Why Do Bigger Firms Pay More For Performance?

Contract Incentives versus Market-induced Incentives

Bo Hu*

Job Market Paper

May 9, 2018

Abstract

This paper assesses the impact of the frictional managerial labor market on managers' compensation growth, career concerns and monetary incentives. I document a new fact that executives in bigger firms are rewarded with higher monetary incentives compared to those who are equally compensated but work in smaller firms. To explain this fact, I develop a novel model featuring direct competition between heterogeneous firms for managerial skills. This simultaneously generates an expected growth in total compensation, and market-induced incentives that substitute the monetary incentives in an optimal contract. Because of the size premium in compensation growth, managers in bigger firms experience lower market-induced incentives. As a substitute, more incentive pays are required in bigger firms. The model is estimated on the data of executives in S&P 1500 firms. I use the model to evaluate several regulations that have been proposed or implemented.

Key Words: executive compensation, firm-size pay premium, dynamic contract, moral hazard, search frictions

(The latest version can be downloaded here.)

*Department of Economics, Vrije University Amsterdam, and Tinbergen Institute, De Boelelaan 1105, 1081 HV Amsterdam. Email: huboshufe@gmail.com. Website: bohuecon.github.io. I am greatly indebted to my PhD supervisors Makoto Watanabe and Pieter Gautier for their guidance and encouragement. All errors are my responsibility.

1 Introduction

Markets forces have increasingly become an important factor in executives incentive package design.[1] According to Apple Inc.'s 2016 proxy statement, "experienced personnel in the technology industry are in high demand, and competition for executive talent is intense." Under this guidance, their executives' contract incentives are designed "to *attract and retain* a talented executive team and align executives interests with those of shareholders". Similarly, the number one compensation philosophy for named executive officers in Amazon is "to *attract and retain* the highest caliber employees by providing above industry-average compensation assuming stock price performance". These examples highlight the impact of market competition on executives' contract incentives. This paper aims to understand the underlying mechanism by linking contract incentives to the "market-induced incentives". I argue that market competition not only revises the participation constraints of executives but also modifies the incentive compatibility constraints.

The model is not merely descriptive, it seeks to explain some stylized facts firstly documented in the paper. It is well known executive compensations are closely tied to firm performance in the form of options and stocks. Called the incentive compensation, it accounts for about 70% of the total compensation in *S&P* 1500 firms. A salient feature of incentive pays is that they increase with firm size. I find such size premium continues to hold after controlling for executives' total compensation or firm related wealth. That is, compared to their equally wealthy counterparts in smaller firms, executives in bigger firms demand higher incentive rewards. This new fact deserves an explanation and understanding the underlying mechanism contributes to the current debate about regulations on executive compensations.

My explanation starts with the simple observation that executives are not only motivated by the incentives offered in the contract, but also motivated by the perspective for higher rewards in the future, called the *market-induced incentives* in this paper. These two sources of incentives substitute each other in the dynamic contract between the firm and the executive. The fact that executives in big and small firms are equally wealthy today does not imply their expected wealth in the future would also be the same. In contrast, firm size represents a "search capital" that executives can explore in the labor market. Consequently, managers in larger firms tend to be wealthier in the future, and by the law of diminishing marginal utility, they are less sensitive to monetary rewards. The market-induced incentives, therefore, are lower for them. As a result, executives in bigger firms require higher incentive pays in the contract.

The heterogeneity in market-induced incentives is driven by the competition for managerial talents in a search market. In the classical career concern models, the labor market is assumed to be frictionless, hence market-induced incentives only depend on planning horizons.[2] In a more realistic search market setting, the market-induced incentives – embedded in wage offers one expect to receive — depend on the position on the job ladder, in particular, the firm size of the current job. While the intuition would hold in any type of search models, I employ the sequential auction framework (Postel-Vinay and Robin 2002) which allows firms to counter outside offers. It is especially relevant because this type of negotiations usually happens in high skills markets such as the academics market as well as the executives market which is the target of this paper.

In this framework, outside offers are used to trigger renegotiations with the current firm, where the current and the outside firms bid sequentially to hire the executive. Essentially they enter into a Bertrand competition. The maximum compensation firms are willing to bid depends on the productivity of the executive. The productivity is persistent, and it depends on his effort in the past. Ex ante, the bidding scheme is an incentive scheme —- the executive works harder in order to increase his expected productivity so that firms are willing to bid more. Because bigger firms are able to bid higher, by the law of diminishing marginal utilities, the market-induced incentives decrease in firm size. Therefore, small firms can make use of market-induced incentives while big firms have to provide more contract incentives.

Formally, I consider a dynamic contracting problem in which a risk-neutral firm hires a risk-averse agent/executive whose productivity is observable and persistent over time. The evolution of the executive's productivity depends on his effort and exogenous, idiosyncratic shocks, both of which are unobservable. That is, there is moral hazard. The output is the executive's productivity scaled by the firm size. To explore the interaction of contract and market-induced incentives, I embed this contracting problem into an equilibrium search model.[3]Executives search on-the-job, and use outside offers to renegotiate with the current firm. Hence, the contract is subject to the two-sided limited commitments — executives and firms can terminate the relationship upon receiving better outside values. The optimal contract exhibits *memory* when the outside offer is not relevant and exhibits the so called *amnesia* when the outside offers are used to renegotiate with the current firm. I further show that the market-induced incentives decrease with firm size when the adjusted concavity measured by relative risk aversion is bigger than 1. This is a very weak requirement in this context.

Solving numerically for this optimal contract becomes difficult in the presence of the incentive compatibility constraint, limited-commitment constraints, together with shocks of large support.[4]The firms problem can be written recursively using promised utility (Spear and Srivastava 1987; Rogerson 1985) but one still needs

[2]Market-induced incentives have been considered as an alternative way to discipline worker's effort since Fama (1980). The concept is similar as career concerns, though the detailed mechanisms can be different in different models, see Grochulski and Zhang (2017) for a discussion.

[3]The real world labor market for executive-level jobs, just as labor markets for ordinary jobs, is full of informational and searching frictions. On the one hand, firms who have executive-level vacancies usually hire head hunters to find candidates. Once contact with the candidate is established, numerous ways including but not restricted to talks and interviews with the candidate and his/her previous colleagues are used to figure out whether the candidate is suitable. On the other hand, top managerial positions are limited, and can only be available upon the current executive leaves; at the same time suitable manager candidates are also limited, and can only be hired after negotiation and competing with the current employer. These are the reasons to describe the executive labor market as a search market.

to solve for the promised value in each state of the world in the next period. Following Marcet and Marimon (2017), Mele (2014), Farhi and Werning (2013), Lamadon (2016), I address this issue using the recursive Lagrangian approach. The Lagrangian multiplier represents the weight of the executive in a constructed Pareto problem, and it keeps track of the incentive compatibility constraint, limited commitment constraints and job-to-job transitions.

I bring my model to the ExecuComp dataset of $S\&P$ 1500 U.S. firms from 1992 to 2016. The sample includes top 5 to 8 executives in each firm year. I supplement the ExecuComp dataset with BoardEX dataset to identify the precise spell start and end dates, and whether each spell ends in a job-to-job transition. I provide direct evidence that executives do transit between firms, they have roughly the same job offer arrival rate and they are more likely to transit to bigger firms. I estimate the model's parameters by matching the process of firm profitabilities, executive job turnovers, sensitivity of executive wealth to performance in the data, etc. I find that the model quantitatively captures the pattern that bigger firm pay executives more for performance. I use the model to evaluate several regulations on executive pays that have been proposed or implemented.

The rest of the introduction is a review on related literature. In Section 2, I present the motivating facts. In section 3, I present my model. I characterize the optimal contract and how incentives change with firm size in Section 4. In Section 5, I present the data and estimation. Section 6 decomposes the total incentives based on the estimates. I also study several counterfactual scenarios to evaluate regulations on executive compensation. And section 7 concludes.

Related Literature

This paper contributes to the literature that explains the scaling of incentive pays with firm size. One line of research starts from Gabaix and Landier (2008) and Tervio (2008), where competitive assignment models of the managerial labor market, absent an agency problem, are presented to explain why total compensation increases with firm size. By assuming CEO's effort has a multiplicative effect on both CEO utility and firm value, Edmans et al. (2009) embed a moral hazard problem into the competitive assignment model. The model quantitatively generates predictions on wealth performance sensitivities that are consistent with the data. Edmans and Gabaix (2011) extend the model further to risk averse executives. This line of explanation relies on that total compensation increase with firm size, and it can not explain why after controlling for total compensation, the incentive pays still increase with firm size, which is the key question of this paper.

In another line of the literature, Margiotta and Miller (2000) derive and estimate a multi-period principal-agent model with moral hazard. Based on this model, Gayle and Miller (2009) show that large firms face a more severe moral hazard problem, hence higher equity incentives are necessary in order to satisfy the incentive compatibility condition. Gayle et al. (2015)) embed the model of Margiotta and Miller (2000) into a Roy model with human capital accumulations, and they find that the quality of the signal is unambiguously poorer in larger firms, and that explains the most of the pay differential between small and big firms. Compared to these papers, my model highlights the role of a frictional labor market with on-the-job search, which generate cross-sectional variations in market-induced incentives. In this way, my explanation of incentive pays does not rely on the heterogeneity of the moral hazard problems across firms, though this heterogeneity can be easily added. For example, I can assume the effort cost or the hazard ratio of working versus shirking changes with firm size in a proper parametric form. Yet, when these heterogeneities are included, how to identify the two sources of variations would be non-trivial. This is left for future research.

In terms of modeling, this paper builds on and links two strands of literature. One strand is the extensive literature on optimal long-term contracts with private information and(or) commitment frictions, e.g., Townsend (1982), Rogerson (1985), Spear and Srivastava (1987), Phelan and Townsend (1991), Harris and Holmstrom (1982), Thomas and Worrall (1990) and Phelan (1995). Compare to this literature, I model and add market-induced incentives in the incentive compatibility constraint and analyze how this new source of incentives influence the optimal contract.

Another important strand of literature uses structural search models to evaluate wage dispersions. For example, Postel-Vinay and Robin (2002), Cahuc et al. (2006), Lise et al. (2016) among others estimate models with risk-neutral workers and sequential auctions. Compare to this literature, I add a dynamic moral hazard problem which allows me to understand how market frictions influence a long-term contract. The extreme case in my model that firms with size below a threshold only pay a flat wage roughly corresponds to this type of models.[5]

This paper is also closely related to Abrahám et al. (2016), which aims to explain wage inequality in general labor market by combining repeated moral hazard and on-the-job search. Other than the differences in topics, there is a critical difference that distinguishes my model from theirs: the productivity (or output) of agents is persistent in my model while is independent in their model, and therefore in my model working hard today rewards the agent in the market tomorrow. This is also where my model is linked to the literature of career concerns (Holmström 1999, Gibbons and Murphy 1992). In the career concern models, the workers productivity level is persistent, yet the market needs to learn it. By exerting effort, the worker can manipulate the market belief and increase the spot wage he receives. In my model, productivity is observed once the executive and the firm meet. Yet due to search frictions, the reward is postponed to the future. Career concern models usually focus on the compensation difference along the time dimension while my focus is cross-sectional. Therefore, I need to model the labor market more realistically as a search market.

[5] In my model, in some parameter set, firms smaller than a threshold will only pay a flat wage and all incentives come from the market. This case, though not very relevant empirically for the market that I am looking at (almost all executives in my sample are provided with some incentive schemes), it is consistent with other theoretical work such a Grochulski and Zhang (2017). This paves the way to study not only top managers' incentive compensation, but also that of mid-level managers and even rank-and-file employees, as suggested by Edmans et al. (2017).

To summarize, the main contribution of this paper is to develop a theory of optimal contract that incorporates a frictional labor market, and apply the theory to quantitatively assess the impact of market competition on contract incentives.

2 Motivating Facts

The analysis of this paper is motivated by two sets of facts. The first piece of fact states that contract incentives increase with firm size even after controlling for total compensations. This serves as a puzzle that urges an explanation by my model. The second set of facts contains some direct evidence in executives' job mobility, motivating the modeling of a frictional managerial labor market. For the first fact, the standard ExecuComp dataset and the contract incentives calculated by Coles et al. (2006) and Coles et al. (2013) are used. For the second set of facts, I supplement ExecuComp with information from the database BoardEX. I only report the main findings here, and postpone the data details and robust checks in later sections and Appendix C.

Contract Incentives

The fact that incentive pays is higher in bigger firms is well documented in the literature and is replicated in table 1 column (1). In the regression, the dependent variable is the log of delta, defined as the dollar change in firm related wealth for a percentage change in firm value. It is a standard measurement of incentives in contract, also known as the "dollar-percentage incentive" or "wealth-performance sensitivity". The independent variable firm size is measured by market capitalization. It shows for 1% increase in firm size, the incentives increase by 0.6%.[6] However, if we plot the magnitude of log(delta) over two dimensions, the total compensation and the firm size, as in figure 1, and compare executives with the same total compensation, still those in big firms get higher contract incentives.[7]

This finding is confirmed in regression analysis. The positive correlation between incentive pays and firm size exists and is significant after controlling for total compensation in various forms, as shown in table 1 from column (2) to column (4). For 1% increase in the firm scale, the incentive measured by delta increases by 0.3%. The result holds after further controlling for age and education, or use alternative proxies for firm size such as $\log(sales)$, or restrict the sample to only CEOs, or restrict the sample only top 200/500/1000 compensated CEOs. This new fact deserves explanations.

[6] Current literature tend to explain this fact by differences in total compensations. Because big firm executives in general have higher total compensations, highly compensated executives require higher incentive pays in dollars to get the same incentives in utiles. Edmans et al.'s (2009) model, for example, is in this spirit. Their model assumes a multiplicative utility for executives. This implies the more executives are paid, the more incentive pays are required to induce their effort.

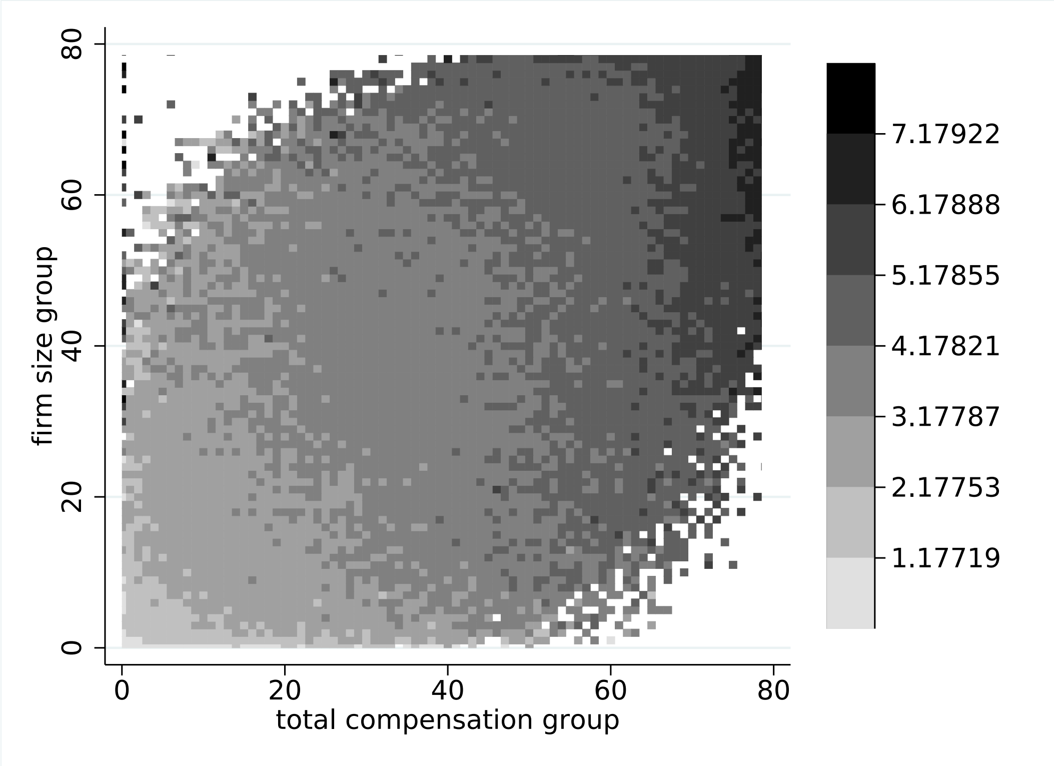

Figure 1: $\log(delta)$ over firm size and total compensation

Note: This is a heatmap of $\log(delta)$ on total compensation and firm size. $delta$ is the wealth-performance sensitivity defined as the dollar change in firm related wealth for a percentage change in firm value. The total compensation is the sum of salary and bonus, the value of restricted stocks and options granted, and value of retirement and long-term compensation schemes. It is the variable `tdc1` in ExecuComp dataset. The firm size is the market capitalization by the end of the fisical year, calculated by `csho` $\times$ `prcc_f` where `csho` is the common shares outstanding and `prcc_f` is the close price by fiscal year. These variables will be used throughout the paper. I divide the whole sample into 80×80 cells according to the total compensation and firm size, and compute the mean of $\log(delta)$ within each cell.

Executive Labor Markets

Job search among executives becomes especially prominent in recent decades. The outside hires for all CEO replacements accounted for only 15 percent in 1970's, rising to more than 26 percent in the 1990's. From 2001 to 2003, large employers in the US used executive search firms to fill 54 percent of jobs paying above $\$150,000$ (IACPR 2003). In my dataset, I observe stable job-to-job transitions, ranges from 3% to 8% from year to year. In particular, it becomes higher after the financial crisis.

There are two features in the data that support a job ladder modeling. First, among the job-to-job transitions that I can observe the size of both the original firm and the target firm, there are significantly larger fraction of executives move from small to bigger firms, as shown in table 2. Among those who do move to smaller firms, I find half of those cases are due to a title change from a non-

Table 1: Incentive Pays Increase with Firm Size

	log(*delta*)			
	(1)	(2)	(3)	(4)
log(Firm Size)	0.578***	0.295***	0.274***	0.273***
	(250.03)	(112.20)	(104.10)	(103.68)
log(tdc1)		0.7159***		
		(176.18)		
tdc1 Dummies (50)			Yes	
tdc1 Dummies (100)				Yes
Year FEs	Yes	Yes	Yes	Yes
Industry FEs	Yes	Yes	Yes	Yes
Year × Industry FEs	Yes	Yes	Yes	Yes
Observations	129458	129184	129185	129185

Note: (a) The t statistics are shown in parentheses, and we denote symbols of significance by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The dependent variable is the log of delta. The independent variable is the log of firm size. The key control variable is total compensation. Refer to note of figure 1 for definitions of these variables. All regressions control for year and industrial dummies and their interaction terms. (b) Column (1) is a regression of log(delta) on log(firm size), which replicates the cross-sectional regression in the literature. From column (2) to column (4), I add log(tdc1), tdc1 evenly grouped into 50 and 100 categories (then transformed into dummies), respectively.

CEO title to a CEO title.[8] The pattern that executives are more likely to transit to bigger firms is also observed across age-groups (table 2) and industries (table 3).

Table 2: Firm Size Changes in Job-to-Job transitions (across age groups)

Age Group	Firm Size Decrease	Firm Size Increase	All
[26, 40)	33	61	94
[40, 50)	259	448	707
[50, 60)	236	394	630
[60, 65)	29	51	80
[65, 70)	8	23	31
[70, 86)	1	4	5
All	681	1229	1910

Second, in a random search model, the observed job-to-job transition rate should decrease with firm size because small firm executives are more likely to accept an offer. To test this, I estimate a Cox model on the duration to job-to-job transitions, controlling for profitability, age, education dummies, executive title dum-

[8] I find 6253 job-to-job transitions from a CompuStat firm, and only 1910 of them that have the size information of original and target firms. The rest firms are private firms and the size information is not disclosed.

Table 3: Firm Size Changes in Job-to-Job Transitions (across Fama & French 12 industries)

Industry	Firm Size Decrease	Firm Size Increase	All
1	12	32	44
2	7	24	31
3	37	59	96
4	13	13	26
5	8	11	19
6	73	104	177
7	4	12	16
8	15	29	44
9	46	111	157
10	27	43	70
11	38	69	107
12	50	52	102

mies, year and industry dummies. Job-to-job transition is defined as the executive leaves the current firm and works in another firm within 30 days. I also construct alternative job-to-job transitions using 15, 60, 90, and 190 days. The result is shown in table 4 column (1). For one percentage increase in firm size, the job-to-job transition hazard rate decreases by 0.04. I add the interaction terms between firm size and age dummies to test if the career concern would decrease as the executive approaches the retire age. The effect is significant as is shown in column (2). The negative correlation between job-to-job transition hazard rate and firm size monotonically decreases with age. I do the same test for alternative job-to-job transition measurement with different day gaps. The result is very robust.

Table 4: Firm Size and EE (30 days)

	(1)	(2)
$\log(\text{Firm Size})$	-0.0416***	-0.0672***
$[30,40) \times \log(\text{Firm Size})$		-0.0792
$[40,45) \times \log(\text{Firm Size})$		-0.00102
$[45,50) \times \log(\text{Firm Size})$		0.00949
$[50,55) \times \log(\text{Firm Size})$		0
$[55,60) \times \log(\text{Firm Size})$		0.123***
$[60,65) \times \log(\text{Firm Size})$		0.180***
$[65,80] \times \log(\text{Firm Size})$		0.118
profit	-0.124****	-0.154***
N	326919	326919

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

3 Model

This section lays out the equilibrium model of the executive labor market.

3.1 Preferences and Technologies

Time is discrete, indexed by t and continues forever. In this economy, there is a fixed measure of individuals who aim at executive level jobs. In the following, I will call individuals who are employed the *executives* or *managers*, and those who are not employed *executive candidates*. These candidates are not necessarily being unemployed from other jobs. They are simply not working as an executive, and aim at finding a managerial job.

Individuals want to maximize expected lifetime utility,

$$\mathbb{E}_0 \Sigma_{t=0}^{\infty} (\beta \times (1-\delta))^t (u(w_t) - c(e_t)),$$

where $\beta \in (0,1)$ is the discount factor, $\delta \in (0,1)$ is the death probability, utility of consumption $u : \mathbb{R} \to \mathbb{R}$ is increasing and concave, $c(\cdot)$ is the dis-utility of effort. Effort e_t takes two values, $e_t \in \{0,1\}$, and cost of effort level 0 is normalized to zero. I denote $c(1)$ by c.

Individuals are heterogeneous in managerial skills. A manager can make use of these skills including his or her own abilities, knowledge base, experiences, and perspectives to increase the productivity of firm. For simplicity, I call them the productivity of the manager which is indexed by $z \in \mathbb{Z} = \{z_1, z_2, ..., z_{n_z}\}$. Individual productivity is observable to himself, to firms that he meets, and can be carried with the individual through job-to-job transitions. Hence, I treat the productivity as general management skills rather than firm-specific skills.[9]The productivity changes over time according to a Markov process $\Gamma_z(z_t, z_{t+1}|e_t)$, which depends on effort e_t. I denote the process by $\Gamma_z(z_t, z_{t+1})$ for $e_t = 1$, and $\Gamma_z^s(|z_t, z_{t+1})$ for $e_t = 0$ (s for shirking).

Once the individual dies, the associated job destroys (if he is matched), and a new executive candidate enters the economy.[10] Executive candidates have constant productivity $z_O \in \mathbb{Z}$, and receive a value W^0. Having a constant productivity during "unemployment phase" is not as restrict as it looks. It merely says when individuals starts their career as an executive, their productivities are drawn from the same distribution $\Gamma_z(z_O, z)$. I further define the likelihood ratio

$$g(z,z') \equiv \frac{\Gamma^s(z,z')}{\Gamma(z,z')}.$$

[9]However, firm-specific skills can be included by a productivity discount upon a job-to-job transition. This is left as a future extension.

[10]In an alternative setting, the executive position continues as a vacancy upon breaking up with the executive. In such case, to endogenize the vacancy value, one need to aggregate number of meetings between vacancies and searching workers by an aggregate matching function. This goes beyond the scope of this paper.

Notice as a likelihood ratio, it satisfies $E[g(z,z')] = \sum_{z'} g(z,z')\Gamma(z,z') = 1$. I assume the Markovian processes are such that

- Taking effort delivers a higher expected productivity:
$$E_\Gamma[z' g(z,z')] < E_\Gamma[z'];$$
- Taking effort is more likely to deliver a higher productivity (Monotone likelihood ratio property, MLRP):
$$g(z,z') \text{ is non-increasing in } z'.$$

On the other side of the labor market are executive jobs (or firms, represented by its board members). Jobs are characterized by the scale of assets that can be controlled by the manager, called firm size, denoted by $s \in \mathbb{S} = [\underline{s}, \bar{s}]$. In empirical sections, I use the firms' market capitalization as a proxy for firm size. Firm size is observable to all matched executives and it is permanent and exogenous.[11]

Executives and jobs are imperfectly informed about executive types and the location of the job, which precludes the optimal assignments as in Gabaix and Landier (2008). Yet, when two agents meet, both are informed about each other's types immediately. Search is random, executives and executive candidates all sample from the same job offer distribution $F(s)$. Executive candidates meet firms with probability λ_0, while on-the-job executives meet other firms with probability λ_1. I treat these parameters exogenous, so we are in a partial equilibrium.

A match between a worker of productivity z and a firm of size s produces a flow of output $f(s,z) = y(s)z$, where $y(s)$ is strictly increasing and concave in s. Later whenever it is needed, I use a particular function form $f(s,z) = \alpha sz$, where the constant $\alpha \in (0,1)$ is the upper bound of the share fraction of the executive. This function form entails that executive's efforts rolled out across the entire firm up to a scale of α and thus have a greater effect in a larger company.[12]

3.2 Timing

An individual is either an employed executive or an executive candidate, and enters each period with the last period productivity, denoted by z. The executive

[11] From the view of labor search literature, one could interpret firm size here as "the productivity of the job" or "firm type". Instead of using total number of employees, I use total asset value as a proxy for firm size since the performance of the firm is usually measured by return on assets. If one interpret firm size as the total number of employees, then it can be endogenized by modeling the labor market of normal workers.

[12] There has been a discussion on the appropriate production function form for executives. Taking s as firm size, and z as the executive's per unit contribution to shareholder values. An additive production function such as $f(s,z) = s + z$ implies the effect of executives on firm value is independent of firm size. This specification is appropriate for a perk consumption. An multiplicative production function such as $f(s,z) = y(s)z$ where $y(s)$ is increasing in s, is appropriate for executives' actions that can roll out across the entire firm to some extent and thus have a greater effect in a larger company. The latter is the function form adopted here.

candidates have a constant productivity $z = z_o$, and with probability λ_0 to meet a firm of size s drawn from the offer distribution $F(s)$. The matched individual and firm bargain on a contract. The firm will offer an contract with a value of least W^0. The individual then enters next period as an employed executive. What is essential for our analysis is the value W^0.

An employed executive starts the period with his current productivity z.

1. **Compensation:** The firm will first pay a wage w for this period, in accordance with the contract.

2. **Production:** Then the executives enter the production phase where he chooses an effort level $e \in \{0, 1\}$. His productivity z' is then realized according to $\Gamma(z, z'|e)$, which depends on his effort choice. The firm only observes the output $y(z, s)$, not the executive's effort. That is, there is moral hazard.

3. **Leave/Die:** After the production phase, with probability δ he dies, otherwise the executive can choose to leave the job voluntarily.

4. **Sequential Auction :** The executive has the a probability λ_1 to sample a job offer of firm size $s' \sim F(s')$ (s' denotes poaching firm size and s is the current firm size.) The current firm can counter the outside offer and make a take-it-or-leave-it contract offer to the executive. The value of the contract to the executive therefore will be determined by a sequential auction between the current and poaching firms as described below. The executive will enter the next period as employed by either current firm or the poaching firm.

The timing line is shown in figure 2.

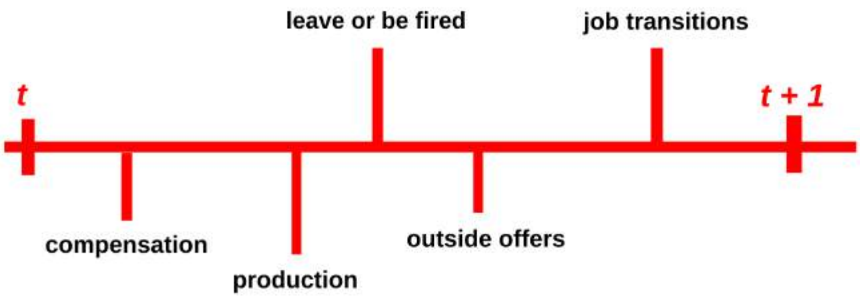

Figure 2: Timing

3.3 Information Structure and Contracts

Call $h_t = (z'_t, s_t, s'_t)$ the state of the match at period t, where z'_t is the realized productivity at t, s_t is the current firm size, s'_t is the size of poaching firm and denote $s'_t = s_o$ if there is no poaching firm. The history of productivity and poaching firms $h^t = (h_1, h_2, .., h_t)$ is common knowledge to the manager and the firm, and is fully contractible. While productivity is contractible, a manager's effort is not and need to be induced by incentives.

I allow the history of poaching firms to be contractible. The advantage is in the optimal contract firms can stipulate how to counter outside offers and they are able to keep the promise. Countering outside offers must be the optimal decision of the firm (or subgame perfect in terms of game terminology). For this, it is necessary to allow firms and managers to have limited commitment, and to choose terminate the contract for a negative surplus. More precisely, a firm will terminate the contract upon negative profits. This happens when the outside offer is so high that a counter offer would break the firm's participation constraint. On the other hand, the manager will terminate the relationship if the current firm can not provide more than the outside value, be the non-executive job value W^0 or the offer of a poaching firm. In the former case, the manager is fired. And in the latter case, there is a job to job transition.

Given the information structure, a feasible contract is a plan that stipulates the compensation $w_t(h^t)$, the recommended effort for the manager $e_t(h^{t-1}) \in \{0,1\}$ and whether to terminate the contract $I_t(h^t)$ at every future history, represented by

$$\{e_t(h^{t-1}), w_t(h^{t-1}), I_t(h^t)\}_{t=0}^{\infty},$$

that satisfied the participation constraints of both sides and incentive compatibility constraint.

I further simplify the contracting problem by two assumptions. Firstly, I assume taking effort $e = 1$ is always optimal. This is in accordance with the fact that almost all executives in my data are provided with some incentive package and holds in all the numerical exercises below. Secondly, I assume a reasonable support of productivity z such that the value of a job is always positive. As a result, the only possibility of terminating the contract is when the poaching firm is larger. In this way, firing is excluded. Thus, $I_t(h^t) = 1$ exclusively implies there is a job-to-job transition and is exogenously determined by the job offer arrival rate and the comparison of s and s'.[13]

In the following, I use the executive's beginning-of-period expected utility, denoted by V, as a co-state variable to summarize the history of productivities and outside offers. A dynamic contract, defined recursively, is

$$\sigma \equiv \{e(V), w(V), W(z', V) | z' \in \mathbb{Z} \text{ and } V \in \Phi\},$$

where e is the effort level suggested by the contract, w is the compensation, and W is the promised value given for productivity z', and Φ is the set of feasible and incentive compatible expected utilities that can be derived following Abreu et al. (1990).[14]

[13] If we allow the domain of z be extended such that for some z the profit is negative even when the firm only offers the non-executive job value W^0, then firing happens. This is an interesting extension that will be analyzed later.

[14] Promised utilities as co-states have been used extensively in models with incentive or participation constraints. Among others, Phelan and Townsend (1991) studied a model of risk-sharing

3.4 The Sequential Auction

Before delving into the contracting problem, I first characterize what happens following an offer from a poaching firm. Here I employ the sequential auction framework (Postel-Vinay and Robin 2002, Cahuc et al. 2006). Let $\Pi(z, s, V)$ denote the discounted profit of a firm with size s, executive productivity of z, and promised value to the executive V. The maximum the firm would like to give to the executive, the maximum bidding value $\overline{W}(z, s)$, is defined by

$$\Pi(z, s, \overline{W}) = 0.$$

The firm would rather fire the executive (and the vacancy value is normalized to 0) if he demands a value higher than $\overline{W}$. I also define $\overline{W}(z, s^o) = W^0$. This is simple: when there is no outside offer, manager's outside value is simply W^0.

The sequential auction works as follows. When a size s firm's executive receives an outside offer from a size s' firm, both firms enter a Bertrand competition won by the largest firm. Consider this sort of auction over an executive z by a firm of size s and one of size $s' > s$. Since it is willing to extract a positive marginal profit out of every match, the best the firm s can do is to provide a promised utility $\overline{W}(z, s)$. Accordingly, the executive accepts to move to a potentially better match with a firm of size s' if the latter offers at least the $\overline{W}(z, s)$. Any less generous offer on the part of the size s' firm is successfully countered by the size s firm.

Now, if s' is less than s, then $\overline{W}(z, s) > \overline{W}(z, s')$, in which case the size s' firm will never raise its offer up to this level. Rather, the executive will stay at his current firm, and be promoted to the continuation value $\overline{W}(z, s')$ that makes him indifferent between staying and joining the size s' firm. The above argument defines outside values of the executive contingent on the state (z, s'), which then serves as the participation constraints in the contracting problem.

What distinguishes this model from the original sequential auction framework is here the wage is not flat. Firms compete on a sequence of wages contingent on all possible future histories, as summarized by $\overline{W}(z, s)$. More importantly, it brings a new source of incentives into the contracting problem. Firms appreciate higher productivities, and are willing to bid more. The *bidding frontier* $\overline{W}(z, s)$ increases in z. The sequential auction therefore begets incentives for managers' effort: if working hard today is not only an input into current production but also an investment in the (inalienable and transferable) human capital, then it is intuitive that the objectives of the firm and the worker become better aligned and the need for short-term compensation incentives decreases.

with incentive constraints, Kocherlakota (1996) analyzed the risk-sharing model with the PC described above, Hopenhayn and Nicolini (1997) studied a model of unemployment insurance and Alvarez and Jermann (2000) studied a decentralized version of the above risk-sharing model with debt constraints.

3.5 The Contracting Problem

We are interested in constrained efficient contracts, that is to say, contracts that satisfy participation and incentive compatibility constraints, and are not Pareto dominated by any other constrained contracts. I now describe the firm's problem in terms of promised utilities. The firm chooses a wage w, a set of promised values $W(z', s')$ depending on the state z' and s'. For the ease of notations, I denote $\tilde{\beta} = \tilde{\beta}$. I write the mixture distribution of outside offers as

$$\tilde{F}(s) = \mathbb{I}(s = s^o)(1 - \lambda_1) + \mathbb{I}(s \neq s^o)\lambda_1 F(s).$$

The expected profit of a match to the firm can be expressed recursively as

$$\Pi(z, s, V) = \max_{w, W(z', s')} \sum_{z' \in \mathbb{Z}} \left[y(s)z' - w + \tilde{\beta} \sum_{s' \in \mathbb{S}} \Pi(z', s, W(z', s'))\tilde{F}(s') \right] \Gamma(z, z'). \tag{BE-F}$$

subject to the promise keeping constraint,

$$V = u(w) - c + \tilde{\beta} \sum_{z' \in \mathbb{Z}} \sum_{s' \in \mathbb{S}} W(z', s')\tilde{F}(s')\Gamma(z, z'), \tag{PKC}$$

the incentive compatibility constraint,

$$\tilde{\beta} \sum_{z' \in \mathbb{Z}} \sum_{s' \in \mathbb{S}} W(z', s')\tilde{F}(s')(1 - g(z, z'))\Gamma(z, z') \geq c. \tag{IC}$$

and the participation constraints of the manager and the firm,

$$W(z', s') \geq \min\{\overline{W}(z', s'), \overline{W}(z', s)\} \tag{PC-E}$$
$$W(z', s') \leq \overline{W}(z', s). \tag{PC-F}$$

The firm chooses the current period wage w and the promised utilities $W(z', s')$ for each state (z', s') at the beginning of tomorrow. These choice variables maximize expected profits subject to the promise keeping constraint (PKC), incentive compatibility constraint (IC), and the participation constraints of the firm and the executive (PC-E and PC-F).

The objective function (BE-F) is the Bellman equation of the firm. The promise keeping constraint makes sure that the choices of the firm honors the promise made in previous periods to deliver the value V to the executive, and V contains all the relevant information in the history. The right hand side of the constraint is the lifetime utility of the executive given the choices made by the firm. (PKC) is also the Bellman equation of an executive with state (z, s, V).

The incentive compatibility of the executive differentiates itself from the promise keeping by the term $(1 - g(z, z'))$. It merely says the continue value of taking effort is higher than not taking effort. This create incentives for the manager to pursue the shareholders' interests rather than his own. As mentioned above, in my data sample, over 99% observations have positive incentive pays, hence, I assume that $e = 1$ is optimal choice of the firm.

Finally, the participation constraints are stated in (PC-E) and (PC-F). The firm commits to the relationship as long as the promised value for the future is not more than $\overline{W}(z', s)$. The sequential auction pins down an outside value of the manager, $\min\{\overline{W}(z', s'), \overline{W}(z', s)\}$.

Before turning to characterize the optimal contract, I define the equilibrium.

3.6 The Equilibrium

An equilibrium is the executive unemployment value W^0, the value function of employed executives W satisfies (PKC), the profit function of the firms Π and an optimal contract policy $\sigma = \{w, e, W(z')\}$ for $z' \in \mathbb{Z}$ that solves the contracting problem (BE-F) with associated constraints (PKC), (IC), (PC-E) and (PC-F), the stochastic process of executive productivity Γ follows the optimal effort choice and a distribution of executives across employment states evolving according to flow equations.

The proof of the existence of the equilibrium is an exercise applying Schauder's fixed point theorem and is detailed in Appendix A.

Proposition 1. *The equilibrium exists.*

4 Contract Characterization

We now go on to derive a characterization of the optimal contract.

Proposition 2. $\Pi(z, s, W)$ *is continuous differentiable, decreasing and concave in* W, *and increasing in* z *and* s. *An optimal contract evolves according to the following updating rule. Given the beginning of the period state* (z, s, V), *the current period compensation is given by* w,

$$\frac{\partial \Pi(z, s, V)}{\partial V} = -\frac{1}{u'(w)}, \tag{1}$$

and the continuation utility follows

$$W(z', s') = \begin{cases} \overline{W}(z', s) & \text{if } \overline{W}(z', s') \geq \overline{W}(z', s) \\ \overline{W}(z', s') & \text{if } \overline{W}(z', s) > \overline{W}(z', s') > W(z') \\ W(z') & \text{if } \overline{W}(z', s) > W(z') \geq \overline{W}(z', s') \end{cases} \tag{2}$$

where $W(z')$ *satisfies*

$$\frac{\partial \Pi(z', s, W(z'))}{\partial W(z')} - \frac{\partial \Pi(z, s, V)}{\partial V} = -\mu(1 - g(z, z')). \tag{3}$$

Proof. See Appendix A. □

Proposition 2 says, abstract from the participation constraints, an optimal contract inherits the essential properties of classical infinite repeated moral hazard model. Equation (1) says the current period compensation w is directed linked to the promised continuation utility V by equating the principal's and agent's marginal rates of substitution between present and future compensations. Equation (3) says, abstract from participation constraints, the continuation utility $W(z')$ only changes to motivate manager's effort. In the extreme case that $\mu = 0$ (μ is the multiplier of IC), $W(z') = V$ keeps constant. Generally, higher V induces a higher $W(z')$. Thus, an optimal dynamic contract has memory.

However, when outside offers are realized that cause the participation constraint to bind, the contract disposes of all history dependence and makes the continuation value depend only on the current state (z, s, s'). This is what Kocherlakota (1996) calls *amnesia*. More precisely, when the outside firm is larger $s' \geq s$, the continuation value equals to the bidding frontier of the current firm $W(z', s') = \overline{W}(z', s)$; when the outside firm is smaller $s' < s$, the continuation value depends on whether the bidding frontier of the outside firm $\overline{W}(z', s')$ can improve upon $W(z')$. Notice these results are not by assumption, but are stipulated in the optimal contract.[15]

What need to be pointed out is the amnesia here is not complete. That is, although $\overline{W}$ does not depend on the previously promised utility V, it does depend on the manager's productivity z'. Since productivity is persistent, taking effort now has influence on the future outside values. This has consequence on the moral hazard problem. It will become clear if we substitute the optimal continuation value into IC constraint. For that, I define two sets of outside firm size s'.[16]

$$\mathcal{M}_1(s) \equiv \{s' \in \mathbb{S} | s' > s\},$$
$$\mathcal{M}_2(z, s, W) \equiv \{s' \in \mathbb{S} | \overline{W}(z, s) > \overline{W}(z, s'), W < \overline{W}(z, s')\}.$$

For $s' \in \mathcal{M}_1$, the executive transits to the outside firm and receives the full surplus of his former job at $\overline{W}(z, s)$. For $s' \in \mathcal{M}_2$, the executive stays with his current firm and uses the outside offer to renegotiate up to $\overline{W}(z, s')$. We can now rewrite IC constraint as follows.

$$\tilde{\beta} \sum_{z'} \Bigg[\lambda_1 \sum_{s' \in \mathcal{M}_1} F(s') \overline{W}(z', s) + \lambda_1 \sum_{s' \in \mathcal{M}_2} \overline{W}(z', s') F(s') \\ + \Big(1 - \lambda_1 \sum_{s' \in \mathcal{M}_1 \cup \mathcal{M}_2} F(s')\Big) W(z')(1 - g(z, z')) \Bigg] \Gamma(z, z') \geq c. \tag{IC$'$}$$

There are two observations. First, compared to an incentive compatibility in a standard moral hazard problem, (IC$'$) has an incentive part offered by the con-

[15]These results build on and extend the dynamic limited commitment literature, pioneered by Thomas and Worrall (1988) and Kocherlakota (1996), and related literature in labor search such as Lentz (2014).

[16]We can similarly rewrite the Bellman equations of firms and managers using the optimal

tract, the *contract incentives*, denoted by Ξ_c,

$$\Xi_c \equiv \tilde{\beta}\Big(1-\lambda_1 \sum_{\mathcal{M}_1\cup\mathcal{M}_2} F(s')\Big)\sum_{z'} W(z')(1-g(z,z'))\Gamma(z,z'), \tag{4}$$

it also has an incentive part due to the sequential auction in the labor market, the *market-induced incentives*, denoted by $\Xi_m(s)$

$$\Xi_m(s) \equiv \tilde{\beta}\lambda_1 \sum_{z'}\left[\sum_{s'\in\mathcal{M}_1} F(s')\overline{W}(z',s) + \sum_{s'\in\mathcal{M}_2} \overline{W}(z',s')F(s')\right](1-g(z,z'))\Gamma(z,z'). \tag{5}$$

This opens the channel that market-induced incentives substitute the contract incentives. Second, $\Xi_m(s)$ depends on the firm size even though the moral hazard problem fundamentally doesn't. I will continue to show that $\Xi_m(s)$ decreases in s. Therefore, in the optimal contract, Ξ_c increases in s. That is, larger firms pay more for performance.

Intuitively, the steeper $\overline{W}(z,s)$ is with respect to z, the larger the incentives are to induce effort. This is rigorously proved in the appendix. Here I take this as given and proceed to a heuristic derivation to show that

$$\frac{\Delta\overline{W}(z,s_1)}{\Delta z} > \frac{\Delta\overline{W}(z,s_2)}{\Delta z} \text{ for } s_2 > s_1.$$

That is the executive in larger firm s_2 receives less market-induced incentives. It follows that

$$\frac{\Delta\overline{W}(z,s)}{\Delta z} = -\frac{\Delta\Pi(z,s,\overline{W})/\Delta z}{\Delta\Pi(z,s,\overline{W})/\overline{W}} = \frac{\tilde{\alpha}\times s}{1/u'(\overline{w})},$$

where $\overline{w}$ is the per-period compensation (wage) corresponds to $\overline{W}$. The first equality follows from implicit differentiation. $\Delta\Pi(z,s,\overline{W})/\Delta z = \tilde{\alpha}\times s$ because keeping the promised value, all increasing output is accrued to the firm. $\tilde{\alpha} = \alpha\,\times$ adjust-factor adjusts for chance that the manager will leave the firm and the job is destructed. $\Delta\Pi(z,s,\overline{W})/\overline{W} = -1/u'(\overline{w})$ follows directly from (1) in proposition 2. It is clear that there are two opposing effect of s. While larger firms are

continuation value, and these equations are consistent with Postel-Vinay and Robin (2002).

$$\begin{aligned}\Pi(z,s,V) = \max_{w,W(z')} \sum_{z'}\Bigg[&y(s)z' - w + \tilde{\beta}\Bigg(\lambda_1 \sum_{s'\in\mathcal{M}_2} F(s')\Pi_1(z',s,\overline{W}(z',s')\\ &+\Big(1-\lambda_1\sum_{s'\in\mathcal{M}_1\cup\mathcal{M}_2} F(s')\Big)\Pi_1(z',s,W(z'))\Bigg)\Bigg].\end{aligned} \tag{BE-F'}$$

$$\begin{aligned}V = u(w) - c + \tilde{\beta}\sum_{z'}\Bigg[&\lambda_1\sum_{s'\in\mathcal{M}_1} F(s')\overline{W}(z',s) + \lambda_1 \sum_{s'\in\mathcal{M}_2} F(s')\overline{W}(z',s')\\ &+\Big(1-\lambda_1\sum_{s'\in\mathcal{M}_1\cup\mathcal{M}_2} F(s')\Big)W(z')\Bigg]\Gamma(z,z'),\end{aligned} \tag{PKC'}$$

able to provide more for a marginal increase in z in terms of dollars (shown in $\tilde{\alpha}s$), the incentives in terms of utilities can actually be lower because the marginal utility for extra one dollar is lower now ($\overline{w}$ increases in s making $u'(\overline{w})$ lower). This happens when the utility function has enough concavity.

Proposition 3 (Contract incentive and Firm size)**.** *Suppose the effort cost $c \leq \bar{c}$ for some $\bar{c} > 0$, then the market-induced incentives $\Xi_m(s)$ decreases with firm size s if*

$$-\frac{wu''(w)}{u'(w)} > 1.$$

Proof. See Appendix A. □

It turns out we have a very neat characterization. The appropriate concavity measurement is the relative risk aversion. As a result, for a CRRA utility form,

$$u(w) = \frac{w^{(1-\sigma)}}{1-\sigma},$$

the requirement is $\sigma > 1$, which is a very weak requirement in this context. The literature almost always estimate/calibrate a higher σ value. For example, a careful calibration study on CEO's by Hall and Murphy (2000) uses σ between 2 and 3. Using an employer-employee matched data from Sweden for the general labor market, Lamadon (2016) estimates a $\sigma = 1.68$.

4.1 Comparative Statics

Next, I examine the model's predictions about market versus contract incentives and the firm size. These predictions hold for a wide range of plausible parameter values. However, because I solve the model numerically, I do not present these predictions as formal propositions. For these practices, I use a model specification the same as that for the estimation in the next section. In particular, the utility function is assumed as in proposition 3. I simulate the model based on the following parameter values: the discount factor $\beta = 0.9$, job arrival rate $\lambda_1 = 0.16$, effort cost $c = 2.5$, and the relative risk averse $\sigma = 2.75$. These parameter values are close to the empirical estimates in the next section. Then I simulate the model and calculate four statistics as follows.

I measure market-induced incentives directly by Ξ_m as defined in (5). This is precisely how market-induced incentives are defined in the model. I calculate the mean of Ξ_m, and how Ξ_m changes with firm size, i.e. the coefficient $\beta_{\Xi-size}$ in the following regression,

$$\Xi_{mit} = \beta_0 + \beta_{\Xi-size}\log(size_{it}) + \beta_1 \log wage_{it} + \epsilon_{0,it}.$$

I measure contract incentives by the wage-performance sensitivity delta, i.e. the dollar change in wage for a percentage change in productivity. This is consistent with the delta in the data as detailed in table 1 and in the next section. However, delta is only a linear measurement of contract incentives while the optimal contract in my model is nonlinear. That is, for every observation in my simulated dataset, there is a nonlinear optimal contract stating a vector of wages in different state of world. To extract the delta information from a nonlinear contract, I regress the vector of log wages on the vector of productivities, and the coefficient of this regression corresponds exactly to the definition of delta. Again, the regression to define delta is done for each observation. I then calculate the mean of delta, and how delta changes with firm size controlling wage level, i.e. coefficient $\beta_{delta-size}$ in the following regression,

$$delta_{it} = \beta_2 + \beta_{delta-size} \log(size_{it}) + \beta_3 \log wage_{it} + \epsilon_{1,it}.$$

Figure 3 plots means of Ξ_m and delta, coefficient $\beta_{delta-size}$ against λ_1. Not surprisingly, as λ_1 increases, the market becomes less frictional, the competition among firms pushes the compensations and induce higher contract incentives measured by delta. At the same time, a higher λ_1 also brings up the market-induced incentives, so Ξ_m increases. What is interesting is how the two sources of incentives change with firm size. With $\lambda_1 > 0$, delta always increases with firm size, and Ξ_m always decreases with firm size. That is, executives in big firms face more contract incentives and less market-induced incentives. Such pattern is amplified as λ_1 increases.

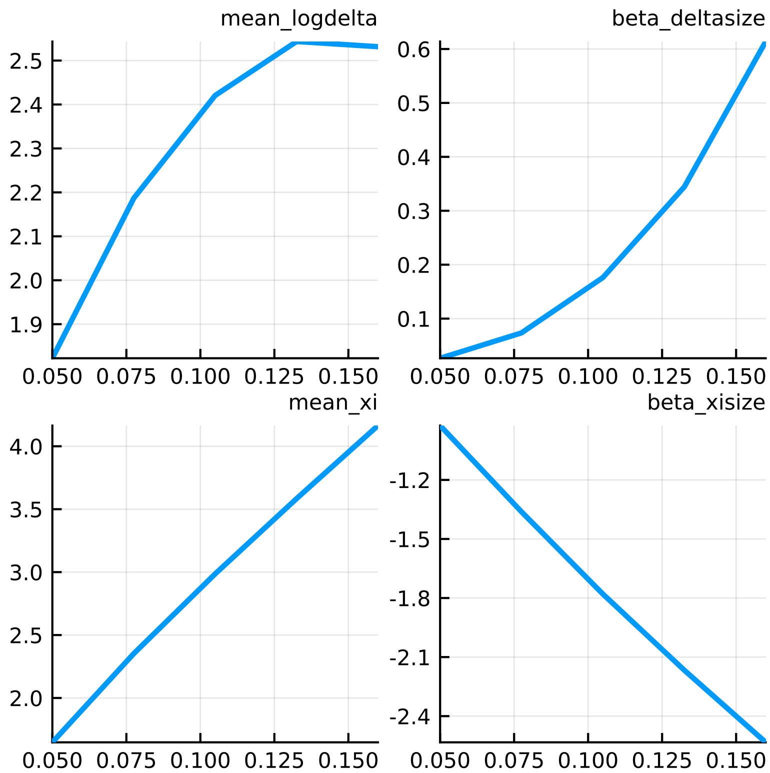

Figure 3: Comparative Statics on λ_1

Figure 4 plots means of Ξ_m and delta, coefficient $\beta_{delta-size}$ against the value of σ. σ can influence both market-induced incentives and contract incentives. On the one hand, a higher σ implies a more concave utility function, thus less variation of wages can generate the same incentive. We observe that delta decreases with σ. On the other hand, the same variation of the bidding frontier is translated into larger market-induced incentives with a higher σ value, making Ξ_m larger. σ also amplifies the differences of market-induced incentives across firm size, therefore, $\beta_{delta-size}$ increases with it.

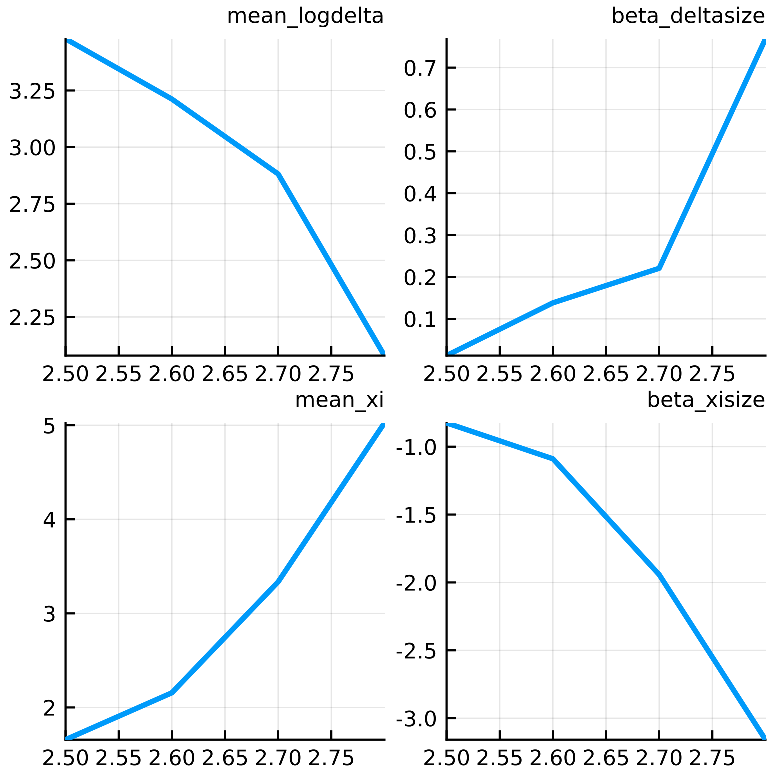

Figure 4: Comparative Statics on σ

Finally, the effort cost c only influences the contract incentives. Therefore, the mean of delta increases with firm size while the mean of Ξ_m and β_{Ξ_m-size} are constant. Yet, a higher c implies a relative lower fraction of market-induced incentives compared to c, we thus observe $\beta_{delta-size}$ increases in c. This figure is shown in the appendix.

5 Estimation

In this section, I bring the model to the data. I first describe how I construct my dataset using information from two databases, and some direct evidence of executives' job-to-job transitions. Then I list model specifications, moments used for estimation and the estimation results. I conclude this section with a comparison of delta in the real data and model-simulated data.

5.1 Data Sources

Data come from ExecuComp (1992 to 2015); Compustat; BoardEX; Coles et al. (2006) and Coles et al. (2013). ExecuComp and Compustat is the standard data source in this literature. They provide rich information on compensations of top five to eight executives in companies included in the S&P 500, S&P 400 MidCap and S&P SmallCap 600 indexes and comprehensive measures of company financial performance. However, this data source does not contain any information of prior employment history. The start and end dates of the current employment ares also not known.

To keep track of the transitions among employment status, I supplement ExecuComp with information from the database BoardEX. The latter contains details of each executive's employment history, including starting and ending dates, firm names and positions. It also provides extra information on education background, social networks, and more accurate records of age. There is no index to directly link ExecuComp and BoardEX. I construct my sample by firstly matching the executives' first, middle and last names, the date of birth and working experiences (in which year the executive worked in which firms). When the three items are consistent between the observations from the two data sources, the executives are identified as the same person. By this way, I am able to identify $33,988$ executives, roughly $2/3$ of the whole ExecuComp sample (from year 1992 to year 2015). The table in the appendix compares the basic descriptive statistics of identified and unidentified executives. There is no obvious sample selection due to my data processing. More details about how I construct the dataset are left in the appendix.

Finally, the data provided by Coles et al. (2006) and Coles et al. (2013) include two important variables on contract incentives: firm related wealth and delta. The former is the dollar value of the executive's stock and option portfolio (in \$000s), and the latter is the Dollar change in wealth associated with a 1% change in the firms stock price (in \$000s).

5.2 Data Description

[A subsection of variables and summary statistics to be added.]

5.3 Numerical Algorithm, Model Specification and Parameters

I estimate the model's parameters using Simulated Methods of Moments. That is, I use a set of moments that are informative for the model's parameters and minimize the distance between data moments and model-generated moments. My moments are partly coefficients from auxiliary regressions, so the approach could alternatively be presented as Indirect Inference.

The main difficulty resides in solving the contracting problem which requires finding the promised utilities $W(z', s')$ in each state of the world for the next period. This becomes infeasible as soon as reasonable supports are considered for $\mathbb{Z}$ and $\mathbb{S}$. However, the first order condition with respect to W reveals that the promised utility in different states are linked to each other and that it is optimal for the firm to promise identical marginal utilities across sates tomorrow as much as possible. The solution is then characterized by how the promised marginal utility evolve over time subject to multipliers of (IC), (PC-E) and (PC-F) I estimated the model fully parametrically and make several parametric assumptions. Being consistent with the analysis before, I use the constant relative risk aversion utility function $u(w) = \frac{w^{1-\sigma}}{1-\sigma}$. The production function is set as $f(z,s) = \alpha sz$ where α is set to be 0.1%. I model the process of productivity by an $AR(1)$ model,

$$z_t = \rho_0(e) + \rho z_{t-1} + \epsilon_t,$$

where ϵ follows a normal distribution $N(0, \sigma_\epsilon)$, and the mean for effort level $e = 0$ is normalized to zero. The process is transformed to a discrete Markovian Chain using Tauchen (1986) on a grid of 5 points.[17] Furthermore, I set the sampling distribution of firm size $F(s)$ a log-normal distribution with expectation of μ_s and standard deviation of σ_s. Finally, the discount rate β is set to be 0.9 (the model is solved annually), the grid number for Lagrangian multiplier λ is 30, and for firm size s is 20.

Table 5: Parameters

Parameters	Description
δ	death probability
λ_1	offer arrival probability when employed
ρ_z	AR(1) coefficient of productivity shock
μ_z^w	mean of disturbance for $e = 1$
σ_ϵ	standard deviation of disturbance
μ_s	mean of F(s)
σ_s	standard deviation of F(s)
c	cost of efforts
σ	relative risk aversion

5.4 Moments and Identifications

Table 5 lists the complete set of parameters I estimate. I next make a heuristic identification argument that justifies the choice of moments used in the estimation. Appendix gives more details on how I construct these moments. The identification can be divided into three parts.

[17] A grid of only 5 points is for the speed of calculation. However, the simulated moments are very robust to this choice.

The first part of the identification is on the exogenous processes, including the executives' productivities, the exit rate and the offer arrival rate. There are direct links between the model and the data. The exit rate directly informs δ. Likewise, the incidence of job-to-job transitions is monotonically related to λ_1. The parameters of the productivity process, ρ_z, μ_z^w and σ_ϵ, are informed directly by the estimates of an AR(1) process on the profitability of each firm-executive match,

$$\texttt{profit}_{it} = \beta_0 + \rho_z \texttt{profit}_{it-1} + \epsilon_{it,0},$$

where i represents the executive-firm match and t represents the year.

The second part of the identification is on the job offer distribution. In particular, μ_s and σ_s are disciplined by the mean of log firm size and the mean of log wage. Given $\lambda_1 > 0$, the higher μ_s is, the more likely that executives can transit to larger firms, and the higher firms can compete for them. Hence, both $\log(wage)$ and $\log(size)$ are higher. σ_s controls the variation of job offers and has a similar effect as μ_s. The shape of job offer distribution can be further disciplined by the relationship between wage and firm size. I regress log wage on log of firm size, controlling for age dummies (divided into 100 groups),

$$\log(wage_{it}) = \beta_1 + \beta_{wage-size} \log(size_{it}) + \text{age dummies} + \epsilon_{it,1}.$$

The coefficient $\beta_{wage-size}$ is included as a moment to be matched.

The final part of the identification concerns the moral hazard problem, in particular two parameters σ and c. To be consistent with the incentive measurement delta in the data, I construct in the simulated data a delta variable defined by the dollar change in wage for a percentage change in productivity. How large delta is, how delta changes with firm size and wage can be used to inform σ and c. I use four moments. The mean of log delta is the first moment. It measures how large the contract incentives are. The next two moments are coefficients in the following two regressions. The first regression is delta on firm size, controlling for wage dummies,

$$\log(delta_{it}) = \beta_2 + \beta_{delta-size} \log(size_{it}) + \text{wage dummies} + \epsilon_{it,2},$$

the second regression is delta on wage (total compensation), controlling for firm size,

$$\log(delta_{it}) = \beta_3 + \beta_{delta-wage} \log(wage_{it}) + \beta_4 \log(size_{it}) + \epsilon_{it,3}.$$

The moments used are coefficient $\beta_{delta-size}$, which informs how the contract incentives change with firm size, and coefficient $\beta_{delta-wage}$, which informs how the contract incentives change with wage. The last moment I use for this part of identification is the fraction of positive delta, which informs the possibility of corner solutions (i.e. flat wage contract).

How do these four moments identify σ and c? First, both σ and c positively influence the mean of log delta. σ impacts delta via both the market-induced incentives and contract incentives. On the one hand, the larger σ is, the more concave the utility function, the higher delta should be. On the other hand, the larger σ is, the smaller market-induced incentives are, and the higher delta should be. Hence, σ and delta are positive correlated. c and delta are also positive correlated, although c only works through the contract incentives: the higher c is, the larger contract incentives are required to satisfy the incentive compatibility constraint. Yet, σ and c have different influence on how delta changes with firm size. Given other parameter values, a higher σ implies higher market-induced incentives hence a larger $\beta_{delta-size}$. A higher c does not change market-induced incentives while making the contract incentives larger, and hence gives a smaller $\beta_{delta-size}$. This difference distinguishes the two parameters. Finally, in the model there are parameter values such that the market-induced incentives are large enough to fulfill the incentive compatibility constraint and therefore delivers $delta = 0$. The fraction of positive delta helps to avoid too many of such corner solution . In all the regressions mentioned above, I also controlled for industry and year dummies for real data.

5.5 Results

Table 6 reports the targeted values of moments in the data and the corresponding values in the estimated model. The last two columns lists the parameter estimates and the standard errors. While I arranged moments and parameters along the identification argument made in the previous subsection, all parameters are estimated jointly.

Table 6: Moments and Estimates

Moments	Target	Model	Estimates	Standard Error
Exit Rate	0.0691	0.0691	$\delta = 0.0691$	0.0012
EE Rate	0.0523	0.055	$\lambda_1 = 0.2759$	0.0017
$\hat{\rho_z}$	0.8111	0.5499	$\rho_z = 0.7$	0.0036
Mean(z)	0.1284	0.1763	$\mu_z^w = 0.06$	0.0006
Var(z)	0.0141	0.0141	$\sigma_z = 0.12$	0.0014
Mean(log(wage))	7.17714	6.5241	$\mu_s = 1.7847$	0.228385
Mean(log(size))	7.44379	8.7934	$\sigma_s = 1.3982$	0.0314657
$\beta_{wage-size}$	0.370295	0.3196		
Mean(log(delta))	4.01842	3.8080		
$\beta_{delta-size}$	0.297673	0.2941	$c = 1.91385$	0.0259
$\beta_{delta-wage}$	0.717209	2.1228	$\sigma = 2.50748$	0.0046
Mean($delta > 0$)	0.994725	0.9844		

Overall, the model provides a decent fit to the data. In particular, it quantitatively captures the negative correlation between delta (the contract incentives) and firm size, which is the key target of the model. For this, a job arrival rate $\lambda_1 = 0.2759$ is required which also matches the job-to-job transition rate in the data. The magnitude of λ_1 indicates that, on average, the executive will receive an outside offer every three years, which seems reasonable. The levels of $\log(wage)$, $\log(size)$ and coefficient $\beta_{wage-size}$ are matched reasonably well, showing the on-the-job search and sequential auction in the model capture the main features of the executive labor market. The magnitude of μ_s and σ_s indicates that most offers are provided by relative small firms though there are high variations in firm size.

While the dynamic contract can provide a reasonable predict on the mean of $\log(delta)$, it misses the target of $\beta_{delta-wage}$. A further inspection of the simulated data and the real world data shows that in the real data, there is a large number of observations which have a small `tdc1` and a high delta, while in the simulated data, a low wage is usually associated with a small delta. Hence, there are some heterogeneity of executive-firm match that are not captured in the model. Finally, the exogenous processes on productivity and outside offers are matched quite well.

5.6 Compare Data and Model-Simulated delta

I conclude this section by providing some visual comparisons of delta from the dataset and the simulated data based on the estimates. The comparison are along two dimensions: wage and firm size. In both the real data and simulated data, I create a variable wage group which divides the sample into 100 groups accroding to the value of wage. Similarly, I create a variable size group based on the firm size. In this way, data is segmented into 100×100 cells. I calculate the average $\log(delta)$ of each cell and exclude those cells with less than 10 observations.

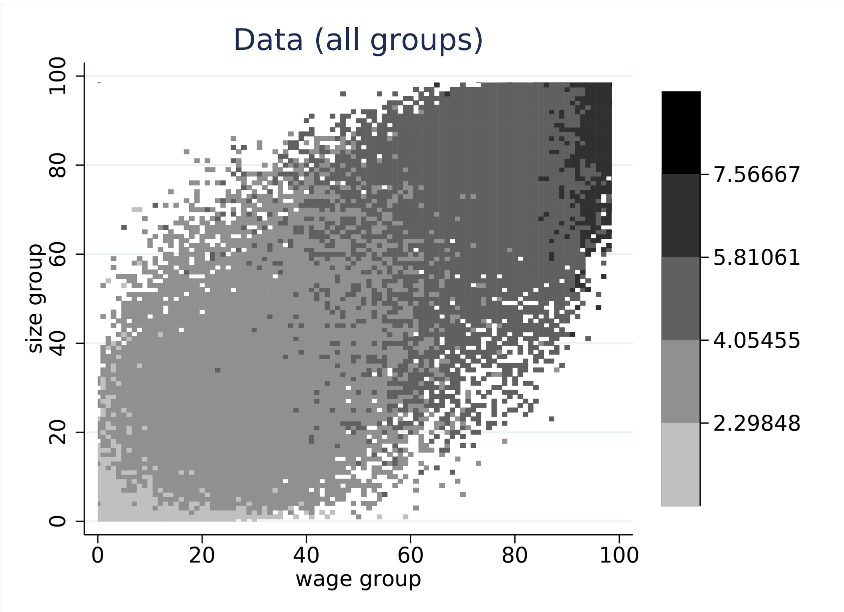

Figure 5: $\log(delta)$ in Data

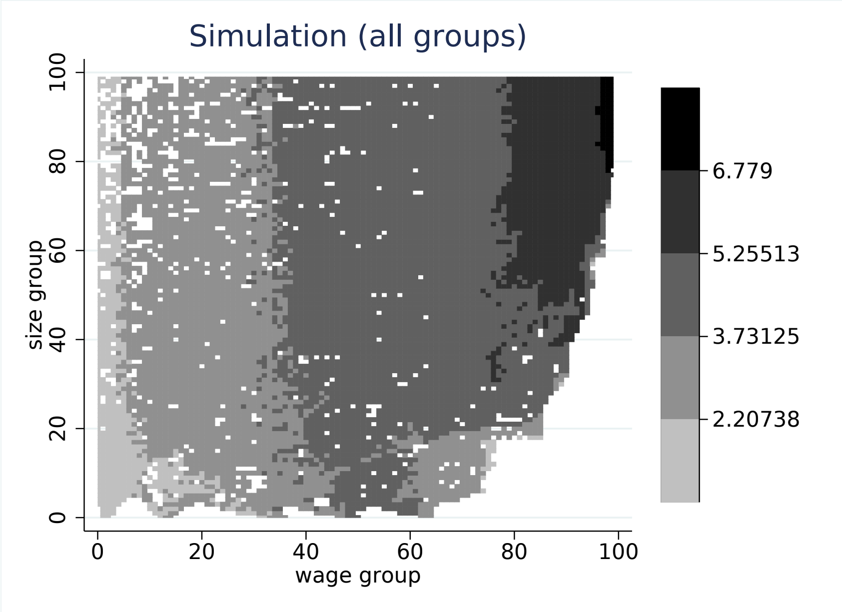

Figure 6: $\log(delta)$ in Simulated Data (with market-induced incentives)

Figure 5 shows $\log(delta)$ from the real data, figure 6 shows $\log(delta)$ of the simulated data, and figure 7 shows $\log(delta)$ of the simulated data when market-induced incentives are ignored.[18] It is clear that in all figures, delta increases with wage in a similar manner. The upper-left corner represents executives in big firms with low total compensations. This part is missing in the real data because only top 5 to 8 executives within the firm are included in the data. Importantly, comparing figures 6 and 7, we find that the simulated data with market-induced incentives have some variations along the size dimension within each wage group while in the model when market-induced incentives are ignored, delta only changes with wage.

How delta changes with size is not so clear in these overview plots. So I zoom into wage groups: wage group 15 to 20, 50 to 55, 80 to 85. They are shown in figure 8, 9, and 10, respectively. In each figure the plots from up to down are $\log(delta)$ from data, simulated data with market-induced incentives, and simulated data without market-induced incentives. As in the real data, the simulated data with market-induced incentives has enough variation of delta along firm size within each wage group. In contrast, in the simulated data ignoring market-induced incentives, there is little variation along firm size within each wage group, though the variation along wage group exists. Moreover, it seems the model (with market-induced incentives) can fit the data better for low and medium wage levels. There are still large variations along the firm size for high wages in the data, while the model does not have enough market-induced incentives to generate that large variations.

[18]Figure 7 is plotted by solving a model where in the contract problem market-induced incentives are removed from the objective and all the constraints.

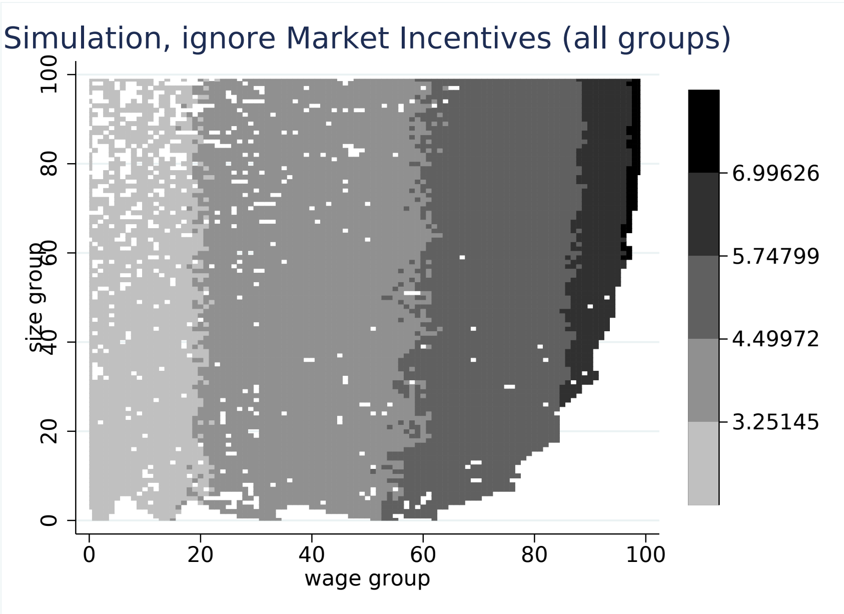

Figure 7: log($delta$) in Simulated Data (without market-induced incentives)

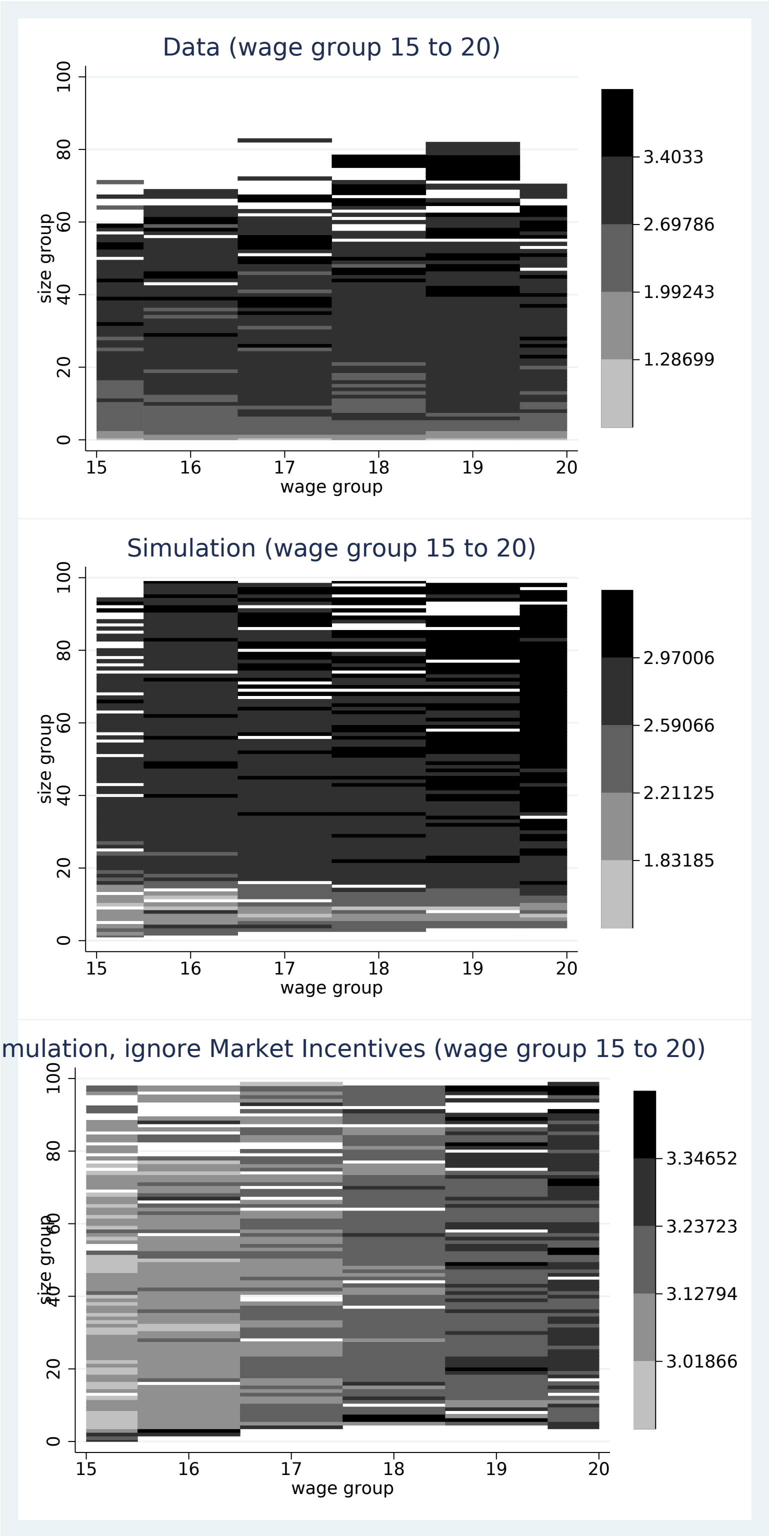

Figure 8: $\log(delta)$ in wage group 15 to 20

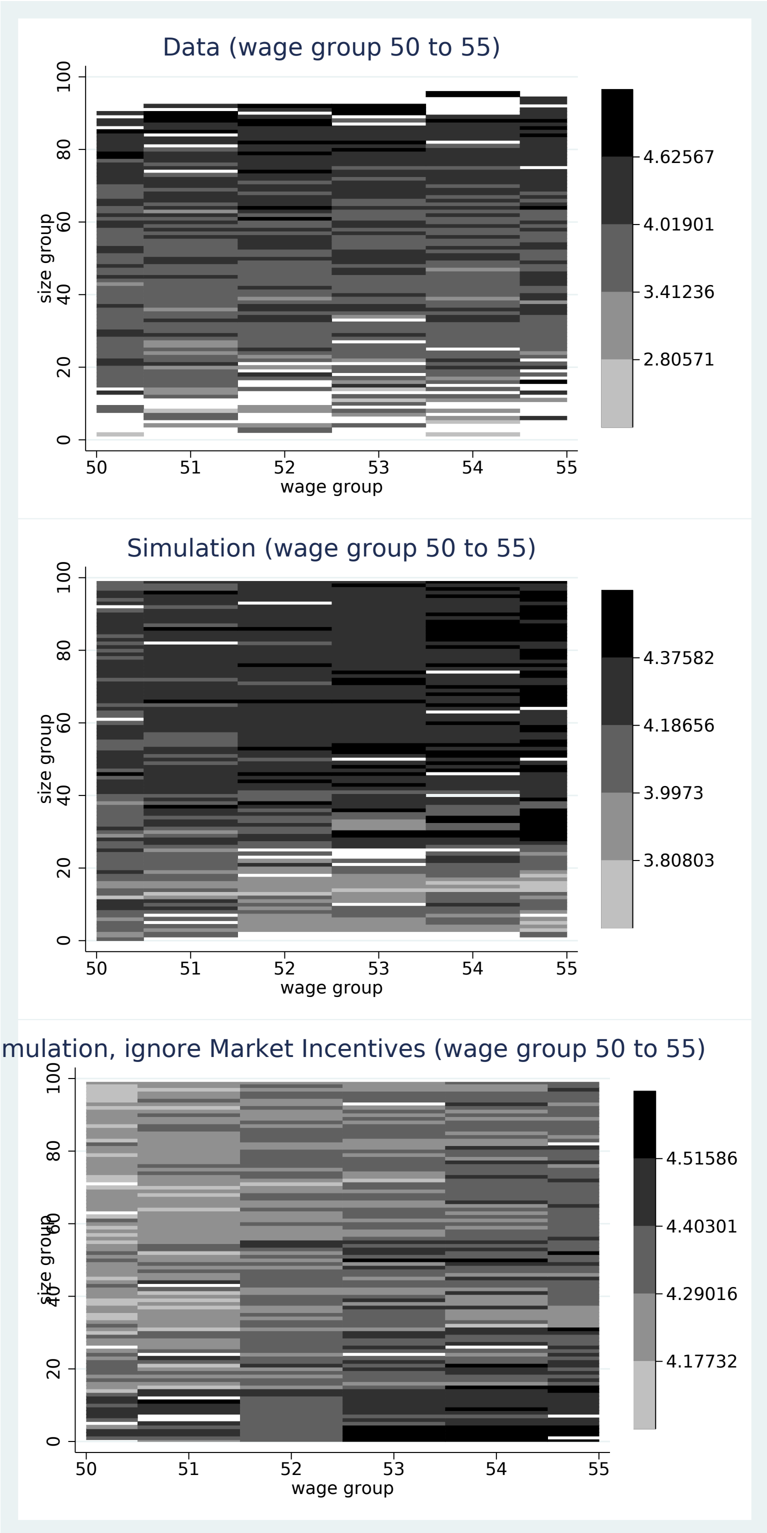

Figure 9: $\log(delta)$ in wage group 50 to 55

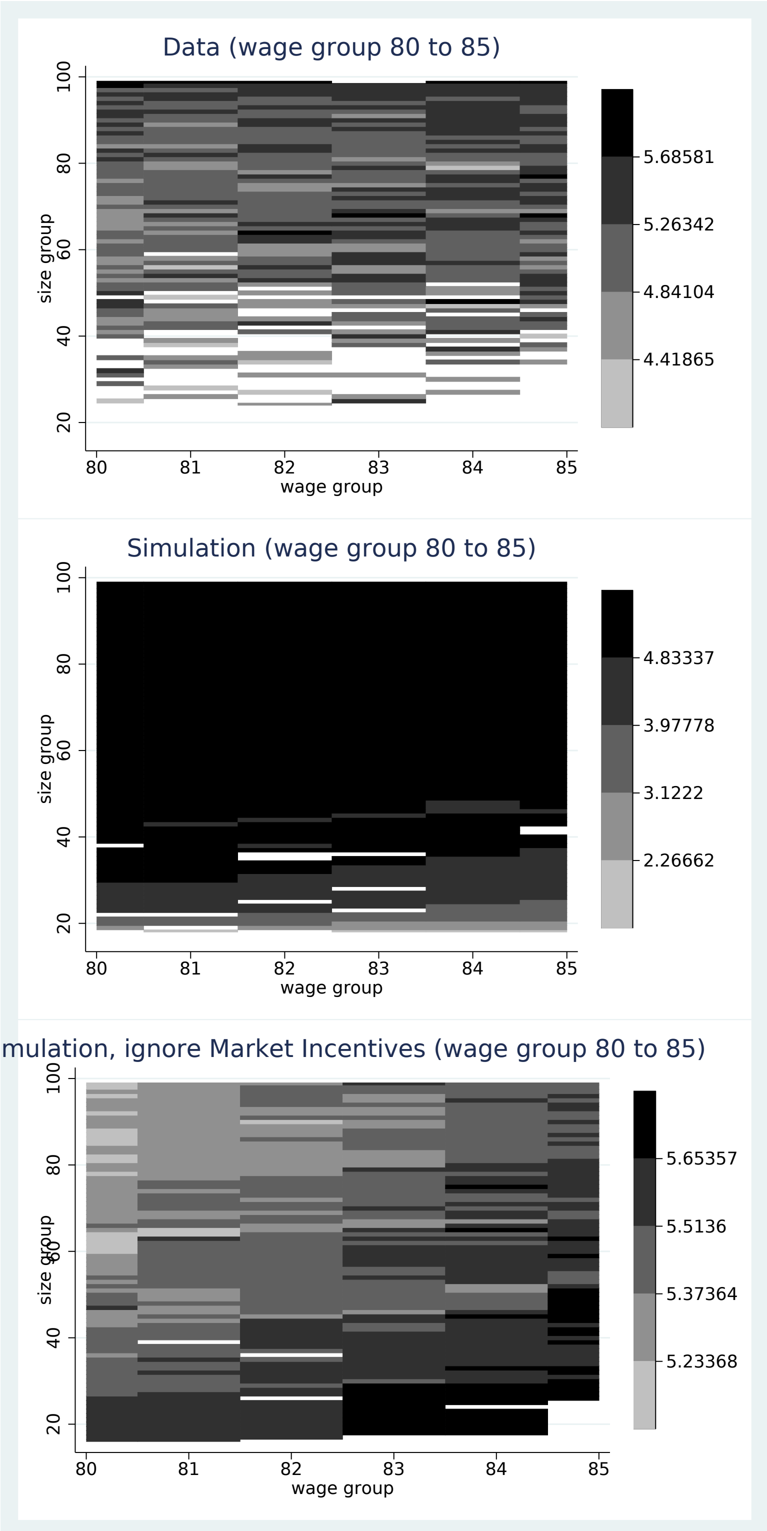

Figure 10: $\log(delta)$ in wage group 80 to 85

6 Quantitative Analysis

6.1 Decomposition

In this section, I do two decompositions. In terms of utilities, I decompose the total incentives into market-induced incentives and contract incentives. In terms of incentive pays delta, I simulate a counterfactual scenario where executives ignore market-induced incentives at all, and then compare the delta with and without market-induced incentives. I show that excluding market-induced incentives will greatly increase the incentive pays in small firms.

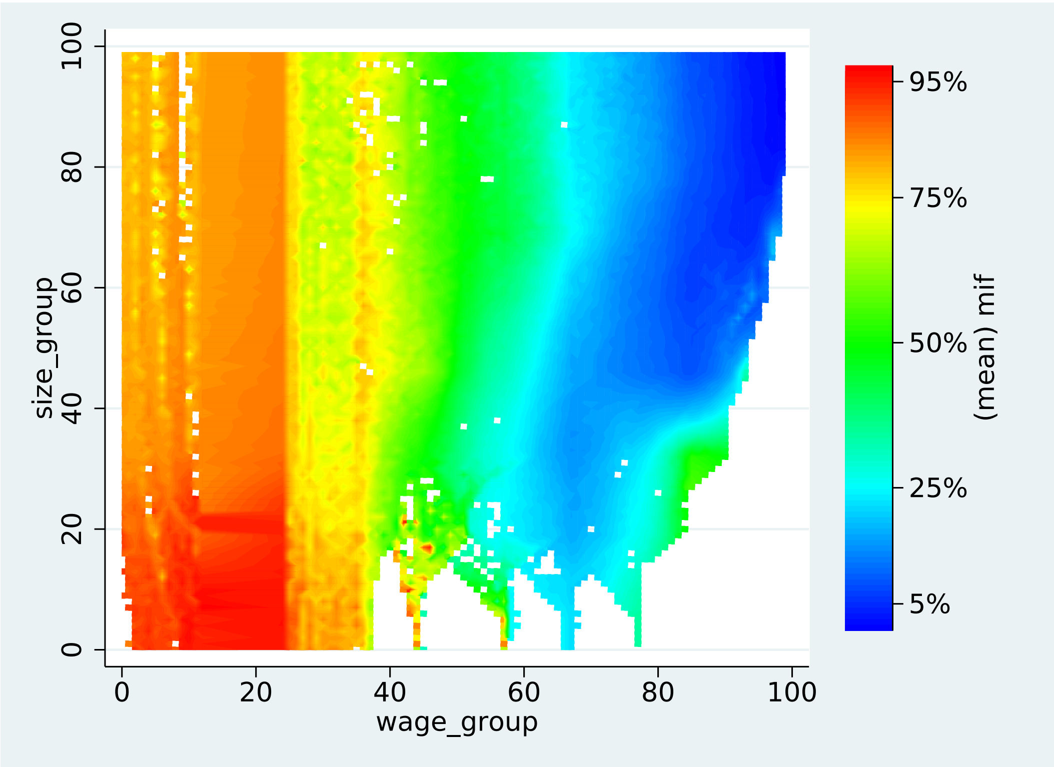

Figure 11: The Fraction of Market Incentives along Firm Size and Wage

Figure 11 plots the fraction of market-induced incentives Ξ_m in total incentives $\Xi_c + \Xi_m$ along the dimensions of firm size and wage. The simulation uses the estimated parameter values. Again, in the simulated data I divide the firm size into 100 groups, divide wage levels into 100 groups, compute the mean of market-induced incentives fraction in each cell and then draw a heat map over the 100×100 cells. The color in the figure represents the value, with red as higher and blue as lower. The blank space indicates a cell with less than 10 observations.

There are two observations. First, market-induced incentives decrease in wage. This is because as wage increases, there are less outside offers that can improve upon the current value. This holds for executives in all firms, though executives in small firms are not likely to get high wages (which gives the blank space in the right-lower corner). Second, within a wage group, the fraction of market-induced incentives decreases in firm size. This is exactly the reason that my model can explain that bigger firms have to provide more contract incentives.

Next, I simulate a counterfactual scenario where market-induced incentives are removed from the contracting problem. The delta simulated from this model is denoted by $delta_c$. This is the delta if all incentives have to be provided by the contract. I then calculate the ratio of $\log(delta)/\log(delta_c)$ and plot it in figure 12. This ratio represents the contribution of the market-induced incentives in $delta$. It shows that almost for each wage group, for the first 30 size groups (out of 100 groups), when market-induced incentives are included, the value of $\log(delta)$ is about 50% to 70% of the $\log(delta_c)$. For size group 30 to 50, the ratio is around 90%. The implication is, if there were no market-induced incentives, small firms would have to pay much higher contract incentives, while the impact on medium and large firms is much smaller.

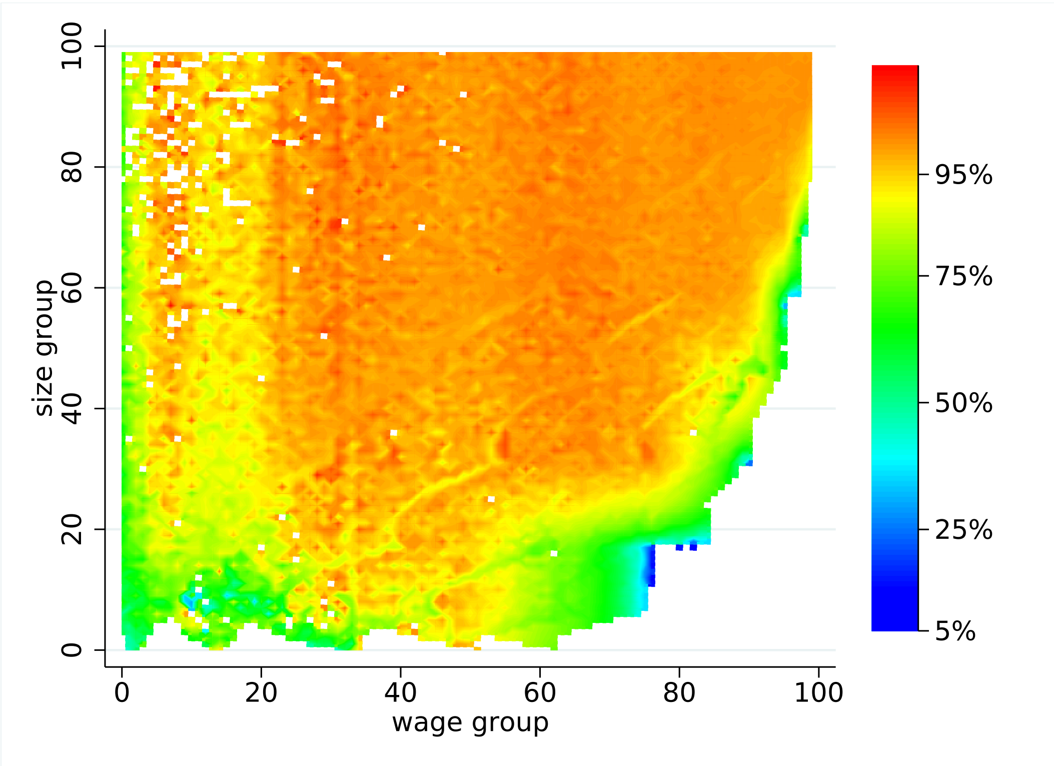

Figure 12: The Ratio of log Delta with and without market-induced incentives along Firm Size and Wage

6.2 Counter-factual analysis

I now use the model to evaluate several specific regulations on executive pays that have been proposed or implemented, including pay ratio restrictions, restrictions on incentive pays.

[to be added]

Since the model involves the incentive problem within each firm, and also connects to the market competition, I explore the contagion effect of good/bad governance. In the model, there are two parameters indicating the quality of governance: effort cost c represents how severe the principle-agent problem is in the firm, and α in the production function represents firm's willingness to pay for the executive, loosely corresponds to the entrenchment.

[to be added]

7 Conclusions

By the law of diminishing utility, wealthy agents are costly to motivate. In a dynamic relationship, the wealth is traced from a life-time utility. Executives in bigger firms are expected to be wealthier, less sensitive to monetary rewards in the future than their equally paid counterparts in smaller firms. Thus, market-induced incentives are lower in bigger firms. This implies higher contract incentive pays are required from big firms. With this intuition, this paper provides a framework for empirically analyzing market-induced and contract incentives in a frictional labor market. I use this framework to investigate why bigger firms pay more for performance. I show that in the counter-factual that when market-induced incentives are ignored, the incentive pays would be much higher in small and medium firms.

Appendix A. Proofs

[to be added]

Appendix B. Computing Algorithm

To further characterize the optimal solution, we resort to the tools developed by Marcet and Marimon (2017, hereafter MM).[19]In dynamic contracting problems with forward looking constraints such as the IC constraint here, the solution does not satisfy the Bellman equation. MM suggest to study a recursive Lagrangian. Under standard general conditions there is a recursive saddle-point functional equation (analogous to a Bellman equation) that characterizes a recursive solution to the planners problem. The recursive formulation is obtained after adding a co-state variable λ_t summarizing previous commitments reflected in past Lagrange multipliers. The time-consistent continuation solution is obtained by using the endogenous λ_t as the vector of weights in the objective function.

Proposition 4 (Marcet and Marimon)**.** *Define Pareto Frontier by*

$$P(z,s,\lambda) = \sup_{W} \Pi(z,s,W) + \lambda W,$$

where Π and W are defined as in (BE-F) and (PKC), and $\lambda > 0$ is a Pareto weight assigned to the executive. Then there exist positive multipliers of $\{\mu, \mu_0(z'), \mu_1(z')\}$ that solve the following problem

$$P(z,s,\lambda) = \inf_{\mu,\mu_0(z'),\mu_1(z')} \sup_{w} h(z,s,\lambda,w) + \hat{\beta} \sum_{z'} P(z',s,\lambda')\Gamma(z,z'),$$

where multiplier μ corresponds to the incentive compatibility constraint, multipliers $\mu_0(z'), \mu_1(z')$ correspond to participation constraints,

$$h(z,s,\lambda,w) = y(s)z' - w + \lambda u(w) - (\lambda + \mu)c,$$

Pareto weight evolves according to

$$\lambda' = \lambda + \mu(1 - g(z,z')) + \mu_0(z') + \mu_1(z'),$$

and

$$\hat{\beta} = \tilde{\beta}(1 - \lambda_1 \sum_{\mathcal{M}_1 \cup \mathcal{M}_2} F(s')).$$

The optimal contract $\{w, W(z')\}$ follows that

$$u'(w) = \frac{1}{\lambda}, \tag{6}$$

$$W(z') = W(z',s,\lambda'). \tag{7}$$

[19]This approach has been used in many applications. A few examples are: growth and business cycles with possible default (Marcet and Marimon (1992), Kehoe and Perri (2002), Cooley, et al. (2004)); social insurance (Attanasio and Rios-Rull (2000)); optimal fiscal and monetary policy design with incomplete markets (Aiyagari, Marcet, Sargent and Seppala (2002), Svensson and Williams (2008)); and political-economy models (Acemoglu, Golosov and Tsyvinskii (2011)).

Proof. See Appendix A. □

Proposition 4 can be illustrated intuitively using the Pareto weight of the executive λ and the multiplier μ of the incentive constraint. Suppose the match starts with a $\lambda^{(0)}$, and assume the participation constraints are not binding so that $\mu_0 = \mu_1 = 0$. $\lambda^{(0)}$ has to satisfy $W(z_O, s, \lambda^{(0)}) = W^0$. To deal with the moral hazard, the optimal contract indicates a $\mu^{(0)}$. Then depending on the realization of z', the weight of the executive will be updated to

$$\lambda^{(i)} = \lambda^{(i-1)} + \mu^{i-1}(1 - g(z, z')) \text{ for } i \text{ in } 1, 2...$$

The evolve of λ continues as such till the match breaks. When there is an outside offer such that the executive moves from his current firm to the outside firm, then the new match starts with a $\lambda^{(n)}$ such that $W(z, s', \lambda^{(n)}) = \overline{W}(z, s)$, where I have denoted the current productivity by z, current firm by s, and the outside firm by s'. It means the new match will assign a new weight to the executive so that he gets the continuation value $\overline{W}(z, s)$. Then the new Pareto weight will evolve again as illustrated above. In a nutshell, proposition 4 allows us to solve the optimal contract in the space of Pareto weight λ instead of in the space of the promised utility. At any moment, we can transfer from the metrics of λ back to the metrics of utilities using (6) and (7).

The advantage of this method is I do not need to find the promised utilities $W(z')$ in each state of the world for the next period. Instead, λ and μ are enough to trace all $W(z')$. Moreover, λ corresponds to the total compensation level (wage level), while μ corresponds to how much contract incentive is provided in the optimal contract. The two multipliers are enough to understand both theoretically and numerically why keeping the same wage level (the same λ), incentive pays increase with firm size (μ increases with firm size).

Appendix C. Data Appendix

[Incomplete Reference]

References

Abrahám, Arpád, Fernando Alvarez-Parra, and Susanne Forstner (2016), "The effects of moral hazard on wage inequality in a frictional labor market."

Abreu, Dilip, David Pearce, and Ennio Stacchetti (1990), "Toward a theory of discounted repeated games with imperfect monitoring." *Econometrica: Journal of the Econometric Society*, 1041–1063.

Alvarez, Fernando and Urban J Jermann (2000), "Efficiency, equilibrium, and asset pricing with risk of default." *Econometrica*, 68, 775–797.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin (2006), "Wage bargaining with on-the-job search: Theory and evidence." *Econometrica*, 74, 323–364.

Coles, Jeffrey L, Naveen D Daniel, and Lalitha Naveen (2006), "Managerial incentives and risk-taking." *Journal of financial Economics*, 79, 431–468.

Coles, Jeffrey L, Naveen D Daniel, and Lalitha Naveen (2013), "Calculation of compensation incentives and firm-related wealth using execucomp: Data, program, and explanation."

Edmans, Alex and Xavier Gabaix (2011), "The effect of risk on the ceo market." *The Review of Financial Studies*, 24, 2822–2863.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter (2017), "Executive compensation: A survey of theory and evidence." Technical report, National Bureau of Economic Research.

Edmans, Alex, Xavier Gabaix, and Augustin Landier (2009), "A multiplicative model of optimal ceo incentives in market equilibrium. review of financial studies."

Fama, Eugene F (1980), "Agency problems and the theory of the firm." *Journal of political economy*, 88, 288–307.

Farhi, Emmanuel and Iván Werning (2013), "Insurance and taxation over the life cycle." *Review of Economic Studies*, 80, 596–635.

Gabaix, Xavier and Augustin Landier (2008), "Why has ceo pay increased so much?" *The Quarterly Journal of Economics*, 123, 49–100.

Gayle, George-Levi, Limor Golan, and Robert A Miller (2015), "Promotion, turnover, and compensation in the executive labor market." *Econometrica*, 83, 2293–2369.

Gayle, George-Levi and Robert A Miller (2009), “Has moral hazard become a more important factor in managerial compensation?” *American Economic Review*, 99, 1740–69.

Gibbons, Robert and Kevin J Murphy (1992), “Optimal incentive contracts in the presence of career concerns: Theory and evidence.” *Journal of political Economy*, 100, 468–505.

Grochulski, Borys and Yuzhe Zhang (2017), “Market-based incentives.” *International Economic Review*, 58, 331–382.

Hall, Brian J and Kevin J Murphy (2000), “Optimal exercise prices for executive stock options.” *American Economic Review*, 90, 209–214.

Harris, Milton and Bengt Holmstrom (1982), “A theory of wage dynamics.” *The Review of Economic Studies*, 49, 315–333.

Holmström, Bengt (1999), “Managerial incentive problems: A dynamic perspective.” *The review of Economic studies*, 66, 169–182.

Hopenhayn, Hugo A and Juan Pablo Nicolini (1997), “Optimal unemployment insurance.” *Journal of political economy*, 105, 412–438.

Kocherlakota, Narayana R (1996), “Implications of efficient risk sharing without commitment.” *The Review of Economic Studies*, 63, 595–609.

Lamadon, Thibaut (2016), “Productivity shocks, long-term contracts and earnings dynamics.” *manuscript, University of Chicago*.

Lentz, Rasmus (2014), “Optimal employment contracts with hidden search.” Technical report, National Bureau of Economic Research.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin (2016), “Matching, sorting and wages.” *Review of Economic Dynamics*, 19, 63–87.

Marcet, Albert and Ramon Marimon (2017), “Recursive contracts.”

Margiotta, Mary M and Robert A Miller (2000), “Managerial compensation and the cost of moral hazard.” *International Economic Review*, 41, 669–719.

Mele, Antonio (2014), “Repeated moral hazard and recursive lagrangeans.” *Journal of Economic Dynamics and Control*, 42, 69–85.

Phelan, Christopher (1995), “Repeated moral hazard and one-sided commitment.” *Journal of Economic Theory*, 66, 488–506.

Phelan, Christopher and Robert M Townsend (1991), “Computing multi-period, information-constrained optima.” *The Review of Economic Studies*, 58, 853–881.

Postel-Vinay, Fabien and Jean-Marc Robin (2002), “Equilibrium wage dispersion with worker and employer heterogeneity.” *Econometrica*, 70, 2295–2350.

Rogerson, William P (1985), "Repeated moral hazard." *Econometrica: Journal of the Econometric Society*, 69–76.

Spear, Stephen E and Sanjay Srivastava (1987), "On repeated moral hazard with discounting." *The Review of Economic Studies*, 54, 599–617.

Tauchen, George (1986), "Finite state markov-chain approximations to univariate and vector autoregressions." *Economics letters*, 20, 177–181.

Tervio, Marko (2008), "The difference that ceos make: An assignment model approach." *American Economic Review*, 98, 642–68.

Thomas, Jonathan and Tim Worrall (1988), "Self-enforcing wage contracts." *The Review of Economic Studies*, 55, 541–554.

Thomas, Jonathan and Tim Worrall (1990), "Income fluctuation and asymmetric information: An example of a repeated principal-agent problem." *Journal of Economic Theory*, 51, 367–390.

Townsend, Robert M (1982), "Optimal multiperiod contracts and the gain from enduring relationships under private information." *Journal of political Economy*, 90, 1166–1186.