Do Incentives or Competition Determine Managers' Wages?*

Hector Chade† and Jan Eeckhout‡

July 17, 2023

Abstract

Does managerial compensation reflect incentive provision or competition for talent by other firms? To answer this question, we study optimal incentive contracts under moral hazard when firms compete for managers. We derive the competitive equilibrium wages that firms offer and isolate the determinants due to contracting, and those due to competition. We then apply our framework using data on executive compensation. We find that for superstar CEOs, competition determines 70% of income, while for the lower ranked CEOs, income predominantly derives from incentive contracts. In absolute terms, the compensation for incentives is remarkably constant across the entire distribution. Thus, both incentives and competition play a key role, but their contribution differs at the top and bottom of the distribution of talent.

Keywords. Competition. Assortative Matching. Contract Theory. Moral Hazard. Incentives. Competitive Equilibrium. Executive Compensation.

*We would like to thank Marco Bassetto, Sergei Guriev, Stephen Hansen, Ilse Lindenlaub, Jeremy Lise, Alejandro Manelli, Giuseppe Moscarini, Costas Meghir, Fabien Postel-Vinay, Andrea Prat and Jean-Marc Robin as well as seminar audiences for insightful discussions and comments. Elena Casanovas, Christoph Hedtrich, and Ian Hsieh provided superb research assistance. Eeckhout gratefully acknowledges support from ERC Advanced Grant 882499, from AEI (Severo Ochoa, Barcelona School of Economics CEX2019-000915-S), and from AEI/FEDER grant, UE - PGC2018-096370-B-I00.

†Arizona State University, hector.chade@asu.edu

‡UPF Barcelona (ICREA-BSE-CREi), jan.eeckhout@upf.edu

1 Introduction

The two dominant paradigms that explain managers' pay are incentive contracts and competitive matching. Both explanations have been influential to further our understanding of managerial compensation. The principal-agent paradigm has had an immense influence in praxis with the implementation of performance-based compensation schemes such as stock options, especially since the 1990s.[1] And researchers have invoked the competition paradigm to rationalize the outlandish pay that superstar executives receive: small differences in the ability of managers magnify enormously when matched with more valuable firms.[2] Yet, despite the success of each of these frameworks, these two paradigms do not talk to each other, and it remains an open question what the quantitative contribution is of each of these two explanations to managerial compensation when they interact with each other.

In this paper, we nest both incentive provision due to moral hazard and competitive matching to determine the productivity and compensation of managers. Firms offer incentive contracts to elicit effort from executives. But rather than offering contracts as in an isolated principal-agent interaction, firms compete with other firms in order to attract the best manager for the job. The novelty of our approach is to derive optimal incentive contracts in such a competitive setting. This implies that optimal contracts now not only take into account the incentive provision for effort (the incentive compatibility constraint), but also the outside option of matching with a competing firm (the individual rationality constraint). The outside option is endogenous and determined in equilibrium.

The model consists of a large number of firms (principals) that differ in productivity and risk-averse managers (agents) who differ in ability. When a firm and a manager match, the firm proposes a binding contract, and if the manager accepts, she then chooses effort. Building on Holmström and Milgrom (1987), we focus on linear incentive schemes that condition on output. We derive the competitive equilibrium of this market, which crucially hinges upon the optimal contracts offered by firms.

We dissect the manager's equilibrium wage into two components. A first component reflects incentive provision due to moral hazard. Since output is noisy, the firm faces a tradeoff between offering incentive pay and base pay. Incentive pay induces higher effort, but exposes the risk

[1] The literature on moral hazard and incentives is vast, and includes the seminal papers by Holmström (1979) and Grossman and Hart (1983).

[2] An important reference on this topic is Gabaix and Landier (2008).

averse manager to higher risk, for which she needs to be compensated and thus is costly to the firm. The higher the noise in output, the harder it is to attribute output to the effort of the worker and the lower the incentive pay in the optimal contract. In our setting, this component is as in Holmström and Milgrom (1987), except for the fact that we allow for heterogeneity in the ability of managers, which means output, effort and wages are all dependent on the manager's characteristic.

A second component reflects competition among firms to obtain the services of the managers. Because we assume that part of the output is deterministic and that there are complementarities between the manager ability and the firm productivity, wages reflect competition as in Rosen (1981), Gabaix and Landier (2008) and Terviö (2008). In order to hire or retain a worker, the firm needs to pay a wage whose rate of change equals the marginal contribution of the manager to the firm's output. This leads to a *superstar effect* at the top: small increases in managerial ability for highly-skill managers translate into drastic increases in compensation at the top.

This second component due to competition consists of two parts. The usual outside option that determines sorting and a part that is novel in this literature. Because optimal contracts depend on the worker and firm characteristics, the outside option also reflects the incentive contracts offered by competing firms. A firm needs to compete to attract and retain its managers. If a competing firm offers more attractive terms along the bilateral contracting dimension, the incumbent firm needs to offer terms that are competitive in order to obtain the services of the manager. The terms may differ, but they should offer at least the same expected utility. For example, if competitors offer a larger performance component, then the incumbent needs to offer either a similar bonus, or increase the base pay of the compensation until the expected utility matches the competing offer. This competition in contractual terms is absent in the standard principal-agent model with moral hazard where the outside option is exogenous.

We then use matched firm-CEO data on executive compensation and firm profits to estimate the primitive parameters of the model using maximum likelihood estimation. The competitive equilibrium allocation that matches the data determines the parameters of the production technology and the resulting managerial compensation. We find that there are complementarities in production between the firm's trait that determines the productivity of its technology, and the manager's trait. As a result, there is positive sorting between the attributes of managers and firms. We also estimate the distribution of the ex-post realization of the firm and manager characteristics.

As predicted by Gabaix and Landier (2008), we find that there are enormous differences in the value of firms, and relatively small differences in the ability of managers. Because of sorting, the firm differences massively magnify the earnings of managers who are only marginally different in ability. But we also find that there is a lot of volatility in both firm and manager characteristics. It is the volatility of firm productivity that pins down the optimal incentive provision, where higher volatility makes it harder to provide incentives to the manager.

Along the distribution, the contribution of the two components — incentive provision and competition — to the estimated wages varies. For the wages of the top managers, the competitive matching component is the dominant contributor to the value of the firm and of wages and accounts for 70%. Incentive contracts contribute the rest. Competition and the superstar effect dominates moral hazard as a determinant of wages at the top of the distribution. Instead, at the bottom of the distribution, incentive contracts account for nearly all of the compensation. In absolute terms, the incentive component is remarkably constant for all managers, but what changes is the competitive component. The firms at the bottom are small and they cannot afford to pay superstars salaries to their managers. Instead, at the top, firms are enormous and tiny differences in the ability of the manager have enormous implications for the firm's profitability. Competition by these larger firms for the managers leads to superstar pay.

In sum, all executives receive an incentive component as compensation that is remarkably constant in levels across the distribution of their abilities. At the bottom of the distribution, this incentive component is virtually *all* of their compensation. As we move up the distribution, executives receive an additional component that firms need to pay to poach the executives from competitors, and for the superstar executives, this *competition component* dominates the incentive provision component.

Key to our approach is that we infer the incentive component from the structure of the model, without using observed incentive pay. An alternative approach to ours is to use observed contracts as our data specifies which part of compensation is variable compensation. However, in practice, there is no immediate mapping of variable pay as observed in the data to incentive pay as understood in the contract theory literature (see for example Bertrand and Mullainathan (2001) among many others). There is a complicated inference problem about what contributes to the fixed component of compensation and what contributes to the variable component. Consider for example an executive who is paid in stock options that at the end of the contract are either worth 10 million or 11 million, say each with equal probability. If we use stock options as a measure to classify pay, we would have 10 and 11 million respectively for the variable component and zero for the fixed component. But this is not the correct classification into the variable and the fixed component of compensation in this example. Because the manager is guaranteed an income of 10 million, which is the fixed component, and the variable component, which is 0 and 1 million respectively. This example illustrates that directly measuring from the data what the stochastic process is of compensation is complicated, and simply using accounting categories to classify pay into fixed compensation and incentive pay is misleading.[3] So instead of trying to measure direcxtly the data generating process of output and compensation based on accounting data, we use the model to infer each component.

RELATED LITERATURE. Our work relates to several strands of literatures. It obviously builds on the large literature on matching and on the principal-agent model with moral hazard. We have already referenced some of the relevant papers in each of these two literature above. Most closely related to our theoretical setting is the work by Edmans, Gabaix, and Landier (2009) who consider a principal-agent problem with a risk-neutral agent with multiplicatively separable preferences in income and effort, and a multiplicative production function. They embed this agency problem in a matching market where firms (principals) and CEOs (agents) match. Their assumptions – most notably, risk neutrality, the functional forms, and the fact that effort is discrete – allow them to explicitly solve the model, and derive several interesting equilibrium properties. In equilibrium, all managers exert high effort and as a result, output is determined exclusively by the match value, which is as in Gabaix and Landier (2008). Our model also has preferences and technology that are multiplicatively separable, but our agents are risk averse, and thus moral hazard distorts effort choice away from first best even for CEOs with high outside options. Moreover, we allow for continuous effort choice that in equilibrium depends on the variance of the shock as in Holmström and Milgrom (1987). This is key for our quantitative analysis since we back out the stochastic component from the data, rather than assume that the incentive component is observed. As we have argued above, the incentive portion of

[3] In addition, firms may know which part of stochastic output is affected by effort, and which part is exogenous (such as oil price fluctuations for example, see Bertrand and Mullainathan (2001)).

the observed contract does not necessarily provide an adequate measure of the risk a manager faces. Only the realized outcomes allow us to measure the risk. As such, our setup allows enough flexibility so that the incentive component can vary with firm size. The model therefore encompasses all potential outcomes, including those where the equity stake of managers in dollar terms is flexibly dependent on firm size, consistent with the findings of Demsetz and Lehn (1985), Jensen and Murphy (1990), Gibbons and Murphy (1992), Hall and Liebman (1998), Schaefer (1998) and Baker and Hall (2004). Our estimation concludes that the incentive component in levels is independent of firm size, consistent with Edmans et al. (2009).

There is also a growing literature that studies matching problems where agents have risk averse preferences, and that thus allow us to embed principal-agent problems in a competitive matching environment. Legros and Newman (2007) provides the first general treatment of models with Imperfectly Transferable Utility (ITU), and other papers study matching between risk averse agents (see Ackerberg and Botticini (2002), Serfes (2005), and Chiappori and Reny (2016)).Our model embeds the Holmström and Milgrom (1987) principal-agent problem in a matching market and we derive its competitive equilibrium in closed form. This explicit solution allows us to apply the framework to estimate the determinants of managerial compensation when there is matching of CEOs to firms in the presence of moral hazard.[4]

Our work also relates to the empirical literature on matching and executive compensation, and labor and personnel economics more broadly. In addition to the already mentioned papers, Bandiera, Guiso, Prat, and Sadun (2015) evaluate which firm and worker characteristics determine matching and the nature of the labor contracts. Using an extremely detailed, novel dataset on the demographics of workers and firms, as well as the terms of worker-firm contracts, they find that the risk profile of workers and their talent are key determinants of the matching pattern. Our focus is on the decomposition of incentives and competition, both in the theory and the structural estimation of the model. And in recent work, Bandiera, Hansen, Prat, and Sadun (2020) estimate a model of CEO performance that allows for mismatch as in our model. They use high frequency survey data and employ a machine learning algorithm to estimate char-

[4]Serfes (2005) also analyzes matching between principals with different output variances (riskiness) and agents with different coefficients of absolute risk aversion under moral hazard and linear contracts. Besides focusing on different characteristics that drive sorting, we characterize the competitive equilibrium of the model and the properties of managerial compensation. For instances of matching principals and agents under moral hazard that do not rely on linear contracts, see the application in Legros and Newman (2007), and that in Chade and Swinkels (2020). For an overview of the literature on matching, including with risk averse agents, see Chade, Eeckhout, and Smith (2017).

acteristics of managers who match with different firms. In our model too, there is mismatch due to the realization of ex-post managerial attribute after matching. We find that ex-post mismatch is sizable. Our results also confirm that minor differences in the manager's ability lead to enormous differences in compensation due to the matching with more productive firms. The ex-post mismatch adds an additional source of variation in compensation.

The paper is organized as follows. In section 2, we lay out the model. We then solve the model and derive the main results in section 3. In section 4, we apply the model using data on executive compensation and estimate the models primitive parameters. Section 5 concludes.

2 The Model

The model builds on the well-known principal-agent problem with moral hazard of Holmström and Milgrom (1987), enriched with a competitive market where heterogeneous managers and firms match (as in Gabaix and Landier (2008)).

We consider a market with two large populations (each of measure one) consisting of heterogeneous firms (principals) and managers (agents) who match pairwise. Each matched pair solves a principal-agent problem under moral hazard, where managerial effort is unobservable to the firm, and where output depends on the pair's characteristics, managerial effort, and a shock. An agent's productivity depends on a parameter that is only revealed to the pair after contracting takes place but before the manager chooses a level effort. This stochastic component in an agent's attribute adds a layer of ex-post heterogeneity that not only allows for mismatch, a realistic feature in labor markets, but it also is essential in the estimation of the model.

Each manager has a trait $x \in [0, 1]$, distributed according to an atomless cumulative distribution function (cdf) Γ; similarly, each firm's trait is a scalar $y \in [0, 1]$ distributed with an atomless cdf Ψ.[5] We call x and y managers' and firms' ex-ante traits, which are observable at the matching stage. There is, however, uncertainty about everybody's ex-post attributes which are payoff-relevant and are realized after matching takes place. A manager with ex-ante trait x draws an ex-post attribute $\omega \geq 0$, from a cumulative distribution $F(\cdot|x)$ with with mean $k(x)$ and variance s_{ω}^2.[6] A firm with ex-ante trait y draws an ex-post attribute σ from a distribution

[5] Although we assume a continuum of firms and managers, it is straightforward to adapt the analysis to the case with a finite number of them. We do this necessarily in the empirical application.

[6] We assume that s_{ω}^2 is independent of x for simplicity. It is straightforward in the analysis below to allow for such dependence as long as it is monotone in x.

$G(\cdot|y)$ with mean $t(y)$ and variance s_σ^2. For now, we only assume that $k(x)$ and $t(y)$ are positive for all x and y. We assume that, conditional on (x, y), ω and σ are independent. We assume throughout that σ is normally distributed; our theoretical results hold generally for any distribution F. Given normality of G, for each firm with ex-ante trait y, we can write $\sigma = t(y) + \varepsilon_\sigma$, where ε_σ is normally distributed with mean zero and variance s_σ^2.

The timing is as follows. There is a competitive market where firms and managers match pairwise, and in which equilibrium contracts are determined. Within each pair, the principal offers a profit-maximizing contract that takes into account the incentive constraints for managerial effort as well as the participation constraint determined by the agent's outside option from matching with a competing principal. Once a match is formed and a contract is accepted, the agent's ex-post characteristic ω realizes and is observed by both matched parties. Then the agent privately chooses a level of effort (moral hazard). Finally, output realizes, which is determined by managerial effort, the parties' characteristics, and the (unobservable) shock to output ε_σ. Payments are then distributed according to the terms stipulated in the contract.[7]

Agents are strictly risk averse and have constant absolute risk aversion (CARA) utility function: an agent's utility of a wage-effort pair $(w, e) \in \mathbb{R} \times \mathbb{R}_+$ is

$$U(w, e) = -\mathrm{e}^{-r\left(w - \frac{e^2}{2}\right)}, \tag{1}$$

where $r > 0$ is the coefficient of absolute risk aversion, and $e^2/2$ is the (monetary) disutility of effort e. An agent with trait x has an outside option that yields the monetary equivalent income $a(x)$ and hence utility $-\mathrm{e}^{-ra(x)}$. In the standard principal-agent problem without competition – in particular, in Holmström and Milgrom (1987) – the outside option is an exogenously given reservation wage. Here, each agent's reservation wage is determined through the competition by firms to attract different managers, and agents with different ex-ante trait x can command different contracts as outside options. Principals are risk neutral and maximize the expected

[7] To simplify the analysis of an already rich model, we do not allow for rematching after the managers' ex-post attributes are realized, although some firms and managers may have incentives to do so. But, incorporating rematching would require an additional friction (for example, search frictions), otherwise there would be continuous rematching in real time. This adds a level of complexity that is beyond the scope of this paper. Note that without rematching, one can interpret a manager's ex-post attribute as either match-specific or permanent (just like the distinction between firm-specific and general human capital).

value of $q - w$, where output q is a random variable given by the production technology

$$q = \omega(e + \sigma) = \omega e + \omega t(y) + \omega \varepsilon_\sigma \tag{2}$$

which exhibits complementarities in (ω, σ) and in (ω, e).[8]

We assume that contracts are linear in output. We appeal to the Holmström and Milgrom (1987) dynamic justification for the optimality of linear contracts, which can be solved "as if" it was a static problem with an ad-hoc restriction to linear contracts and where effort is chosen once and for all. The slope and intercept can be contingent on the agent's ex-post characteristic ω, since ω is realized (and publicly observed) after acceptance of the contract but before the agent chooses effort. Thus, for each realization of $(q, \omega) \in \mathbb{R} \times \mathbb{R}$, the contract pays a wage

$$w(q, \omega) = \beta(\omega) + \alpha(\omega) q. \tag{3}$$

Since the distribution of ω and σ depend on x and y, and since they interact in the production technology, there is scope for sorting on ex-ante traits in the matching market. An agent with a higher ex-post agent characteristic ω produces more output per unit of effort. Whether the assignment of managers to firms exhibits positive or negative assortative matching depends on the properties of the technology.

A matching in our setting is a one-to-one measure-preserving function $\mu : [0, 1] \to [0, 1]$. As is standard in the literature on sorting, we focus on monotone matching. There is positive sorting when μ is increasing, in which case it solves $\Psi(\mu(x)) = \Gamma(x)$ for all x, and negative sorting when μ is decreasing, in which case $1 - \Psi(\mu(x)) = \Gamma(x)$ for all x.[9] Given our assumptions on Ψ and Γ and the assumption of an equal measure of firms and managers, monotonicity of μ is strict in each case.

A competitive equilibrium in this setting consists of a matching function μ, a reservation wage function a, a compensation scheme (contracting wage function) w, and an effort function e, such that: (i) each firm of ex-ante trait y maximizes expected profits by offering w given a to a manager with ex-ante trait x; (ii) each manager with ex-ante trait x and ex-post attribute ω

[8]We show below that these complementarities play a crucial role in the equilibrium sorting pattern.

[9]In principle, μ can be a random variable. Since our interest is in monotone sorting, we define it as a function. Also, note that a matching is a mapping between ex-ante traits instead of firms and managers. But in equilibrium, firms or managers with the same traits behave in the same way. Hence there is no loss of generality in defining a matching between attributes.

chooses effort according to e when matched with a firm with ex-ante trait y that offers w; (iii) given a, each firm with ex-ante trait y maximizes its expected profits by choosing a manager x such that $y = \mu(x)$ and offering w; and (iv) the matching μ clears the market.[10]

3 Equilibrium Analysis

We now solve for the competitive equilibrium. The proof is constructive. We first solve the contracting problem for a given pair consisting of a manager with ex-ante trait x and a firm with ex-ante trait y, which takes as given the outside option $a(x)$ that the manager can obtain in the market. We then pin down the outside option function a and the matching function μ that satisfy the equilibrium requirements described above.

Optimal Contract. Fix a firm-manager pair with ex-ante traits x and y. Since the compensation scheme is characterized by the "slope" and "intercept" functions α and β, respectively, the principal's contracting problem consists of choosing these functions plus a recommended effort function to maximize expected profits:

$$\Pi(y,x,a(x)) = \max_{\alpha,\beta,e} \int_0^\infty \int_{-\infty}^\infty (\omega(e(\omega)+\sigma) - (\beta(\omega)+\alpha(\omega)\omega(e(\omega)+\sigma)))\, dG(\sigma|y)dF(\omega|x) \tag{4}$$

$$\text{s.t.} \int_0^\infty \int_{-\infty}^\infty \left(-\mathrm{e}^{-r\left(\beta(\omega)+\alpha(\omega)\omega(e(\omega)+\sigma)-\frac{e(\omega)^2}{2}\right)}\right) dG(\sigma|y)dF(\omega|x) \geq -\mathrm{e}^{-ra(x)} \tag{5}$$

$$e(\omega) \in \mathrm{argmax}_{\hat{e}} \int_{-\infty}^\infty \left(-\mathrm{e}^{-r\left(\beta(\omega)+\alpha(\omega)\omega(\hat{e}+\sigma)-\frac{\hat{e}^2}{2}\right)}\right) dG(\sigma|y) \quad \forall\, \omega, \tag{6}$$

where the integrand in the objective function is output (recall that $q = \omega(e+\sigma)$) minus the linear wage function, (5) is the participation constraint, and (6) is the incentive constraint.

Note that the participation constraint only needs to hold in expectation, since at the time of contracting ω is not known. In turn, the incentive constraint must hold for each realization of ω, since this is known by the time the agent chooses effort. Despite the apparent complication of the dependence of the payoffs on the attribute ω and the traits (x, y), the solution to the

[10] It is well-known that the competitive equilibrium, the core and the stable matching of an assignment game (Shapley and Shubik (1972)) are "equivalent" in terms of their allocations (see Gretsky et al. (1992)). This equivalence will apply to our model since, as we will see below, it is transferable-utility representable.

problem turns out to be similar to the solution in Section 5 of Holmström and Milgrom (1987):

Proposition 1 (Optimal Contract)**.** *For any firm-manager pair with ex-ante traits (x, y) and manager's outside option $a(x)$, the firm's optimal contract is the triple (α, β, e) given by*

$$\alpha(\omega) = \frac{1}{1 + rs_\sigma^2}, \quad \beta(\omega) = a(x) - \frac{\omega t(y)}{1 + rs_\sigma^2} + \frac{\omega^2}{2\left(1 + rs_\sigma^2\right)^2}\left(rs_\sigma^2 - 1\right), \quad e(\omega) = \frac{\omega}{1 + rs_\sigma^2}, \ \forall \omega,$$

where α is the slope of the contract, β is the intercept, and e the recommended effort function.

The proof is in the appendix. Note that the slope is independent of ω, so all ex-post attributes of the manager are subject to the same constant incentive pressure. In turn, the optimal effort function is increasing in ω: the higher the manager's ex-post attribute is, the larger is the expected impact on output for each unit of effort, and thus the more effort the manager is required to exert. The impact of risk aversion and the variance of σ on the contract slope and effort are as in Holmström and Milgrom (1987). Finally, the intercept function β ensures that the participation constraint holds with equality.

The most important difference with the Holmström and Milgrom (1987) contract is the outside option $a(x)$, which is determined in equilibrium by competitive forces. As we will show below, the equilibrium outside option $a(x)$ depends on the sorting pattern (which manager matches with which firm), and the contracts that competing firms offer.

SORTING PATTERN. Since the wage function is $w(q, \omega) = \beta(\omega) + \alpha(\omega)q$ for each realization of q and ω, we can use the expressions for α and β in Proposition 1 to obtain

$$w(q, \omega) = a(x) + \frac{e^2(\omega)}{2} + re^2(\omega)\frac{s_\sigma^2}{2} + e(\omega)\varepsilon_\sigma.$$

Except for the dependence on ω and the endogenous outside option $a(x)$, this formula is standard. In words, it states that for each ω the compensation scheme consists of the outside option, the disutility of effort, and a risk premium for the uncertainty inherent in the contract due to the random variable σ that the manager is asked to bear, plus a stochastic term due to the shock ε_σ.[11] Define $\pi(q, \omega) \equiv q - w(q, \omega)$ as the firm's profit for each realization of q and ω. Using

[11] If effort was observable (no moral hazard), then the wage would be a constant for each ω and given by $a(x) + (e^2(\omega)/2)$, as intuition suggests.

$q = \omega(e(\omega) + \sigma)$, $\sigma = t(y) + \varepsilon_\sigma$, and the expression for e from Proposition 1, we obtain

$$w(q, \omega) = a(x) + \frac{\omega^2}{2\,(1 + rs_\sigma^2)} + \frac{\omega \varepsilon_\sigma}{1 + rs_\sigma^2}, \tag{7}$$

$$\pi(q, \omega) = \omega t(y) + \frac{\omega^2}{2\,(1 + rs_\sigma^2)} + \frac{rs_\sigma^2 \omega \varepsilon_\sigma}{1 + rs_\sigma^2} - a(x). \tag{8}$$

Note that w depends on x through a and also through the distribution of ω, while π depends also on y through t. The second term in (7) combines both the part of the wage that compensates the manager for the disutility of effort and the part that compensates for the risk the manager bears. In turn, the second term in (8) is explained as follows: effort increases π via q through $\omega e(\omega) = \omega^2/(1 + rs_\sigma^2)$, but decreases π via w by half this amount, so the net effect of e on π is the second term in (8).

Since $\Pi(y, x, a(x)) = \mathbb{E}[\pi(q, \omega)]$, where the expectation is taken with respect to the distribution of ω and σ. Recalling that ε_σ has mean zero and that $k(x) = \mathbb{E}[\omega|x]$ and so $\mathbb{E}[\omega^2|x] = k^2(x) + s_\omega^2$, we obtain

$$\Pi(y, x, a(x)) = k(x)t(y) + \frac{k^2(x) + s_\omega^2}{2\,(1 + rs_\sigma^2)} - a(x) \equiv V(x, y) - a(x), \tag{9}$$

where to simplify the notation we have set

$$V(x, y) \equiv k(x)t(y) + \frac{k^2(x) + s_\omega^2}{2\,(1 + rs_\sigma^2)}. \tag{10}$$

Two implications are evident from equations (9). First, the outside option enters the firm's expected profit $\Pi = V - a$ in an additively separable way and does not depend on y. As a result, matching takes place under transferable utility, and thus can be analyzed as in Becker (1973) with V playing the role of the match output function. The transferable utility representation is due to CARA, normality, and the restriction to linear contracts, which are assumptions inherent in the Holmström and Milgrom (1987) model. Second, Π is supermodular in x and y if and only if V supermodular in x and y, which holds when k and t are increasing in x and y.

In the matching stage, firms choose the manager with ex-ante trait x that maximizes expected profits (9): that is, each firm with ex-ante trait y solves

$$\max_{x \in [0,1]} (V(x,y) - a(x)).$$

Given the additive separability between the outside option $a(x)$ and $V(x,y)$, the equilibrium sorting pattern, that is, whether there is positive or negative sorting, entirely depends on the properties of the function $V(x,y)$. We thus have the following result:

Proposition 2 (Positive Sorting)**.** *Assume that k and t are strictly increasing. Then, positive sorting is optimal, and any equilibrium exhibits positive sorting.*

It is clear from (9) that $V_{xy} > 0$ under the proposition's premise. That positive sorting is optimal then follows from a simple rearrangement argument: if positive sorting fails on a set of (x,y) of positive measure, then rematching these agents positively assortatively increases the aggregate payoff due to $V_{xy} > 0$, which in a transferable utility world leads to a welfare improvement. If Π is supermodular but not strictly so, then there could be other equilibria that fail positive sorting, but that are payoff-equivalent to a positive sorting one (see Legros and Newman (2007)). When k and t are strictly increasing, however, Π is strictly supermodular and thus equilibrium *must* exhibit positive sorting.[12]

COMPETITIVE EQUILIBRIUM. For each firm-manager pair with (x,y) we have obtained the optimal contract that regulates their interaction. Moreover, we have shown that when k and t are strictly increasing, then payoffs exhibit complementarities in ex-ante attributes and thus equilibrium exhibits positive sorting. To complete the derivation of a competitive equilibrium we must compute the equilibrium outside-option function a that clears the market along with the positive sorting matching function.

Consider a firm with y choosing a worker with characteristic x, taking as given $a(x)$. The firm's solution to its maximization problem $\max_x(V(x,y) - a(x))$ satisfies the first-order condition at an interior solution $V_x(x,y) = a'(x)$. Under positive sorting, which obtains if k and t are strictly increasing, this condition holds when $y = \mu(x)$, where $\mu(x) = \Psi^{-1}(\Gamma(x))$ for all x,

[12] If instead of our assumption that k and t are increasing we assume that they are monotone in opposite directions, then one can follow the same steps and derive an analogous result to Proposition 2 entailing negative sorting. We omit the proof.

and thus we have $V_x(x, \mu(x)) = a'(x)$ for all x. It follows that

$$a(x) = a(0) + \int_0^x V_x(s, \mu(s))ds, \ \ \forall x, \tag{11}$$

where $a(0)$ is a constant that reflects the outside option of the lowest ex-ante trait worker.[13] The function a is the "hedonic price" schedule derived from the value of contracts offered by competing firms. As in the standard matching model à la Becker, the rate of change of the outside option of a manager with ex-ante trait x is determined by the marginal contribution she makes to the value of the firm (gross of the outside option) given by V.

We can now write the solution to the outside option more explicitly by using expression (9) for $V(x, y)$ and $V_x(x, y) = k'(x)t(y) + (2k(x)k'(x))/(1 + rs_\sigma^2)$, which yields

$$a(x) = a(0) + \int_0^x \left(k'(z)t(\mu(z)) + \frac{2k(z)k'(z)}{1 + rs_\sigma^2} \right) dz. \tag{12}$$

We thus have the following result:

Proposition 3 (Competitive Equilibrium)**.** *Assume that k and t are strictly increasing functions. Then the following matching function μ, outside-option function a, compensation scheme w, and effort function e constitute a competitive equilibrium:*

$$\mu(x) = \Psi^{-1}(\Gamma(x)), \tag{13}$$

$$a(x) = a(0) + \int_0^x \left(k'(z)t(\mu(z)) + \frac{2k(z)k'(z)}{1 + rs_\sigma^2} \right) dz, \tag{14}$$

$$w(q, \omega) = a(0) + \int_0^x \left(k'(z)t(\mu(z)) + \frac{2k(z)k'(z)}{1 + rs_\sigma^2} \right) dz + \frac{\omega^2}{2\,(1 + rs_\sigma{}^2)} + \frac{\omega \varepsilon_\sigma}{1 + rs_\sigma{}^2}, \tag{15}$$

$$e(\omega) = \frac{\omega}{1 + rs_\sigma^2}, \tag{16}$$

with payoffs $-\mathrm{e}^{-ra(x)}$ for each manager with ex-ante trait x, and $V(\mu^{-1}(y), y) - a(\mu^{-1}(y))$ for each firm with ex-ante trait y. Equilibrium is unique up to the constant $a(0)$.

Intuitively, it is clear that if a firm with ex-ante trait finds it optimal to hire a manager with ex-ante trait $\mu^{-1}(y)$, then by Proposition 1 the firm will offer the contract consisting of (15)–(16). So to complete the proof of Proposition 3, it remains to show that, taking as given the

[13] The constant $a(0)$ satisfies $0 \leq a(0) \leq V(0, 0) = k(0)t(0) + (k^2(0) + \sigma_s^2)/2(1 + rs_\sigma^2)$.

payoff schedule a given by (14), each firm with ex-ante trait y finds it uniquely optimal to hire a manager with ex-ante trait $x = \mu^{-1}(y)$. This follows from the strict supermodularity of V when k and t are strictly increasing. To see this, let $x = \mu^{-1}(y)$ and consider the payoff for a firm with ex-ante trait y from matching instead with a manager with ex-ante trait $x' < x$ (the other case is similar); then

$$\begin{aligned}\Pi(y,x,a(x)) - \Pi(y,x',a(x')) =& \left(V(x,y) - a(0) - \int_0^x V_x(s,\mu(s))ds\right) \\ &- \left(V(x',y) - a(0) - \int_0^{x'} V_x(s,\mu(s))ds\right) \\ =& V(x,y) - V(x',y) - \int_{x'}^{x} V_x(s,\mu(s))ds \\ =& \int_{x'}^{x} V_x(s,y)ds - \int_{x'}^{x} V_x(s,\mu(x'))ds > 0,\end{aligned}$$

where the second equality follows by simplification, and the inequality by μ strictly increasing, V strictly supermodular, and $x' < x$ (which implies that $\mu(x') = y' < y$ under positive sorting). Finally, note that the only multiplicity that can arise is due to the fact that $a(0)$ is not uniquely pinned down.

CONVEXITY OF MANAGERIAL WAGES. In equilibrium, a manager with ex-ante trait x obtains $a(x)$. A natural question that will play an important role below is whether a is convex in x. The following proposition provides a simple answer.

Proposition 4 (Convexity of Wages)**.** *If k is strictly increasing and convex in x and t is strictly increasing in y, then a is convex in x.*

The proof is immediate from (12): simply differentiate a with respect to x. Using the properties of k, t, and positive sorting, the resulting expression is increasing in x.

In words, changes in x leads to increasingly larger changes in compensation a, and this property exacerbates the *superstar effect*.

DECOMPOSITION: INCENTIVES VERSUS SORTING. From the wage equation (15), we can take expectations with respect to both ε_σ and ω and obtain the expected wage for a manager with ex-ante trait x, $W(x, a(x)) \equiv \mathbb{E}[w(q, \omega)]$, given by

$$W(x, a(x)) = a(x) + \frac{k^2(x) + s_\omega^2}{2\,(1 + r s_\sigma^2)}. \tag{17}$$

Since our model combines a principal-agent problem à la Holmström and Milgrom (1987) and a matching problem as in Gabaix and Landier (2008), we can decompose wages and profits in the above expressions as a sum of terms due to moral hazard (bilateral contracting) and to sorting (competition). The expected wage of a manager with ex-ante trait x consists of the outside option $a(x)$ plus an expression that summarizes the value of bilateral contracting. The latter is the compensation for the disutility of effort manager incurs with the matched firm and the risk premium that she demands in order to bear the risk in her compensation (due to moral hazard). That is,

$$W(x, a(x)) = \underbrace{a(0) + \int_0^x k'(z) t(\mu(z)) dz + \int_0^x \frac{2k(z)k'(z)}{1 + r s_\sigma^2} dz}_{\text{Competition}} + \underbrace{\frac{k^2(x) + s_\omega^2}{2(1 + r s_\sigma^2)}}_{\text{Bilateral Contracting}}. \tag{18}$$

Due to sorting in the market, the competition component reflects the marginal contribution of all traits below x. The first integral is due to the marginal contribution to direct output that is complementary in manager type and firm productivity; the second term is also due to competition and complementarity, but it measures the outside option of the manager from getting a better incentive contract with a competing firm. This second integral term in the outside option is due to competition in contracts, and is new. This component is absent in a standard principal-agent problem (as in Holmström and Milgrom (1987)) because there is no competition, and it is absent in the competitive matching setting without moral hazard (as in Gabaix and Landier (2008)). It captures the fact that firms not only need to offer compensation to induce optimal effort, they need to offer a fixed component to stave off the competition and induces the worker to choose to match with the firm and avoiding being poached by a competitor.

In the application below in Section 4, once we have estimated the model, we measure the percentage contribution of each of the two components to managers' wages: the share due to bilateral contracting and the share due to competition.

We can similarly decompose expected profit of a firm with ex-ante trait y matched with a manager with ex-ante trait x:

$$\Pi(y,x,a(x)) = k(x)t(y) + \frac{k^2(x)+s_\omega^2}{2\left(1+rs_\sigma^2\right)} - \left(a(0) + \int_0^x k'(z)t(\mu(z))dz + \int_0^x \frac{2k(z)k'(z)}{1+rs_\sigma^2}dz\right). \quad (19)$$

The first term reflects deterministic output with complementarities in production and which is the key to sorting. The second term emerges as the firm's share of the compensation due to the bilateral contract. The third term in brackets is the competition component that the firm has to pay to compensate for the manager's outside option and consists of the direct outside option as well as the option of a competing contract.

Finally, since for each realization of ω and σ, output is the sum of wages and profits, the expected value of output for a firm with ex-ante trait y that matches with a manager of ex-ante trait x is $Q(x,y) \equiv W(x,a(x)) + \Pi(y,x,a(x))$, and thus

$$Q(x,y) = k(x)t(y) + \frac{k^2(x)+s_\omega^2}{1+rs_\sigma^2}. \quad (20)$$

The first term reflects the complementarities in production of the ex-ante traits of the matched pair, while the second term reflects the effect due to the induced level of effort at an optimal contract.[14]

4 Application: Executive Compensation

In this section, we apply our analytical framework to shed light on executive compensation and the matching of CEOs to firms. We use data on the compensation of CEOs and the profitability of firms to estimate the parameters, given the structure and equilibrium of the model. We obtain estimates for the determinants of wages, profits and output. With the estimated model, we can perform counterfactuals and most importantly, we can determine the extent to which differences in CEO compensation is driven by optimal contracting or by competition.

[14]Note that $Q(x,y) \neq V(x,y)$. Despite the transferable utility feature of our model, the presence of risk aversion and moral hazard makes this model different from the standard matching problem in Becker (1973) with quasilinear utility, where the firm's expected profit would be simply $Q(x,y) - a(x)$, while in our case is $V(x,y) - a(x) = Q(x,y) - W(x,a(x))$.

DATA. We use US Data on Executive Compensation and Firm Profits. We construct data on CEO compensation by imputing the CEO's job performance in the firm's stock market value.[15] To be consistent with our model, we exclusively consider new hires and we do not envisage rematching or separation. We further assume that output, wages and profits are determined as in our principal-agent model.

The data is from the Execucomp (Compustat) database for US publicly traded firms. We construct the sample by selecting all newly hired CEOs during 2015. We then construct wages and profits based on the period 2016-2017. This gives us a sample of $n = 139$ observations.[16] The two-year interval (2016-2017) that we have chosen reflects the tradeoff between choosing a longer interval, which would give us a better estimate of the ex-post outcome of the match, and a shorter interval, which ensures that fewer CEO-firm matches are separated endogenously.[17]

MAPPING THE DATA TO THE MODEL. To construct the firm ex-ante trait y, we rank firms by their sales as of December 31, 2015.[18] We define the ex-ante characteristic y as the rank of market capitalization divided by the number of firms. As a result, the distribution of firms is uniform on $[0, 1]$.

Under the assumption of frictionless matching and positive sorting, we rank workers by the firm they are matched with in the data: $x = y$. As a result, also x is uniformly distributed on $[0, 1]$. This is a normalization, since we use no other ex-ante information on x. It is well known that the sorting pattern of a one-to-one matching model as in Becker (1973) is normalized relative to the distribution of ex-ante traits x and y. Once we have estimated the model, we will

[15] It is well known that data on payoffs to both sides of the market are hard to come by. Even in matched employer-employee data obtained from exhaustive administrative sources there is usually no good information on the productivity at the job level. This has lead to a recent literature identifying sorting from wage data alone (see Eeckhout and Kircher (2011), and Bonhomme et al. (2019)) .

[16] In the entire Execucomp-Compustat sample there are 1,913 firm-CEO pairs in 2016, of which 224 were newly formed during 2015. After dropping matches with missing observations for at least one of the variables we use, with outliers and matches that separated before the end of 2017, we are left with a sample of 139 matched firm-CEO pairs.

[17] Most CEO tenures last well beyond two years. In our sample, there are 33 separations within that two year window. Our model does not incorporate endogenous separations, either due to a bad realization of the CEO performance that leads to his firing, or due to a good one that leads to poaching by a more productive competitor. Since attrition would alter the continuation value and therefore the value of match formation, it could induce a bias in the estimates, especially in the variance of the estimated ex-post heterogeneity (which would be lower in the presence of endogenous separation since extreme realizations are likely to lead to separation). But it is not clear whether and if so in which direction there would be a bias in the mean of the ex-post CEO attribute since both too bad and too good CEOs would separate.

[18] In the Appendix C.1 we report the results where we rank firms by their market capitalization. Our results are robust and we find no substantial difference in our results.

verify that indeed the expected value of a match $V(x,y)$ is supermodular, thus justifying our identifying assumption that there is positive sorting.

For wages we use total compensation – denoted by the variable TDC1 in Execucomp – which includes salary, bonus, restricted stocks, stock options, and long-term incentive payouts. For job performance we use the standard measure of economic profits[19] of the firm (from Compustat) – as measured by Sales minus Cost of Goods Sold (COGS), minus Selling, General & Administrative Expenses (SG&A), minus the user cost of capital over the course of 2016 and 2017.[20] Over this period of two years, our wage variable w is then TDC1(2016)+TDC1(2017), and our profit variable π is Profits(2016)+Profits(2017). Observe that wages are always positive in the sample, and that profits take on both negative and positive values.

ESTIMATION PROCEDURE. The model predicts a relation between observable outcomes (wages w, profits π and matched pairs $y = \mu(x)$) and unobservable primitives (distributions of ex-post attributes ω, σ and technology). As in the model, $G(\cdot|y)$ is normal with mean $t(y)$ and variance s_σ^2. We assume that also $F(\cdot|x)$ is normal with $\omega = k(x) + \varepsilon_\omega$, where ε_ω is normally distributed with mean 0 and variance s_ω^2.[21] It is important to repeat here that each worker ex-ante trait x has a trait-dependent distribution of the ex-post attribute ω. Likewise, each firm with ex-ante trait y has their own trait-dependent distribution of the ex-post attribute σ.

The model also generates an output level q for each (x,y) and realization of (ω,σ), where q is the sum of wages and profits (equations (7) and (8)). Note that once we know $k(x), s_\omega, t(y), s_\sigma$, we know both the distributions F and G and the ex-ante value $V(x,y)$.

Our estimation procedure starts by positing positive sorting. Given we assume Γ and Ψ are distributed identically (uniform on $[0,1]$) this implies that $\mu(x) = x$, or $y = x$. In what follows, we will often express all the primitives along the equilibrium allocation $y = x$ in terms of x:

[19] We have also performed our analysis using the change in market value as a measure of job performance. The results are surprisingly similar. We nonetheless decided to use economic profits because market value includes an important component of expectations of future profits which may be beyond the performance of the CEO.

[20] We construct the user cost of capital, often denoted by RK, following the standard procedure in the literature. We use gross capital (PPEGT) for the capital stock K, and scale the capital stock by R that consists of the nominal interest rate (federal funds rate), the GDP implicit price deflator, and an exogenous depreciation plus risk premium that we set at 12%.

[21] Normality of ω is useful for estimation purposes because it simplifies the algebra considerably. Of course, ω in this case can be negative. As we explain in the appendix after the proof of Proposition 1, we can easily accommodate negative values of ω, for which optimal effort is zero. Moreover, this only occurs in a set of small probability when the mean of ω given x is large for all x. Indeed, in our estimation below, we do find that k is large and that the probability of negative ω is negligible. For robustness, we also estimate the model under log-normal $F(\cdot|x)$, and we find that the results are similar (see Appendix C.3).

$k(x), s_\omega, t(x), s_\sigma$. Once we have obtained the estimates for these primitives, we need to verify whether the obtained function V is indeed supermodular so as to validate the positive sorting assumption.

The challenge is that the primitives are functions of x, and for each x we have one observation for wages and profits for each matched pair. For the estimation, we therefore assume a parametric functional form of the dependence of the distributions on the ex-ante trait x and $y : k(x), s_\omega, t(y), s_\sigma$. We will use the functional form for $k(x)$ and $t(x)$:

$$\theta(x) = \theta_0 \, \mathrm{e}^{\,\theta_1 x}, \;\; \theta \in \{k, t\}. \tag{21}$$

This functional form has the advantage of delivering flexibility in terms of the curvature, yet with limited parameters.[22] In addition to the distribution parameters, we estimate the exogenous preference parameters r (degree of risk aversion) and $a(0)$ (the outside option of the lowest ex-ante trait). We will therefore estimate 8 parameters $\Theta = \{k_0, k_1, s_\omega, t_0, t_1, s_\sigma, r, a(0)\}$ that fully specify the model.

We use maximum likelihood to estimate the model. We invert (7) and (8) to obtain expressions for ω and σ and construct the log-likelihood of the joint normal and lognormal distributions in σ and ω. We report the deviation of the log-likelihood function in the Appendix B. We obtain the parameter estimates from the global maximum estimation that uses the Sobol sequence procedure with 100,000 Sobol points.

ESTIMATION RESULTS. Our objective is to back out the distributions $F(\omega|x)$ for all x and $G(\sigma|y)$ for all y, as well as the technology q, the expected wages $\mathbb{E}w(x)$ and the outside option $a(x)$, the expected profits $\mathbb{E}\pi(y)$ and the ex-ante match surplus function $V(x, y)$. The estimates for the parameters Θ fully determine these underlying primitives and the equilibrium objects derived from them.

Table 1 reports the parameter estimates together with the bootstrap confidence intervals. To better visualize how those parameters change across ex-ante traits of firms y and managers x, we plot the means and standard deviations.

[22]For our purpose, the exponential is the most effective and parsimonious functional form. We have also estimated the model with the polynomial functions $\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$, but the fit is worse and we need more parameters to capture the relevant curvature. See Appendix C.2.

Manager		
k_0	k_1	s_ω
1276154	0.0208	491742
[668640, 1300926]	[0.0165, 0.0298]	[262364, 554376]
Firm		
t_0	t_1	s_σ
15.58	5.76	1028
[11.88, 47.47]	[4.58, 6.40]	[920, 1657]
Risk aversion	Outside option	
r	$a(0)$	
0.0943	480194	
[0.0121, 0.0956]	[99508, 860898]	

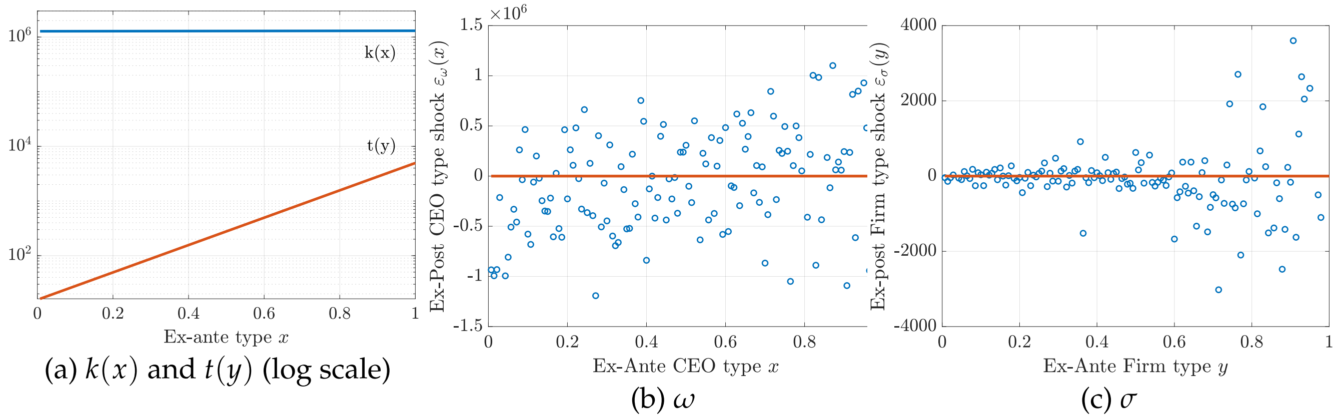
Note: 90% bootstrapped confidence intervals in brackets (1000 iterations).

Table 1: Estimated Parameters Θ.

Figure 1a plots the estimated means $k(x)$ and $t(y)$ of the distributions of ex-post manager and firm characteristics ω and σ. Higher ranked firms generate a lot more match surplus — $t(y)$ is steeply increasing —; the mean of σ of the highest ranked firms is one hundred times larger than that of the lowest ranked firms, going from just above 10 to nearly 1000. This matters predominantly for the competitive component of wages. We turn to this below. The standard deviation of σ is over 210. This indicates first that there is substantial uncertainty regarding the firm productivity, which matters nearly exclusively for the optimal wage contract. The higher the variance, the less the weight on the incentive component of the manager's compensation. Instead, $k(x)$ is increasing but only barely ($k_1 = 0.018$). What the Figure highlights is that the average manager ex-ante trait is substantially higher than the firm ex-ante trait, but it barely changes with x.

Figure 1 shows the backed out values for ω and σ, using the estimated parameters and the implied latent characteristics of managers and firms.

Next, Figure 2a plots the expected wage in expression (7) for the estimated parameters together with the wages observed in the data and on which the estimation is based. The maximum likelihood estimation targets the distributions of ω and σ and not the fit of wages and profits, so it is remarkable how well the model fits the wage data even if not targeted. Figure 2b plots the outside option $a(x)$ from equation (12) and evaluated at the estimated parameters. In line with Proposition 4, a is clearly convex in x, since the estimates of the parameters reveal

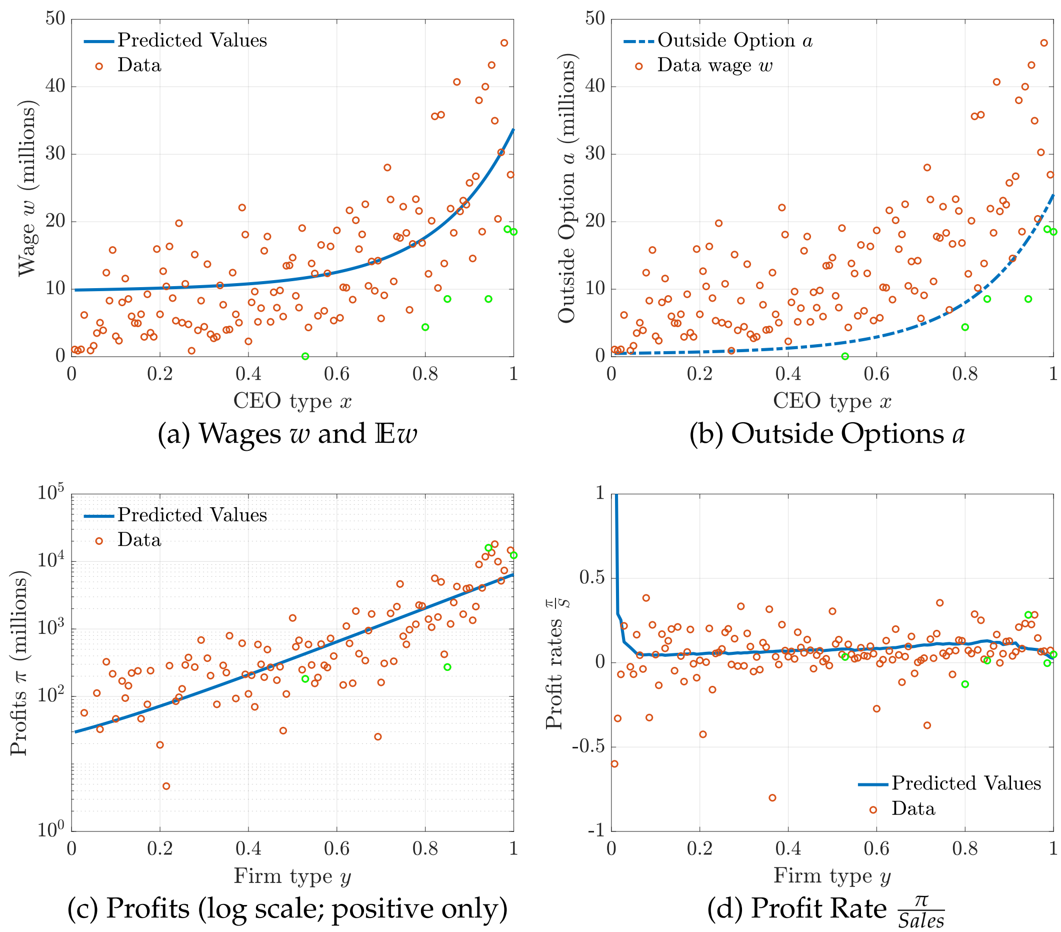


(a) $k(x)$ and $t(y)$ (log scale)

(b) ω

(c) σ

Figure 1: Estimated means $k(x)$ and $t(y)$, and realized values of ε_ω and ε_σ



(a) Wages w and $\mathbb{E}w$

(b) Outside Options a

(c) Profits (log scale; positive only)

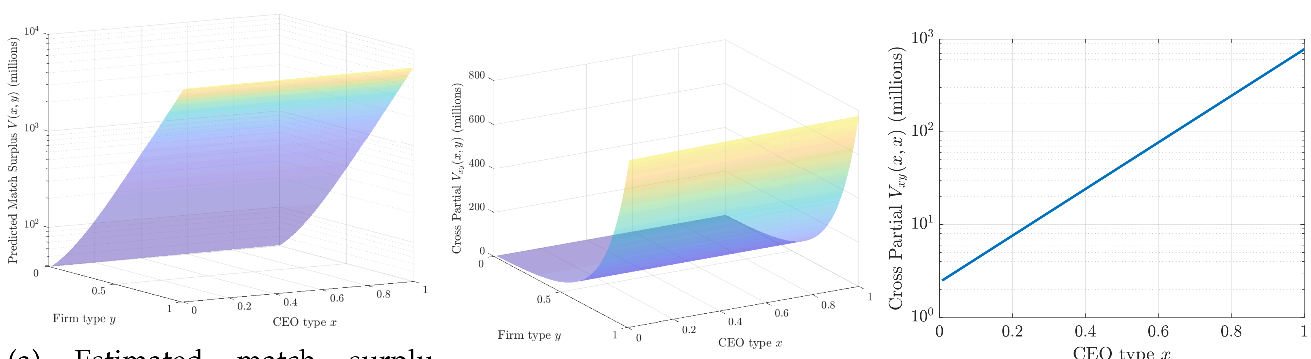
(d) Profit Rate $\frac{\pi}{Sales}$

Figure 2: Wages, Outside Options and Profits: Data and Estimates.

that k and t satisfy the premise of the proposition.

The sharp rise in $t(y)$, the average firm productivity, leads to a sharp increase in wages, even though $k(x)$ barely increases. Mere inspection of the wage formula (7) shows an apparent independence of wages on $t(y)$. But that is misleading since $t(y)$ enters in the outside option $a(x)$ which is determined through competition. This is the standard superstar motive (see Rosen (1981), Gabaix and Landier (2008), Terviö (2008)) that leads to extremely steep wage schedules in worker productivity. Even though expected CEO ex-post attribute ω does not vary much in x, their compensation changes dramatically from less than one million to over 40 million. Given complementarities in productivity between manager and firms, from the superstar effect, small differences in manager abilities ω get blown out of proportion due to the huge differences in the average firm productivity σ.

Figures 2c and 2d plot two different ways to visualize the profits. Because some firms have negative profits – in the data and also in the model since the normal distribution of σ has negative realizations – and profits in levels are extremely disperse, we plot profits in logs (only for those firms with positive profits) in Figure 2c as well as the profit rate as a share of sales in Figures 2d. Again, this moment, which was not targeted, fits the data quite well over the entire distribution y. Except for firms with the lowest y, the expected profit rate is remarkably constant.



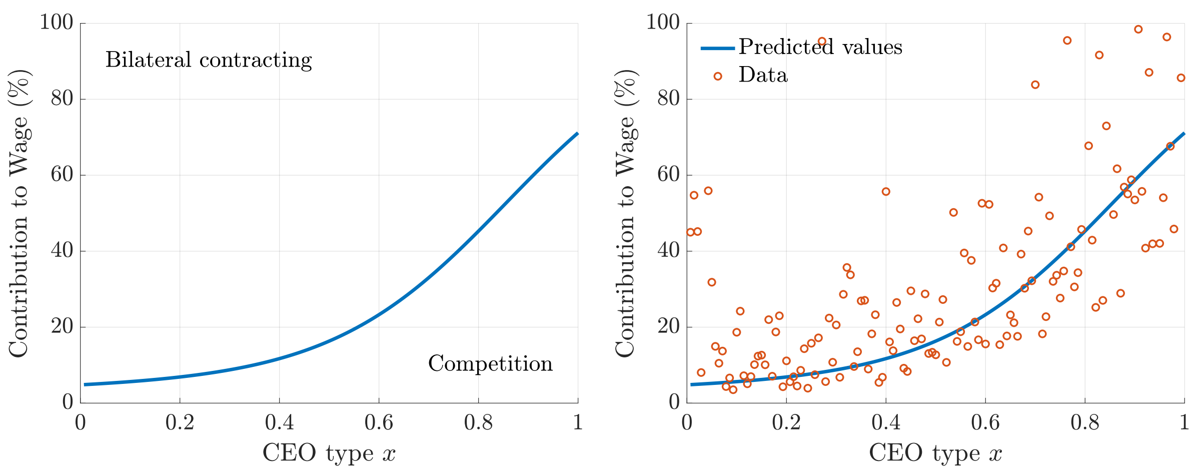
(a) Estimated match surplus $V(x,y)$ (b) Complementarity $V_{xy}(x,y)$ (c) Complementarity at $y = x$

Figure 3: Estimated Match Surplus.

Below we decompose the contribution of competition and incentives to wages, but first we analyze the estimated match surplus. The ex-post output is equal to the sum of wages and profits. With the estimated parameters, we can now also construct the ex-ante firm value V from (see equation (9) for its expression) as well as its cross-partial V_{xy}. While we only have observations along the equilibrium allocation where $x = y$, the estimated model parameters allow us to reconstruct V over the entire domain of $(x,y) \in \mathbb{R}_+^2$. Figure 3 plots both the estimated V and V_{xy} in three dimensions, as well as $V_{xy}(x,x)$ along the equilibrium allocation $x = y$ in two dimensions.

Figure 3a reveals that V is increasing both in CEO ex-ante trait x and in firm ex-ante trait y. Moreover, from the convexity of the plot it appears that there are complementarities and that V is supermodular. To verify this, we evaluate V_{xy} at the estimated parameters. The plot of the cross-partial derivative in Figure 3b confirms that over the entire domain of x and y, V is supermodular, which is also confirmed along the equilibrium allocation (Figure 3c). Supermodularity of V is crucial for our estimation strategy, since we estimated the model *assuming* positive sorting, which allowed us to start from the premise that $x = y$. This identifying assumption is justified so long as the estimated technology is indeed supermodular, as we have just shown is the case.

DECOMPOSITION: INCENTIVES VERSUS COMPETITION. We now turn to the main question that motivates the paper: do incentives or competition determine manager's wages? We use the decomposition in (18) and express the share of competition for each x in percentage terms, while the bilateral contracting share is the residual. Figure 4a plots the contribution to expected wages of the three components, and Figure 4b does the same but for the realized wages.



(a) Share of Competition in Expected Wages (b) Share of Competition in Realized Wages

Figure 4: Decomposition of Wages

We find that the decomposition of wages into the contracting and competition components varies by the ex-ante ranking of the manager. At the top, of the distribution, contracts are predominantly determined by the superstar phenomenon as in Gabaix and Landier (2008). Top firms derive the highest marginal value from hiring the best managers, and they need to pay those managers high wages because the firms's competitors are willing to make competing offers. Not surprisingly, the superstar component dominates the compensation of the superstar managers, and accounts for 70% of the salary. Here we are talking about firms that are extremely valuable. For example, among the firms in our sample that hired a new CEO in 2015 and are at the top include Boeing (sales $200 billion; market value $105 billion; employment 161 thousand) whose CEO Dennis A. Muilenburg received a total compensation over 2016 and 2017 of over 11 million. Other firms at the top in our sample are Hewlett Packard, MdDonald's, and Twenty-first Century Fox. For those top firms, the superstar component due to competition dominates the CEO's income by 70%.

Instead, at the bottom of the distribution, the competition component is much smaller and equal to only 10%. In fact, at the very bottom it is equal to the outside option for the lowest manager ex-ante trait $a(0)$ that we estimate to be equal to $849,000. It is important to keep in mind that we are talking here about very small firms. There are quite a few small publicly traded real estate and health care companies.[23] For those firms the superstar component in compensation is virtually non-existent, simply because those firms are not very valuable and their CEOs are no superstars. Instead, the main component in these executives' compensation is due to incentive provision.

The variation in the superstar component tells us that impact of competition varies enormously across the distribution of firms. In fact, we already knew this from observing the pattern of outside option a as x varies in Figure 2b. But that does not mean that the bilateral contracting component is not important. For one, it accounts for nearly all of the compensation at the bottom of the distribution. But also at the top of the distribution, the monetary value in levels is large and sizable. In fact, the bilateral contracting component is fairly constant along the distribution, and its expected value is equal to around $8 million. Irrespective of their rank x, managers get paid on average a similar incentive portion. The only reason why top CEOs earn higher salaries is because they receive a much higher competition component, which at the top is equal to over $25 million. Firms need to pay managers the wage they could obtain from signing with competing firms or begin poached by them. Due to competitive pressure, firms need to offer terms that match the contracts the managers can obtain from competing

[23]For example, the land development firm St Joe Company (sales $207 million; market value $1.5 billion; 55 employees) is a small company whose CEO received $1.3 million in total compensation over the span of 2016 and 2017. Other firms at the bottom of the distribution of our sample are Prestige Consumer Healthcare (259 employees), the biotech firm Repligen Corporation (168 employees) and EPR Properties (49 employees).

firms. This competing contracts component does not increase the incentives, it merely offers the managers the value of the incentive contract they can obtain elsewhere.

In sum, without moral hazard, total compensation would be due to competition and none would depend on contracting. We find that at the bottom of the distribution, moral hazard is responsible for nearly all compensation, and that the dollar amount of this moral hazard component is constant across the entire distribution. Moving up the distribution, that moral hazard component is supplemented with a component from competition that is increasing in the CEO type. For the superstar executive, that competition component reaches 70% while the bilateral contracting component accounts for 30%.

ROBUSTNESS. Throughout the quantitative exercise, we have made choices on the assumptions of our model that we consider the most natural and that have been used in the literature. Here, we discuss a number of robustness exercises with other specifications of the model. The estimation results are reported in Appendix C.

First, we use a different way to rank firms. Instead of the sales of the firm, we use market values (Section C.1). There may be a concern that sales do not adequately measure the productivity of a firm as technology can generate high value output (and hence high sales) but little value added. The idea is that market value is a better measure of the value of a firm.

Second, instead of the exponential parametric form to discipline the dependence of the functions k and t of ex-ante traits, we use a polynomial of degree two (Section C.2). Clearly, the functional form can matter, but both the exponential and the polynomial have sufficient flexibility to represent similar functional relationships.

Third, we change the parametric distribution of workers ex-post attributes. Instead of the normal distribution, in Section C.3 we estimate the model assuming manager ex-post attributes are lognormally distributed. This permits more skewed realizations and may impact the estimated parameters and therefore the fit of the model. Note that we cannot change the distribution of firm ex-post attribute σ since, as in Holmström and Milgrom (1987), we exploit the explicit solution that is only admissible with normal signals.

In each of these three robustness exercises, we find very similar results to the results in our benchmark model. This can best be seen from inspection of the figures as well as in the parameter estimates. In each case, the fit of the wage and profit distribution is very similar, and most importantly, the main insight of our paper – that the share of competition in the manager's compensation is increasing – remains the same, both qualitatively and quantitatively, with only some minor variations in the magnitude.

Fourth, we also perform a totally different robustness exercise. Instead of using data on wages and profits individually, we use data on only on the sum of wages and profits. With no information on the distribution of this sum between the firm and the manager, we do not expect to adequately estimate a model that is built on managerial incentives in response to contracted wages, that is, the share the manager receives. That is indeed what we find. The estimated compensation is not consistent with the data on wages. This is not surprising since we do not target compensation. This robustness exercise does illustrate that data on wages and profits are crucial to estimate the model and to decompose compensation into the contribution due to incentives and that due to competition.

5 Concluding Remarks

Contracting between a firm and a worker does not occur in isolation. Rather, firms and managers compete before signing the contract. To capture both the contracting component of compensation and the component due to competition, we nest a canonical principal-agent problem with moral hazard in a competitive setting where firms and managers match to jointly produce output. This requires an adjustment of both the contracting problem and the matching problem. In the contracting problem, the outside option is now determined in equilibrium, and the matching problem now features stochastic realizations of the matched attributes as well as risk aversion and moral hazard.

This setting allows us to decompose managers' wages into the contribution from bilateral contracting and from competition. We solve the model and then estimate it using matched firm-manager data on executive compensation. We find that for the top executives in large firms, 70% of the compensation is due to competition. Superstar CEOs get paid superstar wages because competing firms bid up their compensation. Instead, for CEOs at small firms, nearly all the compensation is due to incentive contracts.

We conclude that there is truth in both paradigms that explain managers' pay: both incentive contracts and competition matter. Importantly, however, for the superstar managers the incentive component is substantially smaller than the competition component. Instead, for those managers at the bottom of the distribution, incentive motives dwarf competition. This insight that the contribution of each component varies may help us understand a wide range of markets where compensation is driven by incentives and competition. One such example may be the market for professional sports players. Our results would then suggest that the superstars are paid predominantly based on competition, whereas those at the bottom sign contracts with a dominant incentive component. That may explain why some of the highest paid stars walk leisurely on the field even if their team is losing.

Appendix

A Omitted Proofs

We prove a slightly more general result as we allow the support of ω to be $[\underline{\omega}, \overline{\omega}]$, with $0 \leq \underline{\omega} < \overline{\omega} \leq \infty$. Below we also show how to modify the arguments if $-\infty \leq \underline{\omega} < 0$. It is also obvious from the proof that we do not use that s^2_ω is constant in x.

A.1 Proof of Proposition 1

Recall that if X is normally distributed with mean μ and variance σ^2, then $\mathbb{E}\left[\mathrm{e}^{\kappa X}\right] = \mathrm{e}^{\kappa\mu + \frac{1}{2}\kappa^2\sigma^2}$. Using this fact and integrating with respect to G, one can write problem (4) as

$$\Pi(y, x, a(x)) = \max_{\beta,\alpha,e} \int_{\underline{\omega}}^{\overline{\omega}} \left[(1-\alpha(\omega))\omega(t(y)+e(\omega)) - \beta(\omega)\right] dF(\omega|x) \tag{A1}$$

$$\text{s.t.} \int_{\underline{\omega}}^{\overline{\omega}} -\mathrm{e}^{-r\left(\beta(\omega)+\alpha(\omega)\omega(t(y)+e(\omega)) - \frac{e^2(\omega)}{2} - \frac{r\alpha^2(\omega)\omega^2 s_\sigma^2}{2}\right)} dF(\omega|x) \geq -\mathrm{e}^{-ra(x)} \tag{A2}$$

$$\alpha(\omega) = \frac{e(\omega)}{\omega}, \quad \forall \omega. \tag{A3}$$

Substituting the expression for α in (A3) into (A1)–(A2), and rearranging (A2), we obtain

$$\max_{\beta,e} \int_{\underline{\omega}}^{\overline{\omega}} \left[\left(1 - \frac{e(\omega)}{\omega}\right)\omega(t(y)+e(\omega)) - \beta(\omega)\right] dF(\omega|x) \tag{A4}$$

$$\text{s.t.} \int_{\underline{\omega}}^{\overline{\omega}} \left(1 - \mathrm{e}^{-r\left(\beta(\omega)+e(\omega)(t(y)+e(\omega)) - \frac{e^2(\omega)}{2} - \frac{re^2(\omega)s_\sigma^2}{2} - a(x)\right)}\right) dF(\omega|x) \geq 0. \tag{A5}$$

Denote the multiplier of the constraint by λ. Forming the Lagrangian and optimizing pointwise we obtain the following optimality condition with respect to β:

$$-f(\omega|x) + r\lambda \mathrm{e}^{-rA(\omega)} f(\omega|x) = 0, \ \forall \omega,$$

where to simplify the notation we have set

$$A(\omega) = \beta(\omega) + e(\omega)\left(t(y)+e(\omega)\right) - \frac{e^2(\omega)}{2} - \frac{re^2(\omega)s_\sigma^2}{2} - a(x).$$

Intuitively, the optimality condition implies that $\lambda = \mathrm{e}^{-rA(\omega)}/r > 0$, so the constraint (A5) is binding at the optimum. Also, since the optimality condition must hold for all ω, it follows that A is constant in ω.

The optimality condition with respect e is

$$\left(-(t(y) + e(\omega)) + (\omega - e(\omega)) + r\lambda \mathrm{e}^{-rA(\omega)} \left(t(y) + 2e(\omega) - e(\omega) - re(\omega)s_\sigma^2\right)\right) f(\omega|x) = 0, \;\; \forall\omega.$$

Since $r\lambda\mathrm{e}^{-rA(\omega)} = 1$, we obtain

$$e(\omega) = \frac{\omega}{1 + rs_\sigma^2}, \quad \forall\omega. \tag{A6}$$

Intuitively, the higher the ex-post attribute ω, the higher is the effort required from the manager. Since $\alpha(\omega) = e(\omega)/\omega$, we obtain

$$\alpha(\omega) = \frac{1}{1 + rs_\sigma^2}, \quad \forall\omega. \tag{A7}$$

Inserting (A6) into $A(\omega)$ we obtain

$$A(\omega) = \beta(\omega) + \frac{\omega}{1 + rs_\sigma^2}\left(t(y) + \frac{\omega}{1 + rs_\sigma^2}\right) - \frac{\omega^2}{2(1 + rs_\sigma^2)^2} - \frac{r\omega^2 s_\sigma^2}{2(1 + rs_\sigma^2)^2} - a(x).$$

We claim that $A(\omega)$, which is constant in ω, is equal to zero. Suppose it is strictly positive; then the constraint (A5) holds with strict inequality, contradicting $\lambda > 0$. And if it is strictly negative, the constraint is violated. Hence, $A(\omega) = 0$ and thus $\lambda = 1/r$ and

$$\beta(\omega) = a(x) - \frac{\omega t}{1 + rs^2} - \frac{\omega^2}{\left(1 + rs_\sigma^2\right)^2} + \frac{\omega^2}{2\left(1 + rs_\sigma^2\right)^2} + \frac{r\omega^2 s_\sigma^2}{2\left(1 + rs_\sigma^2\right)^2} \tag{A8}$$

$$= a(x) - \frac{\omega t}{1 + rs_\sigma^2} + \frac{\omega^2}{2\left(1 + rs_\sigma^2\right)^2}\left(rs_\sigma^2 - 1\right), \;\; \forall\omega, \tag{A9}$$

completing the solution of the optimal contract (α, β, e). □

A.2 Allowing for $\underline{\omega} < 0$

The proof of Proposition 1 relies on $F(\cdot|x)$ having support on positive values of ω, since we replaced the incentive constraints by the first-order condition for interior solution (A3).

But we can easily modify the proof in the case in which $\underline{\omega} < 0$. To see this, note that now the incentive constraints reduce to $\alpha(\omega)\omega = e(\omega)$ if $\omega > 0$, and $e(\omega) = 0$ for all $\omega \leq 0$, since the agent's objective function is strictly decreasing in e when $\omega \leq 0$.

But then, for all $\omega < 0$ the cheapest way to provide incentives is to set $\alpha(\omega) = 0$ and $\beta(\omega) = a(x)$. The principal's problem then becomes, after simple algebra,

$$\max_{\beta,e}\left((\mathbb{E}[\omega|\omega\leq 0,x]t(y)-a(x))F(0|x)+\int_0^{\overline{\omega}}\left[\left(1-\frac{e(\omega)}{\omega}\right)\omega(t(y)+e(\omega))-\beta(\omega)\right]dF(\omega|x)\right)$$
$$\text{s.t.}\int_0^{\overline{\omega}}\left(1-\mathrm{e}^{-r\left(\beta(\omega)+e(\omega)(t(y)+e(\omega))-\frac{e^2(\omega)}{2}-\frac{re^2(\omega)s_\sigma^2}{2}-a(x)\right)}\right)dF(\omega|x)\geq 0.$$

Since the first term in the objective function is a constant, this problem is equivalent to

$$\max_{\beta,e}\int_0^{\overline{\omega}}\left[\left(1-\frac{e(\omega)}{\omega}\right)\omega(t(y)+e(\omega))-\beta(\omega)\right]dF(\omega|x)$$
$$\text{s.t.}\int_0^{\overline{\omega}}\left(1-\mathrm{e}^{-r\left(\beta(\omega)+e(\omega)(t(y)+e(\omega))-\frac{e^2(\omega)}{2}-\frac{re^2(\omega)s_\sigma^2}{2}-a(x)\right)}\right)dF(\omega|x)\geq 0.$$

which is the *same* problem as the one solved in Proposition 1 when $\underline{\omega} = 0$. Thus, the optimal contract is the same as in Proposition 1 for $\omega > 0$, and for $\omega \leq 0$ is given by

$$\alpha(\omega)=0,\;\;\beta(\omega)=a(x),\;\;e(\omega)=0.$$

Equations (7)–(8) remain the same for $\omega > 0$ and they are, for $\omega \leq 0$, given by

$$w(p,\omega)=a(x),\;\;\pi(q,\omega)=\omega t(y)-a(x).$$

Thus, the expected value of profits is now

$$\Pi(y,x,a(x))=k(x)t(y)+\frac{\int_0^{\overline{\omega}}\omega^2dF(\omega|x)}{2\left(1+rs_\sigma^2\right)}-a(x),$$

and the expected value of wages is

$$W(x,a(x))=a(x)+\frac{\int_0^{\overline{\omega}}\omega^2dF(\omega|x)}{2\left(1+rs_\sigma^2\right)}.$$

Clearly, if $F(\cdot|x)$ puts small mass on negative values of ω for all x, as will be the case in our estimation (see footnote 21), then $\int_0^{\overline{\omega}} \omega^2 dF(\omega|x) \approx k^2(x) + s_\omega^2$. If this is the case, then one can ignore negative values of ω without much loss.

B Estimation

LIKELIHOOD FUNCTION. With F and G normally distributed, the log-likelihood function to be maximized can be written as:

$$\ln \mathcal{L}(\theta|w,\pi,x) = -\sum_{x\in\mathcal{X}} \frac{\varepsilon_\omega(x)^2}{2s_\omega^2} - \sum_{x\in\mathcal{X}} \frac{\varepsilon_\sigma(x)^2}{2s_\sigma^2} + \sum_{x\in\mathcal{X}} \ln|J| - n\ln s_\omega - n\ln s_\sigma - n\ln(2\pi), \quad \text{(B1)}$$

where $k(x)$ and $t(y)$ are the type-dependent means of ω and σ, and s_ω and s_σ are the standard deviations. The former are evaluated along the equilibrium allocation $y = x$. $|J|$ is the absolute value of the Jacobian of the transformation we derive below. We obtain expressions for ε_ω and ε_σ from solving (7)–(8).

$$\begin{aligned} w &= a + \frac{\omega^2}{2\left(1+rs_\sigma^2\right)} + \frac{\omega}{1+rs_\sigma^2}\varepsilon_\sigma && \text{(B2)} \\ \pi &= \omega t - a + \frac{\omega^2}{2\left(1+rs_\sigma^2\right)} + \frac{rs_\sigma^2}{1+rs_\sigma^2}\omega\varepsilon_\sigma. && \text{(B3)} \end{aligned}$$

substituting one in the other:

$$\pi = \omega t - a + \frac{\omega^2}{2\left(1+rs_\sigma^2\right)} + rs_\sigma^2\left(w - a - \frac{\omega^2}{2\left(1+rs_\sigma^2\right)}\right) \quad \text{(B4)}$$

gives a quadratic equation in $\overline{\omega}$.

$$\begin{aligned} -\pi + \omega t - (1+rs_\sigma^2)a + rs_\sigma^2 w + (1-rs_\sigma^2)\frac{\omega^2}{2\left(1+rs_\sigma^2\right)} &= 0 && \text{(B5)} \\ \frac{1-rs_\sigma^2}{2\left(1+rs_\sigma^2\right)}\omega^2 + \omega t - (1+rs_\sigma^2)a + rs_\sigma^2 w - \pi &= 0 && \text{(B6)} \\ (1-rs_\sigma^2)\omega^2 + 2\left(1+rs_\sigma^2\right)t\omega - 2(1+rs_\sigma^2)^2 a + (rs_\sigma^2 w - \pi)2\left(1+rs_\sigma^2\right) &= 0 && \text{(B7)} \\ (-1+rs_\sigma^2)\omega^2 - 2t\left(1+rs_\sigma^2\right)\omega + 2(1+rs_\sigma^2)\left(\pi + a - rs_\sigma^2(w-a)\right) &= 0 && \text{(B8)} \end{aligned}$$

The Discriminant D of this quadratic form is equal to:

$$\begin{aligned} D &= 4t^2(1+rs_\sigma^2)^2 - 8(1+rs_\sigma^2)(-1+rs_\sigma^2)\left(\pi + a - rs_\sigma^2(w-a)\right) && \text{(B9)} \\ &= 4(1+rs_\sigma^2)\left[t^2(1+rs_\sigma^2) - (-1+rs_\sigma^2)2\left(\pi + a - rs_\sigma^2(w-a)\right)\right] && \text{(B10)} \end{aligned}$$

and then the solutions for $\overline{\omega}$ are:

$$\omega = \frac{t\left(1+rs_\sigma^2\right) \pm \sqrt{D}}{-1+rs_\sigma^2} \tag{B11}$$

and using (7) to solve for σ we obtain:

$$\begin{aligned} \varepsilon_\sigma &= (w-a)\frac{1+rs_\sigma^2}{\omega} - \frac{\omega}{2} && \text{(B12)} \\ &= (w-a)\frac{1+rs_\sigma^2}{\frac{t(1+rs_\sigma^2)\pm\sqrt{D}}{-1+rs_\sigma^2}} - \frac{\frac{t\left(1+rs_\sigma^2\right)\pm\sqrt{D}}{-1+rs_\sigma^2}}{2} && \text{(B13)} \\ &= (w-a)\frac{(1+rs_\sigma^2)(-1+rs_\sigma^2)}{t\left(1+rs_\sigma^2\right)\pm\sqrt{D}} - \frac{t\left(1+rs_\sigma^2\right)\pm\sqrt{D}}{2(-1+rs_\sigma^2)} && \text{(B14)} \end{aligned}$$

So, using $\varepsilon_\omega = \omega - k$, we get:

$$\begin{aligned} \varepsilon_\omega &= \frac{t(1+rs_\sigma^2)\pm\sqrt{D}}{-1+rs_\sigma^2} - k && \text{(B15)} \\ \varepsilon_\sigma &= (w-a)\frac{(1+rs_\sigma^2)(-1+rs_\sigma^2)}{t\left(1+rs_\sigma^2\right)\pm\sqrt{D}} - \frac{t\left(1+rs_\sigma^2\right)\pm\sqrt{D}}{2(-1+rs_\sigma^2)} && \text{(B16)} \\ \varepsilon_\omega + \varepsilon_\sigma &= (w-a)\frac{(1+rs_\sigma^2)(-1+rs_\sigma^2)}{t\left(1+rs_\sigma^2\right)\pm\sqrt{D}} + \frac{t\left(1+rs_\sigma^2\right)\pm\sqrt{D}}{2(-1+rs_\sigma^2)} - k && \text{(B17)} \end{aligned}$$

where $D = \left(r{s_\sigma}^2+1\right)\left(2a+2\pi+t^2-2\pi r{s_\sigma}^2-2r{s_\sigma}^2w-2ar^2{s_\sigma}^4+r{s_\sigma}^2t^2+2r^2{s_\sigma}^4w\right)$, and where abs$|J|$ is the absolute value of the determinant of the Jacobian of the transformation. The Jacobian is given by:

$$\begin{aligned} |J| &= \text{abs}\begin{vmatrix} \frac{\partial\omega}{\partial w} & \frac{\partial\omega}{\partial\pi} \\ \frac{\partial\sigma}{\partial w} & \frac{\partial\sigma}{\partial\pi} \end{vmatrix} && \text{(B18)} \\ &= \frac{\pm\left(rs_\sigma^2+1\right)\left(t+trs_\sigma^2\mp\sqrt{D}\right)}{2\sqrt{D}\left(a+\pi+ars_\sigma^2-wrs_\sigma^2\right)}. && \text{(B19)} \end{aligned}$$

The outside option $a(x)$ solves for equation (12) which is equal to:

$$a(x) = a(0) + \int_0^x \left(k_0 k_1 \mathrm{e}^{k_1 z} t_0 \mathrm{e}^{t_1 z} + \frac{2k_0 \mathrm{e}^{k_1 z} k_0 k_1 \mathrm{e}^{k_1 z}}{1 + r s_\sigma^2} \right) dz \tag{B20}$$

$$= a(0) + \int_0^x \left(k_0 k_1 t_0 \mathrm{e}^{(k_1 + t_1) z} + \frac{2k_0^2 k_1 \mathrm{e}^{2k_1 z}}{1 + r s_\sigma^2} \right) dz \tag{B21}$$

$$= a(0) + \frac{k_0 k_1 t_0}{k_1 + t_1} \left(\mathrm{e}^{(k_1 + t_1) x} - 1 \right) + \frac{k_0^2 \left(\mathrm{e}^{2k_1 x} - 1 \right)}{(1 + r s_\sigma^2)}. \tag{B22}$$

ESTIMATION PROCEDURE We use data on compensation and profits from 2016 and 2017. We start with 1,913 matched CEO-firm pairs during that period, and only keep those whose CEO was newly hired in 2015 and who stayed on at least until the end of 2017. After dropping outliers and missing observations on the variables we use, we are left with 139 matches. Our parameter space is eight-dimensional: $\Theta = \{k_0, k_1, s_\omega, t_0, t_1, s_\sigma, r, a(0)\}$. There are two stages of estimation, both using the Sobol sequence in MATLAB. In the first stage, we compute the average log-likelihood (ln L divided by the number of real matches with a non-negative discriminant) for 100,000 Sobol points. We drop those observations for which the estimated parameters have $D < 0$ and for which there is no real solution. We impose an additional constraint that we must retain at least 90% of the 139 matches. The combination of parameters that yields the highest average log-likelihood in the first stage give us 126 real matches. In the second stage, we use only the 126 matches and compute the log-likelihood (ln L) for 100,000 Sobol points. The combination of parameters that yields the highest log-likelihood in the second stage provides our final estimated parameters.

BOOTSTRAPPING FOR CONFIDENCE INTERVALS

1. Randomly select a subsample of size 100

2. Estimate the parameters $\theta \in \Theta$ with Sobol sequence in the exact same way we did for the full sample

3. Repeat steps 1. and 2. for 1000 times. There will be 1000 sets of Θ's. For each of $\{k_0, k_1, s_\omega, t_0, t_1, s_\sigma, r, a(0)\}$, the 5th and 95th percentiles are respectively the lower and the upper bound of the 90 % confidence interval for that parameter.

C Extensions and Robustness

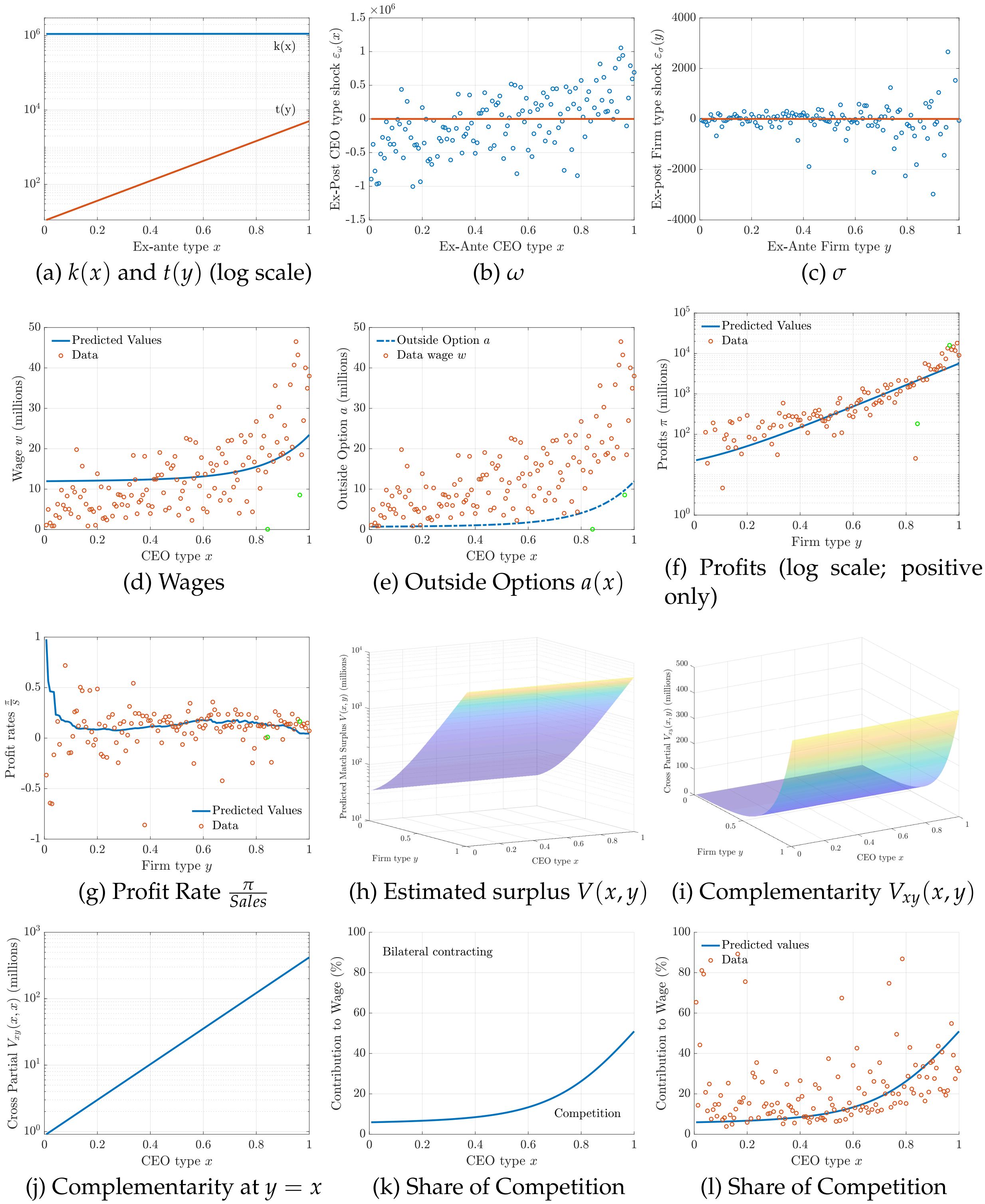
In this section, we report the robustness exercises that we discuss in the main text.

C.1 Using Market Value instead of Sales to Rank Firms

Manager		
k_0	k_1	s_ω
1111748	0.0120	434268
$[690608, 1324537]$	$[0.0102, 0.0166]$	$[268532, 548518]$
Firm		
t_0	t_1	s_σ
10.59	6.16	1105
$[10.44, 38.91]$	$[4.97, 6.62]$	$[936, 1714]$
Risk aversion	Outside option	
r	$a(0)$	
0.0518	711113	
$[0.0102, 0.0952]$	$[670333, 890816]$	

Note: 90% bootstrapped confidence intervals in brackets (1000 iterations).

Table C1: Estimated Parameters Θ.



(a) $k(x)$ and $t(y)$ (log scale)

(b) ω

(c) σ

(d) Wages

(e) Outside Options $a(x)$

(f) Profits (log scale; positive only)

(g) Profit Rate $\frac{\pi}{Sales}$

(h) Estimated surplus $V(x, y)$

(i) Complementarity $V_{xy}(x, y)$

(j) Complementarity at $y = x$

(k) Share of Competition

(l) Share of Competition

Figure C1: Robustness: Using Market Value to rank firms

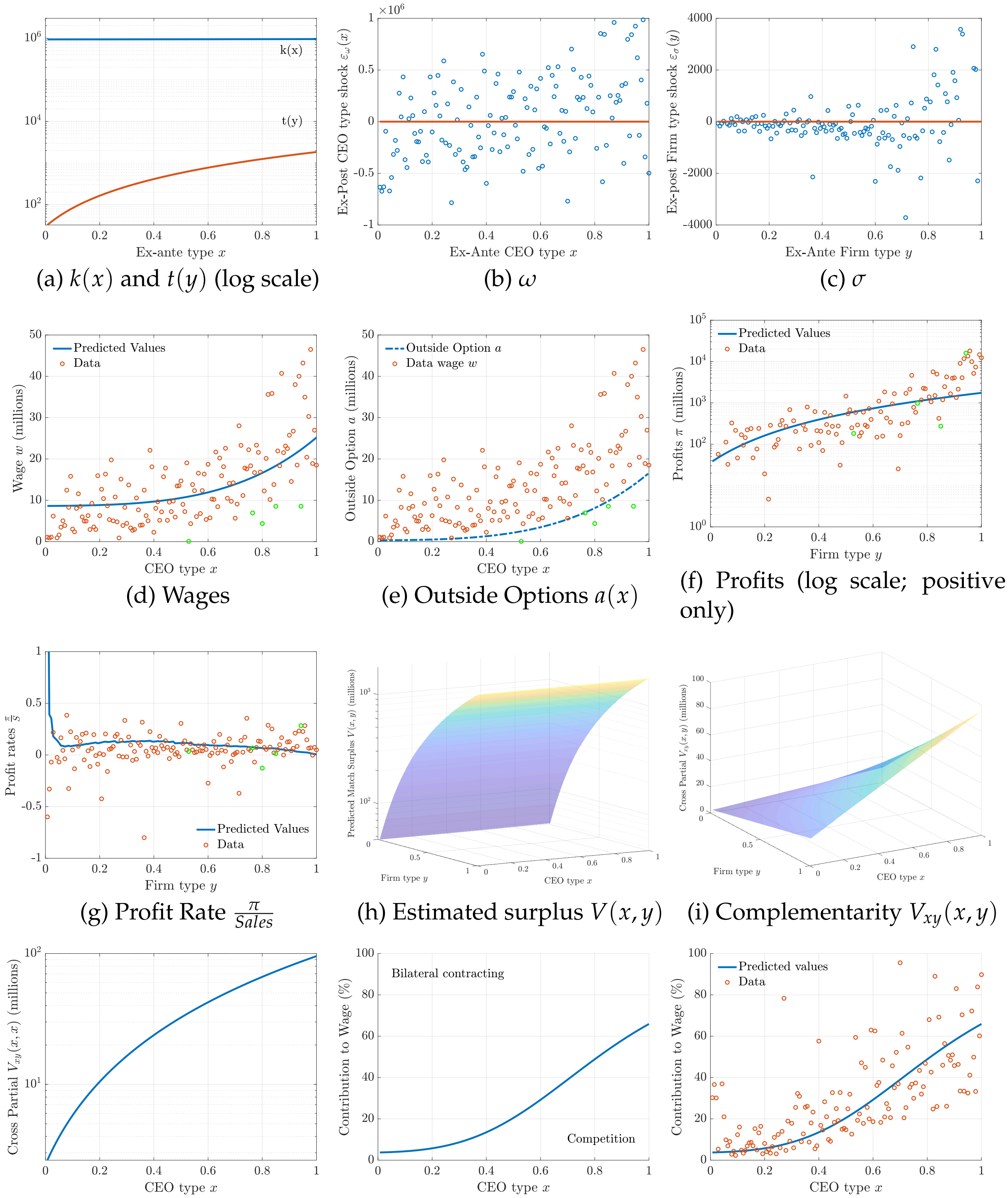
C.2 Using $\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$

Manager			
k_0	k_1	k_2	s_ω
934732	6292	11610	366246
[732097, 1419822]	[1547, 18081]	[1752, 16901]	[308323, 581512]
Firm			
t_0	t_1	t_2	s_σ
29.72	388	1425	1214
[7.09, 47.49]	[23, 1211]	[820, 3347]	[979, 1826]

Risk aversion	Outside option
r	$a(0)$
0.0411	322704
[0.0125, 0.0954]	[40127, 805804]

Note: 90% bootstrapped confidence intervals in brackets (1000 iterations).

Table C2: Estimated Parameters Θ.



(a) $k(x)$ and $t(y)$ (log scale)

(b) ω

(c) σ

(d) Wages

(e) Outside Options $a(x)$

(f) Profits (log scale; positive only)

(g) Profit Rate $\frac{\pi}{Sales}$

(h) Estimated surplus $V(x,y)$

(i) Complementarity $V_{xy}(x,y)$

(j) Complementarity at $y = x$

(k) Share of Competition

(l) Share of Competition

Figure C2: Robustness: Polynomial specification for $k(x)$ and $t(y)$

C.3 Lognormally Distributed ω

C.3.1 Derivations under Lognormality

With F log-normal and G normally distributed, the log-likelihood function can be written as[24]

$$\ln \mathcal{L}(\theta|w,\pi,x) = -\sum_{x\in\mathcal{X}} \ln \omega(x) - \sum_{x\in\mathcal{X}} \frac{[\ln \omega(x)]^2}{2s_\omega^2} - \sum_{x\in\mathcal{X}} \frac{\varepsilon_\sigma(x)^2}{2s_\sigma^2} + \sum_{x\in\mathcal{X}} \ln|J| - n\ln s_\omega - n\ln s_\sigma - n\ln(2\pi), \tag{C1}$$

The derivation of ω and ε_σ is as before. Now the only things that differ are the expected value of wages, profits and output:[25]

$$\begin{aligned} \mathbb{E}[w(q,\omega)] &= a(x) + \int \frac{\omega^2}{2(1+rs_\sigma^2)} dF(\omega|x) \\ &= a(x) + \frac{\mathrm{e}^{2(k+s_\omega^2)}}{2(1+rs_\sigma^2)}, \end{aligned} \tag{C2}$$

$$\begin{aligned} \mathbb{E}\pi(q,\omega) &= \int \left(\omega t - a(x) + \frac{\omega^2}{2\,(1+rs_\sigma^2)}\right) dF(\omega|x) \\ &= \mathrm{e}^{k+\frac{s_\omega^2}{2}} t - a(x) + \frac{\mathrm{e}^{2(k+s_\omega^2)}}{2(1+rs_\sigma^2)}. \end{aligned} \tag{C3}$$

Since $\mathbb{E}\pi(q,\omega) \equiv V(x,y) - a(x)$, it follows that

$$V(x,y) = \mathrm{e}^{k(x)+\frac{s_\omega^2}{2}} t(y) + \frac{\mathrm{e}^{2(k(x)+s_\omega^2)}}{1+rs_\sigma^2}. \tag{C4}$$

Finally, also the outside option $a(x)$ now is different given log-normality of ω (in the estimation we solve this integral numerically):

$$a(x) = a(0) + \int_0^x \left(\mathrm{e}^{k_0\mathrm{e}^{k_1 z}+\frac{s_\omega^2}{2}} k_0 k_1 \mathrm{e}^{k_1 z} t_0 \mathrm{e}^{t_1 z} + \frac{2\mathrm{e}^{2(k_0\mathrm{e}^{k_1 z}+s_\omega^2)} k_0 k_1 \mathrm{e}^{k_1 z}}{1+rs_\sigma^2} \right) dz. \tag{C5}$$

[24] Recall that a random variable X is log-normally distributed with parameters μ and σ^2 if $\log X$ is normally distributed with mean μ and variance σ^2. The mean of X is $e^{\mu+\frac{\sigma^2}{2}}$, and its variance is $(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$; thus, its second moment is $\mathbb{E}[X] = e^{2(\mu+\sigma^2)}$.

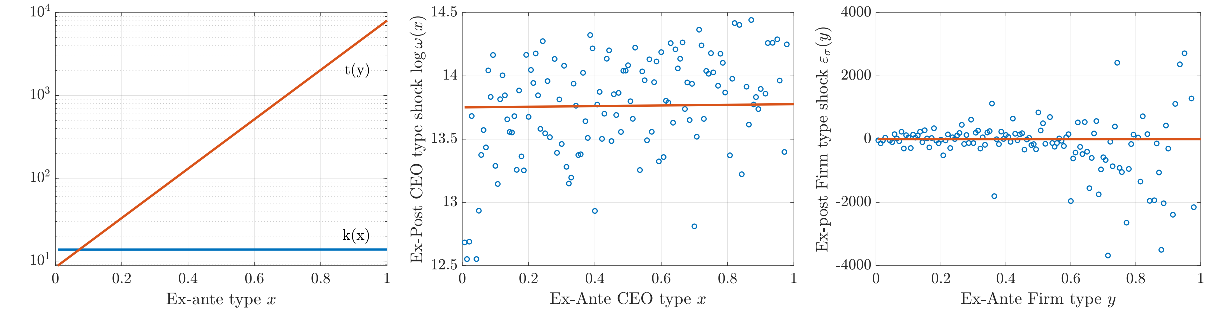
[25] We now assume that, for each x, the distribution of $\log\omega$ is normal with mean $k(x)$ and variance s_ω^2. This implies that the distribution of ω for each x is log-normal with mean and variance determined by $k(x)$ and s_ω^2 as in the previous footnote. With some abuse of notation, we break with previous usage and now denote the mean and the variance of this distribution by expressions that depend on $k(x)$ and s_ω^2 instead of being $k(x)$ and s_ω^2. No confusion should arise since this is used only locally here.

C.3.2 Results under lognormality

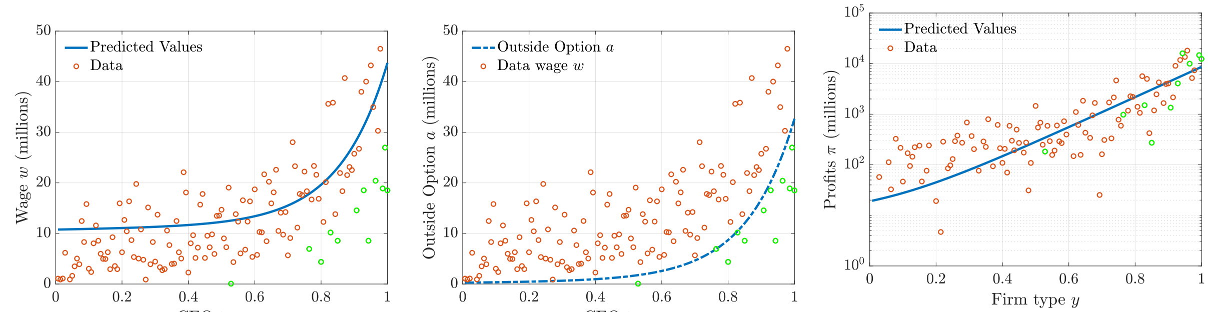
Manager		
k_0	k_1	s_ω
13.75	0.0018	0.47
$[13.19, 14.00]$	$[0.0011, 0.0023]$	$[0.36, 0.56]$
Firm		
t_0	t_1	s_σ
8.38	6.87	1277
$[1.77, 9.67]$	$[6.40, 8.55]$	$[901, 1852]$
Risk aversion	**Outside option**	
r	$a(0)$	
0.0399	269339	
$[0.0064, 0.0944]$	$[30465, 608583]$	

Note: 90% bootstrapped confidence intervals in brackets (1000 iterations).

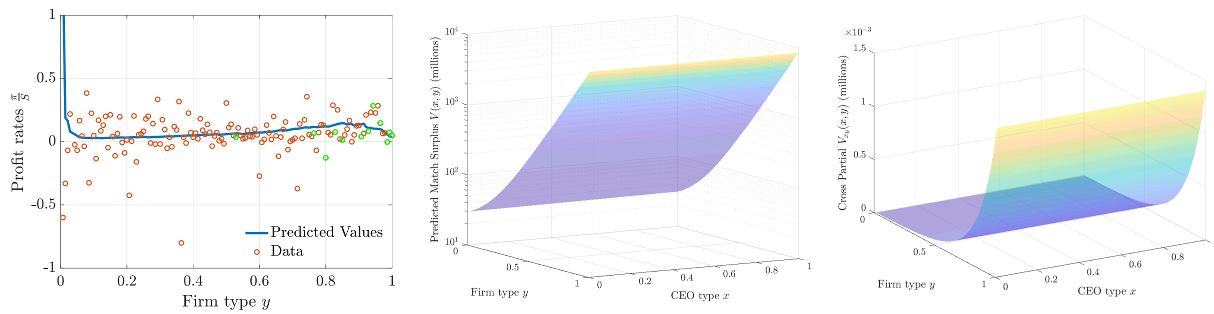
Table C3: Estimated Parameters Θ.



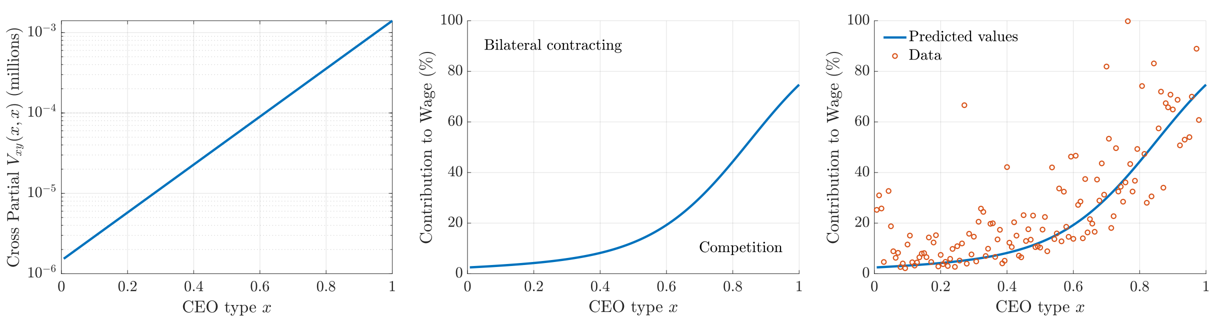
(a) $k(x)$ and $t(y)$ (log scale) (b) $\bar{\omega}$ (c) σ



(d) Wages (e) Outside Options $a(x)$ (f) Profits (log scale; positive only)



(g) Profit Rate $\frac{\pi}{Sales}$ (h) Estimated surplus $V(x,y)$ (i) Complementarity $V_{xy}(x,y)$



(j) Complementarity at $y = x$ (k) Share of Competition (l) Share of Competition

Figure C3: Robustness: lognormally-distributed CEO types

C.4 Estimation using Only Sum of Wages and Profits

C.4.1 Derivations using Sum of Wages and Profits

In this exercise, we abstract from the information contained separately in wages and profits, and use only the information from the sum of wages and profits in each match. Because we only have one degree of freedom, we adjust the problem and reduce the two dimensions of stochastic output (ω and σ) to one (σ only). To that effect, we assume that ω is deterministic with mean $k(x)$ and variance $s_\omega^2 = 0$, i.e. $\varepsilon_\omega = 0$.

The output q is the sum of wages w and profits π. From equations (7) and (8), we obtain, after substituting ω by its deterministic value k:

$$q = w + \pi = kt + \frac{k^2}{1 + rs_\sigma^2} + k\varepsilon_\sigma. \tag{C6}$$

We can now solve explicitly for ε_σ:

$$\varepsilon_\sigma = \frac{q}{k} - t - \frac{k}{1 + rs^2}. \tag{C7}$$

We can now use this expression to estimate the parameters, k, t, r_σ. The log-likelihood function can now be written as

$$\ln \mathcal{L}(\theta|w, \pi, x) = -\sum_{x \in \mathcal{X}} \frac{\varepsilon_\sigma(x)^2}{2s_\sigma^2} + \sum_{x \in \mathcal{X}} \ln |J| - n \ln s_\sigma - \frac{n}{2} \ln(2\pi), \tag{C8}$$

where the new expression for Jacobian is

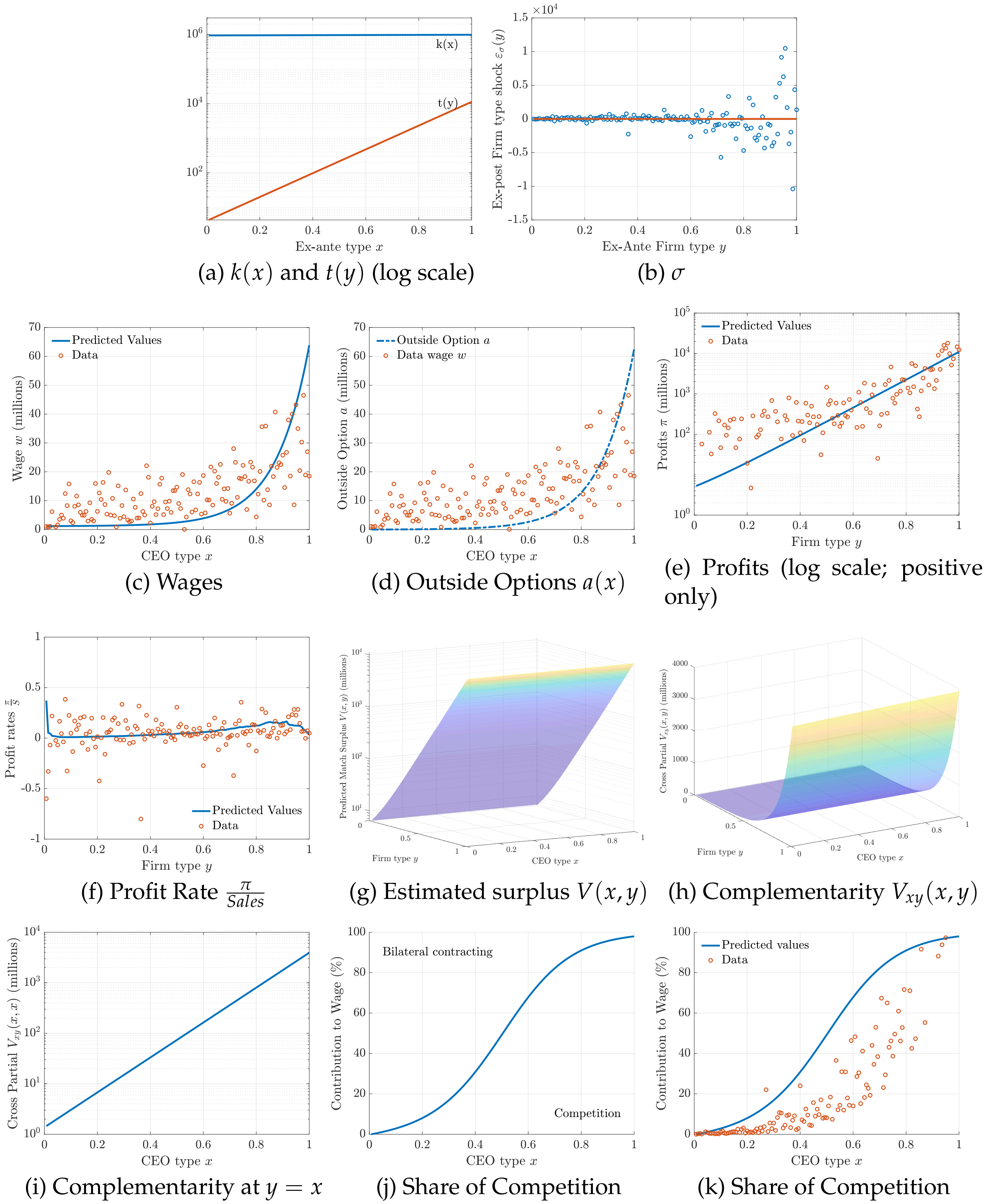
$$|J| = \text{abs}|\frac{\partial \sigma}{\partial q}| = \frac{1}{k}. \tag{C9}$$

C.4.2 Results with Sum of Wages and Profits

	Manager	
k_0	k_1	r
936255	0.0450	0.0764
[560103, 1306025]	[0.0391, 0.0498]	[0.0085, 0.0956]
	Firm	
t_0	t_1	s_σ
4.12	7.92	2205
[3.01, 7.51]	[7.67, 7.99]	[1499, 3396]

Note: 90% bootstrapped confidence intervals in brackets (1000 iterations).

Table C4: Estimated Parameters Θ.



(a) $k(x)$ and $t(y)$ (log scale)

(b) σ

(c) Wages

(d) Outside Options $a(x)$

(e) Profits (log scale; positive only)

(f) Profit Rate $\frac{\pi}{Sales}$

(g) Estimated surplus $V(x,y)$

(h) Complementarity $V_{xy}(x,y)$

(i) Complementarity at $y = x$

(j) Share of Competition

(k) Share of Competition

Figure C4: Robustness: Estimation using surplus data only

References

ACKERBERG, D. AND M. BOTTICINI (2002): "Endogenous matching and the empirical determinants of contract form," *Journal Political Economy*, 110, 564–591.

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

BANDIERA, O., L. GUISO, A. PRAT, AND R. SADUN (2015): "Matching firms, managers, and incentives," *Journal of Labor Economics*, 33, 623–681.

BANDIERA, O., S. HANSEN, A. PRAT, AND R. SADUN (2020): "CEO Behavior and Firm Performance," *Journal of Political Economy*, 128, 1325–1369.

BECKER, G. (1973): "A Theory of Marriage: Part I," *Journal Political Economy*, 81, 831–846.

BERTRAND, M. AND S. MULLAINATHAN (2001): "Are CEOs rewarded for luck? The ones without principals are," *The Quarterly Journal of Economics*, 116, 901–932.

BONHOMME, S., T. LAMADON, AND E. MANRESA (2019): "A distributional framework for matched employer employee data," *Econometrica*, 87, 699–739.

CHADE, H., J. EECKHOUT, AND L. SMITH (2017): "Sorting through search and matching models in economics," *Journal of Economic Literature*, 55, 493–544.

CHADE, H. AND J. SWINKELS (2020): "The Moral Hazard Problem with High Stakes," *Journal of Economic Theory*, 187.

CHIAPPORI, P.-A. AND P. RENY (2016): "Matching to Share Risk," *Theoretical Economics*, 11, 227–251.

DEMSETZ, H. AND K. LEHN (1985): "The Structure of Corporate Ownership: Causes and Consequences," *Journal of Political Economy*, 93, 1155–1177.

EDMANS, A., X. GABAIX, AND A. LANDIER (2009): "A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium," *The Review of Financial Studies*, 22.

EECKHOUT, J. AND P. KIRCHER (2011): "Identifying Sorting – In Theory," *Review of Economic Studies*, 78, 872–906.

GABAIX, X. AND A. LANDIER (2008): "Why has CEO Pay Increased So Much?" *The Quarterly Journal of Economics*, 123, 49–100.

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

GRETSKY, N. E., J. M. OSTROY, AND W. R. ZAME (1992): "The nonatomic assignment model," *Economic Theory*, 2, 103–127.

GROSSMAN, S. AND O. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–46.

HALL, B. AND J. LIEBMAN (1998): "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics*, 113, 653–691.

HOLMSTRÖM, B. (1979): "Moral Hazard and Observability," *The Bell Journal of Economics*, 74–91.

HOLMSTRÖM, B. AND P. MILGROM (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303–28.

JENSEN, M. AND K. J. MURPHY (1990): "erformance Pay and Top-Management Incentives," *Journal of Political Economy*, 98, 225–264.

LEGROS, P. AND A. NEWMAN (2007): "Beauty is a Beast, Frog is a Prince: Assortative Matching with Nontransferabilities," *Econometrica*, 75, 1073–1102.

ROSEN, S. (1981): "The economics of superstars," *The American economic review*, 71, 845–858.

SCHAEFER, S. (1998): "The Dependence of Pay-Performance Sensitivity on the Size of the Firm," *Review of Economics and Statistics*, 80, 436–443.

SERFES, K. (2005): "Risk sharing vs. incentives: contract design under two-sided heterogeneity," *Economics Letters*, 83, 343–349.

SHAPLEY, L. AND M. SHUBIK (1972): "The Assignment Game I: The Core," *International Journal of Game Theory*, 111–130.

TERVIÖ, M. (2008): "The difference that CEOs make: An assignment model approach," *The American Economic Review*, 642–668.