DISCUSSION PAPER SERIES

DP17782

A THEORY OF FAIR CEO PAY

Pierre Chaigneau, Alex Edmans and Daniel Gottlieb

LABOUR ECONOMICS AND BANKING AND CORPORATE FINANCE

ISSN 0265-8003

A THEORY OF FAIR CEO PAY

Pierre Chaigneau, Alex Edmans and Daniel Gottlieb

Discussion Paper DP17782
Published 05 January 2023
Submitted 30 December 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Labour Economics
- Banking and Corporate Finance

A THEORY OF FAIR CEO PAY

Abstract

This paper studies optimal executive pay when the CEO is concerned about fairness: if his wage falls below a perceived fair share of output, the CEO suffers disutility that is increasing in the discrepancy. Fairness concerns do not lead to fair wages always being paid -- to induce effort, the firm threatens the CEO with unfair wages if output is sufficiently low. The optimal contract sometimes involves performance shares: the CEO is paid a constant share of output if it is sufficiently high, but the wage drops discontinuously to zero if output falls below a threshold. Even if the incentive constraint is slack, the optimal contract continues to involve pay-for-performance, to address the CEO's fairness concerns and ensure his participation. Thus, the firm can implement strictly positive levels of effort "for free." This rationalizes pay-for-performance even if the CEO is intrinsically motivated and does not need effort incentives.

JEL Classification: D86, G32, G34, J33

Keywords: Executive compensation

Pierre Chaigneau - pierre.chaigneau@queensu.ca
Queen's University

Alex Edmans - aedmans@london.edu
London Business School, CEPR, and ECGI and CEPR

Daniel Gottlieb - d.gottlieb@lse.ac.uk
London School Of Economics

Acknowledgements
For helpful comments, we thank Simon Gervais, Naveen Gondhi, Moqi Groen-Xu, Dirk Jenter, Wei Jiang, Katya Potemkina, and conference participants at the Finance Theory Group, Financial Management Association, INSEAD Finance Symposium and Rotterdam Behavioral Finance Conference.

A Theory of Fair CEO Pay*

Pierre Chaigneau
Queen's University

Alex Edmans
LBS, CEPR, and ECGI

Daniel Gottlieb
LSE

December 29, 2022

Abstract

This paper studies optimal executive pay when the CEO is concerned about fairness: if his wage falls below a perceived fair share of output, the CEO suffers disutility that is increasing in the discrepancy. Fairness concerns do not lead to fair wages always being paid – to induce effort, the firm threatens the CEO with unfair wages if output is sufficiently low. The optimal contract sometimes involves performance shares: the CEO is paid a constant share of output if it is sufficiently high, but the wage drops discontinuously to zero if output falls below a threshold. Even if the incentive constraint is slack, the optimal contract continues to involve pay-for-performance, to address the CEO's fairness concerns and ensure his participation. Thus, the firm can implement strictly positive levels of effort "for free." This rationalizes pay-for-performance even if the CEO is intrinsically motivated and does not need effort incentives.

KEYWORDS: Executive compensation, fairness, moral hazard.
JEL CLASSIFICATION: D86, G32, G34, J33.

*pierre.chaigneau@queensu.ca, aedmans@london.edu, d.gottlieb@lse.ac.uk. For helpful comments, we thank Simon Gervais, Naveen Gondhi, Moqi Groen-Xu, Dirk Jenter, Wei Jiang, Katya Potemkina, and conference participants at the Finance Theory Group, Financial Management Association, INSEAD Finance Symposium and Rotterdam Behavioral Finance Conference.

Standard moral hazard models assume that the agent cares about pay only for the consumption it enables. As a result, the marginal consumption utility of the additional pay from improving performance must weakly exceed the marginal cost of effort required to do. Such theories aim to capture the general firm-worker relationship, and thus have been applied not only to rank-and-file employees but also CEOs. Indeed, most theories of CEO pay use the standard principal-agent model as the key building block, and have contributed substantially to our understanding of CEO pay.

However, it is not clear that models of the general employment relationship can be automatically applied to CEOs. A crucial difference with other agents is that CEOs are typically wealthy, and nearly all of their consumption needs are already met. Thus, it is not clear that consumption utility is the only, or even the most important, driver of observed contracts. Edmans, Gosling, and Jenter (2022) survey directors and investors on how they set pay contracts. Both sets of respondent highlight how pay is driven not only by the desire to provide consumption incentives, but also the need to ensure the CEO feels fairly treated. This is consistent with prior research suggesting that pay is a hygiene factor – pay above a certain level provides limited additional motivation, but pay below that level is a strong demotivator (e.g. Hertzberg, 1959).

The respondents also suggest that firm value is an important determinant of what directors, investors, and the CEO view to be a fair level of pay. If firm value has increased due to CEO effort, they believe that it is fair to reward the CEO for this increase. If firm value has increased (decreased) due to luck outside the CEO's control, they believe the CEO should share in this good (bad) luck. That a share of firm value is perceived as a fair payment is consistent with the ultimatum game, which has been widely replicated (e.g. Roth et al., 1991). If one party has been gifted an endowment, the other believes it is fair to be offered a sizable share, and will sacrifice his own consumption to punish an unfair offer.

This paper studies optimal CEO pay when the CEO is motivated by both traditional consumption utility and fairness concerns. We model fairness concerns by specifying a perceived fair wage that is increasing in the firm's output, which in turn depends on both CEO effort and luck. The CEO suffers disutility if his wage falls below the fair wage, the magnitude of which is increasing in the discrepancy.

It may seem that fairness concerns should lead to the CEO always receiving a fair wage, but this turns out not to be the case. We start with a linear model that demonstrates the effect of fairness concerns in the most transparent possible setting. The fair wage is linear in output, i.e. the CEO believes that it is fair for him to receive a fixed share of output. If the actual wage is at least the fair wage, his utility equals the wage, as with standard risk neutrality. If the actual wage is below the fair wage, he suffers disutility which is linear in the discrepancy. The CEO's utility function is thus piecewise linear, with a slope of 1 above the fair wage and a slope exceeding 1 below it. The principal is also risk-neutral, and her goal is to find the cheapest contract to induce a given effort level out of a continuum. Both parties are protected by limited liability.

We show that the optimal contract involves a threshold below which the CEO is paid zero, and

above which he receives the fair wage, i.e. a constant share of output. This contradicts the intuition that fairness concerns will lead to the CEO being paid fair wages for all output levels. Instead, fairness concerns mean that *un*fairness can be a powerful motivator. If output is sufficiently low that it is unlikely that the CEO has worked, the firm pays him the most unfair possible wage of zero. Only if output exceeds a lower threshold is the CEO paid the fair wage. Depending on parameter values, there may also be an additional upper threshold above which the CEO is paid the firm's entire output. Thus, our model demonstrates the range of outputs over which the CEO is paid fairly; interestingly it is typically only for an intermediate range, and not for high or low outputs.

Innes (1990) showed that, with standard risk neutrality, the optimal contract is "live-or-die" – the agent receives zero if output is below a threshold, and the entire output above it. The intuition is that it is optimal to concentrate payments in the highest likelihood ratio states, i.e. pay the highest possible amount for sufficiently high outputs. However, such a contract is inefficient under fairness concerns. Even if the CEO works, output may fall below this threshold due to bad luck. If the CEO is paid zero, he suffers significant disutility due to unfairness, which erodes his incentives to work. Thus, it is efficient to offer him a fair wage for intermediate output levels. We show that, if the CEO is not paid zero for outputs with a positive likelihood ratio, the threshold is decreasing in the CEO's fairness concerns – when the CEO is more concerned about fairness, the range of outputs over which he receives a fair wage rises. In addition, this range is increasing in the volatility of the output distribution. The greater this volatility, the likelier it is that output will be moderate even if the CEO works, and so the more important it is to reward him with a fair wage rather than zero.

The contract resembles performance shares, which are frequently offered in reality (see the survey of Edmans, Gabaix, and Jenter (2017)). Standard models, such as Holmström (1979), do not predict discontinuous contracts. Innes (1990) predicts a sharp discontinuity where the CEO's pay increases from zero to the entire output, once output crosses a threshold, but such sharp discontinuities do not exist in reality. To obtain more realistic contracts, Innes (1990) assumes that the principal's payoff cannot be decreasing in output, otherwise she would "burn" output, or the agent would secretly inject his own funds into the company to inflate output. Innes's theory can either be interpreted as a financing model where an entrepreneur (agent) raises funds from an outside investor (principal), or a compensation model where a company (principal) offers a contract to a CEO (agent). While the two justifications for the monotonicity constraint are realistic for the financing application, they may be less relevant for the compensation application. Dispersed shareholders cannot coordinate to burn output, and while the board acts on shareholders' behalf, burning output violates directors' fiduciary duty to the company. Similarly, it would likely be illegal for the CEO to inject his own funds into the company to manipulate the stock price. Indeed, the prevalence of discontinuities in real-life executive compensation contracts suggests that these two justifications are not first-order. Our paper obtains realistic contracts without a monotonicity assumption to rule out discontinuities, and indeed the optimal contract involves a discontinuity.

Performance shares provide fair wages if performance is good and unfair wages if performance is bad, to motivate good performance. Moreover, the discontinuity is milder and thus more realistic – when performance crosses a threshold, the wage jumps from zero, but not to the entire output. Models with a risk-averse agent can also feature moderate discontinuities (paying either zero or the entire output would be inefficient risk-sharing). However, it is unclear how important risk aversion is for CEOs; indeed, Edmans, Gosling, and Jenter (2022) find that CEO risk aversion is the least important out of seven determinants of the sensitivity of pay to performance.

We then extend the model to a non-linear one. Now, the utility function takes a more general form – if the wage is fair, utility is increasing and concave in the wage; if the wage is unfair, utility is a general function of both the wage and output. It is increasing and convex in the former (as in prospect theory) and decreasing in the latter. The fair wage remains increasing in output but need not be linear, and the utility loss from unfair wages need not be linear in the discrepancy. Despite this additional generality, we show that the basic features of the risk-neutral contract remain robust – the payment is zero below a lower threshold, the fair wage above this threshold, and the entire output above a higher threshold. However, there is an additional fourth region, in-between the regions in which the CEO receives the fair wage and the entire output. In this region, his payment exceeds the fair wage, and is generally convex in output. Intuitively, if performance is very strong, the principal wishes to reward the CEO with more than the fair wage. However, since the CEO is risk-averse, it is inefficient to pay him the entire output.

We show that pay is increasing in output even when the incentive constraint is slack. When participation is the only constraint, it may seem that the most efficient way to satisfy this constraint is to pay the CEO a fair wage for all outputs – the prior argument that zero wages are needed to punish low effort no longer applies when the incentive constraint is non-binding. However, guaranteeing the CEO a fair wage for all outputs may lead to the CEO receiving rents. To avoid this, the firm pays him an unfair wage for some output levels. Since the CEO's utility function is convex below the fair wage, if the firm reduces the payment below the fair wage, it is optimal to reduce it all the way to zero. Thus, the firm pays the CEO zero for some output levels, rather than a moderately unfair wage for a greater range of output levels.

That pay is increasing in output even without an incentive constraint means that the firm can induce CEO effort "for free". In a standard moral hazard model, implementing higher effort is always costly to the firm.[1] In our model, since pay is optimally increasing in output to satisfy the participation constraint, lower effort will be more costly to implement than certain higher effort levels, so they will never be induced. A frequent criticism of performance-related pay for CEOs is that it should not be necessary – the CEO should be intrinsically motivated to exert effort, and/or the board should monitor CEO effort. Our model demonstrates that performance-related pay may be optimal not to induce effort, but to secure the participation of a CEO with fairness concerns.

[1]If the CEO is risk-neutral and protected by limited liability, implementing higher effort requires the firm to offer him a higher payment upon success and thus a higher expected wage; if the CEO is risk-averse, this requires the firm to offer him a more sensitive contract and thus a risk premium.

This paper is related to the theoretical literature on executive compensation, recently surveyed by Edmans and Gabaix (2016) and Edmans, Gabaix, and Jenter (2017). While a small number of models focus on adverse selection or retention, the vast majority of these theories feature moral hazard, where pay only matters to the CEO by providing consumption utility. Our paper is also related to CEO pay models that feature reference points. For example, De Meza and Webb (2007) and Dittmann, Maug, and Spalt (2010) study optimal CEO compensation in the presence of loss aversion. In our model, the fair wage can be seen as a reference point; the CEO is also loss-averse as his utility is steeper below the fair wage than above it. The key innovation in our model is that the fair wage depends on output, which leads to a very different optimal contract.

An important literature has studied the effect of fairness concerns in contracts outside the CEO setting. Fehr and Schmidt (1999) study optimal contracts in the presence of inequity aversion, where an agent dislikes another agent receiving less than him, and dislikes even more another agent receiving more than him. Sobel (2005) provides a survey of this literature.[2] In such models, the agents are all paid in the same units, and so it makes sense for agents to compare their consumption. However, these models do not apply to a CEO setting, where the firm's objective function is shareholder value, which is orders of magnitude in excess of CEO pay, and thus in different units. An inequity aversion explanation for performance-sensitive CEO pay is that shareholders feel sorry for a CEO who is not given a share of greater firm value, which seems at odds with real-life perceptions. In our model, it is the CEO who has fairness preferences, rather than shareholders. Moreover, the CEO is only concerned for his own utility, unlike in social preference models where agents are concerned with other agents' utility.

1 The Model

We consider a standard principal-agent model with one added feature: due to fairness considerations, the utility of the agent (manager, "he") depends not only on his pay but also on output.

At time $t = -1$, the principal (firm, "she") offers a contract to the agent. At $t = 0$, if the agent has accepted the contract, he privately chooses an effort level $e \in \mathbb{R}_+$. The agent's cost of exerting effort e is $C(e)$, where $C(\cdot)$ is continuously differentiable with $C'(e) > 0$ for $e > 0$, $C'(0) = 0$, and $\lim_{e\to\infty} C'(e) = \infty$. As is standard, effort can refer not only to working rather than shirking, but also to choosing projects to maximize firm value rather than private benefits or not diverting cash flows. At $t = 1$, output $q \in [0, \overline{q}]$ is realized, where $\overline{q}$ may be finite or infinite, and the agent is paid a wage $w(q)$ that can depend on output. Output is continuously distributed

[2] In Rabin (1993), an agent may put positive or negative weight on the other agent's utility, depending on his assessment of her intentions. Empirically, Fehr, Klein, and Schmidt (2007) show how inequity aversion leads to a principal paying an agent a discretionary bonus upon good performance, even though such a bonus is unenforceable. Fehr, Kirchsteiger, and Riedl (1993) show that workers reciprocate a fair wage with higher effort in the next period, even though this higher effort is unenforceable. Charness and Rabin (2002) use experiments to distinguish between various models of social preferences.

according to a probability density function ("PDF") $\phi(q|e)$ that satisfies the monotone likelihood ratio property ("MLRP") and is continuously differentiable in q and in e with a continuous cross derivative: $\frac{\partial^2}{\partial q \partial e}\phi(q|e)$ is continuous in q. Both the principal and agent are protected by limited liability, so that $0 \leq w(q) \leq q \; \forall q$.

The agent's utility function is $u(w, q)$, which depends on the actual wage w, and may depend on output q if it affects his perceived fair wage. We assume that $u(w, q)$ is increasing and continuously differentiable in w. The agent's reservation utility is given by $\overline{U}$.

As is standard, the principal's problem is to minimize the cost of a contract that induces a target level of effort and is accepted by the agent. Her program is given as follows:

$$\min_{w(q)} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{1}$$

$$\text{s.t. } e^* \equiv \arg\max_e \int_0^{\overline{q}} u(w(q), q)\phi(q|e)dq - C(e) \geq e^T \tag{2}$$

$$\int_0^{\overline{q}} u(w(q), q)\phi(q|e^*)dq - C(e^*) \geq \overline{U} \tag{3}$$

$$0 \leq w(q) \leq q \; \forall q \tag{4}$$

$$w(q) \geq w(q') \; \forall q > q' \tag{5}$$

where (2) is the incentive compatibility constraint ("IC"), (3) is the individual rationality constraint ("IR"), (4) are the limited liability constraints, and (5) is the agent's monotonicity constraint.[3] Note that there is a distinction between the equilibrium level of effort chosen by the manager e^*, and the target level of effort required by the principal e^T. The former will exceed the latter if the incentive constraint is slack.

The above formulation captures fairness concerns in the simplest possible way. We use the standard moral hazard model with continuous effort and continuous output, with the only departure being the specification of the agent's utility function. As a result, any deviation in the optimal contract from standard moral hazard models can be attributed to the utility function. Another formulation, which would follow the ultimatum game more literally, would be to have a single-period model in which the agent first receives his pay and then chooses his effort, or a multi-period model where the agent responds to his first-period pay by choosing effort in the second period. Then, if offered unfair pay, he may withhold effort and reduce total surplus, similar to the respondent in the ultimatum game refusing the proposed share and leading to both parties receiving zero. However, such a formulation would be more ad hoc, as we would need to hard-wire the link between perceived unfairness and next-period effort, rather than fairness entering the utility function. Our framework captures fairness preferences in a standard one-period model in which

[3]Standard models do not require an agent monotonicity constraint since it is a consequence of the MLRP. In our model with fairness concerns, this is not necessarily the case. We assume that the agent's payoff has to be non-decreasing in output otherwise the manager would burn output. Innes (1990) assumed that the principal's payoff has to be non-decreasing in output otherwise the principal would burn output; the agent has more control of output than the principal.

the agent first exerts effort and then receives pay – since he knows the contract when taking his action, he will withhold effort if he anticipates that it will not be fairly rewarded. This allows our results to be compared with standard one-period moral hazard models without fairness concerns, such as Holmström (1979) and Innes (1990). In the model of Akerlof and Yellen (1990), the agent's effort is reduced if his wage is lower than the fair wage, but the wage only takes one value (it is not contingent on the agent's performance).

Define the likelihood ratio $LR(q|e)$ as follows:

$$LR(q|e) \equiv \frac{\frac{\partial \phi}{\partial e}(q|e)}{\phi(q|e)},$$

and let q_0^e be implicitly defined as the output such that the likelihood ratio is zero for effort of e: $LR(q_0^e|e) = 0$ (q_0^e exists and is unique because of MLRP and ϕ being continuously differentiable). To guarantee that an optimal contract exists, we assume:

$$\int_0^{q_0^{e^T}} u(0,q)\frac{\partial}{\partial e}\phi(q|e^T)dq + \int_{q_0^{e^T}}^{\overline{q}} u(q,q)\frac{\partial}{\partial e}\phi(q|e^T)dq \geq C'(e^T)$$

The above inequality means that paying the agent the minimum (zero) for outputs with negative likelihood ratios and the maximum (the entire output) for outputs with positive likelihood ratios will be sufficient to induce effort e^T. If it is not satisfied, then no contract that satisfies bilateral limited liability can implement e^T.

Lemma 1 below derives a sufficient condition for the validity of the first-order approach ("FOA"), which allows us to replace a maximization program in the incentive constraints (equation (2)) by its first-order conditions.[4] Let K_e^+ and K_e^- denote the integral of the positive and negative parts of the second derivative of the joint distribution $\phi(q|e)$ with respect to effort:

$$K_e^+ := \int_0^{\overline{q}} \max\left\{\frac{\partial^2}{\partial e^2}\phi(q|e),\ 0\right\} dq, \tag{6}$$

$$K_e^- := \int_0^{\overline{q}} \min\left\{\frac{\partial^2}{\partial e^2}\phi(q|e),\ 0\right\} dq. \tag{7}$$

Lemma 1 *(First-Order Approach): Suppose that*

$$\int_0^{\overline{q}} \left(K_e^- u(0,q) + K_e^+ u(q,q)\right) dq < C''(e) \tag{8}$$

for all $e \in \mathbb{R}_+$. Then, the FOA is valid.

We henceforth assume that this condition holds. We start in Section 2 by considering a piecewise linear model to demonstrate the effect of fairness concerns in the simplest possible setting. Section

[4]This is related to the condition for the FOA in a model with limited liability in Chaigneau, Edmans, and Gottlieb (2022). The difference is that the utility function also depends on output in Lemma 1.

3 analyzes a non-linear model.

2 Linear Model

The agent's utility function is as follows:

$$u(w,q) \equiv w - \gamma \max\left\{w^*(q) - w, 0\right\}. \tag{9}$$

The first term is the standard risk-neutral utility function. The second term captures the agent's concern for fairness, where $w^*(q)$ is the agent's perceived fair wage for output q, and $\gamma \geq 0$ parametrizes the intensity of his fairness concerns. Thus, if the agent's actual wage falls below his perceived fair wage, he suffers disutility. The fair wage is given by

$$w^*(q) \equiv \rho q, \tag{10}$$

where $\rho \in [0,1]$ is the agent's perceived fair share of output q. One determinant of ρ is the effect of agent effort on output. The survey of Edmans, Gosling, and Jenter (2022) finds that "how much the CEO can affect firm performance" is the main determinant of pay variability and the free-text fields and interviews suggest that fairness is a primary reason: "if the CEO has a greater effect on performance, it is fair to reward her more for good performance." Another potential determinant is ρ in peer firms – the third most popular response in the survey is "the split between fixed and variable pay in peer firms."

With $\gamma = 0$ (no fairness concerns) or $\rho = 0$ (any wage is perceived as fair), the utility function in equation (9) collapses to the standard risk-neutral utility function $u(w,q) = w$. Accordingly, unless otherwise specified, we assume $\gamma > 0$ and $\rho > 0$. The utility function is piecewise linear with a kink at the fair wage. Figure 1 illustrates the utility function for various values of γ and ρ. This piecewise linear utility function is the simplest and most transparent specification for fairness concerns, and allows us to conduct comparative statics with respect to the parameters γ and ρ.

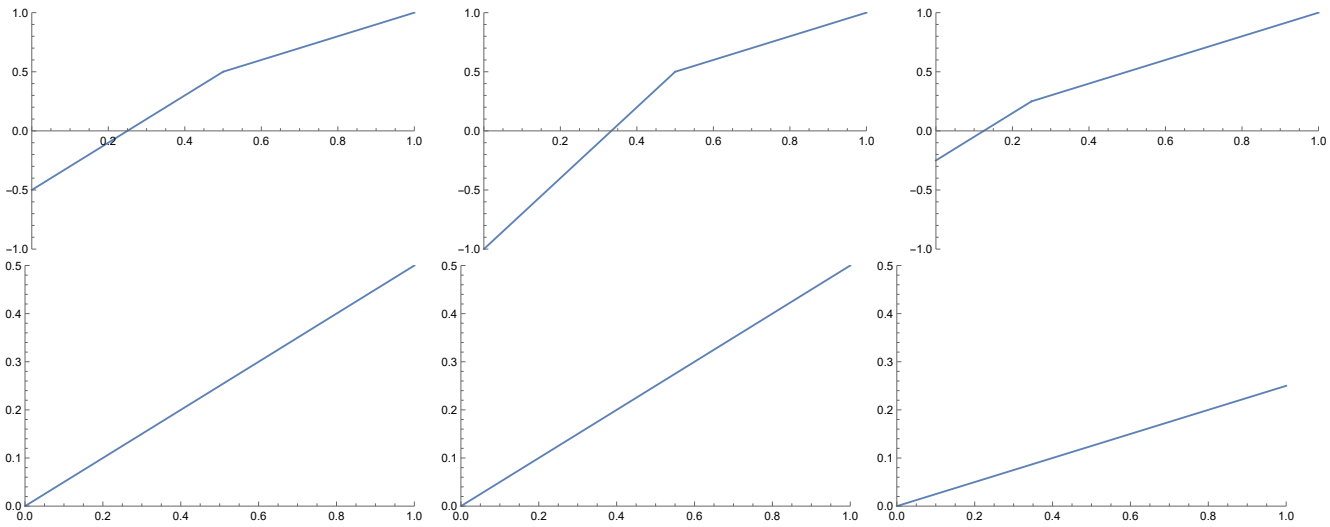

Figure 1: Top row: the function $u(w)$ defined in equation (9) as a function of w for $q = 1$, $\gamma = 1$ and $\rho = 0.5$ on the left; $\gamma = 2$ and $\rho = 0.5$ in the middle; $\gamma = 1$ and $\rho = 0.25$ on the right. Bottom row: the fair wage $w^*(q)$ defined in equation (10) as a function of q for $\gamma = 1$ and $\rho = 0.5$ on the left; $\gamma = 2$ and $\rho = 0.5$ in the middle; $\gamma = 1$ and $\rho = 0.25$ on the right.

As can be seen in Figure 1, the utility function exhibits loss aversion. The agent cares not only about the wage w per se, but also gains and losses relative to the fair wage, with his sensitivity to losses exceeding his sensitivity to gains. De Meza and Webb (2007) and Dittmann, Maug, and Spalt (2010) also study optimal contracts in the presence of loss aversion. The main difference between fairness concerns and standard loss aversion is that, with the former, the fair wage depends on output. A loss aversion model features a reference point that is independent of output. In de Meza and Webb (2007), it is the median of the wage distribution; in Dittmann, Maug, and Spalt (2010) it is last year's salary (they consider an alternative reference point that also includes the market value of the shares and options the agent inherited from the previous year.)

To simplify the analysis, in this section we assume:

$$-\gamma\rho\int_0^{q_0^{e^T}} q\frac{\partial}{\partial e}\phi(q|e^T)dq + \rho\int_{q_0^{e^T}}^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e^T)dq \geq C'(e^T) \tag{11}$$

$$-\gamma\rho\int_0^{\overline{q}} q\phi(q|0)dq < \overline{U} + C(0) \tag{12}$$

$$\rho\int_0^{\overline{q}} q\phi(q|e^*)dq \geq \overline{U} + C(e^*), \text{ where } e^* \text{ satisfies equation (2) with } w(q) = w^*(q)\ \forall q. \tag{13}$$

The assumptions in equations (11)-(13) are not crucial for our results, but reduce the number of cases we need to consider. Inequality (11) ensures that an incentive-compatible contract that elicits effort e^T exists even if the firm never pays more than the fair wage.[5] Inequality (12) implies

[5]This expression considers the contract that provides the highest effort incentives when the principal does not pay more than the fair wage: one that pays zero for outputs that are bad news about effort, and the fair wage for outputs that are good news conditional on effort e^T.

that, even if the marginal cost of effort were zero, an agent who is paid zero for any output would be below his reservation utility and thus reject the contract. Inequality (13) implies that a contract that always pays the fair wage satisfies the participation constraint.[6]

Define $q_m^{\min}$ implicitly as the highest value that satisfies the following equation:

$$-\gamma\rho\int_0^{q_m^{\min}} q\frac{\partial}{\partial e}\phi(q|e^T)dq+\rho\int_{q_m^{\min}}^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e^T)dq\equiv C'(e^T). \tag{14}$$

If a contract that implements e^T exists such that the agent is never paid above the fair wage for any output, $q_m^{\min}$ is the threshold such that the payment is zero below $q_m^{\min}$, and the fair wage above $q_m^{\min}$. Note that, with MLRP, the definition of $q_0^{e^T}$ and equation (14) imply that $q_m^{\min}\geq q_0^{e^T}$.

Proposition 1 studies the case when the target effort level e^T is zero, i.e. the only goal of the contract is to ensure the agent's participation. In this case, the principal is not trying to induce positive effort from the agent, but the agent still chooses effort optimally (see equation (2)) given the contract.

Proposition 1 *(Zero target effort level): When $e^T=0$, if $\overline{U}$ and γ are sufficiently large, $e^*>0$. The following contract is optimal:*

$$w(q)=\begin{cases} w^*(q) & \text{for } q<q_c \\ \rho q_c & \text{for } q\geq q_c \end{cases}, \tag{15}$$

where q_c is set so that the IR in equation (3), with the utility function as in equation (9) and the contract as in equation (15), is satisfied as an equality.

This Proposition considers the case where the target effort level is zero and so the contract is not designed to provide incentives. Perhaps surprisingly, the equilibrium effort is strictly positive even though the principal does not attempt to induce a minimum effort level. Intuitively, a payment increasing in output allows the agent to receive the fair wage for a great range of outputs. This reduces his disutility from receiving unfair wages, and thus allows his IR to be satisfied at lower cost to the principal. Thus, even when the contract is not designed for incentive purposes, the agent's payment can be increasing in output, inducing a positive effort level. One optimal contract pays the agent the fair wage for low outputs ($q<q_c$) to address his fairness concerns, up to a maximum of ρq_c. Since this contract is weakly increasing in output, and strictly increasing for a range of outputs, it induces $e^*>0$. The maximum wage of ρq_c reduces effort incentives, which helps satisfy the IR since the agent has to be compensated for his effort.

Proposition 2 gives the optimal contract when the IC binds. It distinguishes between the cases

[6]Note that we do not need to assume that a contract that always pays the fair wage also satisfies the incentive constraint.

in which the IR is binding or nonbinding. A sufficient condition for the IR to be nonbinding is:

$$\overline{U} + C(e_0) \leq -\gamma\rho \int_0^{\overline{q}} q\phi(q|e_0)dq \quad \text{where } e_0 = \arg\max_e -\gamma\rho \int_0^{\overline{q}} q\frac{\partial\phi(q|e)}{\partial e}dq - C(e), \tag{16}$$

where e_0 is the agent's optimal effort when he is always paid zero.

Proposition 2 *(Binding incentive constraint): When the IC is binding, the optimal contract is given by:*

$$w(q) = \begin{cases} 0 & \text{for } q < q_m \\ w^*(q) & \text{for } q \in [q_m, q_M) \\ q & \text{for } q \geq q_M \end{cases} \tag{17}$$

Moreover:

(a) If $\gamma < \frac{LR(\overline{q}|e^*)}{LR(q_m^{min}|e^*)} - 1$ and the participation constraint is nonbinding, then q_m and q_M are such that the contract is incentive compatible (i.e. satisfy (2) with the utility function as in (9)) and $\frac{LR(q_M|e^*)}{LR(q_m|e^*)} - 1 = \gamma$.

(b) If $\gamma > \frac{LR(\overline{q}|e^*)}{LR(q_m^{min}|e^*)} - 1$ and the participation constraint is satisfied with $q_m = q_m^{\min}$ and $q_M = \overline{q}$, then $q_m = q_m^{\min}$ and $q_M = \overline{q}$;

(c) If the participation constraint is binding, then q_m and q_M satisfy the incentive constraint and the participation constraint in equations (18) and (19) as equalities:

$$-\gamma\rho \int_0^{q_m} q\frac{\partial}{\partial e}\phi(q|e^*)dq + \rho \int_{q_m}^{q_M} q\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e^*)dq = C'(e^*) \tag{18}$$

$$-\gamma\rho \int_0^{q_m} q\phi(q|e^*)dq + \rho \int_{q_m}^{q_M} q\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U}, \tag{19}$$

and are generically such that

$$LR(q_m|e^*)\,(1+\gamma) + \frac{\eta_{IR}}{\eta_{IC}}\gamma = LR(q_M|e^*), \tag{20}$$

where η_{IC} and η_{IR} are the Lagrange multipliers associated with the IC and IR, respectively.

The optimal contract when the incentive constraint is binding is as follows. Without fairness concerns ($\gamma = 0$), the model is similar to the pure moral hazard setting of Innes (1990). The principal generates effort incentives by rewarding high outputs; due to MLRP, it is efficient to concentrate rewards on very high outputs only. Parts (a) and (c)[7] show that, regardless of whether the participation constraint is binding, $\gamma = 0$ leads to $q_m = q_M$: the optimal contract is "live-or-die", involving a single threshold – the agent is paid the minimum possible (zero) below the threshold and the maximum possible (the entire output q) above it. With fairness concerns ($\gamma > 0$), such a contract is suboptimal for two reasons. First, it does not satisfy the participation constraint

[7]Part (b) is inapplicable with $\gamma = 0$ due to MLRP and $q_m^{\min} \leq \overline{q}$.

efficiently, which is a concern if the participation constraint is binding (i.e. part (c) applies). The agent is receiving an unfair wage (zero) for output levels below the threshold, which has a very high disutility cost because of the agent's fairness concerns. Second, it does not satisfy the incentive constraint efficiently. The agent is receiving an unfair wage for some output levels below the threshold, even though these output levels are associated with positive likelihood ratios. Thus, even though these output levels indicate that the agent has worked, he suffers significant disutility for achieving them, reducing his incentives to work. Since the utility function is steeper below $w^*(q)$ rather than above it, it is efficient to increase the rewards for moderately low outputs (that are nevertheless associated with positive likelihood ratios) from 0 to $w^*(q)$, and simultaneously to reduce the rewards for moderately high outputs from q to $w^*(q)$.

Part (a) establishes that, when γ is positive but sufficiently low, the optimal contract has three regions. For outputs below q_m, the agent is paid zero; for outputs above q_M, he is paid the entire output. These regions are similar to Innes (1990). Due to fairness concerns, there is a third region – for intermediate outputs, the agent is paid the fair wage $w^*(q)$. When output hits $q = q_m$, the wage jumps to $w^*(q)$; as output continues to rise, he continues to be paid the fair wage which also rises, since $w^*(q) = \rho q$. Once output reaches $q = q_M$, the wage jumps to the entire output. Thus, fairness concerns cause the principal to depart from the live-or-die contract and offer a fair wage for certain output levels.

Intuitively, the two thresholds q_m and q_M are determined by a trade-off. On the one hand, the principal wishes to concentrate incentives on outputs with the highest likelihood ratios, as in the Innes (1990) model without fairness concerns. On the other hand, the principal wishes to avoid paying zero for outputs with a positive likelihood ratio. A zero payment imposes disutility on the agent due to being unfair, and thus reduce his incentives to take effort e^T. Thus, the principal pays the fair wage, rather than zero, for outputs between q_m and q_M. The distance between them is determined by $(1+\gamma)LR(q_m|e^*) = LR(q_M|e^*)$: with $\gamma = 0$, we have a single threshold; as γ rises, the wider the region in which the agent is paid the fair wage.

Part (b) shows that, when γ is sufficiently high, q_M increases all the way to $\overline{q}$. The highest region disappears, so the agent is never paid the entire output. The optimal contract thus only has two regions – zero for low outputs and the fair wage for high outputs. Decreasing q_m means that the principal pays the fair wage rather than zero for outputs above $q_m^{\min}$, which have a positive likelihood ratio. When fairness concerns are so strong that $\gamma > \frac{LR(\overline{q}|e^*)}{LR(q_m^{\min}|e^*)} - 1$, the penalty γ for payments below the fair wage is sufficiently high for this effect to outweigh the standard desire to concentrate incentives on very high outputs (Innes (1990)). Thus, the optimal contract involves increasing q_M to the highest possible level of $\overline{q}$. Since this means that the agent is never paid above the fair wage for any output, incentive compatibility is achieved by setting $q_m = q_m^{\min}$ as in equation (14).

While the above explains the optimal contract by starting from a model of moral hazard and adding in fairness concerns, another way to view the intuition is to start with a pure fairness model and then add in moral hazard. One may think that fairness concerns would lead to the

agent always being paid the fair wage $w^*(q)$. Since the agent suffers disutility from being paid an unfair wage, the principal provides the agent his reservation utility at minimum cost by always paying a fair wage. However, such a contract does not provide effort incentives efficiently. In particular, since the agent suffers disutility from an unfair wage, it is efficient to "threaten" him with the most unfair possible wage of zero for low output. Thus, fairness concerns do not lead to fair wages for all outputs; in contrast, they can justify unfair wages for some outputs because avoiding unfairness is a motivator. In addition, if output is sufficiently high, the agent is paid the entire output even though this is more than required to meet his fairness constraint. This is because, for incentive provision, it is efficient to concentrate rewards in the highest likelihood ratio states; with a monotone likelihood ratio, this involved paying the agent the maximum possible for high outputs.

In part (b), the contract represents performance shares, where the agent is given shares worth ρq that are forfeited if the output is below a threshold q_m. In standard models where the likelihood ratio is a continuous function of output (as in our setting), such as Holmström (1979), the optimal contract is also a continuous function of output and so does not involve discontinuities. In our model, discontinuities are optimal because, for moderate outputs, it is efficient to pay the agent his fair wage, but if output is sufficiently low, the agent is punished with the most unfair possible wage of zero, and this threat incentivizes effort. In Innes (1990) without a monotonicity constraint for the principal's payoff, the optimal contract is discontinuous but takes a "bang-bang" form where the agent receives either the lowest possible wage or the highest possible wage; we are unaware of cases in which such a contract is offered in reality. Our contract involves discontinuities, but the discontinuity is non-extreme and leads to interior solutions – at q_m, the wage jumps from 0 to a share of output (rather than the entire output), as is the case with performance shares.[8]

Innes (1990) obtains a realistic contract by assuming a monotonicity constraint – that the principal's payoff cannot be decreasing in output. Under such a constraint, the agent receives levered equity and the principal receives debt, and the contract has no discontinuities. Such a contract is realistic for a financing setting, in which the agent is an entrepreneur who is raising financing from the principal, an investor. However, it is less realistic for a contracting setting, in which the agent is a manager who is hired to work for the principal, the firm's shareholders. Hired managers are never given the firm's entire equity. In addition, CEO contracts commonly involve discontinuities (see the survey of Edmans, Gabaix, and Jenter (2017)) – executives are frequently offered bonuses if performance crosses above a certain level, or given equity which they forfeit if performance falls below a certain level.

The two justifications in Innes (1990) for a monotonicity constraint for the principal also may be less applicable to a contracting setting. One justification is that, if the principal's payoff were decreasing, she would "burn" output. This might be possible in a financing setting where there

[8]Chaigneau, Edmans, and Gottlieb (2022) derive conditions under which performance-vesting options are the optimal contract. However, such a contract is continuous as it involves options; the "performance" is a signal separate from output that affects either the number of vesting options or the option strike price.

is a single investor, but is difficult in a large public firm where there are dispersed shareholders. While the board acts on behalf of shareholders, "burning" output would violate fiduciary duty. A second is that, if the agent's payoff increased more than one-for-one with output, he would secretly borrow to increase output. This may be possible in a financing setting in which output is cash flow, but in a contracting setting, the most relevant measure of output is the stock price. An executive secretly taking a loan to boost the stock price would likely be viewed as stock price manipulation. For all of these reasons, the monotonicity constraint for the principal is unlikely to relevant in a contracting setting, which may explain why discontinuities are common in executive contracts. Our model generates an optimal contract that involves discontinuities, but also obtains a realistic contract without requiring a monotonicity constraint for the principal's payoff.

Corollary 1 shows how the contract depends on the intensity of fairness concerns.

Corollary 1 *When the incentive constraint binds, $q_m \leq q_0^{e^T}$ and $q_M < \overline{q}$, the threshold q_m above which the manager is paid a fair wage $w^*(q)$ is decreasing in γ.*

When incentive provision affects the design of the contract and $q_m \leq q_0^{e^T}$, the threshold q_m above which the manager is paid the fair wage is decreasing in fairness concerns γ. Intuitively, the stronger fairness concerns are, the stronger the disutility the agent suffers from receiving zero rather than the fair wage. Thus, stronger fairness concerns unambiguously reinforce effort incentives when the agent is only paid zero for outputs which are bad news for effort ($q_m \leq q_0^{e^T}$), but they also reduce the agent's expected utility from the contract. The principal will therefore change the thresholds q_m and q_M to reduce effort incentives (without violating the incentive constraint) and to increase the agent's expected utility. This is achieved by decreasing q_m, so that the agent receives the fair wage for a larger set of outputs.

The intuition for the condition $q_m \leq q_0^{e^T}$ is as follows. An increase in γ raises the disutility of zero payments. If $q_m \leq q_0^{e^T}$, then zero payments are received only if $q < q_0^{e^T}$ (i.e. for bad news outputs) and so the agent's effort incentives unambiguously rise. To ensure the IC continues to bind, the principal reduces effort incentives. She does so by lowering q_m, because this increases the range of outputs $(q_m, q_0^{e^T})$ over which the agent receives the fair wage even though they are bad news about effort.[9] Thus, an increase in γ unambiguously reduces q_m. If $q_m > q_0^{e^T}$, the effect of γ on effort incentives is ambiguous. Increases γ raises the disutility of receiving zero payments, which arise not only for all bad news outputs ($q < q_0^{e^T}$) but also for some good news outputs $q \in (q_0^{e^T}, q_m)$. Thus, it is unclear whether effort incentives rise or fall, and thus the effect on q_m is ambiguous.

Example 1 illustrates the agent's preferences and the optimal contract for a given parametrization.

Example 1 *The agent's preferences are given by $\gamma = 1$, $\rho = \frac{1}{2}$, and $C(e) = c \times e^2$. Output is lognormally distributed with parameters $e^* = 1$ and $\sigma = 1$. The optimal contract is depicted on the*

[9] While reducing q_m is costly to the principal by reducing the range of outputs over which zero wages are paid, doing so also allows the principal to increase q_M which lowers the cost of the contract.

left of Figure 2 for $c = \frac{1}{2}$ *and* $\overline{U} = 2$*, in the middle of Figure 2 for* $c = \frac{1}{2}$ *and* $\overline{U} = 1$*, and on the right of Figure 2 for* $c = \frac{3}{2}$ *and* $\overline{U} = \frac{3}{2}$*.*

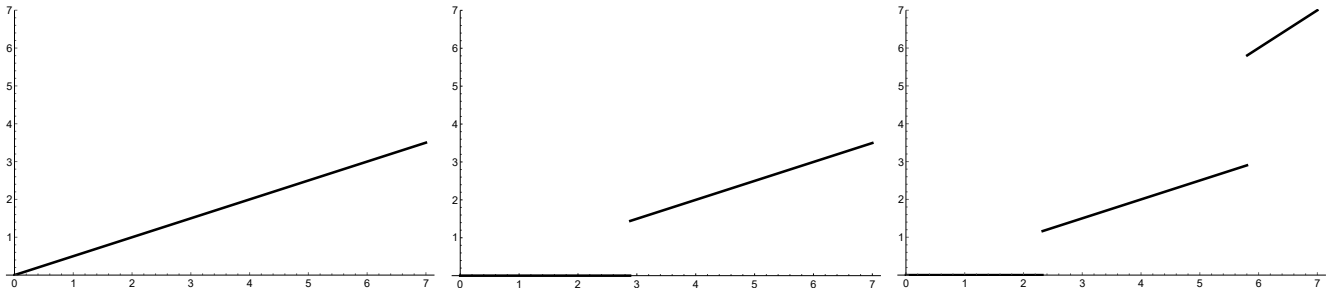

Figure 2: The contract $w(q)$ as a function of q for parameter values described in Example 1.

3 Nonlinear Model

In this section, the utility function is defined as:

$$u(w,q) \equiv \min\{v(w), \nu(w,q)\} \tag{21}$$

where $v(w)$ is the utility over money alone, which is increasing and concave ($v' > 0$, $v'' \leq 0$), with $v(0) = 0$.[10] The term $\nu(w,q)$ is the agent's utility when his payment is below the fair wage, which in turn depends on output. We have $\nu(0,q) \leq 0$, $\nu'_q(w,q) < 0$ (higher output raises the fair wage and thus lowers utility), $\nu'_w(w,q) > 0$, $\nu''_{ww}(w,q) \geq 0$, $\nu''_{wq} = 0$, and $\nu''_{qq} = 0$. For any given q, the two functions $v(w)$ and $\nu(w,q)$ intersect on $(0,\infty)$ at most once.[11] Let this point, if it exists, be denoted by $w^*(q)$, i.e. $v(w^*(q)) \equiv \nu(w^*(q), q)$. At this point, $v'(w) < \nu'(w,q)$, so that there is a kink in $u(w,q)$ as a function of w at $w = w^*(q)$. Thus, $w^*(q)$ captures the agent's perceived fair wage, but we no longer require it to be linear in output ($w^*(q) = \rho q$) as in the linear model of Section 2. This utility function (21) exhibits not only loss aversion, but also concavity above the fair wage $w^*(q)$ and convexity below it, as in prospect theory.[12]

We also assume that for any q,

$$\lim_{w \searrow 0} \nu'(w,q) > \lim_{w \searrow 0} v'(w) \tag{22}$$

so that the utility function is always steeper below the fair wage than above the fair wage.

We assume that an agent who is paid his fair wage for any output is at or above his reservation

[10] This specification for the function $v(w)$ includes CRRA utility with relative risk aversion less than 1, and a "normalized" version of log utility $v(w) = \ln(w+1)$.

[11] Indeed, for $w = 0$ and any q, we have $v(0) \geq \nu(0,q)$. In addition, for any q, $v(w)$ is weakly concave in w whereas $\nu(w,q)$ is weakly convex in w.

[12] However, the model does not exhibit probability weighting as in prospect theory.

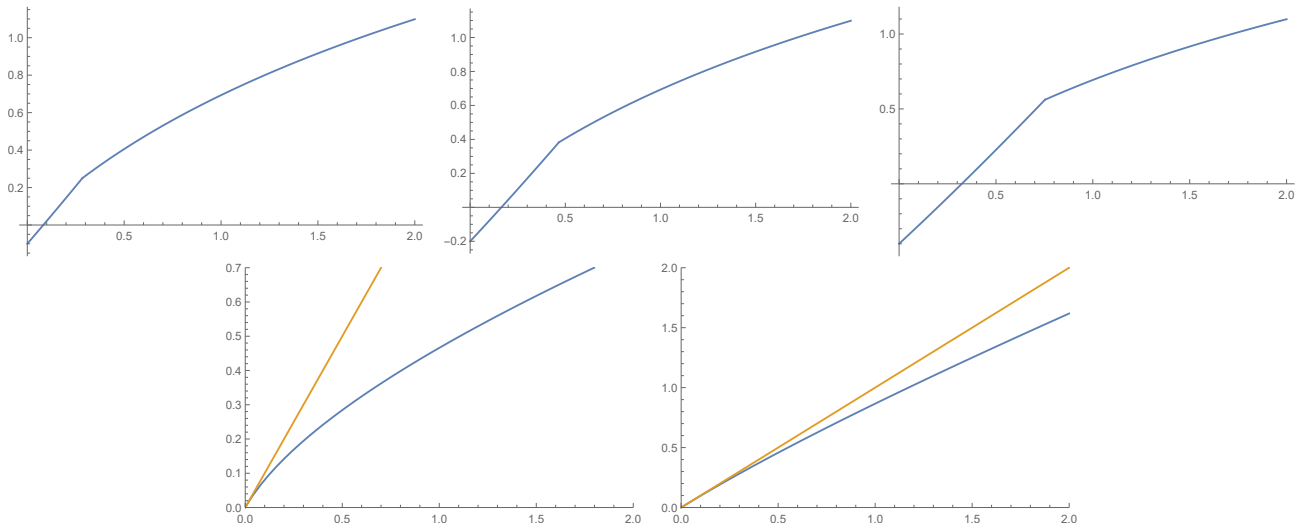

Figure 3: Top row: the function $u(w)$ defined in equation (21) as a function of w for $v(w) = \ln(w+1)$ and $\nu(w,q) = (w+1)^{1.2} - 1 - \frac{1}{5}q$ with $q = 0.5$ on the left, $q = 1$ in the middle, and $q = 2$ on the right. Bottom row: the blue line is the fair wage $w^*(q)$ defined by $v(w^*(q)) \equiv \nu(w^*(q), q)$ as a function of q for $v(w) = \ln(w+1)$ and $\nu(w,q) = (w+1)^{1.2} - 1 - \frac{1}{5}q$ on the left, $v(w) = \sqrt{w+1} - 1$ and $\nu(w,q) = w - \frac{1}{2}q$ on the right. The orange line is principal LL.

utility:

$$\int_0^{\overline{q}} v(w^*(q))\phi(q|e^F)dq - C(e^F) \geq \overline{U}, \quad \text{where } e^F \text{ satisfies equation (2) with } w(q) = w^*(q)\ \forall q, \tag{23}$$

i.e. e^F is the effort level induced when the agent is paid his fair wage for any output. To guarantee the existence of an incentive-compatible contract, we also assume that:

$$\int_0^{q_0^{e^T}} v(w^*(q))\frac{\partial}{\partial e}\phi(q|e^T)dq + \int_{q_0^{e^T}}^{\overline{q}} v(q)\frac{\partial}{\partial e}\phi(q|e^T)dq > C'(e^T) \tag{24}$$

Equation (24) implies that an agent who receives the fair wage for outputs with a negative likelihood ratio and the maximum reward (the whole output) for a positive likelihood ratio will exert effort of at least e^T. The following assumption ensures that a manager who is always paid zero is below his reservation utility:

$$\overline{U} + C(e^*) > 0. \tag{25}$$

Proposition 3 considers the general case in which both the IC and IR may be binding or nonbinding.

Proposition 3 *For the program (1)-(5), the optimal contract is such that:*

$$w(q) = \begin{cases} 0 & \text{for } q \in [0, q_m) \\ w^*(q) & \text{for } q \in [q_m, q_M] \\ v'^{-1}\left(1/\left(\lambda_1 + \lambda_2 LR(q|e^*)\right)\right) & \text{for } q \in [q_M, q_N] \\ q & \text{for } q \in [q_N, \overline{q}] \end{cases}.$$

If $e^T = 0$, then $q_M = q_N = \overline{q}$; if also reservation utility $\overline{U}$ is sufficiently large, then $e^* > 0$.

The optimal contract is given by four regions. As in the linear model, there are three regions in which the agent is paid zero, the fair wage, and the entire output. However, there is an additional region, given by $q \in (q_M, q_N)$. For these output levels, output is sufficiently high that the principal wishes to reward the agent by more than his fair wage ($v'^{-1}\left(1/\left(\lambda_{IR} + \lambda_{IC} LR(q|e^*)\right)\right) > w^*(q)$). It is inefficient to give him the entire output, since the agent exhibits diminishing marginal utility and so does not value this additional reward highly. Thus, unlike in linear model, the optimal contract is continuous at q_M – the principal pays the agent more than his fair wage, but not the entire output. As output rises above q_M, the likelihood ratio increases further and so the actual wage exceeds the fair wage by more. The contract will generally be convex between q_M and q_N.[13] For $q > q_N$, the likelihood ratio is so high that the principal rewards the agent with the entire output. In Lemma 5 in the Online Appendix, we show that, when the IC is not binding and equation (23) holds as an equality, the agent is always paid his fair wage, i.e. $q_m = 0$ and $q_M = q_N = \overline{q}$.

To better understand the effects of incentive provision on the optimal contract, Proposition 3 includes the case in which $e^T = 0$, i.e., the IC is slack. Thus, the principal chooses the cheapest contract that gives the agent an expected utility of $\overline{U}$, and Proposition 3 shows that this involves paying the agent his fair wage $w^*(q)$ for outputs above q_m. Even without moral hazard, the optimal contract does not involve a fixed wage. In turn, a wage that is both increasing in output and also fair provides incentives to exert positive effort, and so the principal obtains effort "for free", as in Proposition 1.[14] Any lower level of effort is suboptimal and will not be induced by the principal. This result is in stark contrast to the case without fairness concerns. In the standard model of Holmström (1979) with a risk-neutral principal and a risk-averse agent, eliciting higher effort is always more costly to the principal. Without an incentive constraint, the optimal contract

[13]However, the contract will be concave if the likelihood ratio is concave, so that very high output is only slightly more indicative of effort, and if risk aversion is sufficiently important compared to prudence (see Chaigneau, Sahuguet and Sinclair-Desgagné, 2017). The latter condition means that protecting the agent against downside risk is relatively unimportant, but providing strong incentives where the agent's marginal utility is high (i.e. for low outputs) is especially important. It will typically not be satisfied for CEOs who have low relative risk aversion due to their wealth.

[14]Note that it is insufficient for the wage to be merely increasing in output for it to provide effort incentives – it may fail to do so if it is also unfair. Indeed, since $\nu(w, q)$ is decreasing in q, a payment schedule which is only slightly increasing in output and is below the fair wage may fail to elicit effort. Intuitively, even though higher output increases the wage, for an increasing wage schedule, it also increases the fair wage. Thus, if the fair wage increases by more than the actual wage, the agent's utility does not increase, and so he is not rewarded for increasing output.

involve a fixed wage for optimal risk-sharing; inducing effort requires an output-contingent wage which leads to inefficient risk-sharing and is thus costly. As a result, any effort level in $\mathbb{R}_+$ can in principle be optimal, depending on model parameters. This is not true with fairness concerns, because the principal will always induce a positive level of effort. Intuitively, providing low effort incentives either requires paying unfair wages for high outputs (which reduces expected utility and fails to satisfy the participation constraint) or paying in excess of fair wages for low outputs (which is unnecessarily costly). Critics of high incentives argue that they are not needed to induce effort, since boards should monitor effort or CEOs should be intrinsically motivated. However, performance-sensitive contracts may be offered not to provide incentives, but to ensure the CEO is fairly paid. A by-product of fair pay is that it incentivizes effort, even if such incentives are unnecessary. Without fairness concerns, it is costly to incentivize high effort levels; with fairness concerns, it is costly to incentivize low effort levels as doing so requires offering unfair pay.

While paying the fair wage for a range of outputs helps satisfy the participation constraint, doing so for all outputs would give the agent rents. The question then becomes: at which outputs does the firm pay below the fair wage, and how much below does it pay? With $\nu'' > 0$, the agent's utility is non-concave below the fair wage. Thus, if the firm pays below the fair wage, it is efficient to pay him zero. Since the fair wage is increasing in output, the disutility from zero wages is also increasing in output, and so it is optimal to pay zero wages for low output levels.

Proposition 3 shows how whether the participation constraint binds affects the optimal contract. When the participation constraint does not bind, we have $q_m > 0$, so that the contract has a discontinuity between zero payments and positive payments. It is optimal to pay the agent the most unfair feasible wage for low outputs, to incentivize him to exert effort and avoid low outputs. However, doing so risks the agent suffering significant disutility and may fail to ensure the agent's participation. Thus, when the participation constraint binds, the contract may not have a discontinuity (see Proposition 3). Overall, the participation constraint binding is a necessary but insufficient condition for pay to be a continuous function of output. As a result, the increasing use of performance shares, which do contain discontinuities, is consistent with the participation constraint no longer binding for many CEOs – that they are willing to accept unfair pay for low output levels suggests that they are above their outside option.

Example 2 *A special case of Proposition 3 is as follows. Consider "normalized" log utility,* $v(w) = \ln(w+1)$,[15] *and an output that follows a truncated normal distribution on* $(0, \infty)$ *with parameters* e^* *and* σ, $LR(q|e^*) \propto q + cste$, *and* $v'^{-1}\left(1/\left(\lambda_1 + \lambda_2 LR(q|e^*)\right)\right)$ *is linear in* q. *The contract is illustrated in Figure 4.*

In panel (a) of Figure 4, the incentive constraint is nonbinding (i.e. $e^* > e^T$*), but the participation constraint is binding. Intuitively, paying the agent zero for low outputs would increase incentives, but significantly reduce the agent's utility due to the perceived unfairness, and may lead to his participation constraint being violated. Since the participation constraint is binding,*

[15] This yields $v'(w) = \frac{1}{w} - 1$, so that $v'^{-1}\left(1/\lambda LR(q|e^*)\right) = LR(q|e^*) - 1$.

the principal pays the agent the fair wage for some low outputs. In panel (b), a lower reservation utility $\overline{U}$ leads to the participation constraint being nonbinding, and the incentive constraint being binding instead (i.e. $e^* = e^T$). As a result, contract design is now driven by its effect on incentives rather than on the agent's utility, and the fair wage is no longer paid for outputs with a negative likelihood ratio.[16] In panel (c), the cost of effort is higher than in panel (b), requiring the principal to increase incentives. She does so by paying the fair wage rather than zero for a larger subset of outputs with a positive likelihood ratio, and payments higher than the fair wage for very high outputs.

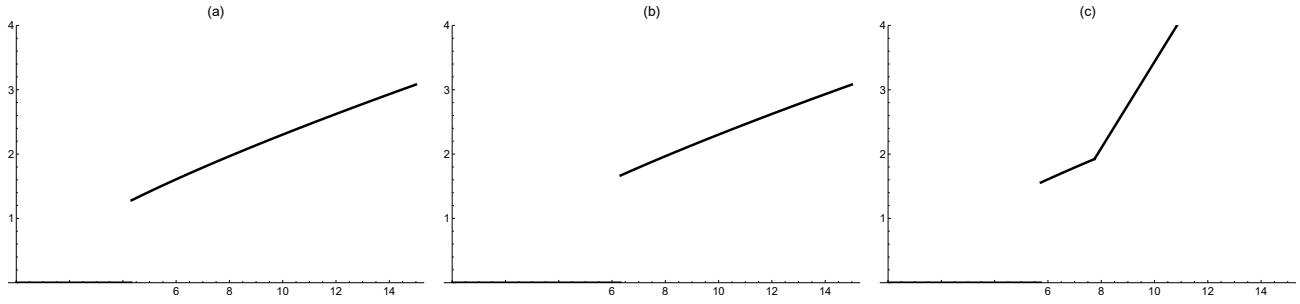

Figure 4: The contract $w(q)$ as a function of q. The agent's preferences are as in Example 2 with $\nu(w,q) = (w+1)^{1.2} - 1 - \frac{1}{5}q$, $C(e) = c \times e^2$, and $e^T = 5$. (a): $\overline{U} = 0$ and $c = 0.02$. (b): $\overline{U} = -2$ and $c = 0.02$. (c): $\overline{U} = -2$ and $c = 0.05$.

4 Conclusion

This paper has studied optimal contracting under fairness preferences, where the agent's perceived fair wage depends on output. We started with a model in which the agent's utility function is piecewise linear – it has a slope of 1 above the perceived fair wage, and in excess of 1 below it. The wedge between the slopes is increasing in the agent's fairness concerns, and the perceived fair wage is itself linear in output. We showed that fairness concerns do not lead to the agent being paid fair wages for all output levels; in contrast, unfair wages can be effective to induce effort. The optimal contract involves two thresholds for output. The agent receives zero below the lower threshold, the entire output above the upper threshold, and the fair wage in between. When fairness concerns are sufficiently strong, the upper region in which the agent receives the entire output disappears, and the contract becomes performance shares. The agent is given shares that pay him his perceived fair share of output, unless output falls below a threshold. The model thus rationalizes the common usage of performance shares in reality; most other contracting theories predict continuous contracts, or extreme discontinuities where the agent's pay switches from zero to the entire output.

[16] When the incentive constraint is binding so that $e^* = e^T$, the likelihood ratio is positive for $LR(q|e^T) > q_0^{e^T} \approx 5$. The approximation is due to the use of the truncated normal distribution with $e = 5$ and $\sigma = 1$.

We then extend the model to a general setting in which the agent is risk-averse, and the perceived fair wage is only increasing in output – it need not be linear. The contract retains the same three regions as in the piecewise linear model, but there is an additional fourth region, in-between the regions in which the CEO receives the fair wage and the entire output. In this region, his payment exceeds the fair wage, and is generally convex in output.

In both models, we show that, even if the incentive constraint is slack, pay is increasing in output – by paying the agent the fair wage over a greater range of outputs, this reduces perceived unfairness and allows the participation constraint to be satisfied at least cost by reducing perceived unfairness. As a result, the firm can induce CEO effort "for free", in contrast to standard risk-neutral models in which inducing higher effort involves paying the agent limited liability rent, and standard risk-averse models in which it requires paying the agent a risk premium. This result may rationalize why performance-related pay is given to agents even if they are intrinsically motivated, or even if there are alternative solutions to the moral hazard problem such as monitoring.

This paper is a first step in modeling CEO pay under fairness preferences, using the standard model to make transparent how fairness concerns affect the optimal contract. For future research, it may be fruitful to explore the other potential determinants of the fair wage suggested by the survey of Edmans, Gabaix, and Jenter (2022), such as peer firm pay in a model of multiple firms, or last year's pay in a dynamic model.

References

[1] Akerlof, George A., and Janet L. Yellen. (1990): "The fair wage-effort hypothesis and unemployment." *Quarterly Journal of Economics* 105, 255–283.

[2] Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb (2022): "How Should Performance Signals Affect Contracts?" *Review of Financial Studies* 35, 168–206.

[3] Chaigneau, Pierre, Nicolas Sahuguet, and Bernard Sinclair-Desgagné (2017): "Prudence and the Convexity of Compensation Contracts." *Economics Letters* 157, 14–16.

[4] Charness, Gary and Matthew Rabin (2002): "Understanding Social Preferences With Simple Tests." *Quarterly Journal of Economics* 117, 817-869

[5] de Meza, David and David C. Webb (2007): "Incentive Design Under Loss Aversion." *Journal of the European Economic Association* 5, 66–92.

[6] Dittmann, Ingolf and Ernst Maug (2007): "Lower Salaries and No Options? On the Optimal Structure of Executive Pay." *Journal of Finance* 62, 303–343.

[7] Dittmann, Ingolf, Ernst Maug, and Oliver Spalt (2010): "Sticks or Carrots? Optimal CEO Compensation when Managers Are Loss Averse." *Journal of Finance* 65, 2015–2050.

[8] Edmans, Alex, Tom Gosling, and Dirk Jenter (2022): "CEO Compensation: Evidence From the Field." Working Paper, London Business School.

[9] Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics* 108, 437–459.

[10] Fehr, Ernst, Alexander Klein, and Klaus M. Schmidt (2007): "Fairness and Contract Design." *Econometrica* 75, 121–154.

[11] Fehr, Ernst and Klaus M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114, 817–868.

[12] Grossman, Sanford J. and Oliver D. Hart (1983): "An Analysis of the Principal-Agent Problem." *Econometrica* 51, 7–45.

[13] Herzberg, Frederick. 1959. The Motivation to Work. New York: Wiley.

[14] Holmström, Bengt (1979): "Moral Hazard and Observability." *Bell Journal of Economics* 10, 74–91.

[15] Innes, Robert D. (1990): "Limited Liability and Incentive Contracting with Ex-Ante Action Choices." *Journal of Economic Theory* 52, 45–67.

[16] Jewitt, Ian, Ohad Kadan, and Jeroen M. Swinkels (2008): "Moral Hazard with Bounded Payments." *Journal of Economic Theory* 143, 59–82.

[17] Rabin, Matthew (1993): "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83, 1281–1302.

[18] Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara and Shmuel Zamir (1991): "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *American Economic Review* 81, 1068–1095.

[19] Sobel, Joel (2005): "Interdependent Preferences and Reciprocity." *Journal of Economic Literature* 63, 392–436.

A Proofs

Proof of Lemma 1:

For a given contract $w(q)$, the effort choice problem of the agent can be written as

$$\max_e \int_0^{\overline{q}} u\left(w(q), q\right) \phi(q|e) dq - C(e).$$

The second derivative of the agent's objective function with respect to e is negative for any e if and only if:

$$\int_0^{\overline{q}} u\left(w(q), q\right) \frac{\partial^2 \phi(q|e)}{\partial e^2} dq < C''(e) \quad \forall e \in (0, \bar{e}). \tag{26}$$

With principal limited liability (see equation (4)), since the utility function increasing in w, the maximum value of u for a given q is $u\left(q, q\right)$. In addition, with agent limited liability (see equation (4)), the minimum payment is $w(q) = 0$; with a utility function increasing in w, this implies that the minimum value of u for a given q is $u\left(0, q\right)$. Therefore, for any given q:

$$u\left(w(q), q\right) \in \left[u\left(0, q\right), u\left(q, q\right)\right].$$

Using notations K_e^+ and K_e^- defined in equations (6) and (7), the expression on the left-hand side ("LHS") of equation (26) can then be rewritten as:

$$\int_0^{\overline{q}} u\left(w(q), q\right) \min\left\{\frac{\partial^2 \phi(q|e)}{\partial e^2}, 0\right\} dq + \int_0^{\overline{q}} u\left(w(q), q\right) \max\left\{\frac{\partial^2 \phi(q|e)}{\partial e^2}, 0\right\} dq. \tag{27}$$

As established above, we have $u\left(w(q), q\right) \geq u(0, q)$ for any q, and $u\left(w(q), q\right) \leq u(q, q)$ for any q. Therefore, for any q such that $\frac{\partial^2 \phi(q|e)}{\partial e^2} \leq 0$ we have $u\left(w(q), q\right) \frac{\partial^2 \phi(q|e)}{\partial e^2} \leq u(0, q) \frac{\partial^2 \phi(q|e)}{\partial e^2}$; and for any q such that $\frac{\partial^2 \phi(q|e)}{\partial e^2} \geq 0$ we have $u\left(w(q), q\right) \frac{\partial^2 \phi(q|e)}{\partial e^2} \leq u(q, q) \frac{\partial^2 \phi(q|e)}{\partial e^2}$. Integrating over q, this implies that expression (27) is less than:

$$\int_0^{\overline{q}} \left(K_e^- u(0, q) + K_e^+ u(q, q)\right) dq,$$

which completes the proof.

Proof of Proposition 1:

We describe the optimal contract when $e^T = 0$, i.e. the IC does not bind for any contract. In the optimization problem with a nonbinding IC, the IR for $e^* \geq 0$ must be binding. Suppose that it is not. Then, the contract that solves the optimization problem in equations (1), (4), and (5) is simply $w(q) = 0$ for any q, which gives utility $u(0, q) = -\gamma \max\{\rho q, 0\} = -\gamma \rho q$ for any $q \in [0, \overline{q}]$, so that:

$$\int_0^{\overline{q}} u(0, q) \phi(q|e^*) dq - C(e^*) = -\gamma \rho \int_0^{\overline{q}} q \phi(q|e^*) dq - C(e^*) < \overline{U},$$

where the inequality follows from equation (12) given that $C' > 0$ and $e^* \geq 0$. This implies that the IR is not satisfied, a contradiction.

The relaxed optimization problem with a nonbinding IC, a binding IR, and the FOA, is:

$$\min_{w(q),e^*} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{28}$$

$$\text{s.t.} \quad \int_0^{\overline{q}} u(w(q),q)\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{29}$$

$$\int_0^{\overline{q}} u(w(q),q)\frac{\partial \phi(q|e^*)}{\partial e}dq = C'(e^*) \tag{30}$$

$$0 \leq w(q) \leq q \tag{31}$$

$$w(q) \geq w(q') \ \forall q > q' \tag{32}$$

At first, we hold effort constant. We then consider the effects of the contract on the effort choice, which matters because it affects the equilibrium effort e^*, and therefore the LHS of equation (29).

Lemma 2 *Let the utility function be as in equation (9) and suppose that effort is constant. On any non-empty subinterval of* $[0,\overline{q}]$*, the optimal contract is such that* $w(q) \leq w^*(q)$*.*

Proof. This proof is by contradiction. Suppose that the contract is not such that by $w(q) \leq w^*(q)$ for all q.

A contract such that $w(q) \geq w^*(q)$ for all $q \in [0,\overline{q}]$ with a strict inequality for some q would not solve the optimization program in equations (28)-(31). Indeed, due to equation (13) the agent would be strictly above his reservation utility so that the payment $w(q)$ could be reduced on some subinterval of $[0,\overline{q}]$, which would decrease the cost of the contract in equation (28) without violating the IR in equation (29) or the limited liability constraints in equation (31), a contradiction.

We now consider a contract which is neither such that $w(q) \leq w^*(q)$ for all q nor such that $w(q) \geq w^*(q)$ for all q with a strict inequality for some q. We show that this contract, which is such that $w(q') > w^*(q')$ for some $q' \in [0,\overline{q}]$ and $w(q) < w^*(q)$ for some $q \in [0,\overline{q}]$ is suboptimal. Denote by Q^+ the subinterval of $[0,\overline{q}]$ such that $w(q) > w^*(q)$. Denote by Q^- the subinterval of $[0,\overline{q}]$ such that $w(q) < w^*(q)$. Consider the following perturbation: for $q \in Q^-$, increase $w(q)$ by $\epsilon/\phi(q|e^*)$, and for $q' \in Q^+$ decrease $w(q')$ by $\epsilon/\phi(q'|e^*)$, where ϵ is positive and arbitrarily small. By construction, this perturbation is cost-neutral for a given effort, i.e. it does not change the principal's objective function. Now consider the effect on the LHS of the IR in equation (29). Since $w(q) \in [0, w^*(q))$ and $w(q') \in (w^*(q'), q']$, the change in the LHS of the IR is: $\epsilon(1+\gamma) - \epsilon = \epsilon\gamma$, which is strictly positive since $\gamma > 0$. Since the LHS of the IR increases and the IR is binding, standard arguments show that it is then possible to construct a contract that leaves the LHS of the IR unchanged relative to the initial contract and reduces the cost of the contract to the principal, which establishes that the initial contract was suboptimal. This rules out any contract such that $w(q) > w^*(q)$ for some q. ■

When the IC is nonbinding, for a given effort, the optimization program in equations (28), (29), and (31) has an infinity of solutions: any contract such that $w(q) \in [0, w^*(q)]\ \forall q$ and equation (29) holds would solve this optimization problem. However, even when the IC is nonbinding, effort is not exogenously given. Moreover, the cost of effort enters the LHS of the IR in equation (29) (by assumption, $C' > 0$). Suppose for now that there exists a contract with $w(q) \in [0, w^*(q)]\ \forall q$ and the monotonicity constraint that induces $e^* = 0$ ($e = 0$ is the minimum effort). Because of the monotonicity assumption in equation (5) and MLRP, the effort minimizing contract such that equation (29) is satisfied, $w(q) \in [0, w^*(q)]\ \forall q$, and the induced effort is $e^* = 0$ takes the form:

$$w(q) = \begin{cases} w^*(q) & \text{if } q < q_c \\ \rho q_c & \text{if } q \geq q_c \end{cases}, \tag{33}$$

where q_c adjust so that the IR in equation (29) is satisfied, and $e^* = e^c$ is the induced effort. If $e^c = 0$, then this contract is optimal since it is indeed the effort minimizing contract even without the restriction $w(q) \in [0, w^*(q)]\ \forall q$.

We now study the case in which $e^c > 0$. For a contract as in equation (33) and an equilibrium effort e^* given by equation (2) with $e^T = 0$, define:

$$A(q_c, e) \equiv \int_0^{q_c} \rho q \frac{\partial}{\partial e}\phi(q|e)dq + \int_{q_c}^{\overline{q}} \left(\rho q_c - \gamma(\rho q - \rho q_c)\right) \frac{\partial}{\partial e}\phi(q|e)dq \tag{34}$$

Given the FOA, for a given q_c the equilibrium effort is zero if and only if $A(q_c, 0) \leq 0$. Since the IR is binding and the FOA applies for any interior solution to the effort choice problem, the optimal value of q_c and the effort e^* induced by a given contract with threshold q_c are implicitly defined by:

$$\int_0^{q_c} \rho q \phi(q|e)dq + \int_{q_c}^{\overline{q}} \left(\rho q_c - \gamma(\rho q - \rho q_c)\right) \phi(q|e)dq - C(e^*) = \overline{U} \tag{35}$$

$$e^* = \begin{cases} 0 & \text{if } A(q_c, 0) \leq 0 \\ C'^{-1}\left(A(q_c, e^*)\right) & \text{if } A(q_c, 0) > 0 \end{cases} \tag{36}$$

For $q_c = \overline{q}$, $A(q_c, e)$ has the same sign as $\int_0^{\overline{q}} q \frac{\partial}{\partial e}\phi(q|e)dq$, which is strictly positive for any e because of $\int_0^{\overline{q}} \frac{\partial}{\partial e}\phi(q|e)dq = 0$ for any e and MLRP. Therefore, for $q_c = \overline{q}$, we have $A(0, e) > 0$ for any e, including $e = 0$, i.e. $e^* > 0$ (see equation (36)). For $q_c = 0$, $A(q_c, e)$ has the same sign as $-\int_0^{\overline{q}} q \frac{\partial}{\partial e}\phi(q|e)dq$, which is strictly negative for any e, i.e. $e^* = 0$ (see equation (36)). Moreover, the derivative of the right-hand side ("RHS") of equation (34) with respect to q_c holding effort constant at $e = e^*$ is:

$$\rho q_c \frac{\partial}{\partial e}\phi(q_c|e^*) + \rho q_c \frac{\partial}{\partial e}\phi(q_c|e^*) + \int_{q_c}^{\overline{q}} \rho\left(1 + \gamma\right) \frac{\partial}{\partial e}\phi(q|e^*)dq,$$

which is strictly positive if $\frac{\partial}{\partial e}\phi(q_c|e^*) > 0$, which by MLRP and definition of $q_0^{e^*}$ is equivalent to $q_c > q_0^{e^*}$.

We now determine for which parameter values this is the case. Consider the effect of a change in q_c on the LHS of the IR in equation (35) when $e^* > 0$ and is therefore given by the first-order condition ("FOC") to the agent's effort choice problem given the FOA. The derivative of the LHS of equation (35) with respect to q_c holding effort constant is:

$$\rho q_c \phi(q_c|e^*) + \rho q_c \phi(q_c|e^*) + \int_{q_c}^{\overline{q}} \rho\,(1+\gamma)\,\phi(q|e^*)dq,$$

which is strictly positive for any e. Moreover, the LHS of the IR in equation (35) is the agent's objective function, and the agent chooses effort e to maximize this objective function. From the envelope theorem, we know that the total effect of a marginal change in a parameter (here q_c) on the objective function is equal to its effect holding effort constant. In sum, the LHS of equation (35) is continuously decreasing in q_c, and given optimal effort choices it is below the RHS for $q_c = 0$ according to equation (12), and it is above the RHS for $q_c = \overline{q}$ according to equation (13). This implies that there exists q_c that satisfies equation (35), and this level of q_c is strictly increasing in $\overline{U}$ according to equation (35). In sum, when $\overline{U}$ is sufficiently high, we have $q_c > q_0^{e^*}$, so that $e^* = e^c > 0$ with a contract as in equation (33) according to the preceding paragraph.

Suppose that there does not exist a contract as in equation (33) that satifies IR in equation (29) and such that $e^* = 0$ (this can happen for example if equation (13) holds as an equality given that the equilibrium effort when $w(q) = w^*(q)\,\forall q$ is positive by MLRP). In addition, suppose that there exists a contract that satifies IR in equation (29) and such that $e^* = 0$ (for example a contract such that $w(q) = q$ for $q \leq q_t$, and $w(q) = q_t$ for $q \geq q_t$, where $q_t \in [0, \overline{q}]$). The latter contract that induces $e^* = 0$ and such that $w(q) > w^*(q)$ for a non-empty subinterval is more costly to the principal when γ is sufficiently high. Thus, even with $e^T = 0$, we can have $e^* > 0$.

Proof of Proposition 2:

We describe the optimal contract when the IC binds. When the condition from Lemma 1 holds so that the FOA applies, a binding IC can be rewritten as:

$$\int_0^{\overline{q}} u(w(q), q)\frac{\partial}{\partial e}\phi(q|e^*)dq = C'(e^*), \tag{37}$$

with $e^* = e^T$ since the IC binds. We now describe the optimal contract when the IC in equation (2) binds.

The first step of the proof establishes that a contract as described in equation (17) is optimal. To this end, we rely on the agent's monotonicity constraint in equation (5) and on Lemma 3 below.

Lemma 3 *Let the utility function be as in equation (9) and suppose that the IC is binding. The optimal contract is such that $w(q) \notin (w^*(q), q)$ for any q.*

Proof. This proof is by contradiction. Suppose that for some q we have $w(q) \in (w^*(q), q)$. Consider any given initial incentive-compatible contract and the following perturbation for any $q > q'$ in this subinterval, increase $w(q)$ by $\epsilon/\phi(q|e^*)$, and decrease $w(q')$ by $\epsilon/\phi(q'|e^*)$, where ϵ is positive and arbitrarily small. By construction, for a given effort this perturbation does not change the principal's or the agent's objective function (note that the agent's objective function is linear in w for any w and q such that $w > w^*(q)$). Now consider the effect on the LHS of the IC in equation (37). With $w(q) \in (w^*(q), q)$ and $w(q') \in (w^*(q'), q')$, the change in the LHS of the IC is:

$$\epsilon \left(LR(q|e^*) - LR(q'|e^*) \right),$$

which is strictly positive by MLRP. Since the LHS of the IC increases and the IC is binding, standard arguments show that it is then possible to construct a contract that leaves the LHS of the IC and IR unchanged compared to the initial contract and reduces the cost of the contract to the principal, which establishes that the initial contract was suboptimal. This rules out any contract such that $w(q) \in (w^*(q), q)$ for any q. ■

The second step of the proof establishes the values of q_m and q_M for a given effort e^* to be induced.

The relaxed optimization problem with $q_m \in [0, \overline{q}]$ and $q_M \in [q_m, \overline{q}]$ is:

$$\min_{q_m, q_M} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{38}$$

$$\text{s.t.} \quad \int_0^{\overline{q}} u(w(q), q) \frac{\partial}{\partial e}\phi(q|e^*)dq = C'(e^*) \tag{39}$$

$$\int_0^{\overline{q}} u(w(q), q)\phi(q|e^*)dq - C(e^*) \geq \overline{U} \tag{40}$$

$$w(q) = \begin{cases} 0 & \text{for } q < q_m \\ w^*(q) & \text{for } q \in [q_m, q_M] \\ q & \text{for } q > q_M \end{cases} \tag{41}$$

With the utility function defined in equation (9), this can be rewritten as, for $q_m \in [0, \overline{q}]$ and $q_M \in [q_m, \overline{q}]$:

$$\min_{q_m, q_M} \int_{q_m}^{q_M} \rho q \phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\phi(q|e^*)dq \tag{42}$$

$$\text{s.t.} \int_0^{q_m} (-\gamma\rho q) \frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_m}^{q_M} \rho q \frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q \frac{\partial}{\partial e}\phi(q|e^*)dq = C'(e^*) \tag{43}$$

$$\int_0^{q_m} (-\gamma\rho q) \phi(q|e^*)dq + \int_{q_m}^{q_M} \rho q \phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) \geq \overline{U} \tag{44}$$

Denote by η_{IC} and η_{IR} the Lagrange multipliers associated with the constraints in equations (43)

and (44), respectively. The FOC for an interior solution are:

$$-\rho q_m \phi(q_m|e^*) - \eta_{IC}\left(-\gamma\rho q_m \frac{\partial}{\partial e}\phi(q_m|e^*) - \rho q_m \frac{\partial}{\partial e}\phi(q_m|e^*)\right) - \eta_{IR}\left(-\gamma\rho q_m \phi(q_m|e^*) - \rho q_m \phi(q_m|e^*)\right) = 0 \tag{45}$$

$$\rho q_M \phi(q_M|e^*) - q_M \phi(q_M|e^*) - \eta_{IC}\left(\rho q_M \frac{\partial}{\partial e}\phi(q_M|e^*) - q_M \frac{\partial}{\partial e}\phi(q_M|e^*)\right) - \eta_{IR}\left(\rho q_M \phi(q_M|e^*) - q_M \phi(q_M|e^*)\right) = 0 \tag{46}$$

which for $q_m \neq 0$ and $q_M \neq 0$ is equivalent to:

$$-1 + \eta_{IC}\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)}(1+\gamma) + \eta_{IR}(1+\gamma) = 0 \tag{47}$$

$$-1 + \eta_{IC}\frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)} + \eta_{IR} = 0 \tag{48}$$

The optimal value of q_m is generically not described by a corner solution. We have $q_m = 0$ in a nongeneric case: when equation (13) is satisfied as an equality at $e^* = e^T$. Now suppose that equation (13) is not satisfied as an equality at $e^* = e^T$, i.e. it is satisfied as a strict inequality. This implies that the IR is nonbinding when $q_m = 0$ (indeed, it is nonbinding for $q_m = 0$ and $q_M = \overline{q}$, and the LHS of the IR in equation (44) is decreasing in q_M). Moreover, a contract with $q_m = 0$ does not provide incentives at the minimum cost, since increasing q_m would increase the LHS of the IC in equation (43) while reducing the cost of the contract in equation (42). Finally, since the IR is nonbinding at $q_m = 0$ and its LHS is continuously differentiable in q_m, the increase in q_m can be small enough that the new contract still satisfies IR. In sum, if equation (13) is not satisfied as an equality at $e^* = e^T$, then we cannot have $q_m = 0$ at the optimal contract.

Likewise, we cannot have $q_m = \overline{q}$, which would imply $q_M = \overline{q}$, at the optimal contract. Indeed, this would violate the IC in equation (43) since the LHS would then be negative and the RHS positive; this would also violate the IR in equation (44) according to equation (12).

Thus, the optimal value of q_m is generically given by the first-order condition in equation (47), which can be rearranged as:

$$LR(q_m|e^*) = \frac{1}{\eta_{IC}}\left(\frac{1}{1+\gamma} - \eta_{IR}\right),$$

where $\eta_{IC} \geq 0$ and $\eta_{IR} \geq 0$.

There are two cases.

Nonbinding IR. In the optimization problem with a nonbinding IR, the IC for $e^T > 0$ must be binding. Suppose that it is not. Then, the contract that solves the optimization problem in equations (1), (4), and (5) is simply $w(q) = 0$ for any q, so that $u(0, q) = -\gamma \max\{\rho q, 0\} = -\gamma\rho q$

for any $q \in [0, \overline{q}]$, and:

$$\int_0^{\overline{q}} u(0,q) \frac{\partial}{\partial e} \phi(q|e) dq = -\gamma\rho \int_0^{\overline{q}} q \frac{\partial}{\partial e} \phi(q|e) dq < 0 < C'(e),$$

for any $e > 0$, i.e. the IC is not satisfied, a contradiction.

If the optimal values of q_m and q_M are interior solutions, equations (47) and (48) with $\eta_{IR} = 0$ (nonbinding IR) and $\eta_{IC} > 0$ (binding IC) immediately give:

$$\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)}(1+\gamma) = \frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)}. \tag{49}$$

With a nonbinding IR, we establish that $q_m \geq q_0^{e^*}$, where $e^* = e^T$. Consider any given initial compensation contract such that $q_m < q_0^{e^*}$ and the following perturbation: increase q_m by an arbitrarily small amount. This perturbation increases the LHS of the IC and reduces the cost of the contract to the principal. Standard arguments show that it is then possible to construct a contract that leaves the LHS of the IC unchanged compared to the initial contract and reduces the cost of the contract to the principal, which establishes that the initial contract was suboptimal.

Denote the subset of values of $\{q_m, q_M\}$ that satisfy the IC by $\mathcal{Q}^{IC}$, and denote the values of $\{q_m, q_M\}$ in this subset by $\{q_m^{IC}, q_M^{IC}\}$. Let q_M^{IC} be a function of q_m^{IC}. This is a continuous function by the implicit function theorem since the LHS of the IC in equation (39) is continuously differentiable in q_m and q_M, and the product of continuous functions is continuous.

Totally differentiating the LHS of the IC with respect to q_m^{IC} and taking into account the effect on q_M^{IC} so that the LHS of the IC remains unchanged gives:

$$\begin{aligned}
\frac{d}{dq_m^{IC}} \int_0^{\overline{q}} u(w(q),q) \frac{\partial}{\partial e}\phi(q|e^*)dq &= \left(u(0,q_m^{IC}) - u(w^*(q_m^{IC}), q_m^{IC})\right) \frac{\partial}{\partial e}\phi(q_m^{IC}|e^*) \\
&\quad + \left(\left(u(w^*(q_M^{IC}), q_M^{IC}) - u(q_M^{IC}, q_M^{IC})\right) \frac{\partial}{\partial e}\phi(q_M^{IC}|e^*) \right) \frac{dq_M^{IC}}{dq_m^{IC}} \\
&= -(1+\gamma) w^*(q_m^{IC}) \frac{\partial}{\partial e}\phi(q_m^{IC}|e^*) - (q_M^{IC} - w^*(q_M^{IC})) \frac{\partial}{\partial e}\phi(q_M^{IC}|e^*) \frac{dq_M^{IC}}{dq_m^{IC}} = 0 \\
\Leftrightarrow \quad \frac{dq_M^{IC}}{dq_m^{IC}} &= -\frac{(1+\gamma) w^*(q_m^{IC})}{q_M^{IC} - w^*(q_M^{IC})} \frac{\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*)}{\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)},
\end{aligned} \tag{50}$$

where both the numerator and the denominator of the second fraction on the RHS are positive since $q_0^{e^*} \leq q_m \leq q_M$.

Now consider the subset $\mathcal{Q}^c$ of values of $\{q_m, q_M\}$, denoted by $\{q_m^c, q_M^c\}$, that leaves the expected

cost of the contract in equation (38) unchanged for the principal. By construction:

$$\frac{d}{dq_m^c}\int_0^{\overline{q}} w(q)\phi(q|e^*)dq = -w^*(q_m^c)\phi(q_m^c|e^*) - (q_M^c - w^*(q_M^c))\phi(q_M^c|e^*)\frac{dq_M^c}{dq_m^c} = 0$$
$$\Leftrightarrow \quad \frac{dq_M^c}{dq_m^c} = -\frac{w^*(q_m^c)}{q_M^c - w^*(q_M^c)}\frac{\phi(q_m^c|e^*)}{\phi(q_M^c|e^*)}. \tag{51}$$

Because of MLRP, for $q_0^{e^*} \leq q_m \leq q_M$, we have:

$$\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)} \leq \frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)} \quad \Leftrightarrow \quad \frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\frac{\partial}{\partial e}\phi(q_M|e^*)} \leq \frac{\phi(q_m|e^*)}{\phi(q_M|e^*)}, \tag{52}$$

with strict inequalities for $q_M > q_m$.

For any given element in $\mathcal{Q}^{IC}$, there are two possible cases:

1) For values of q_m^{IC} and q_M^{IC} such that $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*)}{\phi(q_m^{IC}|e^*)} < \frac{\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)}{\phi(q_M^{IC}|e^*)}$ and $q_m \geq q_0^{e^*}$, a marginal increase in q_m and associated decrease in q_M (since $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$) that satisfies incentive compatibility as in equation (50) results in a lower cost to the principal because of equations (51) and (52).

2) For values of q_m^{IC} and q_M^{IC} such that $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*)}{\phi(q_m^{IC}|e^*)} > \frac{\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)}{\phi(q_M^{IC}|e^*)}$ and $q_m \geq q_0^{e^*}$, a marginal increase in q_m and associated decrease in q_M (since $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$) that satisfies incentive compatibility as in equation (50) results in a higher cost to the principal because of equations (51) and (52).

Consider the smallest value for q_m^{IC} and corresponding highest value for q_M^{IC} in the subset $\mathcal{Q}^{IC}$, and denote them by $q_m^{\min}$ and $q_M^{\max}$. We can show by construction that $q_M^{\max} = \overline{q}$: according to equations (11), an incentive-compatible contract such that $q_m \geq q_0^{e^*}$ and $q_M = \overline{q}$ exists; by definition of $\mathcal{Q}^{IC}$ and $q_M^{\max}$, this means that $q_M^{\max} = \overline{q}$. Since IR is nonbinding and the cost of a contract is decreasing in q_m, all else equal, $q_m^{\min}$ is implicitly defined by incentive compatibility with $q_M^{\max} = \overline{q}$ in equation (14). From equations (11) and (14), we have $q_m^{\min} \geq q_0^{e^*}$. There are two cases.

First, if $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{\min}|e^*)}{\phi(q_m^{\min}|e^*)} > \frac{\frac{\partial}{\partial e}\phi(\overline{q}|e^*)}{\phi(\overline{q}|e^*)}$, then due to MLRP and $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$, for any element of $\mathcal{Q}^{IC}$, we have $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*)}{\phi(q_m^{IC}|e^*)} > \frac{\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)}{\phi(q_M^{IC}|e^*)}$, so that case 2) described above is relevant for any element of $\mathcal{Q}^{IC}$. Therefore, the optimal values of q_m and q_M are respectively $q_m^{\min}$ and $\overline{q}$. That is:

$$w(q) = \begin{cases} 0 & \text{for } q \in [0, q_m^{\min}) \\ w^*(q) & \text{for } q \in [q_m^{\min}, \overline{q}] \end{cases}, \tag{53}$$

where $q_m^{\min}$ is defined in equation (14).

Second, if $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{\min}|e^*)}{\phi(q_m^{\min}|e^*)} < \frac{\frac{\partial}{\partial e}\phi(\overline{q}|e^*)}{\phi(\overline{q}|e^*)}$, then for elements in the subset $\mathcal{Q}^{IC}$, for low enough values of q_m^{IC} and high enough values of q_M^{IC}, case 1) described above is relevant. Moreover, since $\gamma > 0$ and the likelihood ratio $LR(q|e)$ is continuous in q by assumption, for elements in the subset

$\mathcal{Q}^{IC}$, for high enough values of q_m^{IC} and low enough values of q_M^{IC} (since $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$), case 2) described above is relevant. In sum, the optimal values of q_m and q_M belong to the subset $\mathcal{Q}^{IC}$ and satisfy the following equation:

$$(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)} = \frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)}. \tag{54}$$

Binding IR. When both the IC and IR are binding, q_m and q_M must satisfy:

$$-\gamma\rho\int_0^{q_m} q\frac{\partial}{\partial e}\phi(q|e^*)dq + \rho\int_{q_m}^{q_M} q\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e^*)dq = C'(e^*) \tag{55}$$

$$-\gamma\rho\int_0^{q_m} q\phi(q|e^*)dq + \rho\int_{q_m}^{q_M} q\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{56}$$

We also know that the optimal value of q_m is generically an interior solution, so we have three cases.

1. First, if the optimal values of q_m and q_M are interior solutions, equations (47) and (48) with $\eta_{IR} > 0$ and $\eta_{IC} > 0$ immediately give:

$$\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)}(1+\gamma) + \frac{\eta_{IR}}{\eta_{IC}}\gamma = \frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)} \tag{57}$$

   Because of MLRP, the LHS of equation (57) is strictly increasing in q_m, and the RHS is strictly increasing in q_M. Thus, for any pair $\{q_m, q_M\}$ that satisfy this equation, q_M is strictly increasing in q_m.

2. If $q_M = q_m$, then q_m must satisfy:

$$-\gamma\rho\int_0^{q_m} q\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e^*)dq = C'(e^*) \tag{58}$$

$$-\gamma\rho\int_0^{q_m} q\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{59}$$

   The LHS of the IR in equation (59) is strictly decreasing in q_m. Thus, there exists at most one value of q_m such that equation (59) holds, and this value is strictly decreasing in $\overline{U}$. The derivative of the LHS of IC in equation (58) with respect to q_m is $q_m\frac{\partial}{\partial e}\phi(q_m|e^*)(-\gamma\rho - 1)$, which by MLRP and definition of $q_0^{e^*}$ is positive if and only if $q_m < q_0^{e^*}$. Thus, there exists at most two values of q_m such that equation (59) holds, and these values are independent of $\overline{U}$. In sum, generically we cannot have IC and IR binding with $q_M = q_m$.

3. If $q_M = \overline{q}$, then q_m must satisfy:

$$-\gamma\rho \int_0^{q_m} q \frac{\partial}{\partial e}\phi(q|e^*)dq + \rho \int_{q_m}^{\overline{q}} q \frac{\partial}{\partial e}\phi(q|e^*)dq = C'(e^*) \tag{60}$$

$$-\gamma\rho \int_0^{q_m} q\phi(q|e^*)dq + \rho \int_{q_m}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{61}$$

The LHS of the IR in equation (61) is strictly decreasing in q_m. Thus, there exists at most one value of q_m such that equation (61) holds, and this value is strictly decreasing in $\overline{U}$. The derivative of the LHS of IC in equation (60) with respect to q_m is $q_m \frac{\partial}{\partial e}\phi(q_m|e^*)\left(-\gamma\rho - \rho\right)$, which by MLRP and definition of $q_0^{e^*}$ is positive if and only if $q_m < q_0^{e^*}$. Thus, there exists at most two values of q_m such that equation (61) holds, and these values are independent of $\overline{U}$. In sum, generically we cannot have IC and IR binding with $q_M = \overline{q}$.

Proof of Corollary 1:

When the IC is binding with a contract as in Proposition 2, the IC can be written as in equation (55). In this equation, only the first term on the LHS depends on γ. Furthermore, with $q_m \leq q_0^{e^T}$ and a binding IC which implies $e^* = e^T$, we have $\frac{\partial}{\partial e}\phi(q|e^*) < 0$ for any $q < q_m$, so that the first term on the LHS of equation (55) is strictly positive. Moreover, with $q_m \leq q_0^{e^T}$ the IR must be binding as established in the proof of Proposition 2.

With a contract as in Proposition 2, the first derivatives of the LHS of the IC and IR in equations (55) and (56) with respect to γ are respectively:

$$-\rho \int_0^{q_m} q \frac{\partial}{\partial e}\phi(q|e^*)dq > 0 \qquad \text{and} \qquad -\gamma\rho \int_0^{q_m} q\phi(q|e^*)dq < 0.$$

Thus, following a marginal change in γ, q_m and q_M must change in a way that decreases the LHS of the IC and increases the LHS of the IR.

With a contract as in Proposition 2, the first derivatives of the LHS of the IC in equation (55) with respect to q_m and q_M are respectively:

$$-(1+\gamma)\rho q_m \frac{\partial}{\partial e}\phi(q_m|e^*) > 0 \qquad \text{and} \qquad (\rho - 1)q_M \frac{\partial}{\partial e}\phi(q_M|e^*) < 0.$$

With a contract as in Proposition 2, the first derivatives of the LHS of the IR in equation (56) with respect to q_m and q_M are respectively:

$$-(1+\gamma)\rho q_m \phi(q_m|e^*) < 0 \qquad \text{and} \qquad (\rho - 1)q_M \phi(q_M|e^*) < 0.$$

In sum, following a marginal increase in γ an increase in both q_m and q_M would strictly decrease the LHS of the IR, while an increase in q_m and a decrease in q_M would strictly increase the LHS of the IC. Therefore, the only changes in q_m and q_M that leave the LHS of both the IC and IR unchanged overall following an increase in γ involve a decrease in q_m.

Proof of Proposition 3:

At first, we describe the optimal contract when the IC does not bind, which in particular is the case for $e^T = 0$. We then verify when the IC does not bind. In the optimization problem with a nonbinding IC, the IR for $e^* \geq 0$ must be binding. Suppose that it is not. Then, the contract that solves the optimization problem in equations (1), (4), and (5) is simply $w(q) = 0$ for any q, so that, using equation (25) at any effort e^* with $\nu(0, q) \leq 0$ by assumption:

$$\int_0^{\overline{q}} u(0,q)\phi(q|e^*)dq = \int_0^{\overline{q}} \nu(0,q)\phi(q|e^*)dq \leq 0 \quad \Rightarrow \quad \int_0^{\overline{q}} u(0,q)\phi(q|e^*)dq - C(e^*) < \overline{U}$$

i.e. IR is not satisfied, a contradiction.

The relaxed optimization problem with a nonbinding IC, a binding IR, and the FOA, is:

$$\min_{w(q)} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{62}$$

$$\text{s.t.} \quad \int_0^{\overline{q}} u(w(q),q)\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{63}$$

$$\int_0^{\overline{q}} u(w(q),q)\frac{\partial\phi(q|e^*)}{\partial e}dq = C'(e^*) \tag{64}$$

$$0 \leq w(q) \leq q \tag{65}$$

$$w(q) \geq w(q') \; \forall q > q' \tag{66}$$

Consider the subset of contracts that induce a given effort e^*, i.e. such that equation (64) holds (given the FOA). In this subset of contracts, we will show that any contract that satisfies IR as an equality as in equation (63) and the constraints on contracting in equations (65) and (66) is associated with a higher expected cost (as specified in equation (62)) than a contract of the following form:

$$w(q) = \begin{cases} 0 & \text{if } q < q_m \\ w^*(q) & \text{if } q \geq q_m \end{cases} . \tag{67}$$

This in turn implies that the optimal contract is as in equation (67).

The proof is as follows. Given the constraints on contracting in equations (65) and (66), a contract as in equation (67), denoted by $w^O(q)$, and an alternative contract, denoted by $w^A(q)$, can differ in the following subintervals, where some of these subintervals can be empty. First, for outputs such that $w^O(q) = 0$ but $w^A(q) > 0$; denote the union of these subintervals of outputs by $\mathcal{Q}_0$. Second, for outputs such that $w^O(q) = w^*(q)$ but $w^A(q) < w^*(q)$; denote the union of these subintervals of outputs by $\mathcal{Q}_-$. Third, for outputs such that $w^O(q) = w^*(q)$ but $w^A(q) > w^*(q)$; denote the union of these subintervals of outputs by $\mathcal{Q}_+$.

Since contracts $w^O(q)$ and $w^A(q)$ both satisfy equation (63), as already specified, we have:

$$\int_{\mathcal{Q}_0} \left(\nu(w^A(q),q) - \nu(0,q)\right) \phi(q|e^*)dq + \int_{\mathcal{Q}_+} \left(v(w^A(q)) - v(w^*(q))\right) \phi(q|e^*)dq = \int_{\mathcal{Q}_-} \left(\nu(w^*(q),q) - \nu(w^A(q),q)\right) \phi(q|e^*)dq \quad (68)$$

Because of equation (68), we can split the subset $\mathcal{Q}_-$ into two subsets, $\mathcal{Q}_-^1$ and $\mathcal{Q}_-^2$, such that:

$$\begin{cases} \int_{\mathcal{Q}_-^1} \left(\nu(w^*(q),q) - \nu(w^A(q),q)\right) \phi(q|e^*)dq = \int_{\mathcal{Q}_0} \left(\nu(w^A(q),q) - \nu(0,q)\right) \phi(q|e^*)dq \\ \int_{\mathcal{Q}_-^2} \left(\nu(w^*(q),q) - \nu(w^A(q),q)\right) \phi(q|e^*)dq = \int_{\mathcal{Q}_+} \left(v(w^A(q)) - v(w^*(q))\right) \phi(q|e^*)dq \end{cases} . \quad (69)$$

We now use Taylor expansions, properties of the functions ν and v (including $\nu''_{ww}(w,q) > 0$, $v''(w) < 0$, and $\nu''_{wq}(w,q) = 0$ so that $\nu'_w(w,q_1) = \nu'_w(w,q_2)$ for $q_1 \neq q_2$), and $q < q_m$ for any $q \in \mathcal{Q}_0$, to write:

$$\begin{array}{ll} \nu(w^A(q),q) - \nu(0,q) < \nu'_w(w^A(q_m),q_m)w^A(q) & \text{for } q \in \mathcal{Q}_0 \\ \nu(w^*(q),q) - \nu(w^A(q),q) > \nu'_w(w^A(q_m),q_m)\left(w^*(q) - w^A(q)\right) & \text{for } q \in \mathcal{Q}_- \\ v(w^A(q)) - v(w^*(q)) < v'(0)\left(w^A(q) - w^*(q)\right) & \text{for } q \in \mathcal{Q}_+ \end{array} \quad (70)$$

Use the inequalities in equation (70) to get:

$$\int_{\mathcal{Q}_0} \left(\nu(w^A(q),q) - \nu(0,q)\right) \phi(q|e^*)dq < \nu'_w(w^A(q_m),q_m) \int_{\mathcal{Q}_0} w^A(q)\phi(q|e^*)dq \quad (71)$$

$$\int_{\mathcal{Q}_-^1} \left(\nu(w^*(q),q) - \nu(w^A(q),q)\right) \phi(q|e^*)dq > \int_{\mathcal{Q}_-^1} \nu'_w(w^A(q_m),q_m)\left(w^*(q) - w^A(q)\right) \phi(q|e^*)dq \quad (72)$$

$$\int_{\mathcal{Q}_+} \left(v(w^A(q)) - v(w^*(q))\right) \phi(q|e^*)dq < v'(0) \int_{\mathcal{Q}_+} \left(w^A(q) - w^*(q)\right) \phi(q|e^*)dq \quad (73)$$

$$\int_{\mathcal{Q}_-^2} \left(\nu(w^*(q),q) - \nu(w^A(q),q)\right) \phi(q|e^*)dq > \nu'_w(w^A(q_m),q_m) \int_{\mathcal{Q}_-^2} \left(w^*(q) - w^A(q)\right) \phi(q|e^*)dq \quad (74)$$

Combining the equalities in equation (69) and the inequalities in equations (72)-(74), we have:

$$\nu'_w(w^A(q_m),q_m) \int_{\mathcal{Q}_-^1} \left(w^*(q) - w^A(q)\right) \phi(q|e^*)dq < \nu'_w(w^A(q_m),q_m) \int_{\mathcal{Q}_0} w^A(q)\phi(q|e^*)dq \quad (75)$$

$$\nu'_w(w^A(q_m),q_m) \int_{\mathcal{Q}_-^2} \left(w^*(q) - w^A(q)\right) \phi(q|e^*)dq < v'(0) \int_{\mathcal{Q}_+} \left(w^A(q) - w^*(q)\right) \phi(q|e^*)dq \quad (76)$$

Moreover, because of assumption (22) and properties of the function v and ν, we have $\nu'_w(w,q) >$

$v'(w)$ for any $\{w, q\}$. Combining with equations (75) and (76), this gives:

$$\int_{\mathcal{Q}_-^1} \left(w^*(q) - w^A(q)\right) \phi(q|e^*)dq \;<\; \int_{\mathcal{Q}_0} \left(w^A(q) - 0\right) \phi(q|e^*)dq \tag{77}$$

$$\int_{\mathcal{Q}_-^2} \left(w^*(q) - w^A(q)\right) \phi(q|e^*)dq \;<\; \int_{\mathcal{Q}_+} \left(w^A(q) - w^*(q)\right) \phi(q|e^*)dq \tag{78}$$

By definition of the subintervals and of the contracts w^O and w^A, combining inequalities in equations (77) and (78) implies:

$$\int_0^{\overline{q}} w^O(q)\phi(q|e^*)dq < \int_0^{\overline{q}} w^A(q)\phi(q|e^*)dq. \tag{79}$$

This completes this part of the proof.

The next part of the proof considers when effort e^* is strictly positive even when the IC does not bind ($e^* < e^T$) and for when the IC does not bind. With a contract as in equation (67), the FOC to the effort choice problem can be rewritten as:

$$\int_0^{q_m} \nu(0, q)\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w^*(q))\frac{\partial}{\partial e}\phi(q|e^*)dq = C'(e^*) \tag{80}$$

With $q_m = 0$, we have $u(w^*(q), q) = v(w^*(q))$ by definition of the fair wage and the utility function, and $w^{*\prime}(q) > 0$ and $v' > 0$. Combining this with MLRP shows that the LHS of equation (80) is strictly positive. Finally, by assumption, $C'(0) = 0$ and $C'(e) > 0$ for any positive e, so that the equilibrium effort e^* is strictly positive when $q_m = 0$.

The derivative of the LHS of equation (80) with respect to q_m holding effort constant at $e = e^*$ is:

$$\nu(0, q_m)\frac{\partial}{\partial e}\phi(q_m|e^*) - v(w^*(q_m))\frac{\partial}{\partial e}\phi(q_m|e^*). \tag{81}$$

By definition we have $v(w^*(q_m)) = \nu(w^*(q_m), q_m)$, so that $\nu(0, q_m) - v(w^*(q_m)) < 0$ since $\nu(w, q)$ is strictly increasing in w. Therefore, the expression in equation (81) is strictly positive if and only if $\frac{\partial}{\partial e}\phi(q_m|e^*)$, which by MLRP and definition of $q_0^{e^*}$ is equivalent to $q_m < q_0^{e^*}$. In sum, there is $\hat{q}_m$ such that $e^* > 0$ if $q_m \in [0, \hat{q}_m)$.

We now determine for which parameter values this is the case. Consider the effect of a change in q_m on the LHS of the IR in equation (63) when $e^* > 0$ and is therefore given by the first-order condition (FOC) to the agent's effort choice problem given the FOA. The derivative of the LHS of equation (63) with respect to q_m holding effort constant is:

$$\nu(0, q_m)\phi(q_m|e^*) - v(w^*(q_m))\phi(q_m|e^*). \tag{82}$$

The expression in equation (82) is strictly negative for any e (see the preceding paragraph). More-

over, the LHS of the IR in equation (63) is the agent's objective function, and the agent chooses effort e to maximize this objective function. From the envelope theorem, we know that the total effect of a marginal change in a parameter (here q_m) on the objective function is equal to its effect holding effort constant. In sum, the LHS of equation (63) is continuously decreasing in q_m, and given optimal effort choices it is below the RHS for $q_m = \overline{q}$ according to equation (25), and it is above the RHS for $q_m = 0$ according to equation (23). This implies that there exists q_m that satisfies equation (35), and this level of q_m is strictly decreasing in $\overline{U}$ according to equation (63). In sum, when $\overline{U}$ is sufficiently high, we have $q_m < \hat{q}_m$, so that $e^* > 0$ according to the preceding paragraph.

With the contract in equation (67), where q_m is set to solve the IR in equation (63) as an equality, when the induced e^* is greater than e^T, the IC does not bind. On the contrary, when the induced e^* is strictly less than e^T, the IC binds, and the principal's problem is solved in the next part of the proof below.

In this second part of the proof, we describe the optimal contract when the IC binds. By Lemma 1, when the IC in equation (2) is binding, it can be replaced by the FOC:

$$\int_0^{\overline{q}} u(w(q), q) \frac{\partial}{\partial e} \phi(q|e^*) dq = C'(e^*). \tag{83}$$

This part of the proof has two steps.

Lemma 4 *On any non-empty subinterval of* $[0, \overline{q}]$*, we have* $w(q) \notin (0, w^*(q))$.

Proof. An optimal contract is nondecreasing in q according to the agent's monotonicity constraint in equation (5). Accordingly, consider any given initial contract that satisfies the constraints in the optimization program with $w(q) \geq w(q')$ for two given outputs $q > q'$, $w(q) \in (0, w^*(q))$, $w(q') \in (0, w^*(q'))$, and the following perturbation: increase $w(q)$ by $\epsilon/(\nu'(w(q), q)\phi(q|e^*))$, and decrease $w(q')$ by $\epsilon/(\nu'(w(q'), q')\phi(q'|e^*))$, where ϵ is positive and arbitrarily small. The change in the LHS of the IR in (3) holding effort constant is:

$$\frac{\epsilon}{(\nu'(w(q), q)\phi(q|e^*))} \nu'(w(q), q)\phi(q|e^*) - \frac{\epsilon}{(\nu'(w(q'), q')\phi(q'|e^*))} \nu'(w(q'), q')\phi(q'|e^*) = 0.$$

Consider the effect of this perturbation on the cost of the contract holding effort constant:

$$\frac{\epsilon}{\nu'(w(q), q)\phi(q|e^*)} \phi(q|e^*) - \frac{\epsilon}{\nu'(w(q'), q')\phi(q'|e^*)} \phi(q'|e^*) = \epsilon \left(\frac{1}{\nu'(w(q), q)} - \frac{1}{\nu'(w(q'), q')} \right) \tag{84}$$

With $\nu''_{ww} \geq 0$, $\nu''_{wq} = 0$, and $w(q) \geq w(q')$, we have $\nu'(w(q), q) \geq \nu'(w(q'), q')$, so that the expression in equation (84) is negative. Consider the effect on the LHS of the IC in (83) with a utility function

as in equation (21). The change in the LHS of the IC for this perturbation is:

$$\frac{\epsilon}{\nu'(w(q),q)\phi(q|e^*)}\nu'(w(q),q)\frac{\partial\phi(q|e^*)}{\partial e} - \frac{\epsilon}{\nu'(w(q'),q')\phi(q'|e^*)}\nu'(w(q'),q')\frac{\partial\phi(q'|e^*)}{\partial e}$$
$$= \epsilon\left(LR(q|e^*) - LR(q'|e^*)\right) \tag{85}$$

By MLRP we have $LR(q|e^*) > LR(q'|e^*)$. Thus, the LHS of the IC increases, which relaxes a binding constraint. Given that the agent's problem is concave in effort given assumptions on the FOA, the agent will optimally exert higher effort than with the initial contract. The IR remains satisfied (if it is satisfied at effort e_1 but the agent is better off under effort e_2, then it is satisfied at effort e_2 as well). This contradicts the optimality of the initial contract. ■

Combining the agent's monotonicity constraint ($w(q)$ is nondecreasing in q) in equation (5) and Lemma 4, we have $w(q) = 0$ for $q \in [0, q_m]$, for some $q_m \in [0, \overline{q}]$. For now, consider a given $q_m \in [0, \overline{q}]$. Using Lemma 4, the agent's monotonicity constraint and principal limited liability, there is q_m such that, for $q > q_m$, the payment is $w(q) \in [w^*(q), q]$. This implies that $u(w, q) = v(w)$ for $q > q_m$. That is, for a given q_m, the relaxed optimization problem that gives the optimal contract to induce effort $e^* = e^T$ can be rewritten as:

$$\min_{w(q)} \int_{q_m}^{\overline{q}} w(q)\phi(q|e^*)dq \tag{86}$$

$$\text{s.t.} \quad \int_0^{q_m} u(0,q)\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w(q))\frac{\partial}{\partial e}\phi(q|e^*)dq = C'(e^*) \tag{87}$$

$$\int_0^{q_m} u(0,q)\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w(q))\phi(q|e^*)dq \geq \overline{U} \tag{88}$$

$$w(q) \in [w^*(q), q] \ \forall q \tag{89}$$

We henceforth consider the subset of values of q_m such that the optimization problem in equations (86)-(89) has a solution (the optimization problem has a solution for some q_m because of equations (23) and (24)). Using the notation in Jewitt, Kadan, and Swinkels (2008), we have $\underline{m}(q) = w^*(q)$ and $\overline{m}(q) = q$. We can apply Proposition 1 in their paper to derive the optimal contract on $[q_m, \overline{q}]$ given that the payment $w(q)$ is 0 on $[0, q_m)$ (note that the first terms on the LHS of equations (87) and (88) are independent of $w(q)$ and can therefore be treated as constants in the optimization problem in equations (86)-(89)). In sum, the optimal contract is defined implicitly by:

$$\frac{1}{u'_w(w(q),q)} = \begin{cases} \frac{1}{u'_w(0,q)} & \text{for } q \leq q_m \\ \frac{1}{v'(w^*(q))} & \text{for } q > q_m \text{ and } \lambda_{IR} + \lambda_{IC}LR(q|e^*) < \frac{1}{v'(w^*(q))} \\ \lambda_{IR} + \lambda_{IC}LR(q|e^*) & \text{for } q > q_m \text{ and } \frac{1}{v'(w^*(q))} < \lambda_{IR} + \lambda_{IC}LR(q|e^*) < \frac{1}{v'(q)} \\ \frac{1}{v'(q)} & \text{for } q > q_m \text{ and } \frac{1}{v'(q)} < \lambda_{IR} + \lambda_{IC}LR(q|e^*) \end{cases}$$

with $\lambda_{IR} \geq 0$ and $\lambda_{IC} > 0$, which are the Lagrange multipliers associated respectively with the

constraints (88) and (87), and which therefore depend on q_m (in general, these are not the Lagrange multipliers associated with the IR and IC of the original optimization problem). Equivalently:

$$w(q) = \begin{cases} 0 & \text{for } q \leq q_m \\ w^*(q) & \text{for } q > q_m \text{ and } \lambda_{IR} + \lambda_{IC} LR(q|e^*) < \frac{1}{v'(w^*(q))} \\ v'^{-1}\left(1/\left(\lambda_{IR} + \lambda_{IC} LR(q|e^*)\right)\right) & \text{for } q > q_m \text{ and } \frac{1}{v'(w^*(q))} < \lambda_{IR} + \lambda_{IC} LR(q|e^*) < \frac{1}{v'(q)} \\ q & \text{for } q > q_m \text{ and } \frac{1}{v'(q)} < \lambda_{IR} + \lambda_{IC} LR(q|e^*) \end{cases}.$$

Online Appendix
A Theory of Fair CEO Pay

Lemma 5 *When IC is nonbinding and equation (23) holds as an equality, the optimal contract is* $w(q) = w^*(q)\ \forall q$.

Proof. Assume that IC is nonbinding, equation (23) holds as an equality, and let e^F be the induced effort when the agent is paid the fair wage for any output:

$$\int_0^{\overline{q}} v(w^*(q)) \frac{\partial}{\partial e} \phi(q|e^F) dq = C'(e^F) \tag{90}$$

We now show by contradiction that the contract $w^*(q)$ is the least costly way of satisfying the IR constraint in equation (63). If another contract $\hat{w}(q)$ induces the same effort and satisfies the IR at a smaller cost than $w^*(q)$, then we must have $\hat{w}(q) < w^*(q)$ for some $q \in \mathcal{Q}_-$ (otherwise this contract would not be associated with a lower cost), and $\hat{w}(q) > w^*(q)$ for some $q \in \mathcal{Q}_+$ given equation (23) satisfied as an equality (otherwise this contract would not provide expected utility $\overline{U}$). Now consider the agent's expected utility under contract $\hat{w}(q)$:

$$\begin{aligned} \int_0^{\overline{q}} u(\hat{w}(q), q) \phi(q|e^F) dq &= \int_{[0,\overline{q}]\setminus(\mathcal{Q}_- \cup \mathcal{Q}_+)} u(w^*(q), q) \phi(q|e^F) dq \\ &+ \int_{\mathcal{Q}_-} u(\hat{w}(q), q) \phi(q|e^F) dq + \int_{\mathcal{Q}_+} u(\hat{w}(q), q) \phi(q|e^F) dq \geq \overline{U}, \end{aligned} \tag{91}$$

where the inequality is because the contract $\hat{w}(q)$ must satisfy the IR.

For $q \in \mathcal{Q}_-$, by the intermediate value theorem we know that for any given q there exists $\check{w}(q) \in (\hat{w}(q), w^*(q))$ such that:

$$\begin{aligned} u(\hat{w}(q), q) &= u(w^*(q), q) + u'(\check{w}(q), q) \left(\hat{w}(q) - w^*(q)\right) \\ &= u(w^*(q), q) + \nu'(\check{w}(q), q) \left(\hat{w}(q) - w^*(q)\right) \end{aligned} \tag{92}$$

Integrating:

$$\int_{\mathcal{Q}_-} u(\hat{w}(q), q) \phi(q|e^F) dq = \int_{\mathcal{Q}_-} \left(u(w^*(q), q) + \nu'(\check{w}(q), q) \left(\hat{w}(q) - w^*(q)\right)\right) \phi(q|e^F) dq$$

Let $\nu'_{\mathcal{Q}_-}$ be implicitly defined by:

$$\int_{\mathcal{Q}_-} u(\hat{w}(q), q) \phi(q|e^F) dq = \int_{\mathcal{Q}_-} u(w^*(q), q) \phi(q|e^F) dq + \nu'_{\mathcal{Q}_-} \int_{\mathcal{Q}_-} \left(\hat{w}(q) - w^*(q)\right) \phi(q|e^F) dq, \tag{93}$$

where $\nu'_{\mathcal{Q}_-} \geq \nu'^*)(q), q)$ since $\nu'' \geq 0$.

Likewise, for $q \in \mathcal{Q}_+$, by the intermediate value theorem we know that for any given q there exists $\check{w}(q) \in (w^*(q), \hat{w}(q))$ such that:

$$\begin{aligned} u(\hat{w}(q), q) &= u(w^*(q), q) - u'(\check{w}(q), q)\left(\hat{w}(q) - w^*(q)\right) \\ &= u(w^*(q), q) - v'(\check{w}(q), q)\left(\hat{w}(q) - w^*(q)\right) \end{aligned} \tag{94}$$

Integrating:

$$\int_{\mathcal{Q}_+} u(\hat{w}(q), q)\phi(q|e^F)dq = \int_{\mathcal{Q}_+} \left(u(w^*(q), q) + v'(\check{w}(q), q)\left(\hat{w}(q) - w^*(q)\right)\right)\phi(q|e^F)dq$$

Let $v'_{\mathcal{Q}_+}$ be implicitly defined by:

$$\int_{\mathcal{Q}_+} u(\hat{w}(q), q)\phi(q|e^F)dq = \int_{\mathcal{Q}_+} u(w^*(q), q)\phi(q|e^F)dq + v'_{\mathcal{Q}_+}\int_{\mathcal{Q}_+}\left(\hat{w}(q) - w^*(q)\right)\phi(q|e^F)dq, \tag{95}$$

where $v'_{\mathcal{Q}_+} \leq v'^*(q), q)$ since $v'' \leq 0$.

In sum, substituting in the RHS of equation (91):

$$\begin{aligned} &\int_{[0,\overline{q}]\setminus(\mathcal{Q}_-\cup\mathcal{Q}_+)} u(w^*(q), q)\phi(q|e^F)dq + \int_{\mathcal{Q}_-} u(\hat{w}(q), q)\phi(q|e^F)dq + \int_{\mathcal{Q}_+} u(\hat{w}(q), q)\phi(q|e^F)dq \\ &= \int_{[0,\overline{q}]\setminus(\mathcal{Q}_-\cup\mathcal{Q}_+)} u(w^*(q), q)\phi(q|e^F)dq \\ &+ \int_{\mathcal{Q}_-} u(w^*(q), q)\phi(q|e^F)dq + \nu'_{\mathcal{Q}_-}\int_{\mathcal{Q}_-}\left(\hat{w}(q) - w^*(q)\right)\phi(q|e^F)dq \\ &+ \int_{\mathcal{Q}_+} u(w^*(q), q)\phi(q|e^F)dq + v'_{\mathcal{Q}_+}\int_{\mathcal{Q}_+}\left(\hat{w}(q) - w^*(q)\right)\phi(q|e^F)dq \\ &= \int_0^{\overline{q}} u(w^*(q), q)\phi(q|e^F)dq + \nu'_{\mathcal{Q}_-}\int_{\mathcal{Q}_-}\left(\hat{w}(q) - w^*(q)\right)\phi(q|e^F)dq + v'_{\mathcal{Q}_+}\int_{\mathcal{Q}_+}\left(\hat{w}(q) - w^*(q)\right)\phi(q|e^F)dq, \end{aligned} \tag{96}$$

where $v'_{\mathcal{Q}_+} < \nu'_{\mathcal{Q}_+}$ due to the assumptions on the first derivatives with respect to w of the function v and ν. Since the contract $\hat{w}(q)$ must provide expected utility of at least $\overline{U}$ to satisfy IR, equations (23) satisfied as an equality and (96) imply:

$$\nu'_{\mathcal{Q}_-}\int_{\mathcal{Q}_-}\underbrace{\left(\hat{w}(q) - w^*(q)\right)}_{<0}\phi(q|e^F)dq + v'_{\mathcal{Q}_+}\int_{\mathcal{Q}_+}\underbrace{\left(\hat{w}(q) - w^*(q)\right)}_{>0}\phi(q|e^F)dq \geq 0 \tag{97}$$

The inequalities under braces in equation (97) combined with $v'_{\mathcal{Q}_+} < \nu'_{\mathcal{Q}_+}$ give:

$$\int_{\mathcal{Q}_+} \underbrace{(\hat{w}(q) - w^*(q))}_{>0} \phi(q|e^F)dq \geq \int_{\mathcal{Q}_-} \underbrace{(w^*(q) - \hat{w}(q))}_{>0} \phi(q|e^F)dq$$

$$\Leftrightarrow \quad \int_{\mathcal{Q}_- \cup \mathcal{Q}_+} \hat{w}(q)\phi(q|e^F)dq \geq \int_{\mathcal{Q}_- \cup \mathcal{Q}_+} w^*(q)\phi(q|e^F)dq \qquad (98)$$

By definition of the subsets $\mathcal{Q}_-$ and $\mathcal{Q}_+$, this implies that the expected cost of the contract $\hat{w}(q)$ is higher than the expected cost of the contract $w^*(q)$. This concludes the proof of Lemma 5. ■