risk in their compensation, driven by moral hazard and hidden information in this framework, when stakeholder claims on firm assets are more concentrated among those most affected by firm performance.

In general, CEOs of firms with more employees are paid a higher τ_5, controlling for total assets and debt-equity ratio. This relationship holds in both the PMH and HMH models, with only two common exceptions in the primary sector: (L,L,L) versus (L,S,L) and (L,L,S) versus (L,S,S). The relationship between firm assets and τ_5 is somewhat weaker, but larger firm assets are generally linked to higher τ_5.

The estimation results illustrate the empirical content of the HMH and PMH models, as discussed in Section 6. The HMH model is never rejected, whereas the PMH model is rejected except for the unrestricted PMH where heterogeneity in risk-aversion parameters is allowed across managers based on firm types and their accounting reports. A particularly surprising finding is that, unlike the HMH model, the unrejected PMH model suggests managers would be willing to pay shareholders for the privilege of holding their positions in bad accounting states, with their risk preferences varying based on the state of the accounting report. These unexpected results likely indicate model misspecification. Future research could address this within the PMH framework by explicitly modeling selection in the executive labor market when managers' risk attitudes are heterogeneous. Alternatively, the HMH model provides plausible estimates with a parsimonious model, supporting the widely held view that managers are capable of manipulating accounting reports and that shareholders deter such behavior through internalizing the incentives for managers to report truthfully in compensation schemes.

3 Impacts of the Sarbanes–Oxley Act

This Section examines how the Sarbanes–Oxley Act of 2002 (SOX) affected executive compensation and corporate governance by analyzing changes in the structure and level of executive pay before and after its implementation.

The passage of SOX serves as a pivotal case for understanding how regulation can shape executive compensation, particularly by imposing stringent penalties for financial misreporting. Using the panel data constructed from S&P 1500 firms, Gayle *et al.*, 2022, estimate the impact of SOX on CEO compensation by applying the PMH and HMH models before and after SOX. Their analysis highlights a methodological framework for integrating a Difference-in-Difference (DID) research design with structural econometrics. They find that SOX: (1) reduced the conflicts of interest between shareholders and CEOs, mainly by lowering the gross loss to shareholders if the CEO shirks; (2) increased the cost of agency, primarily through the risk premium paid to CEOs to align their incentives with the shareholders' objective.

The term SOX is often used to refer collectively to both the Sarbanes–Oxley (SOX) Act in 2002 (which applies to publicly listed companies in the US) and contemporary listing rules changed by the NYSE and NASDAQ.[6] The catalyst for SOX was a failure in corporate governance that led to the dismissal of executives and, in some cases, subsequent prosecution for fraud, conviction, and imprisonment. These executives violated legal constraints that were subject to audit. Enacting SOX brought greater accountability to financial statements, more rigorous enforcement of property rights in governance, and higher penalties for fraud, factors that discourage managers from breaking the law.

Although SOX primarily targeted illegal managerial actions, it also influenced the incentives of law-abiding managers: to act in the firm's interest or their own interest; to report unverifiable

[6]This convention for the term of SOX follows from Zhang, 2007, Carter *et al.*, 2009, Hart, 2009, Hochberg *et al.*, 2009, Linck *et al.*, 2009, Bargeron *et al.*, 2010, and Cohen *et al.*, 2013, among others.

financial information truthfully, or not. Using the PMH and HMH framework, Gayle *et al.*, 2022, quantify, before and after the implementation of SOX: (1) how much the goals of a CEO diverge from those of the shareholders measured by τ_1 and τ_2 as in (3.18) and (3.19), and (2) the costs shareholders incur to incentivize their CEOs, measured by τ_4 and τ_5 as in (3.21) and (3.22). These measures are further aggregated across the good and bad accounting reports based on the probability of those states.

Their research design demonstrates how difference-in-difference can be adapted to a structural econometric framework. To separate changes due to the implementation of SOX from other aggregate factors, they follow the literature to compare a designated control group to a treatment group to gauge the impact of SOX. Previous reduced-form studies use firms that were compliant with the SOX provisions of board structure before SOX as the control group in their difference-in-differences (DID) design and firms that did not comply with the measure until after enactment as the treatment group. [7] If SOX affects the primitives in the contracting environment of CEO compensation mainly

[7] Previous literature examine the impact of SOX on corporate governance (Linck *et al.*, 2009), CEO compensation (Chhaochharia and Grinstein, 2009, Cohen *et al.*, 2013, Guthrie *et al.*, 2012), investment (Cohen *et al.*, 2013, Banerjee *et al.*, 2015, Lu and Wang, 2015), firm performance (Duchin *et al.*, 2010, Bargeron *et al.*, 2010), and earnings management (Chen *et al.*, 2015, Joo and Chamberlain, 2017).

through changing board structure, Gayle *et al.*, 2022, conjecture that noncompliant firms experience more changes in their model estimates after SOX than the compliant firms.

The main sample covers a pre-SOX era during 1993–2001 and a post-SOX era including 2004 and 2005. Their empirical analysis omits data on the two years 2002 and 2003, when legislation was in flux. They define noncompliant (treated) firms as those that missed at least one of the three following criteria: (1) an entirely independent compensation committee before July 25, 2002, when SOX was approved, (2) an independent majority board before February 13, 2002, when the SEC asked NYSE and NASDAQ to review their corporate governance requirements, and (3) an entirely independent audit committee before December 31, 2000[8]. The remaining firms for which they have data on these criteria are denoted as compliant firms and used as the control group. Because the control group complied with SOX before it was implemented, at least on these three criteria, they conjecture the response to SOX would be more pronounced within the treatment group. To ensure a sufficient number of observations by firm type, they classify the DID sample only by sector and firm size.

Table 8.4 summarizes the structural-DID results. The es-

[8]This last criterion follows Duchin *et al.*, 2010, which suggest that the exchanges adopted the recommendations of the independence of audit committee as early as the end of 1999.

timated confidence intervals are presented for τ_1, τ_2, τ_4 and τ_5, and the effect of SOX on them based on DID. The DID estimates are calculated as the difference between the changes in the noncompliant (NC) firms and those in the compliant (C) firms. For example, the impact of SOX on τ_1, the gross loss of shareholders from the manager shirking, is shown below, and similarly for the other measures

$$\tau_1^{\text{DID}} \equiv \Delta\tau_1^{\text{NC}} - \Delta\tau_1^{\text{C}} \equiv \left(\tau_1^{\text{NC, post}} - \tau_1^{\text{NC, pre}}\right) - \left(\tau_1^{\text{C, post}} - \tau_1^{\text{C, pre}}\right). \quad (8.19)$$

Panel A of Table 8.4 shows that SOX significantly reduced the potential losses shareholders would face if the CEO shirks; this reduction was particularly evident in noncompliant firms. Panel B shows that the benefits that CEOs from deviating from the firm's objectives became less heterogeneous across firm type: by introducing uniform regulations, SOX curtailed opportunities for malfeasance and reduced reliance on firm-specific internal penalties.

In equilibrium, shareholders resolve the conflicts of interest of the CEO at a cost. These costs include two main components: the certainty-equivalent wage for working under perfect monitoring—the fixed wage that would induce the CEO to work if the shareholders could perfectly monitor his actions, and a risk premium, compensating the CEO for bearing risk in his compensation from incentivizing him to act in the shareholders' interest. Panel C shows that, following the implementation of SOX, certainty-equivalent wages for working under perfect monitoring increased in the primary sector for almost all types of firms, although the effects varied across other sectors depending on the type of firm. However, agency costs increased universally. Firms became more dependent on incentive-based compensation, and evidence suggests that the stricter regulation under SOX made it more expensive to effectively align CEO incentives with shareholder interests.

Table 8.4
Impact of SOX on Agency Costs

	Size	Compliant (C)		Noncompliant (NC)		NC versus C
Panel A: Gross Loss to Shareholders (%)		$\tau_1^{\text{C, pre}}$	$\Delta\tau_1^{\text{C}}$	$\tau_1^{\text{NC, pre}}$	$\Delta\tau_1^{\text{NC}}$	τ_1^{DID}
Primary	S	$(9.62, 9.65)$	$(-3.18, -3.17)$	$(14.14, 14.19)$	$(-4.45, -4.04)$	$(-1.27, -0.86)$
	L	$(4.84, 4.89)$	$(-0.60, -0.60)$	$(7.45, 7.51)$	$(-1.40, -1.38)$	$(-0.80, -0.78)$
Consumer	S	$(14.48, 14.56)$	$(-8.18, -8.11)$	$(16.44, 16.45)$	$(-8.46, -8.38)$	$(-0.35, -0.21)$
	L	$(5.91, 6.01)$	$(1.73, 1.89)$	$(9.61, 10.20)$	$(-4.86, -4.39)$	$(-6.74, -6.12)$
Service	S	$(20.94, 20.98)$	$(-9.13, -9.07)$	$(18.07, 18.24)$	$(-6.08, -6.01)$	$(3.05, 3.06)$
	L	$(10.96, 11.05)$	$(-5.93, -5.83)$	$(12.60, 12.75)$	$(-9.83, -9.78)$	$(-4.01, -3.85)$
Panel B: Benefit to CEO ($thousands)		$\tau_2^{\text{C, pre}}$	$\Delta\tau_2^{\text{C}}$	$\tau_2^{\text{NC, pre}}$	$\Delta\tau_2^{\text{NC}}$	τ_2^{DID}
Primary	S	$(1610, 1699)$	$(668, 691)$	$(3281, 3542)$	$(-496, -382)$	$(-1187, -1049)$
	L	$(780, 830)$	$(1335, 1456)$	$(2541, 2719)$	$(1024, 1069)$	$(-387, -311)$
Consumer	S	$(4403, 4795)$	$(-668, -586)$	$(6556, 7224)$	$(-3090, -2644)$	$(-2504, -1977)$
	L	$(2473, 2843)$	$(911, 1501)$	$(5831, 6715)$	$(-937, -745)$	$(-2246, -1848)$
Service	S	$(5013, 5522)$	$(-1102, -918)$	$(2824, 3188)$	$(2213, 2602)$	$(3131, 3703)$
	L	$(6988, 7673)$	$(-4640, -4465)$	$(5887, 6857)$	$(-3104, -3070)$	$(1370, 1570)$
Panel C: Certainty-Equivalent Wage for Working ($thousands)		$\tau_4^{\text{C, pre}}$	$\Delta\tau_4^{\text{C}}$	$\tau_4^{\text{NC, pre}}$	$\Delta\tau_4^{\text{NC}}$	τ_4^{DID}
Primary	S	$(1393, 1456)$	$(2988, 3060)$	$(1917, 2095)$	$(1358, 1417)$	$(-1702, -1572)$
	L	$(3614, 3674)$	$(3424, 3526)$	$(4423, 4566)$	$(3686, 3777)$	$(251, 263)$
Consumer	S	$(977, 1232)$	$(410, 480)$	$(668, 1117)$	$(-1279, -1063)$	$(-1760, -1473)$
	L	$(5160, 5607)$	$(277, 720)$	$(6285, 6999)$	$(-1158, -1071)$	$(-1792, -1435)$
Service	S	$(3481, 3882)$	$(-2458, -2305)$	$(2650, 2959)$	$(1074, 1298)$	$(3379, 3755)$
	L	$(9732, 10212)$	$(-1667, -1561)$	$(9335, 10058)$	$(-2862, -2753)$	$(-1202, -1192)$
Panel D: Cost of Moral Hazard ($thousands)		$\tau_5^{\text{C, pre}}$	$\Delta\tau_5^{\text{C}}$	$\tau_5^{\text{NC, pre}}$	$\Delta\tau_5^{\text{NC}}$	τ_5^{DID}
Primary	S	$(32, 95)$	$(37, 108)$	$(93, 270)$	$(-90, -32)$	$(-199, -69)$
	L	$(31, 91)$	$(53, 155)$	$(73, 216)$	$(45, 136)$	$(-19, -8)$
Consumer	S	$(133, 388)$	$(45, 115)$	$(237, 687)$	$(-333, -117)$	$(-448, -161)$
	L	$(233, 681)$	$(240, 683)$	$(374, 1088)$	$(58, 144)$	$(-539, -182)$
Service	S	$(209, 610)$	$(-232, -80)$	$(161, 470)$	$(126, 349)$	$(206, 582)$
	L	$(248, 728)$	$(47, 153)$	$(380, 1104)$	$(75, 185)$	$(28, 38)$

Note: Excerpt from Gayle *et al.*, 2022, Table 3, 4, 5, 6. They use the intersection of the identified sets of the risk aversion parameter across the pre- and post-SOX periods to calculate the confidence intervals for the welfare measures. This table reports the estimates for the treatment group and control group in the DID analysis. The non-compliant firms are the treatment group, including firms who missed at least one of the following criteria before SOX: (1) a majority independent board, (2) an entirely independent audit committee, and (3) an entirely independent compensation committee. The rest of their sample that contains those indicators is denoted as compliant firms and used as the control group.

$$\tau_1 \equiv \sum_{s=1}^{2} \varphi_s E_s \left\{x\left[1 - g_s(x)\right]\right\}$$

$$\tau_2 \equiv b_{t+1}\left[\left(b_t - 1\right)\gamma\right]^{-1} \ln\left(\alpha/\beta\right)$$

$$\tau_4 \equiv \gamma^{-1}\frac{b_{t+1}}{b_t - 1}\ln\alpha$$

$$\tau_5 \equiv \sum_{s=1}^{2} \varphi_s, E_s, \left[w_s, (x)\right] - \tau_4$$

9

Human Capital and Promotion

A well-established stylized fact in labor economics—pay increases with firm size (Oi and Idson, 1999)—holds true in the executive labor market as well. Human capital plays a fundamental role in explaining the firm-size pay premium for executives. As Section 2 illustrates, executives in large firms are older, more educated, but have less executive experience and less tenure than those in smaller firms; Work experience as executives in more firms increases executive compensation at higher ranks in the hierarchy. Moral hazard is also at play: Top executives are paid a significant portion of their total compensation in stock and options; The composition of firm-denominated securities varies substantially across ranks and executives at different points in their lifecycle.

This Section presents a generalized Roy's 1951 job-sorting model with human capital accumulation and moral hazard in a dynamic setting that was developed in Gayle *et al.*, 2015. The model allows compensation to depend on experience from jobs in the past, which creates a trade-off in job choice between firm-specific tenure and this form of general human capital. In addition, nonpecuniary utility varies by firm–rank and effort choice, which enables the model to account for different levels of moral hazard between large and small firms and among ranks and industries. In the basic framework below, human capital is modeled extensively, but independently of effort. Subsection 2 relaxes this to allow human capital accumulation to depend on effort, reflecting executives' career concerns as another motivation for effort despite moral hazard.

1 The basic model

We begin with the dynamic PMH model from Section 3, and adapt the framework to incorporate heterogeneity of jobs by firm and rank. There are J firms indexed by $j \in \{1, \dots, J\}$. Each firm holds K positions, indexed by $k \in \{1, \dots, K\}$ and ranked in hierarchical order.

Preferences and Choices Apart from CARA utility, executives are characterized by age t and a vector of human capital h_t, which captures demographics and indexes work experience. They retire upon reaching or before age $T < \infty$, and stay retired thereafter. At the beginning of period t, an executive chooses her consumption c_t, and for any $t \leq T$, makes her employment or retirement choices. Let $d_{jkt} \in \{0, 1\}$ indicate the executive's choice of rank k in firm j at age t, and let d_{0t} indicate retirement. She then chooses her effort level, working or shirking, which is unobserved to firms. Let l_t be the indicator for working. With discount factor δ, the lifetime utility of an executive with human capital $\{h_t\}_{t=0}^{\infty}$ is given by

$$- \sum_{t=1}^{\infty} \delta^t \exp(-\gamma c_t) \Bigg[d_{0t} \exp(-\varepsilon_{0t}) + \sum_{j=1}^{J} \sum_{k=1}^{K} d_{jkt} [\alpha_{jkt}(h_t) l_t + \beta_{jkt}(h_t)(1 - l_t)] \exp(-\varepsilon_{jkt}) \Bigg], \qquad (9.1)$$

where $\alpha_{jkt}(h_t)$ and $\beta_{jkt}(h_t)$ are the nonpecuniary utility costs from working and shirking, which depend on the firm, rank, age and experience. We assume that $\alpha_{jkt}(h) > \beta_{jkt}(h)$ for any h, so managers prefer shirking to working, creating a conflict of interest with shareholders maximizing profit.

The taste shocks $\varepsilon_t \equiv (\varepsilon_{0t}, \varepsilon_{11t}, \ldots, \varepsilon_{JKt})$ are firm- and rank-specific, and independent of effort and idiosyncratic across executives. ε_{0t} is the shock from choosing retirement, and ε_{jkt} is the taste shock from working in firm j at rank k. The taste shocks are modeled to discern the unobserved factors that affect the executive choices to work, shirk, or retire. The taste shocks are independently and identically distributed, for example, according to a Type-1 Extreme Value distribution (T1EV).[1]

Human Capital Human capital is modeled extensively, incorporating skills that depend on education and work experience. We define a vector of time-invariant attributes and skills, h_1, which includes gender and education dummies. We further introduce a vector, $h_{2t} = (h_{211t}, \ldots, h_{2JKt})$, to capture the individual's history of rank-firm choices, including retirement. We denote a transition function of h_{2t} by $\overline{H}_{jk}(\cdot)$, depending on firm and rank jk, to capture the evolution of human capital. Assume the function is deterministic, human capital follows the law of motion

$$h_{2t+1} = \sum\nolimits_{j=1}^{J} \sum\nolimits_{k=1}^{K} d_{jkt} \overline{H}_{jk}(h_{2t}). \tag{9.2}$$

This specification of human-capital accumulation captures where human capital is acquired (firm and rank) and allows the transition of human capital to depend on firm and rank.

To illustrate how human capital accumulation works in our model, suppose that h_{2t} is a three-dimensional vector, $h_{2t} \equiv \left(h_{2t}^{(1)}, h_{2t}^{(2)}, h_{2t}^{(3)}\right)$, and that each firm is indexed by two

[1]Using the T1EV distribution facilitates formulating the model equilibrium in terms of conditional choice probability following Hotz and Miller, 1993, which greatly simplifies the computation in discrete choice models (such as work, shirk, or retire in this context).

dimensions. The first, $j_1 \in \{0,1\}$, denotes whether the executive is newly hired by this firm, $j_1 = 0$, or the executive worked for this firm last period, $j_1 = 1$. The second, $j_2 \in \{1,2,\ldots,J_2\}$, denotes firm size and industrial sector; hence, $j = j_1 \otimes j_2 \in \{0,1\} \otimes \{1,2,\ldots,J_2\}$. Let $h_{2t}^{(1)}$ measure the tenure of the executive in the current firm, capturing firm-specific human capital. Let $h_{2t}^{(2)}$ measure the number of years of executive experience, and let $h_{2t}^{(3)}$ measure the number of different firms the executive has worked in since becoming an executive. The last two, $h_{2t}^{(2)}$ and $h_{2t}^{(3)}$ intend to capture the years of general human capital. The second, $h_{2t}^{(2)}$, is standard in the learning-by-doing human-capital-accumulation literature; however, the third, $h_{2t}^{(3)}$, is meant to capture the idea that management may require many different skills, and the greater the number of firms an executive worked in, the better she may be when she becomes a top executive. We then specify the transition function of human capital as

$$\overline{H}_{jk}(h_{2t}) = h_{2t} + \Delta_{jkt}, \tag{9.3}$$

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

The additional element we add to the standard model of human capital accumulation, captured by $h_{2t}^{(3)}$ in the illustration, implies that younger executives may change firms more often than otherwise to gain this dimension of human capital and acquire skills from working in different organizations. This element appeals to the prediction of the experimentation literature on human capital (Miller, 1984, Antonovics and Golan, 2012, Sanders, 2013) except that the experimentation literature requires learning about skills unknown to executives, whereas we get the same prediction if the upper level of the hierarchy values a skill set that can be acquired through managerial experience in multiple firms.

Firm Technology Each firm employs multiple executives. We decompose each firm's output (revenue) into the sum of the productivity of multiple executives, aggregate technological shocks, and firm-idiosyncratic excess returns. With slight abuse of notation, let τ denote the calendar time. Denote by $t(\tau, n)$ the age of the n^{th} executive and her human capital at τ by $h_{t(\tau,n)}$. Let $F_{jk}(h_{t(\tau,n)})$ denote the executive's contribution to the j^{th} firm's revenue in τ if she chooses the k^{th} job with that firm by setting $d_{jkt(\tau,n)} = 1$. Let $\pi_{\tau+1}$ denote a return from an exogenous aggregate productivity shock that affects every firm, and $\pi_{j,\tau+1}$ denote the (net) excess return to the j^{th} firm. Let $\mathcal{E}_{j\tau}$ denote the value of firm j at the beginning of calendar time τ. Finally, denote by $w^{(n)}_{jk,\tau+1}$ the firm's compensation to executive n if she was employed at rank k in period τ. We assume the revenue of firm j at τ is decomposed into three components

(i) Executive Contributions: $\sum_{n=1}^{N}\sum_{k=1}^{K} d_{jkt(\tau,n)}F_{jk}(h_{t(\tau,n)})$ captures the productivity (additively separable) of all of firm j's executives depending on their respective human capital, independent of effort.

(ii) Aggregate Productivity: $\mathcal{E}_{j\tau}(\pi_{\tau+1}-1)$ represents the firm's revenue attributable to changes in aggregate productivity.

(iii) Firm-Level Excess Returns: $\mathcal{E}_{j\tau}\pi_{j,\tau+1}$ reflects the value added through excess returns, which depends on the efforts of all the executives.

Assuming all dividends are paid only when the firm is liquidated, the equity of the firm evolves according to the law of motion[2]

$$\mathcal{E}_{t+1} \equiv \sum_{i=1}^{N}\sum_{k=1}^{K} d_{jk}(t+1)[F_{jk}(h_{it+1})-w_{jk,t+1}^{(n)}]+\mathcal{E}_{jk}(\pi_{t+1}+\pi_{jt+1}). \quad (9.4)$$

$\pi_{j,\tau+1}$, the excess return, is the rate of increase in the value of the firm net of the executives' contributions, and in excess of the return of the aggregate market.

Span of Control The efforts of executives only affect the firm through $\pi_{j,\tau+1}$, which is determined stochastically. If all the executives in the firm work diligently, the value of $\pi_{j,\tau+1}$ is drawn from a distribution with probability density function $f_j(\pi)$. If everyone except the k^{th} ranked executive works, conditional on any level of human capital h, the value of $\pi_{j,\tau+1}$ is drawn from the distribution $f_j(\pi)g_{jk}(\pi \mid h)$. We analyze an equilibrium where everyone works, since all the executives in the data set receive incentive pay.[3] In equilibrium, the distribution when an executive is not observed empirically but estimated based on model assumptions. We assume that the regularity conditions (3.1), (3.2), (3.3) hold for all firms and executives.

---

[2]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.

[3]As discussed in Section 3, based on the one-shot deviation principle in game theory (Fudenberg and Tirole, 1991), we only need to consider individual deviations where one executive shirks while every other executive within the firm works. Therefore, we can disregard the case where more than one executive shirks within a firm.

Executives contribute to firm productivity through two separable components, human capital and effort. The distortion in the firm productivity distribution caused by an executive who shirks, $g_{jk}(\pi \mid h)$, depends on firm and executive characteristics. Thus, this model captures the variations in the agency problem across different firms and executives. However, the likelihood ratio $g_{jk}(\pi \mid h)$ does not depend on the number of executives in their firm or other managers' human capital; relaxing this assumption would endogenize the optimal number of executives and the configuration of human capital within the team, a challenge for future research. Since $g_{jk}(\pi \mid h)$ measures the extent to which executive effort affects firm returns, it can be interpreted as a measure of shareholders' span of control.

The estimates of $g_{jk}(\pi \mid h)$ provide some insight regarding the role of rank in the firm. For example, we can test whether our measure of span of control declines with rank, which would be consistent with Williamson, 1967.[4] Although the assumption of constant returns to scale precludes us from making predictions about the size distribution of firms, we can, however, test whether the span of control increases with firm size. If so, then

[4]In Williamson's 1967 hierarchical model of firms, there is 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 managers' shirking causes larger distortions in higher ranks. In contrast, Mirrlees, 1999, offers an alternative view of a firm as a decentralized contractual organization.

using the utility-function estimates of the costs of shirking, we can calculate whether the costs of agency increase in firm size. This might provide one justification to a diminishing returns to scale in firm size as postulated in Lucas, 1978.

Information Environment and Timeline Each executive privately observes their own taste shocks, effort level, and outside wealth. Similarly, consumption choices made by executives remain private. All other information is public. The market publicly observes each executive's human capital, rank, firm assignments, and compensation from the previous period's employment. Although the productivity of each executive $F_{jk}(h_{t(\tau,n)})$ cannot be separately observed, it is considered public knowledge based on the executives' human capital, and independent of executive effort.

Specifically, at the beginning of each period τ, the market observes

$$\left(h_{t(\tau,n)}, d_{t(\tau,n)}\right) \quad \text{and} \quad \sum_{j=1}^{J}\sum_{k=1}^{K} d_{jk,t(\tau-1,n)} w_{jk\tau}^{(n)}$$

for all N executives, as well as the aggregate market return π_τ, the initial equity $\mathcal{E}_{j\tau}$, and the excess returns $\pi_{j\tau}$ of all J firms. In addition, each executive privately observes their own outside wealth $\xi_\tau^{(n)}$, idiosyncratic taste shocks $\varepsilon_\tau^{(n)}$, and recalls their own effort history $\{l_s^{(n)}\}_{s=0}^{t-1}$.

At the beginning of each period, executives receive compensation from the previous period of employment based on their contracts. After observing her own taste shock vector, each executive privately chooses her consumption and asset portfolio allocation. Simultaneously, she decides whether to retire; and if she decides not to retire, which firm to be employed in, and the desired rank and effort level. She approaches the firm to negotiate with the shareholders. We assume the executive makes an ultimatum offer that the shareholders can only accept or reject. If no agreement is reached, the executive does not work during that period, and there is no additional hiring by the firm.

Capital Markets Following Margiotta and Miller, 2000, we assume that executives have sufficient access to financial markets to smooth their outside wealth. In this model, this assumption implies the existence of a complete contingent-claims market for consumption, such that executives can insure themselves against all publicly disclosed events. The price measure is denoted as Λ_τ and its derivative as λ_τ at date τ. For each $\tau \in \{0, 1, 2, \ldots\}$, Λ_τ represents the price at time 0 of claims contingent on consumption delivered at date τ. For example, $E[\lambda_\tau]$ indicates the number of consumption units forgone at time 0 to secure a guaranteed unit of consumption at time τ. The τ-period interest rate can then be expressed as $\{E[\lambda_\tau]\}^{-1} - 1$.

We measure $w_{jkt(\tau)+1}$, the executive's compensation for holding position k at a firm of type j at the beginning of age $t+1$, in terms of current consumption units. Since an executive's wealth is determined endogenously through their compensation, it cannot be fully insured if it is tied to the firm's returns $\pi_{j,\tau+1}$. Value-maximizing banks naturally avoid providing insurance to executives against fluctuations in their own firm's excess returns, as such insurance could disincentivize executives from working, leading to expected losses for the bank. Additionally, public disclosure laws require top executives to report their financial holdings in securities issued by their own firm. With these disclosures, banks can easily safeguard against potential insider trading, ensuring they are not exposed to this type of risk. Therefore, while there is a complete market against public events, the market cannot insure executives from the firm-related risk in their compensation. This market incompleteness arises due to executives' private information about their own effort, i.e. moral hazard.

We will solve for the equilibrium contract in three steps. First, we derive an individual executive's conditional choice probabilities for retiring or where to work (firm and rank), taking the compensation schedule as given (depending on human capital and age, and firm and rank). We consider an executive who works each period until retiring. Second, we find an individual firm's cost-minimizing contract to elicit work from a given executive, where incentive compatilibility and participation constraints are established. Finally, we find the equilibrium contract such that, in aggregate, the demand for executives equals the supply.

Intertemporal Consumption and Employment Choices We first derive the value function for an executive who works diligently each period before retiring. The separability of preferences, the executives' absolute risk aversion, and the completeness of the capital market allow us to focus on individual executive's indirect utility function. This function expresses the executive's expected lifetime utility in two multiplicative parts, the utility from post-retirement, and an index of human capital that reflects the utility value associated with human capital.

First, we derive the indirect utility function for an executive retiring at age t. Upon retirement, the executive faces a single budget constraint with wealth ξ_t. Second, we recursively solve the problem backwards for the same executive at age $t-1$ using Bellman's 1957 principle, where the value function at age t is the indirect utility function derived in the first step. We define a few notations below to express the lifetime utility. Details of the derivation are provided in Gayle *et al.*, 2015.

The consumption choice problem solved in the first step is standard in the asset-pricing literature (Debreu, 1959, Chap

7). After retirement, executives face a complete contingent claims market to smooth their consumption. As Rubinstein, 1981, shows, the CARA assumption implies very few securities are required to characterize the optimal financial portfolio. Following Margiotta and Miller, 2000, let b_τ denote the price of a perpetual bond that, contingent on the history through date τ, pays a unit of consumption from period τ onward in perpetuity, in period τ prices

$$b_\tau \equiv E_\tau \left[\sum\nolimits_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau} \right]. \tag{9.5}$$

Similarly, let a_τ denote the price of a risky security with a random payoff of $(\ln \lambda_s - s \ln \delta)$ units of consumption in each period s from period τ to perpetuity, in period-t prices

$$a_\tau \equiv E_\tau \left[\sum\nolimits_{s=\tau}^{\infty} \frac{\lambda_s}{\lambda_\tau} (\ln \lambda_s - s \ln \delta) \right]. \tag{9.6}$$

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

$$v_{jkt(\tau)+1} \equiv \exp\left(-\gamma w_{jkt(\tau)+1}(h, \pi_t)/b_{\tau+1}\right). \tag{9.7}$$

Let $p_{jkt}(h)$ denote the probability of choosing (j, k) at age t conditional on h. Similarly, we denote the retirement probability by $p_{0t}(h)$. We define an index of human capital for a t-year-old executive with characteristics h who always works as

$$A_t(h) = p_{0t}(h)E\left[\exp\left(-\frac{\varepsilon^*_{0t}}{b_\tau}\right)\right] + \sum_{j=1}^{J}\sum_{k=1}^{K} p_{jkt}(h)\alpha_{jkt}(h)^{\frac{1}{b_\tau}} E\left[\exp\left(-\frac{\varepsilon^*_{jkt}}{b_\tau}\right)\right]\left\{A_{t+1}\left[\overline{H}_{jk}(h)\right] E_t[\upsilon_{jkt(\tau)+1}]\right\}^{1-\frac{1}{b_\tau}}. \quad (9.8)$$

The index $A_t(h)$ is a choice-probability-weighted average of expected outcomes from making different (j, k) choices, including retirement. By inspection, the index is strictly positive, and lower values of $A_t(h)$ are associated with higher values of human capital. Thus, increasing expected compensation reduces $E_t[\upsilon_{jkt(\tau)+1}]$ and $A_t(h)$. Similarly, $A_t(h)$ is monotonically increasing in $\alpha_{jkt}(h)$, the nonpecuniary losses of the executive, weighted across different (j, k) choices. The lifetime utility from age $t < R$ is higher (less negative) if this index of human capital is lower, as shown in Lemma 9.1.

Lemma 9.1 (Lemma 4.1 of Gayle *et al.*, 2015)**.** Let $V_t\left(h, \xi_t, a_\tau, b_\tau\right)$ denote the discounted sum of expected utility from age $t < R$ onward. For an executive with characteristics h and wealth ξ_t who has not yet observed ε_t and will make optimal consumption and job-match choices thereafter, subject to never shirking,

$$V_t\left(h, \xi_t, a_\tau, b_\tau\right) = -\lambda_\tau b_\tau \exp\left(-\frac{a_\tau + \gamma\xi_t}{b_\tau}\right) A_t(h). \quad (9.9)$$

The term $-\lambda_\tau b_\tau \exp\left[-\left(a_t + \gamma\xi_t\right)/b_\tau\right]$ is the value function for a retiree defined above. Thus, (9.9) 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 9.1 in the lifetime utility formulation, (9.1), to solve for their employment choice. This is summarized in Theorem 9.2.

Theorem 9.2 (Theorem 4.2 of Gayle *et al.*, 2015)**.** If $t \leq R$ and $l_s = 1$ for all $s \in \{t, \ldots, R\}$, then job choices d_t (including retirement) are picked to sequentially maximize

$$\begin{aligned} d_{0t}\varepsilon_{0t} + \sum_{j=1}^{J}\sum_{k=1}^{K} d_{jkt}\Big\{&\varepsilon_{jkt} - \ln\alpha_{jkt}(h) \\ &- (b_\tau - 1)\ln A_{t+1}\left[\overline{H}_{jk}(h)\right] - (b_\tau - 1)\ln E_t[\upsilon_{jkt(\tau)+1}]\Big\}. \end{aligned} \tag{9.10}$$

The above formulation builds on several models of labor market sorting, augmented with the value of human capital. First, it is a generalized Roy model: Based on human capital, executives sort into jobs and firms. The generalized form includes a Roy model and a compensating-differentials model: As in Roy models, the current compensation is given as $\upsilon_{jkt(\tau)+1}$, and as in hedonic-price models the nonpecuniary benefits from the job are given by $\alpha_{jkt}(h)$ and ε_{jkt}. The generalized dynamic Roy model is augmented by an additional component, $A_{t+1}[\overline{H}_{jk}(h)]$, the future expected utility attached to a job. This index of utility includes pecuniary compensation, expected earnings growth, as well as expected nonpecuniary utility from future jobs and ranks, and can be interpreted as the value of human capital acquired in different ranks and jobs. Thus, working and acquiring experience in some firms and ranks is associated with differential pay, nonpecuniary benefits, expected earnings growth, or promotion probability over the executive's career.

With the representation of the lifetime utility in (9.10), we now characterize the firm- and rank-choice probabilities and how they change over the lifecycle in an equilibrium in which all executives work diligently. These choice probabilities will map the model's parameters and the observed choice probabilities in the data, and therefore play an important role in the estimation strategies. The vector of conditional-choice probability functions, $p_t(h) \equiv (p_{11t}(h), \ldots, p_{JKt}(h))$, that are used to compute $A_t(h)$ in (9.8) are precisely the probability functions that characterize an executive's choices when solving the optimization function described by (9.10). Appealing to Proposition 1 of Hotz and Miller, 1993, a mapping exists, $q(p) \equiv (q_{11}\left[p_t(h)\right], \ldots, q_{JK}\left[p_t(h)\right])$, between the conditional choice probabilities and the ex-ante value from (9.10)

$$q_{jk}\left[p_t(h)\right] = \ln \alpha_{jkt}(h) + (b_\tau - 1) \ln A_{t+1}\left[\overline{H}_{jk}(h)\right] + (b_\tau - 1) \ln E_t[\upsilon_{jkt(\tau)+1}]. \tag{9.11}$$

(9.11) characterizes the executive supply by relating the choice probabilities of the different rank–firm combinations and retirement to the utility from the compensation $(E_t\left[\upsilon_{jkt(\tau)+1}|h\right])$. While no shirking occurs in equilibrium, every job history is possible providing a sufficiently high realization of taste shock ε_{jkt}. Underlying this result is our assumption that ε_{jkt} has full support and is privately known to only the executive.

For an individual executive, given h, the solution to her optimization problem in (9.10) rests on the relative differences between the realization of taste shocks $(\varepsilon_{11t}-\varepsilon_{0t},\ldots,\varepsilon_{JKt}-\varepsilon_{0t})$ rather than their levels, ε_t.[5] Substituting (9.11) into (9.10), we see that if position (j,k) is the optimal employment choice, then $\varepsilon_{jkt}-\varepsilon_{0t}>q_{jk}\left[p_t(h)\right]$ and

$$(j,k)=\underset{(j',k')}{arg\,max}\left\{\varepsilon_{j'k't}-q_{j'k'}\left[p_t(h)\right]\right\}. \tag{9.12}$$

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

$$(\varepsilon_{11t}-\varepsilon_{0t},\ldots,\varepsilon_{JKt}-\varepsilon_{0t})\equiv q\left[p_t(h)\right]\equiv(q_{11t},\ldots,q_{JKt}). \tag{9.13}$$

It follows that, for any level of taste shock for retirement ε_{0t}, the vector of idiosyncratic shocks $(\varepsilon_{0t},q_{11t}+\varepsilon_{0t},\ldots,q_{JKt}+\varepsilon_{0t})=\varepsilon_t$ makes an executive indifferent between accepting any of the JK positions or retiring.

[5] This becomes apparent from substituting out $d_{0t}=1-\sum_{j=1}^{J}\sum_{k=1}^{K}d_{jkt}$ in (9.10), collecting terms involving d_{jkt}, and noting that the additive constant, ε_{0t}, has no effect on the optimal choices.

Example We show the form (9.11) takes if the distribution of ε_t is Type I extreme value and the transition function of human capital is as defined in (9.3)

$$\ln\left(\frac{p_{jkt}(h)}{p_{0t}(h)}\right) = -\ln\alpha_{jkt}(h) - (b_\tau - 1)\left(\frac{1}{b_{\tau+1}}\ln p_{0,t+1}(h+\Delta_{jkt}) + \ln\Gamma\left[1+\frac{1}{b_{\tau+1}}\right] + \ln E_t[\upsilon_{jkt(\tau)+1}]\right), \tag{9.14}$$

where $\Gamma[\cdot]$ is the complete gamma function. In this particular case, the model delivers a log-linear equilibrium sorting function in the log-odds ratio. (9.14) characterizes the supply for any given market rental rate of human capital in different ranks and firms, $\upsilon_{jkt+1}(h)$. Although there is no wealth effect due to the CARA-utility assumption, there is an intertemporal substitution effect captured by the relative price of a bond today versus tomorrow, b_τ and $b_{\tau+1}$. If the probability of retirement next period increases for some exogenous reason, this would decrease the probability of choosing any job today.

(9.14) shows that executives trade off jobs based on three dimensions: nonpecuniary benefit, $\alpha_{jkt}(h)$; human-capital accumulation, Δ_{jkt}; and expected utility from compensation, $E_t[\upsilon_{jkt(\tau)+1}]$. There is a clear trade-off between the pecuniary and nonpecuniary benefits. However, it is ambiguous how two jobs compare when they differ in the human-capital accumulation among other aspects, because we are not yet able to assert the sign of the effect of human capital h on the probability of retiring next period $p_{0,t+1}(\cdot)$ as compensation is determined in equilibrium.

Labor Demand and Optimal Contract Our analysis of executive supply hinges on the assumption that executives work diligently given the compensation in equilibrium. We turn to an analysis of how shareholders optimally contract with executives to deter them from shirking. If one executive were to shirk privately, shareholders cannot infer that from the firm's output because every firm outcome that might occur when one executive shirks could also occur when every executive works. Technically, the likelihood ratio, $g_{jkt(\tau)}(\pi \mid h)$, is bounded, so shareholders cannot assert whether shirking occurs or not. In the basic model, we preserve the deterministic transition to human capital which is not affected by shirking. But incentives need to be in place to induce executives to work. As all executives employed work diligently in equilibrium, it suffices to look at an optimal contract in which shareholders deter each executive from shirking individually given that the rest of the team works diligently. We focus on the one-period spot contract because the optimal long-term contract is a repetition of the one-period contract, as shown in Margiotta and Miller, 2000. Intuitively, this is a result of executives having access to a complete market for public events across periods and the firm's outcome only depending on the manager's concurrent actions.

The shareholders' objective is to maximize the output subtracting the total expected compensation to the executive team. Assuming the regularity condition (3.1) holds that working creates higher expected output than shirking, this amounts to minimizing $E_t[w_{jkt(\tau)+1}(h,\pi) \mid h]$ or, equivalently, $E_t[\ln v_{jkt(\tau)+1} \mid h]$, subject to a market-participation constraint and an incentive compatibility constraint.

The market-participation, characterized by the executive-employment-decision rule in (9.11), relates the participation $p_t(h)$ to the ex-ante expected utility $E_t[v_{jk,t+1} \mid h]$ before executives observe their taste shocks in the next period. In other words, it does not ensure that an individual executive will accept the offer, as it depends on her taste shock. However, given the distribution of taste shocks, the market-wide participation $p_t(h)$ is consistent with the expected utility $E_t[v_{jk,t+1} \mid h]$. Since executives are risk averse, the lowest wage to deliver the expected utility $E_t[v_{jk,t+1} \mid h]$ is the certainty-equivalent. Define the certainty-equivalent compensation as the fixed amount $w^*_{jkt(\tau)+1}(h)$ satisfying

$$\begin{aligned} \ln E_t[v_{jk,t+1} \mid h] &= \ln E_t\left[(-\gamma w_{jkt(\tau)+1}(h)/b_{\tau+1})\right] \\ &= -\gamma w^*_{jkt(\tau)+1}(h)/b_{\tau+1}. \end{aligned} \tag{9.15}$$

The certainty-equivalent wage solving the market-participation constraint (9.11) is

$$w^*_{jkt(\tau)+1}(h) = \gamma^{-1} b_{\tau(t)+1} \Big\{ \left(b_{\tau(t)} - 1\right)^{-1} \ln \alpha_{jkt}(h)$$
$$+ \ln A_{t+1}\left[\overline{H}_{jk}(h)\right] - \left(b_{\tau(t)} - 1\right)^{-1} q_{jk}\left[p_t(h)\right] \Big\}. \quad (9.16)$$

If effort could be costlessly monitored, $w^*_{jkt(\tau)+1}(h)$ would attract the executives with probability $p_t(h)$. More precisely, a cohort of executives aged t all with human capital h confronted with job opportunities across K ranks offered in J firms would sort into the jobs following the probability distribution $p_t(h)$. The equilibrium compensation schedule must conform with the market-participation constraint implied by the certainty-equivalent wage to attract the mix of executives dictated by the conditional-choice probabilities and does not depend on the firm's excess return. However, there is little reason to presume that a contract subject to the participation constraint induces working over shirking. Because firms do not observe the effort expended, they resort to embedding incentives in executive contracts to induce working.

To elicit diligence from the executives, the shareholders must offer a contract that gives a higher expected utility than enticed in the participation constraint (9.11). In this version of the model, the incentive-compatibility constraint that guarantees that executives get a higher expected utility than shirking provides simplifies to

$$\alpha_{jkt}(h)^{1/(b_\tau-1)}E_t[\upsilon_{jkt(\tau)+1}]$$
$$\leq \beta_{jkt}(h)^{1/(b_\tau-1)}E_t[\upsilon_{jkt+1}g_{jkt(\tau)}(\pi \mid h)]. \tag{9.17}$$

The compensation schedule to induce work, which minimizes expected wage payments from employment subject to the market-participation and incentive-compatibility constraints, can be decomposed into two parts, a fixed component (the certainty equivalent (9.16)) and a variable component. We define the variable component by

$$r_{jkt(\tau)+1}(h,\pi) \equiv \gamma^{-1}b_{\tau(t+1)}\ln\left[1-\eta(h,b_\tau)\left\{g_{jkt(\tau)}(\pi \mid h)\right.\right.$$
$$\left.\left.-\left[\alpha_{jkt}(h)/\beta_{jkt}(h)\right]^{1/(b_\tau-1)}\right\}\right], \tag{9.18}$$

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

$$\int\left[\eta^{-1}+[\alpha_{jkt}(h)/\beta_{jkt}(h)]^{1/(b_{\tau(t)}-1)}-g_{jkt(\tau)}(\pi \mid h)\right]^{-1}f_j(\pi)\mathrm{d}\pi = 1. \tag{9.19}$$

From (9.18), a greater $g_{jkt(\tau)}(\pi \mid h)$, which implies that the outcome π is relatively more likely to occur when there is shirking, leads to a lower $r_{jkt(\tau)+1}(h,\pi)$. Contracting to pay less in states that are relatively more likely to occur when shirking encourages the executive to work. In contrast, if the firm's return is very high, based on the regularity condition (3.3) that $g_{jkt(\tau)}(\pi \mid h) \to 0$ as $\pi \to \infty$, there is almost zero chance that shirking could generate such an extraordinary return. It is also when the firm contracts to pay the most. Hence, the variable component of the compensation, $r_{jkt(\tau)+1}(h,\pi)$, has a finite upper bound of

$$\overline{r}_{jkt(\tau)+1}(h) \equiv \gamma^{-1} b_{\tau(t+1)} \ln \left[1 + \eta(h, b_\tau) \left[\alpha_{jkt}(h)/\beta_{jkt}(h)\right]^{1/(b_\tau - 1)}\right]. \quad (9.20)$$

Theorem 9.3 states that the optimal contract is the sum of the compensating equivalent wage and the variable component defined in the optimal contract.

Theorem 9.3 (Theorem 4.3 of Gayle *et al.*, 2015)**.** The cost-minimizing one-period contract that attracts a executive of age t with experience h to select the k^{th} position in the j^{th} firm with probability $p_t(h)$ and work is

$$w_{jkt(\tau)+1}(h,\pi) = w^{*}_{jkt(\tau)+1}(h) + r_{jkt(\tau)+1}(h,\pi). \quad (9.21)$$

Theorem 9.3 characterizes the cost-minimizing contract subject to the market-participation and incentive-compatibility constraints. The market-participation constraint relates the certainty-equivalent wage required to attract any type of executive with characteristics h at a certain probability for each job. In equilibrium, the perceived probability of attracting an executive is the choice probability derived from the executive's utility-maximization problem; the market-participation constraint derives from the supply (9.11) ensuring this condition. Additionally, the incentive-compatibility constraint is satisfied so the executive works diligently. The expectation of the variable component given by (9.18 is the risk premium, which reflects that in addition to the certainty equivalent, shareholders incur an additional cost is needed to compensate the executive for exposing him to risk from the firm's output through his compensation. The optimal long-term contract can be implemented by a sequence of the one-period contract defined in (9.21), based on the assumptions that executives have access to a complete market and that firm returns are independent of manager's actions more than one-period ago.

Example We end this Section by showing the form the equilibrium takes in this model in the case where the distribution of ε_t is Type I extreme value and the transition function of human capital is as defined in (9.2). Following Hotz and Miller, 1993, the equilibrium in this example can be calculated. The value of each job choice to the executive is now given by the sum of ε_{jkt} and the following deterministic term

$$\begin{aligned} W_{jkt}(h, b_\tau) = & -\ln \alpha_{jkt}(h) - (b_\tau - 1)\Bigg(\frac{1}{b_{\tau+1}} \ln p_{0,t+1}(h + \Delta_{jkt}) \\ & + \ln \Gamma\left[1 + \frac{1}{b_{\tau+1}}\right]\Bigg) + (b_\tau - 1)\Bigg[\frac{\gamma}{b_{\tau+1}} F_{jkt(\tau)}(h) \\ & - E \ln\left(1 - \eta(h, b_\tau)\left\{g_{jkt(\tau)}(\pi \mid h) - [\alpha_{jkt}(h)/\beta_{jkt}(h)]^{\frac{1}{b_\tau - 1}}\right\}\right)\Bigg]. \end{aligned} \tag{9.22}$$

Then, the equilibrium choice probabilities take the logit form

$$p_{jkt}(h) = \frac{\exp[W_{jkt}(h, b_\tau)]}{1 + \sum_{j=1}^{J}\sum_{k=1}^{K}\exp[W_{jkt}(h, b_\tau)]} \quad \text{if } j = 1, \ldots, J \tag{9.23}$$

$$p_{0t}(h) = \frac{1}{1 + \sum_{j=1}^{J}\sum_{k=1}^{K}\exp[W_{jkt}(h, b_\tau)]}. \tag{9.24}$$

On the right-hand side of (9.22), the first three components are the same as the expression for the log-odds ratio in (9.14). The remaining two components in (9.22) express the log of the expected utility from compensation, $E_t[\upsilon_{jkt(\tau)+1}]$, measured as the difference between the executive's productivity and the risk premium needed for the variable part of the executive's compensation. Thus, higher productivity for an executive at a specific firm and rank is associated with higher probability for the executive to select that firm and rank, assuming all other factors remain constant. In this sense, the sorting of executives into firms and ranks is efficient. However, increased moral hazard within a particular firm and rank reduces the probability of choosing that firm and rank, all else being equal. Hence, the agency problem introduces inefficiencies in the matching between executives and firms.

The optimal contract simplifies to

$$\begin{aligned} w_{jkt+1}(h, \pi) &= F_{jkt(\tau)}(h) \\ &- \frac{b_{\tau+1}}{\gamma} E \ln\left(1 - \eta(h, b_\tau)\left\{g_{jkt(\tau)}(\pi \mid h) - [\alpha_{jkt}(h)/\beta_{jkt}(h)]^{\frac{1}{b_\tau - 1}}\right\}\right) \\ &+ \frac{b_{\tau+1}}{\gamma} \ln\left(1 - \eta(h, b_\tau)\left\{g_{jkt(\tau)}(\pi \mid h) - [\alpha_{jkt}(h)/\beta_{jkt}(h)]^{\frac{1}{b_\tau - 1}}\right\}\right). \end{aligned} \tag{9.25}$$

The expression in (9.25) for $w_{jkt+1}(h, \pi)$ depends only on the primitives of the model, but the equilibrium sorting probabilities in (9.23)(9.24) also depend on next-period's retirement probabilities—which is an equilibrium object—along with the model's primitives. Since executives must retire when $t = R$, i.e. $p_{0R}(h) = 1$, the equilibrium can be calculated using backward induction:

1. Solve for $\eta(h, b_\tau)$ using (9.19) and use it to compute $w_{jkt+1}(h, \pi)$ using (9.25).

2. For each executive, set $t = R-1$ and compute $W_{jkR-1}(h, b_{\tau(R-1)})$ and $p_{0R-1}(h)$, which will be functions of only the model primitives and $\eta(h, b_{\tau(R-1)})$ calculated in Step 1.

3. Form $W_{jkR-2}(h, b_{\tau(R-2)})$ using model primitives, $p_{0R-1}(h)$ from Step 2 and $\eta(h, b_{\tau(R-2)})$ calculated in Step 1.

4. Recursively repeat Step 3 for $R - 3, \ldots t$.

We defer the identification analysis because the next model extension nests the basic model.

2 Extension with career concerns

The basic model above assumes that executives' effort does not affect the transition of human capital on a given job. This Section builds on the idea of Gibbons and Murphy, 1992, that career concerns—concerns about the effects of current performance on future compensation–affect effort choice. Thus, we extend the basic model to allow executives to internalize career concerns as an incentive for effort. This is done by relaxing the assumption that human capital evolves independently of the executives' effort. This implies that human capital is now the executives' private information and is unobserved by the firms and markets. Executives have full knowledge of their own productivity, which evolves deterministically according to their choices. Effort affects productivity, providing implicit incentives because current effort may impact future employment choices, promotions, and pay. The private-human-capital model nests the public-human-capital model.[6] Hence, we retain the notations and assumptions of the basic model on executives and firms, choices, and preferences. Below are the parts that will be modified.

Human-Capital Accumulation and Effort As in the basic model, human capital is captured by h_1, a vector of time-invariant components, and h_{2t}, a vector of evolving components. The evolution of h_{2t} is modified to be dependent on effort. If

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

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

$$h_{2t+1} = \sum_{j=1}^{J} \sum_{k=1}^{K} d_{jkt} \left[l_t \overline{H}_{jk}(h_{2t}) + (1 - l_t) \underline{H}_{jk}(h_{2t}) \right]. \tag{9.26}$$

If $l_t = 1$, human capital evolves in the same pattern as in the basic version of the model, (9.2). However, if $l_t = 0$ (shirking), the law of motion of human capital becomes $h_{2t+1} = \sum_{j=1}^{J} \sum_{k=1}^{K} d_{jkt} \underline{H}_{jk}(h_{2t})$. If $\overline{H}_{jk}(h_{2t}) = \underline{H}_{jk}(h_{2t})$ for all (j, k, h), the effort choice l_t drops out of (9.26). Therefore, the law of motion (9.26) nests (9.2).

Example 9.1. Suppose the example in (9.3) still holds for $\overline{H}_{jk}(h_{2t})$ and we specify a similar equation for $\underline{H}_{jk}(h_{2t})$

$$\underline{H}_{jk}(h_{2t}) = h_{2t} + \underline{\Delta}_{jkt}, \tag{9.27}$$

where $\underline{\Delta}_{jkt} \equiv \left(\underline{\Delta}_{jkt}^{(1)}, \underline{\Delta}_{jkt}^{(2)}, \underline{\Delta}_{jkt}^{(3)} \right)$. Suppose the executive shirks. If she is employed at a new firm, then she loses all her firm-specific human capital, $\underline{\Delta}_{jkt}^{(1)} = -h_{2t}^{(1)}$; if she remains employed at the same firm or retires, her firm-specific human capital does not change, $\underline{\Delta}_{jkt}^{(1)} = 0$. Furthermore, $\underline{\Delta}_{jkt}^{(2)} = 0$ and $\underline{\Delta}_{jkt}^{(3)} = 0$ as long as she shirks, meaning that she does not gain an additional year of executive experience or an increase in the number of firms she has worked in.

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

(i) Human capital accumulation through effort, affects individual output $F_{jkt(\tau)}(h)$, more than joint output $g_{jkt(\tau)}(\pi \mid h)$

$$\begin{aligned} & F_{jkt(\tau)}\left(\overline{H}_{jk}(h_{2t-1})\right) - F_{jk(\tau)}\left(\underline{H}_{jk}(h_{2t-1})\right) \\ & \quad > \int \pi \left[g_{jkt(\tau)}\left(\pi \middle| \overline{H}_{jk}(h_{2t-1})\right) - g_{jkt(\tau)}\left(\pi \middle| \underline{H}_{jk}(h_{2t-1})\right)\right] f_j(\pi)\, \mathrm{d}\pi, \\ & \quad \text{for all } (j, k, t, h_{2t-1}). \end{aligned} \tag{9.28}$$

(ii) If $l_0 = 0$, then $F_{jkt(\tau)}(h) \equiv \underline{F}$ for all h.

Assumption (i) is a regulatory assumption that ensures managers' incentive to overreport their human capital. Assumption (ii) is an initial condition that places an upper bound on output and simplifies the off-equilibrium-path analysis by ensuring that a firm does not benefit from employing an executive who has shirked in the initial period.[7]

Capital Markets, Timing and Information The capital market and timing assumptions of the extended and basic models are the same; however, the information structure of the extended model is a bit more complicated. Since human capital is executives' private information, $F_{jkt(\tau)}(h_t)$ is private information and cannot be separately observed. Instead, firms observe the aggregate output of the executive team. We further assume that firms observe all accepted and rejected contracts and employment histories, to simplify the off-equilibrium analysis. Firms cannot ascertain whether an executive shirked based on the output, nor from her job history, by the assumptions that individual productivity is private information and that the private taste shocks ε_{jkt} has full support.

Intertemporal Consumption and Employment Choices The managers' intertemporal consumption choices are unchanged from the basic model, but employment choice needs some additional notations and concepts. Let $h'_t = (h'_1, h'_{2t})$ denote shareholders' belief about a manager's human capital—so called, the manager's reputation—while $h_t = (h_1, h_{2t})$ continues to

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

denote the manager's actual human capital. The contract is based on the manager's reputation, h_t', not the manager's actual human capital, h_t. However, if the executive shirks, firm returns are related to the manager's actual human capital and drawn from $g_{jkt(\tau)}(\pi \mid h) f_j(\pi)$, not $g_{jkt(\tau)}(\pi \mid h') f_j(\pi)$. Consequently, the conditional-choice probabilities depend on both the manager's actual human capital, h_t, and the manager's reputation, h_t'. Therefore, when contracts are only offered for diligent work, shareholders believe that h_t' follows the law of motion $h_{t+1}' = \overline{H}_{jk}(h_t')$ in any given history. In truth, if a manager deviates and shirks at age t, her next-period human capital is $h_{t+1} = \underline{H}_{jk}(h_t)$.

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

$$\upsilon'_{jkt(\tau)+1} \equiv \exp\Big(-\gamma w_{jkt(\tau)+1}(h_t', \pi_t)/b_{\tau(t+1)}\Big). \qquad (9.29)$$

Analogous to the definition of $A_t(h)$ in the basic model, here we define an index of human capital as the recursion

$$B_t\left(h,h'\right) = p_{0t}\left(h,h'\right) E_t\left[\exp\left(\frac{-\varepsilon_{0t}^*}{b_\tau}\right)\right] + \sum_{j=1}^{J}\sum_{k=1}^{K}\left\{p_{jkt}\left(h,h'\right) E_t\left[\exp\left(\frac{-\varepsilon_{jkt}^*}{b_\tau}\right)\right] V'_{jkt}(h,h')\right\}. \tag{9.30}$$

where

$$V'_{jkt}(h,h') \equiv \min\left[\alpha_{jkt}(h)^{\frac{1}{b_\tau}}\left\{B_{t+1}\left[\overline{H}_{jk}(h),\overline{H}_{jk}\left(h'\right)\right] E_t\left[v'_{jkt(\tau)+1}\right]\right\}^{1-\frac{1}{b_\tau}}, \beta_{jkt}(h)^{\frac{1}{b_\tau}}\left\{B_{t+1}\left[\underline{H}_{jk}(h),\overline{H}_{jk}\left(h'\right)\right] E_t\left[v'_{jkt+1}g_{jkt(\tau)}(\pi|h)\right]\right\}^{1-\frac{1}{b_\tau}}\right]. \tag{9.31}$$

The difference in $A_t(h)$ and $B_t\left(h,h'\right)$, as in (9.8) and (9.30), stems from $V'_{jkt}(h,h')$ as defined in (9.31). It arises because shareholders write the contract intended to incentivize working based on reputation h', while a manager with an actual human capital level h that differs from h' may opt to shirk rather than work. Thus, the manager compares the value (disutility) from working another period with that of shirking (netting out the value from retiring then), summarized in the minimization problem in (9.31). The resulting $V'_{jkt}(h,h')$ is the conditional valuation function for match (j,k) for a manager with demographics (t,h) and reputation h'. Suppose the manager is perfectly monitored and thus could never shirk, then her reputation would equal her human capital, $h'=h$, and $B_t(h,h')$ would simplify to $A_t(h)$. When firms cannot monitor effort, h can deviate from h', and the manager could choose shirking over working because the contract is designed to incentivize a manager with human capital h', not h, rendering $B_t(h, h')$ different from $A_t(h)$. When the manager shirks, she reaps the nonpecuniary benefit from shirking since $\beta_{jkt}(h) < \alpha_{jkt}(h)$, but firm returns are drawn from $g_{jkt(\tau)}(\pi \mid h) f_j(\pi)$ rather than $f_j(\pi)$, affecting the probability distribution of her compensation. Subsequent, her reputation $\overline{H}_{jk}(h')$ diverges further from her human capital $\underline{H}_{jk}(h)$. Theorem 9.4 now extends the employment and job choice problem in (9.10) to include the choice of effort, which involves shirking potentially when $h' \neq h$.

Theorem 9.4 (Theorem 5.1 in Gayle *et al.*, 2015)**.** If $h'_{t+1} \equiv \overline{H}_{jk}(h'_t)$, then job matches d_t and effort levels l_t are picked sequentially in time to maximize

$$\varepsilon_{0t} d_{0t} + \sum_{j=1}^{J} \sum_{k=1}^{K} d_{jkt} \left[\varepsilon_{jkt} - \ln V'_{jkt}(h, h') \right]. \tag{9.32}$$

A comparison between (9.32) and (9.10) begins with an intuitive result that $B_t(h, h') \leq A_t(h)$ for any given compensation schedule. That can be shown by induction (omitted) because the option to shirk can lower the disutility in any period. Thus, the maximization value of (9.32) exceeds that of (9.10), and is equal if and only if $\underline{H}_{jk}(h) = \overline{H}_{jk}(h)$ for all (j, k, t, h), so that $B_t(h, h) = A_t(h)$ for all (t, h). In this way, the case of private information of human capital nests the public-information case. If $t \leq R$ and $l_s = 1$ for all $s \in \{t, \ldots, R\}$, we obtain the same characterization of the conditional probabilities as in the basic model with public information, (9.11). Furthermore, the rest of the employment-choice analysis carries through to the private-information model.

Example 9.2. Suppose that ε_{jkt} is independently and identically distributed as a Type I extreme value with location and scale parameters $(0, 1)$. Denote the probability of retirement for a manager with demographics (t, h) and reputation h' as $p_{0t}(h, h')$, then $B_t(h, h')$ simplifies to

$$B_t\left(h, h'\right) = \Gamma\left(\frac{b_\tau + 1}{b_\tau}\right) p_{0t}\left(h, h'\right)^{\frac{1}{b_\tau}}, \tag{9.33}$$

where $p_{0t}\left(h, h'\right) = \left[1 + \sum_{j=1}^{J} \sum_{k=1}^{K} V'_{jkt}(h, h')^{-b_\tau}\right]^{-1}$. (9.33) has the same form as (9.14), the definition of $A_t(h)$, except it depends on $p_{0t}\left(h, h'\right)$ instead of $p_{0t}(h)$, reflecting the role of executives' reputation apart from their human capital.

Labor Demand and Optimal Contract The main difference between the labor demand and contracts in the basic and extended models is that career concerns may ameliorate the divergence of incentives between managers and shareholders in the extended model. Based on the definition of $V'_{jkt}(h, h')$ in (9.31), in equilibrium where $h' = h$, the compensation schedule must satisfy the incentive-compatibility constraint

$$\alpha_{jkt}(h)^{1/(b_\tau-1)}E_t[\upsilon_{jkt(\tau)+1}]B_{t+1}\left[\overline{H}_{jk}(h),\overline{H}_{jk}(h)\right]$$
$$\leq \beta_{jkt}(h)^{1/(b_\tau-1)}E_t[\upsilon_{jkt(\tau)+1}g_{jkt(\tau)}(\pi \mid h)]B_{t+1}\left[\underline{H}_{jk}(h),\overline{H}_{jk}(h)\right]. \quad (9.34)$$

Thus, whenever $B_{t+1}\left[\overline{H}_{jk}(h),\overline{H}_{jk}(h)\right] < B_{t+1}\left[\underline{H}_{jk}(h),\overline{H}_{jk}(h)\right]$, the investment value of human capital from working alleviates the agency problem. In the case of constant compensation, condition (9.34) simplifies to

$$\ln\alpha_{jkt}(h) + (b_\tau - 1)\ln B_{t+1}\left[\overline{H}_{jk}(h),\overline{H}_{jk}(h)\right]$$
$$\leq \ln\beta_{jkt}(h) + (b_\tau - 1)\ln B_{t+1}\left[\underline{H}_{jk}(h),\overline{H}_{jk}(h)\right]. \quad (9.35)$$

So, when the investment value of human capital from working exceeds the disutility relative to shirking, the incentive-compatibility constraint becomes non-binding, eliminating the need for performance-based compensation which incurs a risk premium. Thus, career concerns provide implicit incentives that substitute explicit incentives embedded in contracts. Since implicit incentives are larger when executives are young, explicit incentives increase as managers approach retirement age.

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

$$\begin{aligned} r_{jkt(\tau)+1}(h,\pi) \quad &\equiv \frac{b_{\tau+1}}{\gamma}\ln\Bigg[1-\eta(h,b_\tau)\Bigg[g_{jkt(\tau)}(\pi\mid h) \\ &- \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{\frac{1}{b_{\tau(t)}-1}}\frac{B_{t+1}\left[\overline{H}_{jk}(h),\overline{H}_{jk}\left(h\right)\right]}{B_{t+1}\left[\underline{H}_{jk}(h),\overline{H}_{jk}(h)\right]}\Bigg]\Bigg]. \end{aligned} \tag{9.36}$$

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

$$\begin{aligned} &\int\Bigg[\eta^{-1} \quad + \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{\frac{1}{b_\tau-1}}\left[\frac{B_{t+1}\left[\overline{H}_{jk}(h),\overline{H}_{jk}(h)\right]}{B_{t+1}\left[\underline{H}_{jk}(h),\overline{H}_{jk}(h)\right]}\right] \\ &- g_{jkt(\tau)}(\pi\mid h)\Bigg]^{-1} f_j(\pi)\,\mathrm{d}\pi = 1. \end{aligned} \tag{9.37}$$

For a manager who worked diligently up to period $t-1$, the difference between the risk premiums in the basic and the extended models is the value of human capital attained by diligent work relative to the value of human capital attained if the manager had shirked. Theorem 9.5 states that the optimal contract is the sum of the compensating-equivalent wage and the variable component defined in the optimal contract.

Theorem 9.5 (Theorem 5.2 of Gayle *et al.*, 2015)**.** If $h' = h$, then the cost-minimizing one-period contract that attracts a manager of age t with experience h to select the k^{th} position in the j^{th} firm with probability $p_t(h)$ and work is

$$w_{jkt(\tau)+1}(h,\pi) = w^*_{jkt(\tau)+1}(h) + r_{jkt(\tau)+1}(h,\pi). \tag{9.38}$$

The difference between the cost-minimizing contracts in the basic and extended models is the risk premium, which is weakly smaller when there are career concerns.

Example 9.3. Another divergence from the basic model concerns the case of a constant compensation to elicit work rather than shirk, if (9.35) holds. As $w^*_{jkt(\tau)+1}(h)$ is the same under the basic and the extended models, the only difference is in $r_{jkt(\tau)+1}(h, \pi)$. There, the main difference between (9.36) (9.37) and (9.18) (9.19) is $\frac{B_{t+1}[\overline{H}_{jk}(h), \overline{H}_{jk}(h)]}{B_{t+1}[\underline{H}_{jk}(h), \overline{H}_{jk}(h)]}$ which simplifies in our extreme-value example to

$$\frac{B_{t+1}\left[\overline{H}_{jk}(h), \overline{H}_{jk}(h)\right]}{B_{t+1}\left[\underline{H}_{jk}(h), \overline{H}_{jk}(h)\right]} = \left(\frac{p_{0t+1}[h + \Delta_{jkt}, h + \Delta_{jkt}]}{p_{0t+1}[h + \underline{\Delta}_{jkt}, h + \Delta_{jkt}]}\right)^{\frac{1}{b_{\tau+1}}}. \tag{9.39}$$

If (9.35) holds, the incentive compatibility constraint would be automatically satisfied with a constant compensation, so $\eta(h, b_\tau) = 0$ if

$$\begin{aligned} &\frac{(b_\tau - 1)}{b_{\tau+1}} \ln\Big(p_{0t+1}[h + \Delta_{jkt}, h + \Delta_{jkt}] - p_{0t+1}[h + \underline{\Delta}_{jkt}, h + \Delta_{jkt}]\Big) \\ &\quad \leq \ln\left[\beta_{jkt}(h) - \alpha_{jkt}(h)\right]. \end{aligned} \tag{9.40}$$

In that case, the optimal contract that elicits diligent effort is given by

$$\begin{aligned} w_{jkt(\tau)+1}(h, \pi) = \frac{b_{\tau+1}}{\gamma}\Bigg\{&\frac{\ln \alpha_{jkt}(h)}{b_\tau - 1} \\ &+ \frac{\ln p_{0t+1}[h + \underline{\Delta}_{jkt}, h + \Delta_{jkt}]\Gamma\left[1 + \frac{1}{b_{\tau+1}}\right]}{b_{\tau+1}} \\ &+ \frac{1}{b_\tau - 1} \ln\left(\frac{p_{jkt}(h, h)}{p_{0t}(h, h)}\right)\Bigg\}. \end{aligned} \tag{9.41}$$

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

In addition, the result that the optimal long-term contracts can be implemented as a sequence of short-term contracts does not apply in the extended model. In the extended model shirking executives affect the firm's future returns, both directly through $F_{jkt(\tau)}$, and also, since $h \neq h'$ for shirking executives, indirectly through the cost of achieving incentive compatibility. Thus, a long-term contract that promises to punish managers for poor firm performance several periods from now has a current deterrent effect, and when used in conjunction with immediate punishment is potentially cheaper to implement because more than one signal is used to achieve incentive compatibility in any given period. We interpret the optimal one-period contract in the extended model as an economically meaningful departure from the null hypothesis that the data can be rationalized by a sequence of short-term contracts replicating an optimal long-term contract.

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

Theorem 9.6 (Theorem 5.3 of Gayle *et al.*, 2015)**.** A sequential equilibrium with one-period contracts exists where expected compensation equals the worker's marginal productivity

$$E_t \left[w_{jkt(\tau)+1}(h, \pi) \right|h] = F_{jkt(\tau)}(h). \tag{9.42}$$

At the offer stage, a manager with any level of human capital h offers the cost-minimizing contract specified in (9.41) for the beliefs h'. These offers are accepted if the manager has never deviated from making these equilibrium offers in the past. Any other offer is rejected. Firms believe that all managers making offers deviating from the above contract have shirked and may shirk further. In equilibrium, no executive shirks and $h' = h$.

The full description of strategies and beliefs on and off the equilibrium path and a proof is in the appendix of Gayle *et al.*, 2015. We establish by construction the existence of a sequential equilibrium in which managers sequentially expropriate all the rent that can be extracted from one-period contracts. Along the equilibrium path, managers work every period, so $h = h'$ for all t. If the manager shirks, then $h \neq h'$, and the variable pay components, designed for reputation h', do not necessarily align the incentives of shareholders with those of the manager who is off the equilibrium path. Having deviated from the equilibrium path by shirking once, the manager may shirk further, as (9.30) indicates. If a manager who has shirked offers any contract other than $w_{jkt(\tau)+1}(h', \pi)$, shareholders interpret this deviation from the equilibrium as an indication that the manager has shirked. Whenever a manager shirks, we impose an upper bound of $\underline{F}$ on productivity, such that no firm would retain such a manager because no manager who has shirked through t could have accumulated sufficient wealth to compensate the firm for the expected productivity losses from shirking. This assumption effectively truncates behavior off the equilibrium path because, given the shareholders' beliefs, it is a best response of the manager who has optimally selected (j, k) to demand $w_{jkt(\tau)+1}(h', \pi)$ and follow the continuation path implied by $B_t(h, h')$.[8]

Example 9.4. We end this Section by showing how one would calculate the equilibrium in the extended model if the distribu-

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

tion of ε_t is Type I extreme value and the transition function of human capital is as defined in Equations (9.3) and (9.27). We assume that $\underline{\Delta}_{jkt} = 0$ and, when $h = h'$, we compress the double argument to one for illustrative purposes. Redefine $W_{jkt}(h, b_\tau)$ to be inclusive of both the basic and extended models

$$\begin{aligned}
W_{jkt}(h, b_\tau) &= -\ln \alpha_{jkt}(h) \\
&- (b_\tau - 1)\left(\tfrac{1}{b_{\tau+1}} \ln p_{0t+1}[h + \Delta_{jkt}] + \ln \Gamma\left[1 + \tfrac{1}{b_{\tau+1}}\right]\right) \\
&+ (b_\tau - 1) \times \left[\tfrac{\gamma}{b_{\tau+1}} F_{jkt(\tau)}(h) - E \ln\left(1 - \eta(h, b_\tau) \cdot \right.\right. \\
&\left.\left.\left\{ g_{jkt(\tau)}(\pi \mid h) - \left[\tfrac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{\frac{1}{b_\tau - 1}} \left(\tfrac{p_{0t+1}[h+\Delta_{jkt}]}{p_{0t+1}[h, h+\Delta_{jkt}]}\right)^{\frac{1}{b_{\tau+1}}} \right\}\right)\right]. \qquad (9.43)
\end{aligned}$$

The equilibrium ex-ante choice probabilities have the same form as in (9.23) using the new definition for $W_{jkt}(h, b_\tau)$ in (9.43). The optimal contract simplifies to

$$\begin{aligned}
w_{jk,+1}(h, \pi) &= F_{jkt(\tau)}(h) - \frac{b_{\tau+1}}{\gamma} E \ln\left(1 - \eta(h, b_\tau)\left\{ g_{jkt(\tau)}(\pi \mid h)\right.\right. \\
&\left.\left. - \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{\frac{1}{b_\tau - 1}} \left(\frac{p_{0t+1}[h + \Delta_{jkt}]}{p_{0t+1}[h, h + \Delta_{jkt}]}\right)^{\frac{1}{b_{\tau+1}}} \right\}\right) \\
&+ \frac{b_{\tau+1}}{\gamma} \ln\left(1 - \eta(h, b_\tau)\left\{ g_{jkt(\tau)}(\pi \mid h)\right.\right. \\
&\left.\left. - \left[\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right]^{\frac{1}{b_\tau - 1}} \left(\frac{p_{0t+1}[h + \Delta_{jkt}]}{p_{0t+1}[h, h + \Delta_{jkt}]}\right)^{\frac{1}{b_{\tau+1}}} \right\}\right). \qquad (9.44)
\end{aligned}$$

There is one major difference between the sorting probabilities in the extended and basic models: Human capital has two different effects on the sorting patterns in equilibrium. The explicit human-capital motive captured by $p_{0t+1}[h + \Delta_{jkt}]$ in

(9.43), is already in the basic model; the implicit effect of human capital comes from the career concerns' incentive effect, captured by $p_{0t+1}[h + \Delta_{jkt}]/p_{0t+1}[h, h + \Delta_{jkt}]$ in (9.43). This is because career concerns reduce the risk premium that must be paid to an executive. Therefore, if the executive is comparing two jobs with the same productivity technology (i.e., $F_{jkt(\tau)}(h)$, $f_j\pi$) and $g_{jkt(\tau)}(\pi \mid h)$), nonpecuniary benefits (i.e., $\alpha_{jkt}(h)$ and $\beta_{jkt}(h)$) and human-capital accumulation potential (i.e., Δ_{jkt}), but different career concerns, the executive has a higher probability of choosing the job with the greater career concerns because the certainty-equivalent wage would be higher there. Therefore, career concerns ameliorate the inefficiencies introduced into the sorting and assignment problem by the agency problem. Also, $w_{jkt(\tau)+1}(h, \pi)$ depends not only on primitives as in the basic model, but also on the next-period retirement probability, which is an equilibrium object. Therefore, the objects of the equilibrium must be calculated recursively. Note that in period $t = R - 1$ the contract is the same as in the basic model and therefore the equilibrium can be calculated with the following steps:

1. For each executive, set $t = R - 1$.

   (a) Solve for $\eta(h, b_{\tau(R-1)})$ using (9.19) and use it to compute $w_{jk,R-1}(h, \pi)$ using (9.25).

(b) Compute $W_{jkR-1}(h, b_{\tau(R-1)})$ and $p_{0R-1}(h, h')$, which will be a function of only the model's primitives and $\eta(h, b_{\tau(R-1)})$ calculated in Step 1(a).

2. For each executive, set $t = R - 2$.

   (a) Solve for $\eta(h, b_{\tau(R-2)})$ using (9.37) and use it to compute $w_{jk,R-2}(h, \pi)$ using (9.44) with $p_{0R-1}(h, h')$ calculated in Step 1(b).

   (b) Compute $W_{jkR-2}(h, b_{\tau(R-2)})$ and $p_{0R-2}(h, h')$ using the primitives of the model, $p_{0R-1}(h, h')$ from Step 1(a) and $\eta(h, b_{\tau(R-2)})$ calculated in Step 2(a).

3. Recursively repeat Step 3 for $R - 3, \ldots, t$.

3 Identification

The extended model nests the basic model, so it suffices to analyze identification in the extended model. The model is characterized by the preference parameters, γ, $\alpha_{jkt}(h_t)$, $\beta_{jkt}(h_t)$ and $G(\varepsilon_{11t}, \ldots, \varepsilon_{JKt})$; the technology parameters, $F_{jkt(\tau)}(h)$, $f_j(\pi)$ and $g_{jkt(\tau)}(\pi_{j,\tau+1} \mid h_t)$; and the human-capital transition functions, $\underline{H}_{jk}(h)$ and $\overline{H}_{jk}(h)$. Our data consist of matched panel data on firms and their managers in different time periods, $(w_{njk\tau}, d_{njk\tau}, \pi_{j\tau}, h_{n\tau}, t_{n\tau}, b_\tau)$ where $n = 1, \ldots, N$ indexes the individual executives, $j = 0, \ldots, J$ indexes the firms, $k = 1, \ldots, K$

indexes the rank and $\tau = 1, \ldots \Upsilon$ indexes the time periods. There are two potential cases to consider: when it is optimal to work, and when it is optimal to shirk. While Example 9.3 shows that working could be compatible with a constant compensation, since all the contracts in our data contain a variable component, we focus on what can be identified when it is optimal to work and the incentive compatibility constraint binds.

When the data are generated by an equilibrium where managers work, $F_{jkt(\tau)}(h)$, $f_j(\pi)$ and $\overline{H}_{jk}(h)$ are immediately identified from the data, $F_{jkt(\tau)}(h)$ is identified from the conditional expectation of $w_{njk\tau}$ on $h_{n\tau}$, $t_{n\tau}$ and $d_{njk\tau}$ using the rent-extraction condition in (9.42); $f_j(\pi)$ is identified from observations on $\pi_{j\tau}$; while $\overline{H}_{jk}(h)$ is identified from the empirical distribution of $h_{n\tau+1}$ at $t_{n\tau} + 1$ conditional on $d_{njk\tau}$ and $h_{n\tau}$ at $t_{n\tau}$. As shown in Magnac and Thesmar, 2002, the distribution of the unobserved taste shocks $G(\varepsilon_{11t}, \ldots, \varepsilon_{JKt})$ is not identified nonparametrically. Thus, we assume the econometrician knows $G(\varepsilon_{11t}, \ldots, \varepsilon_{JKt})$, and analyze the identification of γ, plus the semiparametric identification of $\alpha_{jkt}(h_t)$, $\beta_{jkt}(h_t)$, $g_{jkt(\tau)}(\pi_{j,\tau+1} \mid h_t)$ and $\underline{H}_{jk}(h)$. It is instructive to highlight the differences between the basic and extended models, by letting $1\{\text{private}\}$ denote an indicator function taking a value of one if human capital is private and zero if not, and defining a virtual shirking parameter as

$$\beta^*_{jkt}(h) \equiv \beta_{jkt}(h) \left\{ \frac{B_{t+1}\left[\underline{H}_{jk}(h), \overline{H}_{jk}(h)\right]}{A_{t+1}\left[\overline{H}_{jk}(h)\right]} \right\}^{1\{\text{private}\}(b_t - 1)}. \quad (9.45)$$

We proceed in three steps: first the identification of $\alpha_{jkt}(h_t)$, $\beta^*_{jkt}(h_t)$ and $g_{jkt(\tau)}(\pi_{j,\tau+1} \mid h_t)$ is considered when γ is known. Then we explore conditions under which γ is identified. The third step establishes conditions under which $\beta_{jkt}(h)$ and $\underline{H}_{jk}(h)$ are identified from the knowledge of $\beta^*_{jkt}(h)$.

Step 1: The finite-upper-bound property of $r_{jkt(\tau)+1}(h, \pi)$ as established in (9.20) and the optimal compensation schedule in (9.38) imply that compensation is bounded and the manager's maximum compensation is

$$\lim_{\pi \to \infty} w_{jkt(\tau)+1}(h, \pi) = w^*_{jkt(\tau)+1}(h) + \overline{r}_{jkt(\tau)+1}(h) \equiv \overline{w}_{jkt(\tau)+1}(h). \quad (9.46)$$

Suppose γ is known and define the mappings for $g_{jkt(\tau)}(\pi_{j,\tau+1} \mid h_t, \gamma)$, $\alpha_{jkt}(h_t, \gamma)$ and $\beta^*_{jkt}(h_t, \gamma)$ as

$$g_{jkt(\tau)}(\pi_{j,\tau+1} \mid h_t, \gamma) = \frac{e^{\gamma \overline{w}_{jkt(\tau)+1}(h_t)/b_{\tau+1}} - e^{\gamma w_{jkt+1}(h_t, \pi_{j,\tau+1})/b_{\tau+1}}}{e^{\gamma \overline{w}_{jkt+1}(h_t)/b_{\tau+1}} - E[e^{\gamma w_{jkt(\tau)+1}(h,\pi)/b_{\tau+1}} \mid h_t, j]} \quad (9.47)$$

$$\alpha_{jkt}(h_t, \gamma) = \frac{\exp(q_{jk}\left[p_t(h)\right])}{A_{t+1}\left[\overline{H}_{jk}(h)\right]^{b_\tau - 1}} E\left[e^{-\gamma w_{jkt(\tau)+1}(h,\pi)/b_{\tau+1}} \mid h_t, j\right]^{1-b_\tau} \quad (9.48)$$

$$\beta^*_{jkt}(h_t, \gamma) = \frac{\exp(q_{jk}[p_t(h)])}{A_{t+1}\left[\overline{H}_{jk}(h)\right]^{b_\tau - 1}} E\left[e^{\gamma w_{jkt(\tau)+1}(h,\pi)/b_{\tau+1}} g_{jkt(\tau)}(\pi | h, \gamma) | h_t, j\right]^{1-b_\tau}. \quad (9.49)$$

These mappings are derived from Equations (9.38), (9.11) and (9.34). They are similar to the mappings derived in Gayle and Miller, 2015, except that $g_{jkt(\tau)}(\pi_{j,\tau+1} \mid h_t, \gamma)$ is conditional on h_t, while $\alpha_{jkt}(h_t, \gamma)$ and $\beta^*_{jkt}(h_t, \gamma)$ are scaled by $\exp(q_{jk}\left[p_t(h)\right])$

$A_{t+1}\left[\overline{H}_{jk}(h)\right]^{1-b_\tau}$ to reflect the equilibrium sort and the dynamic human-capital accumulation. Additionally, instead of $\beta_{jkt}(h)$, (9.49) refers to $\beta^*_{jkt}(h)$ which is not a primitive of the model, but instead an equilibrium object. The equilibrium compensation schedule, $w_{jkt+1}(h_t, \pi_{j,\tau+1})$, is identified by the conditional expectation of $w_{njk\tau}$ on $(d_{njk\tau}, \pi_{j\tau}, h_{n\tau}, t_{n\tau}, b_\tau)$; therefore, $\overline{w}_{jkt(\tau)+1}(h_t)$ is also identified by the maximum of $w_{njk\tau}$ conditional on $(d_{njk\tau}, h_{n\tau}, t_{n\tau}, b_\tau)$. Therefore, the likelihood ratio, $g_{jkt(\tau)}(\pi_{j,\tau+1} \mid h_t, \gamma)$, is identified if γ is known.

Appealing to Proposition 1 of Hotz and Miller, 1993, a mapping $q_{jk}[\cdot]$ exists conditional on $G(\varepsilon_{11t}, \ldots, \varepsilon_{JKt})$. Consider the exponent version of (9.11) raised to the power of $1/b_\tau$

$$\alpha_{jkt}(h)^{\frac{1}{b_\tau}}\left\{E_t\left[\upsilon_{jkt+1}\right]A_{t+1}\left[\overline{H}_{jk}(h)\right]\right\}^{1-\frac{1}{b_\tau}} = \exp\left[\frac{q_{jk}(p_t[h])}{b_\tau}\right]. \tag{9.50}$$

Substituting (9.50) into (9.8) gives

$$\begin{aligned} A_t(h) \;&= p_{0t}(h)E\left[\exp\left(\frac{-\varepsilon^*_{0t}}{b_{\tau(t)}}\right)\right] \\ &+ \sum_{j=1}^{J}\sum_{k=1}^{K} p_{jkt}(h)E\left[\exp\left(\frac{-\varepsilon^*_{jkt}}{b_{\tau(t)}}\right)\right]\exp\left[\frac{q_{jk}(p_t[h])}{b_\tau}\right]. \end{aligned} \tag{9.51}$$

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

$$A_t(h) = \varphi(p_t[h], h, b_\tau), \tag{9.52}$$

where $\varphi(\cdot)$ is a known function. The choice probability, $p_t(h)$, is identified by the conditional expectation of $d_{njk\tau}$, on $(h_{n\tau}, t_{n\tau}, b_\tau)$ and therefore $A_t(h)$ is identified. It follows immediately from Equations (9.48) and (9.49) that $\alpha_{jkt}(h_t, \gamma)$ and $\beta^*_{jkt}(h_t, \gamma)$ are identified up to γ, since $\exp(q_{jk}[p_t(h)])$, $A_{t+1}\left[\overline{H}_{jk}(h)\right]^{1-b_\tau}$ and $g_{jkt(\tau)}(\pi \mid h_t, \gamma)$ are identified. To summarize, if the risk-aversion parameter is known, then $\alpha_{jkt}(h)$, $g_{jkt(\tau)}(h)$ and $\beta^*_{jkt}(h)$ are semiparametrically identified.

Step 2: Gayle and Miller, 2015, show that in a moral hazard model with neither turnover, promotion, nor human capital accumulation, the risk-aversion parameter, γ, is only set-identified. Their analysis exploits conditions derived from both cost minimization and profit maximization—that in equilibrium the (shareholder) principal only offers work contracts if it is more profitable than paying (executive) agents to shirk. Their analysis proves that any positive value of the risk aversion parameter can be rationalized by the cost minimization conditions; the profit maximization condition is necessary to obtain an inequality that defines an interval for the identified set of γ. Introducing turnover, promotion and human capital yields additional moments for identification. Viewing the compensation schedule offered in different ranks and firms as a lottery, we use the equilibrium sorting condition over ranks and firm types to point-identify γ.

The equilibrium sorting condition identifies γ when exclusion restrictions exist that limit the dependence of the taste parameters on variables that help determine the contract. Substituting for $A_{t+1}\left[\overline{H}_{jk}(h)\right]$ in (9.50) using (9.52), and rearranging, we obtain

$$\begin{aligned}
&\alpha_{jkt}(h)^{\frac{1}{b_\tau - 1}} E_t[\upsilon_{jkt(\tau)+1}] \\
&\quad = \frac{\exp\left[q_{jk}(p_t[h])/(b_\tau - 1)\right]}{\varphi\left(p_{t+1}[\overline{H}_{jk}(h)], \overline{H}_{jk}(h), b_{\tau+1}\right)} \\
&\quad \equiv z_{jkt}(h, b_\tau, b_{\tau+1}). \qquad (9.53)
\end{aligned}$$

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

Let x denote a vector of instruments constructed from (h, j, k, b_τ) for each observation, and define the unconditional density of π as $f(\pi)$. Applying the law of iterated expectations to (9.53) implies

$$\begin{aligned}
&E[z_{jkt}(h, b_\tau, b_{\tau+1}) \mid x] \\
&\quad = E\left[\alpha_{jkt}(h)^{\frac{1}{b_\tau - 1}} \exp\left(\frac{-\gamma w_{jkt(\tau)+1}(\pi, h)}{b_{\tau+1}}\right) \frac{f_j(\pi)}{f(\pi)}\middle| x\right]. \qquad (9.54)
\end{aligned}$$

Thus γ and $\alpha_{jkt}(h)$ are identified off (9.54).

Step 3: Using (9.45), we rewrite (9.49) as

$$\begin{aligned}
&\beta_{jkt}(h) B_{t+1}\left[\underline{H}_{jkt}(h), \overline{H}_{jkt}(h)\right]^{b_\tau - 1} \\
&\quad = \beta^*_{jkt}(h) A_{t+1}\left[\overline{H}_{jkt}(h)\right]^{b_\tau - 1} E\left[e^{-\gamma w_{jkt(\tau)+1}(h,\pi)} \mid h_t, j\right]^{1-b_\tau}. \qquad (9.55)
\end{aligned}$$

The product $\beta_{jkt}(h) B_{t+1}\left[\underline{H}_{jkt}(h), \overline{H}_{jkt}(h)\right]$ is identified from (9.55) because the right side that equation is identified from the previous steps. However, no further headway can be made without adding restrictions to the model. Imagine that the data is generated by the extended model and substitute the virtual parameter $\beta^*_{jkt}(h)$ defined in (9.45), the incentive-compatibility constraint for the extended model, into (9.34). This gives the incentive-compatibility constraint for the basic model, (9.17), with $\beta^*_{jkt}(h)$ replacing $\beta_{jkt}(h)$. Neither (9.10) nor (9.11) depend on $\beta_{jkt}(h)$ or the information structure because the manager works in the equilibrium of both models. Therefore, the solution to the optimal-contract problem given by Equations (9.18), (9.19) and (9.21) for the private-information model is obtained by replacing $\beta_{jkt}(h)$ with $\beta^*_{jkt}(h)$.

These arguments suggest that models with private information about human capital, and induce career concerns, are observationally equivalent models with public information about human capital. Specifically, $\beta^*_{jkt}(h)$ indexes observationally equivalent models that differ only in their specification of $\underline{H}_{jkt}(h)$ and $\beta_{jkt}(h)$. We formally state this result as follows. For the case where bond prices are constant over time.[9]

Theorem 9.7 (Theorem 6.2 of Gayle *et al.*, 2015)**.** Let Θ denote the class of models under consideration, consisting of elements

$$\theta \equiv (\alpha_{jkt}(h), \beta^*_{jkt}(h), \gamma, f_j(\pi), g_{jkt(\tau)}(\pi \mid h), G(\varepsilon)).$$

Suppose $b_\tau = b$ for all τ and $(w_{njk}, d_{njk}, \pi_j, h_n, t_n)$ is generated by $\tilde{\theta}$. For every $\hat{\gamma} > 0$ and all proper probability distribution functions $\hat{G}(\varepsilon)$ defined on the same support as $\tilde{G}(\varepsilon)$, there exists a unique $\hat{\theta}$ solving Equations (9.32), (9.38), (9.42), (9.47), (9.48) and (9.49) that is observationally equivalent to $\tilde{\theta}$.

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

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

(i) We could specialize $\beta_{jkt}(h)$, by assuming it does not depend on the executive's age, $\beta_{jkt}(h) = \beta_{jk}(h)$ for all t, and assume there is a maximal age of retirement R. Recall $A_{t+1}\left[\overline{H}_{jk}(h)\right] = B_{t+1}\left[\underline{H}_{jkt}(h), \overline{H}_{jkt}(h)\right] = 1$ for all $t \geq R$, because there is no incentive effect from career concerns in the period immediately preceding retirement. Consequently the shirking parameter is identified for executives at age $R-1$ from (9.45) as

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

Having identified $\beta_{jkt}(h)$ the continuation value associated with shirking the first time is then identified off (9.49) for all $t \leq R-2$ as

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

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

(ii) Similarly, suppose $\beta_{jkt}(h)$ is independent of the aggregate prices in the economy, as summarized in our model by b_τ. Similar to Subsection 4 in Section 6, assume there are two distinct bond prices b_τ and $b_{\tau'}$, and then from (9.55), we can show that $B_{t+1}\left[\underline{H}_{jkt}(h), \overline{H}_{jkt}(h), b_\tau\right]$[10] is identified relative to a normalization that $B_{t+1}\left[\underline{H}_{jkt}(h), \overline{H}_{jkt}(h), b_{\tau'}\right]$ $= 1$.

(iii) Finally, if we assume that the off equilibrium belief about the law of motion of human-capital accumulation, $\underline{H}_{jkt}(h)$, is known, we can calculate $B_{t+1}\left[\underline{H}_{jkt}(h), \overline{H}_{jkt}(h), b_\tau\right]$ numerically from $t = R$ backward using Equation (9.30).

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

10

Estimating the Models with Human Capital

This Section is an empirical application by Gayle *et al.*, 2015, for the model with public and private human capital. They assume throughout that ε_t is distributed as a Type 1 extreme value, which makes it easier to compute $A_t(h)$ and $B_t(h, h')$ by (9.33) and $q_{jk}[p_t(h)]$ by Equation (9.14). The human-capital transition functions are deterministic, on and off the equilibrium path, denoted by $\overline{H}_{jkt}(h)$ and $\underline{H}_{jkt}(h)$, respectively, in Equations (9.3) and (9.27).

Their approach for the estimation and testing of the model is summarized in four steps, with correspondence to the identification conditions in Section 3:

1. Flexibly estimate $w_{jkt}(\pi, h)$, $\overline{w}_{jkt}(h)$, $f_j(\pi)$, $f(\pi)$, $\overline{H}_{jkt}(h)$, $p_{jkt}(h)$ and $p_{0t}(h)$.

2. Estimate γ and $\alpha_{jkt}(h)$ based on sample analogs of population moments implied by (9.54), plugging in the estimated parameters obtained from Step 1.

3. Use the formulas from Equations (9.47) and (9.49) to estimate $g_{jkt(\tau)}(\pi \mid h)$ and $\beta^*_{jkt}(h)$, using the estimates of γ from Step 2 and the estimated parameters from Step 1.

4. Recursively calculate $B_{t+1}\left[\underline{H}_{jkt}(h), \overline{H}_{jkt}(h)\right]$ numerically, assuming that $\beta_{jkt}(h)$ is independent of b_τ and that $\underline{H}_{jkt}(h)$ is known, and test the implied overidentifying restrictions.

As an alternative, we could utilize the equilibrium computing algorithm outlined in Example 9.4 to estimate the model. That involves implementing a nested fixed-point algorithm, where an inner loop calculates $\eta(h, b_\tau)$ and $B_{t+1}\left[\underline{H}_{jk}(h), \overline{H}_{jk}(h)\right]$ for different values of the primitives, while an outer loop uses the results from the inner loop to estimate the primitives of the model, which are repeated until it converges. However, this estimation strategy is not only computationally intensive, but also somewhat opaque in practical applications. It also requires a fully parametric specification of $f_j(\pi)$, $g_{jkt(\tau)}(\pi \mid h)$ and $F_{jkt(\tau)}(\pi \mid h)$. Instead, the empirical approach above employs nonparametric estimates for these parameters. It also imposes identification restrictions only when needed: for example, the restrictions needed to identify $B_{t+1}\left[\underline{H}_{jkt}(h), \overline{H}_{jkt}(h)\right]$ are only applied when estimating the effects of career concerns.

Step 1: The state space for the dynamic system is the Cartesian product of several dimensions:

- the manager's age t
- her firm and rank last period, $j_{t-1} \in \{1, \ldots, 36\}$ and $k_{t-1} \in \{0, 1, \ldots, 5\}$
- her personal background, $h_t \in \{1, \ldots, H\}$, including fixed components (such as cohort, gender and education) and other variable components of human capital (such as measures of executive experience).

Job matches in their model follow a stochastic law of firm-rank transitions and retirement, denoted by $p_{jkt}(h_t)$ and $p_{0t}(h_t)$. They estimate a multinomial logit model of firm type and position transitions with some (but not all) interactions for exit, promotions, and turnover. In estimation, they exploit Bayes' rule: Given background h_t, the (joint) probability, $p_{jkt}(h_t)$, is the product of the probability of choosing the j^{th} firm conditional on choosing the k^{th} rank, and the (marginal) probability of choosing Rank k. The compensation schedule, $w_{jkt(\tau)}(\pi, h)$, is estimated using a polynomial, and the boundary condition, $\overline{w}_{jkt(\tau)}(h)$, is estimated using the maximum of $w_{jkt(\tau)}(\pi, h)$ over π. Finally, $f_j(\pi)$ and $f(\pi)$ are estimated using kernel-density estimators with normal kernel and the Silverman rule of thumb for the bandwidth.

Step 2: The moment conditions to estimate γ and $\alpha_{jkt}(h)$ are obtained based on the exclusion restrictions (9.54) discussed in Section 6

$$E[z_{jkt}(h, b_\tau, b_{\tau+1})x] = E\left[\alpha_{jkt}(h)^{\frac{1}{b_\tau - 1}} \exp\left(\frac{-\gamma w_{jkt(\tau)+1}(\pi,h)}{b_{\tau+1}}\right) \frac{f_j(\pi)}{f(\pi)} x\right]. \tag{10.1}$$

Upon substituting (9.11)(9.33) into (9.53), $z_{jkt}(h)$ simplifies to

$$z_{jkt}(h) \equiv \Gamma\left[\frac{b_{t(\tau)+1}+1}{b_{t(\tau)+1}}\right]^{-1} p_{0,t+1}\left(\overline{H}_{jk}(h)\right)^{\frac{-1}{b_{t(\tau)+1}}} \left[\frac{p_{0t}(h)}{p_{jkt}(h)}\right]^{\frac{1}{(b_{t(\tau)}-1)}}. \tag{10.2}$$

A sample analog of the moment condition is obtained using the estimates from Step 1. Consistent estimates of γ and $\alpha_{jkt}(h)$ are then obtained from the sample moments along with standard errors adjusted for the two-step estimation.

They estimate $\alpha_{jkt}(h)$ using a log-linear regression on 16 variables involving age t and human capital h, and interactions of these variables with both firm and rank. They also permit the risk-aversion parameter to vary by the 36 firm types, but not by rank. This allows them to test whether risk aversion γ depends on firm size, raising the possibility of misspecification of absolute-risk-aversion (Baker and Hall, 2004). They use a moment condition for each rank and firm combination. Additional moment conditions involve instrument variables, including explanatory variables for $\alpha_{jkt}(h)$, bond prices, and the lag of

Ranks 1 through 4. However, the null hypothesis that γ varies with firm size is rejected, hence they re-estimate a restricted version where there is a common γ for all firm types and additive effects of rank and firm type in $\alpha_{jkt}(h)$. They obtain similar results from both the restricted and unrestricted versions, and hence only the restricted version is reported.

Step 3: Estimation of $g_{jkt(\tau)}(\pi \mid h)$ is based on the identification (9.47), which involves a few terms estimated as follows. They compute $\overline{w}_{jkt(\tau)+1}(h)$ using the maximum of $\hat{w}\,(h_t, \pi)$, estimated as a polynomial expansion from Step 1 for each value of (j, k, t, h). They estimate $E[e^{\gamma w_{jkt(\tau)+1}(h,\pi)/b_{\tau+1}} \mid h_t, j]$ by integrating over the nonparametrically estimated density of π for a given j from Step 1, while substituting the estimates $\hat{\gamma}$ from Step 2 and $\hat{w}\,(h_t, \pi)$ from Step 1. A similar procedure is used in the estimation of $\beta^*_{jkt}(h)$ based on (9.49).

Step 4: In the extended model, substituting the Type I extreme-value functional form of $q_{jk}\,[p_t(h)]$ into (9.49) and rearranging gives

$$\begin{aligned}\beta_{jkt}(h) \equiv \frac{p_{0t}(h)}{p_{jkt}(h)} B_{t+1}\left[\underline{H}_{jk}(h), \overline{H}_{jk}(h)\right]^{1-b_{t(\tau)}} \\ \cdot \left\{\frac{E_t[v_{jkt(\tau)+1}] - \overline{v}^{-1}_{jkt(\tau)+1}}{1 - \overline{v}_{jkt(\tau)+1} E_t[v^{-1}_{jkt(\tau)+1}]}\right\}^{1-b_{t(\tau)}}.\end{aligned} \tag{10.3}$$

for all (j, k, t, h). Estimates of $\beta_{jkt}(h)$ and $B_t(h, h')$ are obtained recursively. Noting that $B_{T+1}(h, h') \equiv 1$ and substituting our estimated risk-aversion parameter and conditional-choice probabilities into (10.3) yields $\beta_{jkT}(h)$. Substituting $\beta_{jkT}(h)$ into (9.31) yields $V'_{jkT}(h, h')$ and hence $B_T(h, h')$, using (9.33). More generally, given $B_{t+1}\left[\underline{H}_{jk}(h), \overline{H}_{jk}(h)\right]$, $\beta_{jkt}(h)$ is obtained from (10.3), and hence estimates of $V'_{jkt}(h, h')$ and $B_t(h, h')$ are produced from Equations (9.31) and (9.33), respectively.

1 Pay differentials

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

Expected Compensation Decomposition The expected compensation can be broken down into two components: the certainty-equivalent wage and the risk premium. The certainty-equivalent wage, which represents the value of compensation adjusted for risk preferences, can be expressed further as the sum of three distinct factors

$$w^*_{jkt(\tau)+1}(h) = \underbrace{\tfrac{b_{t(\tau)+1}}{\gamma(b_{t(\tau)}-1)} \ln \alpha_{jkt}(h)}_{\Delta^{\alpha}_{jkt}(h)} + \underbrace{\tfrac{b_{t(\tau)+1}}{\gamma} \ln A_{t+1}\left[\overline{H}_{jk}(h)\right]}_{\Delta^{A}_{jkt}(h)} - \underbrace{\tfrac{b_{t(\tau)+1}}{\gamma(b_{t(\tau)}-1)} q_{jk}[p_t(h)]}_{\Delta^{q}_{jkt}(h)}. \tag{10.4}$$

Each of these terms captures a different economic rationale underlying executive pay:

1. Nonpecuniary benefits: The first component, denoted $\Delta^{\alpha}_{jkt}(h)$, is the compensating differential that accounts for the nonpecuniary benefits of working in (j, k) relative to the outside option. This term arises in a static model as well.

2. Investment in human capital: The second term, $\Delta^{A}_{jkt}(h)$, captures how employment in (j, k) contributes to compensating differentials through the accumulation of human capital.

3. Idiosyncratic preference shocks (Rosen, 1974) and demand for executives: The third component, $\Delta^{q}_{jkt}(h)$, reflects the variation in compensation based on an executive's idiosyncratic preference for (j, k) and the prevailing demand for executive talent at a given firm and rank, $q_{jk}[p_t(h)]$ 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.

Beyond the certainty-equivalent wage, executives also receive compensation for bearing risk, which is reflected in the risk premium. The risk premium is defined as the difference between the expected total compensation and the certainty-equivalent wage

$$\Delta^{r}_{jkt}(h) \equiv E\left[r_{jkt(\tau)+1}(h,\pi)\right] = F_{jkt(\tau)}(h) - w^{*}_{jkt(\tau)+1}(h). \tag{10.5}$$

The risk premium $\Delta^{r}_{jkt}(h)$ measures the cost of agency, the compensating differential to a risk-averse executive for bearing firm-related risk by holding firm-denominated securities. It is computed in the same way in the public- and private-human capital models, since the executives never shirk in equilibrium.

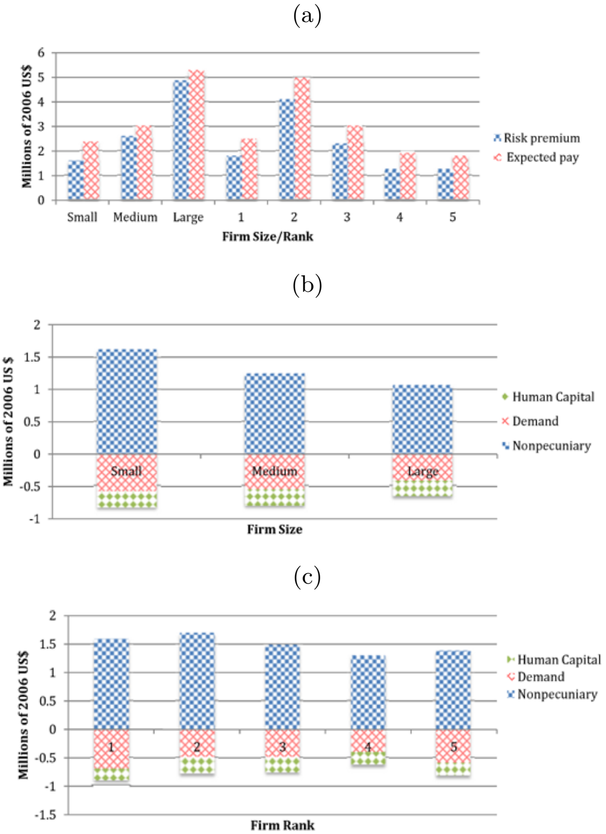
Compensating Differentials by Firm Size and Ranks Figure 10.1 presents the components of the expected pay by firm size and rank. Figure 10.1a shows that executive expected pay is greater in large firms and in higher ranks (up to Rank 2). In stark contrast, 10.1b shows that the certainty-equivalent wage decreases with firm size. The average certainty-equivalent wage of an executive in a small firm is \$780,000, falling to \$430,000 for a medium-size firm, and to \$390,000 for a large firm. Examining each compensating differential in the certainty-equivalent wage in (10.4), executives are willing to forgo some wage for the investment value of human capital. The discount for the value of human capital accumulation remains largely the same across firm size. Meanwhile, larger firms have a higher demand for executives, reflected in greater (less negative) compensating differentials for marginal hires. However, a third factor dominates: small firms must compensate executives more for the nonpecuniary losses from working.

In addition to the negative relationship between firm size and nonpecuniary benefit from working, the distribution of ranks across firm size, as demonstrated in Figure 2.1b, contributed to the difference between the average compensation and the certainty equivalent by firm size. The certainty-equivalent wage exhibits a hump-shaped pattern over ranks, based on Figure 10.1c, starting at \$570,000 in Rank 5, increasing monotonically to \$900,000 in Rank 2, before declining to \$690,000 in Rank 1. Rank 3 executives have a higher certainty-equivalent compensation, \$730,000, than Rank 1 executives, but Rank 1 executives have a slightly higher certainty-equivalent compensation than Rank 4 executives, \$660,000. A similar pattern exists in Table 2.9 for the average total compensation by rank, which ranges from \$1,269,000 (for Rank 5) to \$4,794,000 (for Rank 2). The compression of the certainty equivalent pay at the top ranks is largely due to the rise of risk premium.

In larger firms, the risk premium increasingly outweighs the certainty-equivalent wage, as Figure 10.1a shows. This aligns with the observed pattern that the variance of compensation rises with firm size.[1] In principle, the higher variability of compensation in large firms could be due to volatility in abnormal returns

[1] See Table 2A in the online appendix of Gayle *et al.*, 2015.

Figure 10.1
Rank and Firm-Size Pay Decomposition



Note: Excerpt from Gayle *et al.*, 2015, Figure 3. This figure presents the rank and firm-size pay decomposition, illustrating how different factors contribute to compensation across firm sizes and executive ranks. The figure consists of three bar charts: (a) compares risk premium and expected pay across different firm sizes (Small, Medium, and Large) and executive ranks (1 to 5), showing that most of expected pay is composed of the risk premium, increases with firm size, but exhibit a hump shape over rank, with rank 2 getting the highest pay. The stacked bars in (b) and (c) decompose the compensation into three components–from bottom to top–Human Capital, Demand, and Nonpecuniary benefits, by firm sizes in (b) and by rank in (c). In both (b) and (c), nonpecuniary benefits make up the largest share of total compensation, while demand and human capital offset part of the compensating differential from nonpecuniary benefits. Small firms have the largest compensating differential from nonpecuniary benefits. The compensating differential for human capital increases as executives rise through ranks 5 to 2, and decreases from rank 2 to rank 1.

that factors into compensation packages and are accounted for by the risk premium, or to other forms of heterogeneity, both observed and unobserved. Ultimately, a risk premium designed to solve agency problems reconciles how expected compensation rises with firm size whereas the certainty-equivalent wage declines as Figure 10.1b shows.

Figure 10.1c shows that executives give up more compensation for human capital investment as they rise in ranks until Rank 1, where the trend reverses. The lifecycle theory of human capital predicts that the investment value of human capital declines with age, a pattern supported by Table 2.9 in Section 2, which shows that higher ranks are occupied by older executives with greater executive experience. However, in our model, the value of human capital is inversely related to the probability of retirement. Thus, the inverted pattern in Figure 10.1c reflects the exit probability, which is lowest in Rank 2, highest in Rank 1, and lower in larger firms, as shown in Table 2.10 in Section 2. The value of human capital is remarkably high as a fraction of the certainty equivalent wage, ranging from a quarter to a half approximately. The major new finding on human capital investment is that even late in the career cycle, variety in job experience adds to human capital, and that the value of human capital is higher in large firms. This finding suggests that top-ranking executive positions may demand general human capital accumulated through experience in various firms.[2] Overall, although human capital accumulation is important, the risk premium is the largest component explaining variation in pay across firm sizes and ranks. We further discuss the sources of this variation and the nature of the agency problem in different ranks and firms in the next subsection.

The Risk Premium Crucial to the estimation of the risk premium is the risk-aversion measure; hence, they first examine the robustness of their estimates of risk-aversion. They estimate the risk aversion parameter[3] to be 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 one million dollars. This is similar to the results of Gayle and Miller, 2009a, presented in Section 5, where the executive's risk-aversion parameter is found to be 0.501 for the period 1944-1978 and 0.519 for the period 1993-2004. Their estimate of risk aversion is generally lower than that found in laboratory experiments and field studies,[4] which

[2]Table 9A in the online appendix of Gayle *et al.*, 2015, showing that the value of human capital increases with turnover by roughly $13K supports this hypothesis.

[3]Initially, they specified the risk-aversion parameter as a function of gender and firm size, but found no evidence against an identical coefficient of risk aversion across gender and firm size.

[4]E.g. Holt and Laury, 2002, Holt and Laury, 2005, Harrison *et al.*, 2005, Harrison *et al.*, 2007, Andersen *et al.*, 2008, Dohmen *et al.*, 2010.

is reasonable given that they are studying executives. The low risk-aversion parameter accompanied with high risk premium further supports that the risk premium is the major component explaining the variation of compensation over rank and firm.

Table 10.1 displays their estimates of the risk premium $\Delta^r_{jkt}(h)$, showing that, at Ranks 4 and 5, the cost of agency, measured by $\Delta^r_{jkt}(h)$, is small and insignificant in small firms, but it amounts to \$1.5 million, \$3.3 million and \$1 million for Ranks 3, 2, and 1. Roughly 82% of the compensation of a CEO (Rank 2), versus 72% for Rank 1, 76% for Rank 3, 65% for Rank 4, and 69% for Rank 5, is due to the risk premium. The service sector pays a higher risk premium than the other two, a factor which helps close the gap between the considerably higher levels of average compensation paid in that sector and those reported in Table 10.1.

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

Table 10.1
Risk Premium from Agency

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)

Note: Excerpt from Table 3, Gayle *et al.*, 2015. 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.

In this framework, expected compensation is the executive's marginal product. Thus, executives with a Ph.D., who receive an average expected compensation of $3.0 million, are more productive than those with an MBA, $2.7 million, and those without either, $2.8 million. An executive with a Ph.D. receives a higher risk premium, $2.3 million, than one with an MBA, $2.1 million, but an executive with an MBA has a higher fraction of expected compensation, 78%, than one with Ph.D., 76%, as risk premium. There is a $362,000 spike in the risk premium for new executives, but it declines by $65,000 with each extra year of tenure and age. Consequently, the lower certainty-equivalent wage offered to first-year executives is partially hidden by data on their average compensation. Given that larger firms have more executives with MBA degrees and fewer tenured executives, the above two findings both work to overstate the firm-size pay premium in the raw data. The overall effect of the interaction with firm size and rank is ambiguous: For example, the effect of Rank 1 overstates the effect of firm size while the effect of Rank 5 understates it. After controlling for the effect of rank and human capital, they find a negative relationship between firm size and certainty-equivalent wage, the main cause of which is the positive relationship between firm size and the risk premium.

2 Agency-cost decomposition

The risk premium is the agency cost in a moral-hazard model. As in all previous models, they compare the risk premium paid to the manager with the gross loss from shirking to shareholders and the executive's net benefit from shirking. Career concern contributes to the last welfare measure, which is new in this model.

The net benefit from shirking to the executive is denoted by $\Delta_{jkt}^{\beta *}(h)$, it is the sum of two components: The first, denoted $\Delta_{jkt}^{\beta}(h)$, is the compensating differential for the current disutility from working against shirking; it measures the misalignment of incentives from the executive's perspective. The second, denoted $\Delta^{B}_{jkt}(h)$, measures the difference in the conditional continuation values from working in the current period t versus shirking, due to the investment value of human capital. Thus, the second component captures how career concern ameliorates the agency problem. The definitions of the two components are

$$\Delta^{\beta}_{jkt}(h) \equiv \frac{b_{t(\tau)+1}}{\gamma(b_{t(\tau)}-1)} \ln\left(\frac{\alpha_{jkt}(h)}{\beta_{jkt}(h)}\right), \tag{10.6}$$

and

$$\Delta^{B}_{jkt}(h) \equiv \frac{b_{t(\tau)+1}}{\gamma} \ln\left(\frac{A_{t+1}\left[\overline{H}_{jk}(h)\right]}{B_{t+1}\left[\underline{H}_{jk}(h), \overline{H}_{jk}(h)\right]}\right). \tag{10.7}$$

The net benefit from shirking, $\Delta^{\beta*}_{jkt}(h)$, is identified from data on choice probabilities, compensation schedule, the abnormal return distribution, the risk-aversion parameter and the likelihood ratio (see (9.49)) and is therefore identified without appealing to the functional-form assumptions on career concerns (i.e., $\underline{H}_{jk}(h)$) or exclusion restrictions.

The identification of the second component, $\Delta^{B}_{jkt}(h)$, relies on the functional-form assumptions on $\underline{H}_{jk}(h)$ given in (9.27). Aside, the identification of $\beta_{jkt}(h)$ uses the exclusion restriction that $\beta_{jkt}(h)$ is independent of the aggregate conditions in the economy, i.e., bond prices. These are the only places where these conditions are required for identification.

Firm Size and Rank Table 10.2 shows the estimates of the gross loss from shirking to shareholders, denoted by $\Delta^{g}_{jkt}(h)$. Table 10.3 reports the estimates of $\Delta^{\beta^*}_{jkt}(h)$ and Table 10.4 reports the estimates of the career concern component, $\Delta^{B}_{jkt}(h)$. Figure 10.2 summarizes all the estimates across firm size. Figure 10.2 (and Table 10.2) shows that small, consumer-sector firms lose much more of their equity value when a Rank-5 executive shirks, 33.6%, compared to that of large firms, 8%. This is in contrast to Baker and Hall, 2004, who find constant loss across firm size. Intuitively, shirking executives in small firms have higher marginal impact on firm performance than they would in large firms on each unit of equity. This finding also implies that distinguishing shirking from working is more difficult in larger firms; in other words, the signal quality is lower.

Figure 10.2 and Table 10.3 illustrate that $\Delta^{\beta}_{jkt}(h)$ decreases with firm size, declining by \$3.1 million for medium firms and \$4.5 million for large firms. Sectoral variation is also evident, with $\Delta^{\beta}_{jkt}(h)$ being \$3.8 million higher in the service sector compared to the consumer sector and \$2.6 million lower in the primary sector. Additionally, Figure 10.2 and Table 10.4 show that $\Delta^{B}_{jkt}(h)$ and career concerns remain unchanged across firm sizes. The estimates of $\Delta^{\beta}_{jkt}(h)$ and $\Delta^{B}_{jkt}(h)$ suggest that the risk premium is at most weakly decreasing with firm size. However, since signal quality is consistently poorer in larger firms, this ultimately leads to an increasing risk premium as firm size grows. Furthermore, there is a positive correlation between firm size and the expected gross equity loss due to shirking. When the estimates are scaled by the average equity value, the resulting gross equity losses amount to \$102 million for small firms, \$203 million for medium firms, and \$393 million for large firms. This indicates that while the absolute gross equity loss from shirking is greater in larger firms, the agency cost exhibits a concave increasing relationship with firm size.

Across ranks, the most surprising result is that the gross loss (Table 10.2), $\Delta^{g}_{jkt}(h)$, declines in rank, in contrast to the conventional wisdom that shareholders risk more from chairmen and CEOs who shirk than lower-ranked managers. Instead, these results support the view that managers directly overseeing the firm's operations can affect firm returns the most. In turn, the return of the firm is a better signal of these managers' effort.

The nonpecuniary benefit to an executive from shirking, $\Delta^{\beta*}_{jkt}(h)$, as Table 10.3 shows, is about \$10 million for a 50-year-old Rank 5 executive in a small firm in the consumer sector. As rank increases, the compensating differential is positive and economically significant although not statistically significant. This is consistent with a model with private human capital where significant career concerns exist at all ranks. Career concerns offset the differential for diligent work versus shirking (Table

Table 10.2
Gross Loss to Shareholders from Not Providing Incentives

	$E(x(1-g(x)))$	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)				

Note: Excerpt from Table 4, Gayle *et al.*, 2015. 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.

Table 10.3
Net Compensating Differentials for Working versus Shirking

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)

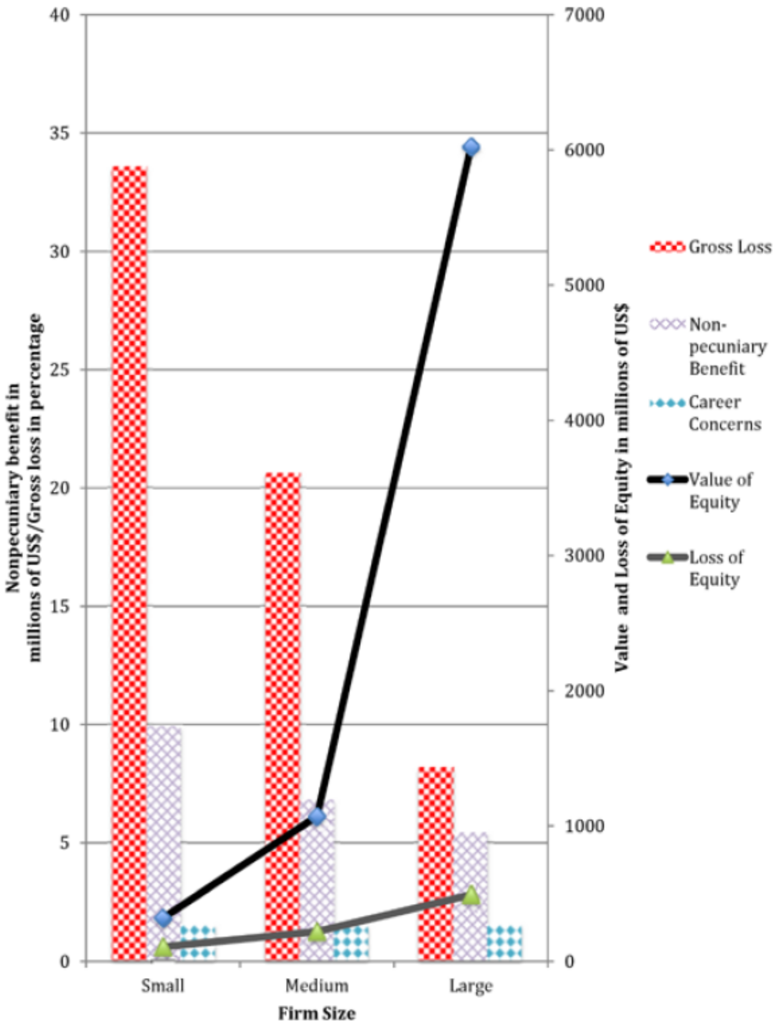
Note: Excerpt from Table 5, Gayle *et al.*, 2015. Compensation is measured in millions of 2006 US$; standard errors are listed in parentheses; tenure and executive experience (Exec. exp.) are measured in years; NBE (NAE) is the number of times the executive changed firms before (after) entering one of the ranks in our sample.

Table 10.4
Career Concern Amelioration of Agency Cost

Variable	Constant	Age-50	Age-50 Squared	Tenure	Exec. Exp.	NBE	NAE	Female
Constant	-1.547	0.006	0.001	0.009	0.015	0.059	0.050	0.154
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
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)

Note: Excerpt from Table 6, Gayle *et al.*, 2015. 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.

Figure 10.2
Agency Cost Decomposition by Firm Size



Note: Excerpt from Figure 4, Gayle *et al.*, 2015. This figure illustrates the agency cost decomposition across firm size. The x-axis represents firm size (Small, Medium, and Large), while the left y-axis measures nonpecuniary benefits to the manager (in millions of US$) or the gross loss to the firm (in percentage) if the manager shirks, and the right y-axis measures value and loss of equity (in millions of US$). Gross loss is highest for small firms as a percentage of firm size and declines with firm size, as the red bars show. The light purple bars represent nonpecuniary benefits, while the blue dotted bars reflect career concerns, both of which are relatively consistent across firm sizes. A black solid line with diamond markers tracks value of equity, showing a steep increase as firm size grows. Additionally, a gray solid line marks the loss of equity in green triangles, which remains relatively small compared to other components.

10.4), $\Delta^B_{jkt}(h)$, by between 15% and 22%; as a percentage of the gross compensating differential. Career concerns is most prevalent in Rank 3, which is just below CEO. It is least effective in Rank 1, the top rank. Likewise, the role of career concerns declines with age, tenure, executive experience, and experience in different firms.

3 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 sorting (Gabaix and Landier, 2008), and agency costs (Gayle and Miller, 2009a) in explaining why executive pay increases with firm size. The equilibrium framework by Gayle *et al.*, 2015, incorporates both sorting and agency considerations, and additionally, human capital and career concerns.

Most of the variation in pay across firm size is attributed to the agency problem, as the risk premium increases by firm size. Using a hierarchy (constructed in Gayle *et al.*, 2012) of executive ranks, they find that the risk premium also increases with rank. This suggests that higher-ranked executives or those in larger firms have less span of control on the firm returns, contrary to the conventional wisdom that shareholders risk more from chairmen and CEOs who have greater latitude to shirk than lower-ranked officers. This is evident in that the expected gross loss shareholders would incur from a shirking executive declines significantly with firm size and rank. As executives in larger firms and higher ranks are harder to monitor, this in turn leads the risk premium to increase with firm size and rank. This also speaks to the view of a firm as an organization (Alchian and Demsetz, 1972; Mirrlees, 1976), rather than a chain of command (Williamson, 1967; Calvo and Wellisz, 1980).

Finally, they examine the role of human capital in providing implicit incentives that ameliorate the moral-hazard problem. While the costs and benefits of shirking are identified as before, to identify the disutility of shirking and continuation value, both off-the-equilibrium-path, they use functional-form assumptions on human capital acquisition through shirking. Their findings show that the explicit incentives in the contract increase with age because career concerns decline over the executive lifecycle. Another finding is that both the CEO and executives just below CEO invest more heavily in human capital than managers in lower ranks, counter to the conventional understanding in textbook labor economics that executives in top ranks invest less in human capital.

Acknowledgements

We thank Jonathan Glover (the Editor) and the anonymous referee for their helpful comments. We thank the coauthors for their collaboration on the published work this manuscript is based on: Sumru Altug, George-Levi Gayle, Limor Golan, V. Joseph Hotz, Natalia Khorunzhina, Chen Li, Mary M. Margiotta, and Holger Sieg.

References

Abowd, J. M. and D. S. Kaplan. (1999). "Executive compensation: six questions that need answering". *Journal of Economic Perspectives*. 13(4): 145–168.

Alchian, A. A. and H. Demsetz. (1972). "Production, Information Costs, and Economic Organization". *The American Economic Review*. 62(5): 777–795.

Altug, S. and R. A. Miller. (1990). "Household Choices in Equilibrium". *Econometrica*. 58: 543–570.

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.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström. (2008). "Eliciting risk and time preferences". *Econometrica*. 76(3): 583–618.

Andrews, D. W. and P. J. Barwick. (2012). “Inference for parameters defined by moment inequalities: A recommended moment selection procedure”. *Econometrica*. 80(6): 2805–2826.

Andrews, D. W. and P. Guggenberger. (2009). “Hybrid and size-corrected subsampling methods”. *Econometrica*. 77(3): 721–762.

Andrews, D. W. and G. Soares. (2010). “Inference for parameters defined by moment inequalities using generalized moment selection”. *Econometrica*. 78(1): 119–157.

Antle, R. and A. Smith. (1985). “Measuring executive compensation: Methods and an application”. *Journal of Accounting Research*. 23(2): 296–325.

Antle, R. and A. Smith. (1986). “An empirical investigation of the relative performance evaluation of corporate executives”. *Journal of Accounting Research*: 1–39.

Antonovics, K. and L. Golan. (2012). “Experimentation and Job Choice”. *Journal of Labor Economics*. 30(2): 333–366.

Armstrong, C. S., M. E. Barth, A. D. Jagolinzer, and E. J. Riedl. (2010a). “Market reaction to the adoption of IFRS in Europe”. *The Accounting Review*. 85(1): 31–61.

Armstrong, C. S., A. D. Jagolinzer, and D. F. Larcker. (2010b). "Chief executive officer equity incentives and accounting irregularities". *Journal of Accounting Research*. 48(2): 225–271.

Baker, G. P. and B. J. Hall. (2004). "CEO Incentives and Firm Size". *Journal of Labor Economics*. 22(4): 767–798.

Banerjee, S., M. Humphery-Jenner, and V. Nanda. (2015). "Restraining Overconfident CEOs Through Improved Governance: Evidence from the Sarbanes–Oxley Act". *Review of Financial Studies*. 28: 2812–2858.

Bargeron, L. L., K. M. Lehn, and C. J. Zutter. (2010). "Sarbanes–Oxley and Corporate Risk-Taking". *Journal of Accounting and Economics*. 49: 34–52.

Bellman, R. (1957). *Dynamic Programming*. Princeton, NJ: Princeton University Press.

Berle, A. A. and G. C. Means. (1932). *The Modern Corporation and Private Property*. New York: Harcourt, Brace, & World.

Bertomeu, J., A. Beyer, and D. J. Taylor. (2022). "Moral hazard and the information content of prices". *Review of Economic Dynamics*. 46: 1–24.

Bertomeu, J., Y. Liang, and I. Marinovic. (2023). "A Primer on Structural Estimation in Accounting Research". *Foundations and Trends® in Accounting*. 18(1–2): 1–137.

Bertomeu, J., P. Ma, and I. Marinovic. (2020). “How Often Do Managers Withhold Information?” *The Accounting Review*. 95(4): 73–102.

Bertrand, M. and K. F. Hallock. (2001). “The gender gap in top corporate jobs”. *ILR Review*. 55(1): 3–21.

Calvo, G. A. and S. Wellisz. (1980). “Supervision, Loss of Control, and the Optimum Size of the Firm”. *Journal of Political Economy*. 86(5): 943–952.

Carter, M. E., L. J. Lynch, and S. L. C. Zechman. (2009). “Changes in Bonus Contracts in the Post-Sarbanes–Oxley Era”. *Review of Accounting Studies*. 14: 480–506.

Carter, M. E., L. J. Lynch, and I. Tuna. (2007). “The role of accounting in the design of CEO equity compensation”. *The Accounting Review*. 82(2): 327–357.

Chen, X., Q. Cheng, and X. Wang. (2015). “Does increased board independence reduce earnings management? Evidence from recent regulatory reforms”. *Review of Accounting Studies*. 20: 899–933.

Chernozhukov, V., H. Hong, and E. Tamer. (2007). “Estimation and confidence regions for parameter sets in econometric models 1”. *Econometrica*. 75(5): 1243–1284.

Chhaochharia, V. and Y. Grinstein. (2009). “CEO compensation and board structure”. *The Journal of Finance*. 64(1): 231–261.

Cohen, D. A., A. Dey, and T. Z. Lys. (2013). “Corporate governance reform and executive incentives: Implications for investments and risk taking”. *Contemporary Accounting Research*. 30(4): 1296–1332.

Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Vol. 17. Yale University Press.

DeMarzo, P. M. and Y. Sannikov. (2017). “Learning, Termination, and Payout Policy in Dynamic Incentive Contracts”. *The Review of Economic Studies*. 84(1): 182–236.

Dewatripont, M., I. Jewitt, and J. Tirole. (1999). “The economics of career concerns, part I: Comparing information structures”. *The Review of Economic Studies*. 66(1): 183–198.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde. (2010). “Are risk aversion and impatience related to cognitive ability?” *American Economic Review*. 100(3): 1238–1260.

Duchin, R., J. G. Matsusaka, and O. Ozbas. (2010). “When are outside directors effective?” *Journal of Financial Economics*. 96(2): 195–214.

Edmans, A., X. Gabaix, and D. Jenter. (2017). “Executive compensation: A survey of theory and evidence”. *The handbook of the economics of corporate governance*. 1: 383–539.

Edmans, A., X. Gabaix, T. Sadzik, and Y. Sannikov. (2012). "Dynamic CEO compensation". *The Journal of Finance*. 67(5): 1603–1647.

Fee, C. E., C. J. Hadlock, and J. R. Pierce. (2013). "Managers with and without style: Evidence using exogenous variation". *The Review of Financial Studies*. 26(3): 567–601.

Finnerty, J. E. (1976). "Insiders and market efficiency". *The Journal of Finance*. 31(4): 1141–1148.

Frydman, C. and D. Jenter. (2010). "CEO compensation". *Annual Review of Financial Economics*. 2(1): 75–102.

Frydman, C. and R. S. Molloy. (2011). "Does tax policy affect executive compensation? Evidence from postwar tax reforms". *Journal of Public Economics*. 95(11-12): 1425–1437.

Fudenberg, D., B. Holmstrom, and P. Milgrom. (1990). "Short-term contracts and long-term agency relationships". *Journal of Economic Theory*. 51(1): 1–31.

Fudenberg, D. and J. Tirole. (1991). "Perfect Bayesian equilibrium and sequential equilibrium". *Journal of Economic Theory*. 53(2): 236–260.

Gabaix, X. and A. Landier. (2008). "Why Has CEO Pay Increased So Much?" *The Quarterly Journal of Economics*. 123(1): 49–100.

Gayle, G.-L., L. Golan, and R. A. Miller. (2012). "Gender differences in executive compensation and job mobility". *Journal of Labor Economics*. 30(4): 829–872.

Gayle, G.-L., L. Golan, and R. A. Miller. (2015). "Promotion, turnover, and compensation in the executive labor market". *Econometrica*. 83(6): 2293–2369.

Gayle, G.-L., C. Li, and R. A. Miller. (2022). "Was Sarbanes–Oxley costly? Evidence from optimal contracting on CEO compensation". *Journal of Accounting Research*. 60(4): 1189–1234.

Gayle, G.-L. and R. A. Miller. (2009a). "Has moral hazard become a more important factor in managerial compensation?" *American Economic Review*. 99(5): 1740–1769.

Gayle, G.-L. and R. A. Miller. (2009b). "Insider information and performance pay". *CESifo Economic Studies*. 55(3-4): 515–541.

Gayle, G.-L. and R. A. Miller. (2015). "Identifying and testing models of managerial compensation". *The Review of Economic Studies*. 82(3): 1074–1118.

Gibbons, R. and K. J. Murphy. (1992). "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence". *Journal of Political Economy*. 100(3): 468–505.

Goolsbee, A. (2000). “What happens when you tax the rich? Evidence from executive compensation”. *Journal of Political Economy*. 108(2): 352–378.

Guthrie, K., J. Sokolowsky, and K.-M. Wan. (2012). “CEO compensation and board structure revisited”. *The Journal of Finance*. 67(3): 1149–1168.

Hall, B. J. and J. B. Liebman. (1998). “Are CEOs really paid like bureaucrats?” *The Quarterly Journal of Economics*. 113(3): 653–691.

Harris, M. and B. Holmstrom. (1982). “A theory of wage dynamics”. *The Review of Economic Studies*. 49(3): 315–333.

Harrison, G. W., E. Johnson, M. M. McInnes, and E. E. Rutström. (2005). “Risk aversion and incentive effects: Comment”. *American Economic Review*. 95(3): 897–901.

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(2): 433–458.

Hart, O. (2009). “Regulation and Sarbanes–Oxley”. *Journal of Accounting Research*. 47: 437–445.

Hayes, R. M., M. Lemmon, and M. Qiu. (2012). “Stock options and managerial incentives for risk taking: Evidence from FAS 123R”. *Journal of Financial Economics*. 105(1): 174–190.

Hayes, R. M. and S. Schaefer. (2000). “Implicit contracts and the explanatory power of top executive compensation for future performance”. *The RAND Journal of Economics*. 31(2): 273–293.

Hochberg, Y. V., P. Sapienza, and A. Vissing-Jørgensen. (2009). “A Lobbying Approach to Evaluating the Sarbanes–Oxley Act of 2002”. *Journal of Accounting Research*. 47: 519–583.

Holmström, B. (1999). “Managerial incentive problems: A dynamic perspective”. *The Review of Economic Studies*. 66(1): 169–182.

Holt, C. A. and S. K. Laury. (2005). “Risk aversion and incentive effects: New data without order effects”. *American Economic Review*. 95(3): 902–904.

Holt, C. A. and S. K. Laury. (2002). “Risk aversion and incentive effects”. *American Economic Review*. 92(5): 1644–1655.

Hotz, V. J. and R. A. Miller. (1993). “Conditional choice probabilities and the estimation of dynamic models”. *The Review of Economic Studies*. 60(3): 497–529.

Jaffe, J. F. (1974). “Special information and insider trading”. *The Journal of Business*. 47(3): 410–428.

Jensen, M. C. and K. J. Murphy. (1990). “Performance pay and top-management incentives”. *Journal of Political Economy*. 98(2): 225–264.

Joo, J. H. and S. L. Chamberlain. (2017). “The effects of governance on classification shifting and compensation shielding”. *Contemporary Accounting Research*. 34(4): 1779–1811.

Khorunzhina, N. and R. A. Miller. (2022). “American Dream Delayed: Shifting Determinants of Homeownership”. *International Economic Review*. 63(1): 3–35.

Li, C. (2024). “Are Top Management Teams Compensated as Teams? A Structural Modeling Approach”. *Management Science*. 70(12): 8753–8771.

Linck, J. S., J. M. Netter, and T. Yang. (2009). “The effects and unintended consequences of the Sarbanes-Oxley Act on the supply and demand for directors”. *The Review of Financial Studies*. 22(8): 3287–3328.

Lorie, J. H. and V. Niederhoffer. (1968). “Predictive and statistical properties of insider trading”. *The Journal of Law and Economics*. 11(1): 35–53.

Lu, J. and W. Wang. (2015). “Board independence and corporate investments”. *Review of Financial Economics*. 24: 52–64.

Lucas, R. E. (1978). “On the size distribution of business firms”. *The Bell Journal of Economics*. 9(2): 508–523.

Magnac, T. and D. Thesmar. (2002). “Identifying dynamic discrete decision processes”. *Econometrica*. 70(2): 801–816.

Malcomson, J. M. and F. Spinnewyn. (1988). “The multiperiod principal-agent problem”. *The Review of Economic Studies*. 55(3): 391–407.

Margiotta, M. M. and R. A. Miller. (2000). “Managerial compensation and the cost of moral hazard”. *International Economic Review*. 41(3): 669–719.

Masson, R. T. (1971). “Executive motivations, earnings, and consequent equity performance”. *Journal of Political Economy*. (6): 1278–1292.

Miller, R. A. (1984). “Job Matching and Occupational Choice”. *Journal of Political Economy*. 92(6): 1086–1120.

Miller, R. A. and A. Nguyen. (2024). “Optimal long term executive contracts”. *Working Paper*.

Miller, R. A. and H. Sieg. (1997). “A Microeconometric Comparison of Household Behavior Between Countries”. *Journal of Business and Economic Statistics*. 15(2): 237–254.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. Boston, MA: NBER Books.

Mirrlees, J. A. (1976). “The Optimal Structure of Incentives and Authority within an Organization”. *The Bell Journal of Economics*. 7(1): 105–131.

Mirrlees, J. A. (1999). “The Theory of Moral Hazard and Unobservable Behaviour: Part I”. *The Review of Economic Studies*. 66(1): 3–21.

Murphy, K. J. (2013). “Executive compensation: Where we are, and how we got there”. In: *Handbook of the Economics of Finance*. Vol. 2. Elsevier. 211–356.

Murphy, K. J. and J. Zabojnik. (2004). “CEO Pay and Appointments: A Market-Based Explanation for Recent Trends”. *American Economic Review*. 94(2): 192–196.

Murphy, K. J. and J. Zabojnik. (2006). “Managerial Capital and the Market for CEOs”. *Working Paper*. University of Southern California (USC).

Myerson, R. B. (1982). “Optimal coordination mechanisms in generalized principal–agent problems”. *Journal of Mathematical Economics*. 10(1): 67–81.

Oi, W. Y. and T. L. Idson. (1999). “Firm size and wages”. *Handbook of Labor Economics*. 3: 2165–2214.

Oyer, P., S. Schaefer, N. Bloom, J. V. Reenen, W. B. MacLeod, M. Bertrand, S. E. Black, P. J. Devereux, D. Almond, and J. Currie. (2011). “Personnel Economics: Hiring and Incentives”. In: *Handbook of Labor Economics*. Ed. by O. Ashenfelter and D. Card. Vol. 4. Elsevier. Chap. 20. 1769–1823.

Pakes, A., J. Porter, K. Ho, and J. Ishii. (2005). “The method of moments with inequality constraints”. *Working Paper*. Harvard University.

Pan, Y. (2017). “The determinants and impact of executive-firm matches”. *Management Science*. 63(1): 185–200.

Pan, Y., T. Y. Wang, and M. S. Weisbach. (2015). "Learning about CEO ability and stock return volatility". *The Review of Financial Studies*. 28(6): 1623–1666.

Rey, P. and B. Salanié. (1990). "Long-term, short-term and renegotiation: On the value of commitment in contracting". *Econometrica: Journal of the Econometric Society*: 597–619.

Romano, J. P. and A. M. Shaikh. (2010). "Inference for the identified set in partially identified econometric models". *Econometrica*. 78(1): 169–211.

Rose, N. L. and C. Wolfram. (2002). "Regulating executive pay: Using the tax code to influence chief executive officer compensation". *Journal of Labor Economics*. 20(S2): S138–S175.

Rosen, A. M. (2008). "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities". *Journal of Econometrics*. 146(1): 107–117.

Rosen, S. (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition". *Journal of Political Economy*. 82(1): 34–55.

Roy, A. D. (1951). "Some Thoughts on the Distribution of Earnings". *Oxford Economic Papers*. 3(2): 135–146.

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

Sanders, C. (2013). "Skill Uncertainty, Skill Accumulation and Occupational Choice".

Sannikov, Y. (2008). "A Continuous-Time Version of the Principal-Agent Problem". *The Review of Economic Studies*. 75(3): 957–984.

Seyhun, H. N. (1986). "Insiders' profits, costs of trading, and market efficiency". *Journal of Financial Economics*. 16(2): 189–212.

Seyhun, H. N. (1992a). "The effectiveness of the insider-trading sanctions". *The Journal of Law and Economics*. 35(1): 149–182.

Seyhun, H. N. (1992b). "Why does aggregate insider trading predict future stock returns?" *The Quarterly Journal of Economics*. 107(4): 1303–1331.

Strebulaev, I. A. and T. M. Whited. (2012). "Dynamic models and structural estimation in corporate finance". *Foundations and Trends® in Finance*. 6(1-2): 1–163.

Taylor, L. A. (2010). "Why are CEOs rarely fired? Evidence from structural estimation". *The Journal of Finance*. 65(6): 2051–2087.

Taylor, L. A. (2013). "CEO wage dynamics: Estimates from a learning model". *Journal of Financial Economics*. 108(1): 79–98.

Terviö, M. (2008). "The difference that CEOs make: An assignment model approach". *The American Economic Review*. 98(3): 642–668.

Williamson, O. E. (1967). "Hierarchical control and optimum firm size". *Journal of Political Economy*. 75(2): 123–138.

Zakolyukina, A. A. (2018). "How common are intentional GAAP violations? Estimates from a dynamic model". *Journal of Accounting Research*. 56(1): 5–44.

Zhang, I. X. (2007). "Economic Consequences of the Sarbanes–Oxley Act of 2002". *Journal of Accounting and Economics*. 44: 74–115.