How should performance signals affect contracts?

Article:

Chaigneau, Pierre, Edmans, Alex and Gottlieb, Daniel ORCID: 0000-0002-0555-6185 (2022) How should performance signals affect contracts? Review of Financial Studies, 35 (1). 168 - 206. ISSN 0893-9454

https://doi.org/10.1093/rfs/hhab026

How Should Performance Signals Affect Contracts?*

Pierre Chaigneau — Queen's University
Alex Edmans — LBS, CEPR, and ECGI
Daniel Gottlieb — LSE

Review of Financial Studies, forthcoming

Abstract

The informativeness principle demonstrates that a contract should depend on informative signals. This paper studies *how* it should do so. Signals that indicate the output distribution has shifted to the left (e.g. weak industry performance) reduce the threshold for the manager to be paid; those that indicate output is a precise measure of effort (e.g. low volatility) decrease high thresholds and increase low thresholds. Surprisingly, "good" signals of performance need not reduce the threshold. Applying our model to performance-based vesting, we show that performance measures should affect the strike price rather than the number of vesting options, contrary to practice.

Keywords: informativeness principle, limited liability, option repricing, pay-for-luck, performance-based vesting, performance-sensitive debt

JEL Classification: D86, G32, G34, J33.

*pierre.chaigneau@queensu.ca, aedmans@london.edu, d.gottlieb@lse.ac.uk. We thank three anonymous referees, the Editor (Holger Mueller), Sabrina Buti, Alan Douglas, James Dow, Sergei Glebkin, Barney Hartman-Glaser, Steven Matthews, Sebastian Pfeil, Nicolas Sahuguet, Ming Yang, and conference participants at the AFA, Econometric Society World Congress, HEC-McGill Spring Finance Workshop, NFA, Rotterdam Executive Compensation Conference, SFS Cavalcade, Université Paris Dauphine, and the University of Waterloo for valuable comments. Gottlieb thanks the Dorinda and Mark Winkelman Distinguished Scholar Award for research support.

Shareholders would like managers to maximize the equity value of the firm. Therefore, it seems that executive compensation contracts should pay the manager only according to the stock price. However, real-life contracts are typically based on additional signals of performance. For example, Bettis et al. (2018) find that 70% of large U.S. firms pay their executives with performance-vesting equity, where the number of securities granted depends on performance relative to a threshold. Out of these firms, 86% use at least one accounting threshold, and so the CEO's pay depends on more than just the stock price. Similarly, Murphy's (2013) survey reports that companies use a variety of financial and non-financial performance measures when determining CEO bonuses.

The main theoretical justification for including additional performance measures is Holmström's (1979) informativeness principle. This principle states that any signal should be included in a contract if it provides incremental information about the agent's effort, over and above that already conveyed in output. Incorporating such a signal allows the principal to provide the same incentives at lower cost. While the principle has been highly influential, it does not show *how* informative signals should be used. Two questions are critical for principals to be able to incorporate signals into real-life contracts. First, *which* dimensions of the contract should a signal affect? For example, with performance-vesting options, a signal could affect either the strike price, the number of vesting options, or both. In practice, however, only the number of options and not the strike price depend on performance. Is this practice optimal? Second, *which way* should performance affect the relevant dimensions? In practice, "good" performance measures that indicate high effort always increase the number of vesting options. While intuitive, is this practice actually optimal?

This paper addresses these open questions in a general optimal contracting model with risk aversion and continuous effort. We study how a contract based on output (such as the stock price) should be modified when the principal also has access to an additional signal (such as accounting performance). In order to have practical implications for real-life contracts, we extend the original Holmström (1979) model to incorporate limited liability, which applies to almost all compensation contracts – the payment made to a manager cannot be negative. Explicitly modelling this restriction is needed, otherwise the model might indicate that a signal should affect the contract for outputs where limited liability binds, and so the contract cannot actually respond to the signal.

In the most general model, where the only assumption on the output distribution is the monotone likelihood ratio property, the optimal contract involves a threshold output level. The manager receives zero for outputs below the threshold and a strictly positive amount above it. This amount is increasing in output but will be typically nonlinear. Indeed, many real-life contracts involve a threshold, such as performance-vesting stock or options, bonuses, or dismissal contracts where the agent is fired if output falls below a certain level. This is a generalization of Innes (1990) who showed that, with risk neutrality, limited liability, and a monotonicity constraint, the optimal contract pays zero below a threshold and increases linearly above it.

We use our model to understand how the signal realization optimally affects the threshold. It may seem intuitive that signals that indicate managerial effort – such as high accounting performance – should reduce the threshold and thus increase the payment (the "individual informativeness effect"). However, this is not necessarily the case because the signal also has a second effect. Not only may a signal itself indicate high or low effort, but it can also affect the information the principal infers about effort from the output level (the "output inference effect").

To analyze the output inference effect in more detail, we assume that the output distribution has both a location parameter (such as the mean of the distribution) and also a scale parameter (such as the volatility). Many standard output distributions have both parameters, such as the normal, skew-normal, logistic, Cauchy, and uniform distributions. Doing so allows us to model the signal realization as affecting these parameters, and thus study how the threshold varies with these parameters. We can now decompose the output inference effect into two components. The first is the "location effect" – the signal indicates that the location of the output distribution has shifted, i.e. that the entire distribution has moved to the left or to the right. For example, consider the contract for the manager of an industry incumbent, and let the signal be the number of competitors in that industry. A low number of competitors is individually a good signal of effort, because it indicates that the incumbent has developed good products that have driven out competition or deterred entry. However, this consideration may be outweighed by the location effect – more competitors cause a leftward shift in the distribution of the firm's output, and so a given output is a more positive signal of effort.

The second component of output inference is the "precision effect". The signal also indicates how precise output is as a measure of effort. This precision, in turn, depends on two parameters – the aforementioned scale parameter, which captures the distribution's volatility, and the impact parameter, which captures the extent to which effort increases the distribution's location or mean. For example, a signal that indicates low volatility, or that effort has a high impact on output, suggests high precision.

Unlike the location effect, where the directional impact on the threshold is unambiguous (signals that indicate a leftward shift in the distribution decrease the threshold), signals that indicate high precision may increase or decrease it. The direction of the effect depends on how high the “original” threshold was. If it is high, the manager is paid only if output is sufficiently high to be good news about effort. If output is a precise measure of effort, even moderately high output is a sufficiently strong indicator of effort to justify the manager being paid – the threshold falls. In contrast, if the original output threshold is low, the agent is paid unless output is sufficiently low to be bad news about effort. If output is a precise measure of effort, even moderately low output is a sufficiently weak indicator of effort to justify the manager not being paid – the threshold rises.

Overall, the precision effect leads to less extreme thresholds: low thresholds rise and high thresholds fall. The output threshold for a given signal realization in turn depends on the exogenous parameters of the model. Moreover, contrary to intuition, we also show that the impact and scale parameters do not always have the opposite effect on the threshold.

Having studied how the signal affects the threshold, we then study how it affects the slope of the contract above the threshold. So that we can define the contract’s slope in an unambiguous manner, we specialize the model to the case in which the optimal contract is linear above the threshold. Doing so allows our model to have implications for performance-vesting options where the number of options ultimately received by the manager depends on performance. The contract is now an option; the threshold is the strike price, and the (linear) slope is the number of vesting options. Despite its popularity, we are unaware of any theories that study under what conditions performance-based vesting is optimal, and what performance signals should be used. We start by deriving the first set of sufficient conditions for options to be the optimal contract when the agent is risk-averse – log utility, normally-distributed output, limited liability on the manager, and a sufficiently convex cost of effort.

We show that the effect of a signal on the number of vesting options depends only on the precision effect, and not the individual informativeness or location effects. This is because the number of vesting options determines the slope of the contract. In turn, differences in slopes across signal realizations depend only on how precise output is as an indicator of effort. This result suggests that the common practice of making vesting depend on signals such as accounting and stock price performance (either for the firm in question or peer firms) may be suboptimal. Such signals are informative about either the manager’s effort and/or the location of the output distribution but are unlikely to affect output precision. Thus, they should optimally affect the overall level of pay, not the sensitivity of pay to performance, and change the strike price rather than the number of vesting options.

In practice, option strike prices are sometimes reset, although on a discretionary basis rather than the resetting being specified in the contract. Such resetting typically involves a lowering of the strike price, and follows poor stock price performance (Brenner, Sundaram, and Yermack (2000)). Acharya, John, and Sundaram (2000) theoretically justify such practices based on the need to restore *future* effort incentives when options fall out of the money. Our paper shows that contractual repricing of stock options can be optimal to reward or punish *past* effort. Unlike the number of vesting options, which depends only on the precision effect, the strike price depends also on the individual informativeness and location effects. The location effect means that it may sometimes be optimal to lower the strike price upon a signal that individually conveys bad news about effort (such as a high number of competitors), contrary to conventional wisdom that such repricing necessarily results from rent extraction.

This paper is related to the theoretical literature on pay-for-performance, surveyed by Holmström (2017). A number of papers extend the original Holmström (1979) informativeness principle and thus study *whether*, but not how, signals should be incorporated into optimal contracts. Examples include Townsend (1979), Gjesdal (1982), Gale and Hellwig (1985), Amershi and Hughes (1989), Allen and Gale (1992), Kim (1995), and Chaigneau, Edmans, and Gottlieb (2019). Similarly, the applied literature on pay-for-performance studies questions such as whether agents should be paid for luck – i.e. *whether* signals of peer performance should affect the contract – such as Oyer (2004), Axelson and Baliga (2009), Gopalan, Milbourn, and Song (2010), Hoffmann and Pfeil (2010), and Hartman-Glaser and Hébert (2020)[1] – rather than *how* signals in general (not just signals of peer performance) should affect contracts.

More closely related is Marinovic and Varas (2019), who show that, if the manager can manipulate output, performance-based vesting is always optimal to deter such manipulation. We show that, even if the manager is unable to manipulate output, performance-based vesting is optimal if the signal affects the precision of output as an effort measure, but not if the signal is informative only along other dimensions. Many common performance signals should affect the strike price rather than the number of vesting options. Also related is Chaigneau, Edmans, and Gottlieb (2018), who study how volatility (although not individual informativeness, location, or impact) affects the optimal contract. In their setting, there are no additional signals and the contract is written after volatility is realized, so volatility affects the contract that is offered. In contrast, we study how incentives are allocated across signal *realizations* – i.e. states of nature – that are associated with different volatility. The contract is written before volatility is realized and the contract is contingent on the signal realization, as is the case for performance-vesting equity.

[1]Dittmann, Maug, and Spalt (2013) calibrate the cost savings from incorporating peer performance in executive contracts, and Johnson and Tian (2000) compare the incentives provided by indexed and non-indexed options.

1 The Model

1.1 Setup

There are two parties, a principal (firm), and an agent (manager). The manager exerts an unobservable effort level $e \in [0, \bar{e}]$ and is protected by limited liability. Effort entails any action that improves output but is costly to the manager, such as working rather than shirking, choosing projects that generate cash flows rather than private benefits, or not extracting rents. The manager's cost of effort $C(\cdot)$ is strictly increasing, strictly convex, twice continuously differentiable in $[0, \bar{e})$, with $C'(0) = 0$ and $\lim_{e \nearrow \bar{e}} C'(e) = +\infty$. His utility over money $u(\cdot)$ is strictly increasing, weakly concave, and twice differentiable. The manager has outside wealth $\bar{W} > 0$ and reservation utility $\overline{u}$.

Effort affects the probability distribution of output q and a signal s, which are both observable and contractible. Output is continuously distributed with full support on $(\underline{q}, +\infty)$, where $\underline{q}$ is either $-\infty$ or 0. To ensure that an optimal contract exists, we assume that the signal is discrete, $s \in \{s_1, ..., s_S\}$. Note that the signal can have one or multiple dimensions.

The signal is distributed according to the probability mass function $\phi_e^s := \Pr(\tilde{s} = s | \tilde{e} = e)$, which is strictly positive and twice continuously differentiable in e. Output is distributed according to the cumulative distribution function $F(q|e, s)$, which is twice continuously differentiable in q and e and has a strictly positive density $f(q|e, s)$. The joint distribution of output and the signal is $f(q, s|e) := \phi_e^s f(q|e, s)$. The likelihood ratio is defined as:

$$LR_s(q|e) := \frac{\frac{\partial f}{\partial e}(q, s|e)}{f(q, s|e)} = \frac{\partial \phi_e^s / \partial e}{\phi_e^s} + \frac{\frac{\partial f}{\partial e}(q|e, s)}{f(q|e, s)}, \tag{1}$$

which we assume to be strictly increasing in output q ("MLRP"). We call $\frac{\partial \phi_e^s / \partial e}{\phi_e^s}$ the likelihood ratio of the signal, and $\frac{\frac{\partial f}{\partial e}(q|e,s)}{f(q|e,s)}$ the likelihood ratio of output (conditional on the signal). For simplicity, for all s we assume that $\lim_{q \nearrow +\infty} LR_s(q|e) = \infty$, and $\lim_{q \searrow -\infty} LR_s(q|e) = -\infty$ when the support is unbounded below.[2]

The firm has full bargaining power and offers the manager a schedule of payments $\{w_s(q)\}$ conditional on each realization of (q, s). It maximizes expected output q minus the expected payment. We follow Grossman and Hart (1983) and separate the principal's problem into two stages. The first stage determines the cost of implementing each effort. Given this cost, the second stage determines which effort to implement.

[2]This assumption simplifies expressions by ruling out corner solutions, but is not important for any of our results.

The optimal schedule of payments $\{w_s(q)\}$ that implements effort $\hat{e}$ solves:

$$\min_{\{w_s(q)\}} \sum_s \phi_{\hat{e}}^s \int_{\underline{q}}^{+\infty} w_s(q)\, f(q|\hat{e},s)dq \tag{2}$$

$$\text{subject to} \qquad \sum_s \phi_{\hat{e}}^s \int_{\underline{q}}^{+\infty} u\left(\bar{W} + w_s(q)\right) f(q|\hat{e},s)dq - C(\hat{e}) \geq \overline{u}, \tag{3}$$

$$\hat{e} \in \arg\max_e \sum_s \phi_e^s \int_{\underline{q}}^{+\infty} u\left(\bar{W} + w_s(q)\right) f(q|e,s)dq - C(e), \tag{4}$$

$$w_s(q) \geq 0 \ \ \forall q, s. \tag{5}$$

The firm minimizes the expected payment (2) subject to the manager's individual rationality ("IR") (3), incentive compatibility ("IC") (4), and limited liability constraints ("LL") (5). We say that the manager has "worked" if he exerts effort $\hat{e}$. Otherwise, we say that he has "shirked".

1.2 The Optimal Contract

The first-order approach ("FOA") simplifies the analysis of moral hazard models with a continuum of efforts by allowing a continuum of incentive constraints (equation (4)) to be replaced by the local incentive constraint that prevents local deviations.[3] Appendix B explains how standard conditions used to justify the FOA with commonly-used output distributions cannot be used in models with limited liability, and derives a new sufficient condition for the validity of the FOA under limited liability. Henceforth, we assume that this condition holds; we also consider contracts that implement $\hat{e} > 0$ (and thus the IC binds); if the principal wishes to implement $\hat{e} = 0$, she trivially offers a flat wage. Let $\lambda \geq 0$ and $\mu > 0$ denote the Lagrange multipliers associated with the IR (3) and IC (4), respectively. The optimal contract is given by Lemma 1 below.

Lemma 1 *(Optimal contract): Suppose the FOA is valid. The optimal contract to implement* $\hat{e} > 0$ *satisfies:*

$$w_s(q) = \begin{cases} u'^{-1}\left(1\Big/\left(\lambda + \mu\left[\frac{\partial \phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)\right) - \bar{W} & \text{if } \lambda + \mu\left[\frac{\partial \phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right] \geq \frac{1}{u'(\bar{W})} \\ 0 & \text{if } \lambda + \mu\left[\frac{\partial \phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right] < \frac{1}{u'(\bar{W})} \end{cases}. \tag{6}$$

[3]See Chaigneau, Edmans, and Gottlieb (2019) for the informativeness principle without the FOA.

Note that the IR may not bind in the optimal contract, in which case we have $\lambda = 0$. Due to MLRP, the payment is increasing in the likelihood ratio, but cannot fall below zero due to limited liability. Thus, the manager is only paid when the likelihood ratio exceeds a cutoff. For each signal realization s, this cutoff for the likelihood ratio is associated with a threshold output q_s^*.[4] If output exceeds the threshold, it is sufficiently likely that the manager has worked that the firm pays him a strictly positive amount. Again due to MLRP, the payment is monotonically increasing in output, but it will typically be nonlinear.

Since the signal realization cannot affect the contract below the threshold, it can affect the contract in two ways: it can affect the threshold, and it can affect the slope of the contract above the threshold. Section 2 now analyzes these two channels in turn.

2 How Performance Signals Affect Contracts

2.1 Performance Signals and the Threshold

This section studies how the signal realization affects the threshold. Recall from equation (1) that the likelihood ratio can be decomposed into two components: the likelihood ratio of the signal ($\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s}$) and the likelihood ratio of output ($\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$). A signal can therefore affect the likelihood ratio, and thus the threshold, in two ways. First, it can be individually informative about effort (through the likelihood ratio of the signal) – the "individual informativeness effect". Second, it can affect the information the principal infers about effort from output (through the likelihood ratio of output) – the "output inference effect". For example, even if effort does not affect economic conditions, and so economic conditions are uninformative about effort, economic conditions will affect the likelihood ratio if they affect the information output provides about effort – e.g., if a given output level is more indicative of effort in recessions than booms.

To further analyze these effects, we now parametrize the output distribution. This allows us to model the signal realization as affecting the distribution's parameters, and thus study how the threshold varies with these parameters. Specifically, we consider output distributions in the location-scale family. Such distributions have a scale parameter σ_s which can be interpreted as the distribution's volatility, and a location parameter h_s which gives the "location" of the distribution – for example, for the normal distribution, the location parameter is its mean. We assume that the location parameter depends on effort, where $h_s'(e) > 0$ for all e (higher effort shifts the distribution rightward). Thus, there exists a function $g(\cdot)$ such that we can rewrite

[4] This threshold may be zero if output has support on $[0, \infty)$.

the density as:

$$f(q|e,s) \equiv \frac{1}{\sigma_s} g\left(\frac{q-h_s(e)}{\sigma_s}\right). \tag{7}$$

Without loss of generality, let $h_s(e) = \xi_s + \zeta_s \Upsilon(e)$ and normalize $\Upsilon(\hat{e}) = 0$ and $\Upsilon'(\hat{e}) = 1$, so that $h_s(\hat{e}) = \xi_s$ and $h_s'(\hat{e}) = \zeta_s > 0$. We refer to ξ_s as the equilibrium location parameter and ζ_s as the impact parameter; the latter captures the effect of effort on output. Using equation (7), the likelihood ratio of output can be rewritten:

$$\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)} = -\frac{\zeta_s}{\sigma_s} \frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)}. \tag{8}$$

We assume that the likelihood ratio of output in equation (8) is continuously differentiable in q, σ_s, ξ_s, and ζ_s. An output distribution with location and scale parameters that satisfies MLRP has a single peak (see the proof of Proposition 1), which we denote by q_s^P. As can be seen in equation (8), the likelihood ratio of output is zero at the peak.[5]

Proposition 1 studies how the signal realization affects the threshold q_s^* above which the manager gets paid. It holds "all else equal across signals" – we are comparing the threshold under two different signal realizations s_i and s_j that differ along only one dimension (e.g. the impact parameter ζ_s); all other dimensions are constant. Note that we are not undertaking comparative statics (e.g. changing ζ_s across all signals) that would change the contracting environment.

Proposition 1 *(Effect of signal on threshold): All else equal across signals:*

(i) If $\frac{\partial \phi_{\hat{e}}^{s_i}/\partial e}{\phi_{\hat{e}}^{s_i}} > \frac{\partial \phi_{\hat{e}}^{s_j}/\partial e}{\phi_{\hat{e}}^{s_j}}$*, then* $q_{s_i}^* \leq q_{s_j}^*$*. Higher individual informativeness decreases the threshold.*

(ii) If $\xi_{s_i} > \xi_{s_j}$*, then* $q_{s_i}^* \geq q_{s_j}^*$*. Higher equilibrium location parameter increases the threshold.*

(iii) If $\zeta_{s_i} > \zeta_{s_j}$ *and* $q_s^* > q_s^P$ *(*$q_s^* < q_s^P$*) for* $s \in \{s_i, s_j\}$*, then* $q_{s_i}^* \leq q_{s_j}^*$ *(*$q_{s_i}^* \geq q_{s_j}^*$*). A higher impact parameter decreases the threshold if thresholds are above the peak, and increases the threshold if they are below the peak.*

(iv) If $\sigma_{s_i} > \sigma_{s_j}$ *and* $q_s^* < \min\left\{q_s^P, \xi_s\right\}$ *(*$q_s^* > \max\left\{q_s^P, \xi_s\right\}$*) for* $s \in \{s_i, s_j\}$*, then* $q_{s_i}^* \leq q_{s_j}^*$ *(*$q_{s_i}^* \geq q_{s_j}^*$*). For symmetric distributions, we have* $q_s^P = \xi_s$*, so the conditions simplify to*

[5]At the peak of the distribution, we have $f'(q|\hat{e},s) = 0$. From equations (7) and (8), this is where the likelihood ratio is zero; by MLRP, this point is unique.

$q_s^* < (>) q_s^P$. *A higher scale parameter decreases the threshold if thresholds are below the peak, and increases the threshold if they are above the peak.*

Part (i) is the individual informativeness effect. If $\frac{\partial \phi_{\hat{e}}^{s_i}/\partial e}{\phi_{\hat{e}}^{s_i}} > \frac{\partial \phi_{\hat{e}}^{s_j}/\partial e}{\phi_{\hat{e}}^{s_j}}$, then signal realization s_i is individually more indicative of high effort than s_j. Thus, the threshold should be lower under s_i than s_j. This is the intuitive effect mentioned previously – if the signal is individually good news about effort, a lower minimum output is required for the manager to be paid. For example, signal s could be accounting profits, with s_i representing higher profits than s_j. High profits indicate high effort and thus increase the payment.

The output inference effect is given by parts (ii)-(iv) and can be decomposed into two components. The first is the "location effect" and captured by part (ii). If s_j is associated with a lower equilibrium location parameter ξ_{s_j} than s_i, then it indicates that the output distribution has shifted to the left. Due to MLRP, this shift means that achieving any given output level is more indicative of working than shirking. Thus, a lower minimum output is required for the manager to be paid. For example, let s be industry performance where s_j represents a downturn and s_i an upswing. If achieving an output level is more indicative of effort in a downturn, the threshold should be lower – the intuition behind relative performance evaluation.

The second component of output inference is the "precision effect". This effect is driven by two parameters of the output distribution: the impact parameter ζ_s (captured by part (iii)) and the scale parameter σ_s (captured by part (iv)). While relative performance evaluation would suggest that individually uninformative signals only have value if they shift the output distribution (i.e. have a location effect), the precision effect means that, surprisingly, a signal can affect the contract even if it is neither individually informative nor shifts the output distribution. However, unlike the location effect, where the impact of ξ_s is independent of the level of the thresholds, the precision effect depends critically on the thresholds.

The intuition is as follows. Low σ_s and high ζ_s mean that output is driven more by effort than luck. Thus, they are good news about effort when combined with high output, as they indicate that this high output is likely due to high effort rather than good luck. In contrast, they are bad news when combined with low output, as they indicate that this low output is likely due to low effort rather than bad luck. This contrasts the location effect, where a signal associated with a low equilibrium location parameter ξ_s is unambiguously good news about effort.

When the output threshold is high across signal realizations, the manager is paid only if output is sufficiently high to be good news about effort. When output is driven more by effort than luck, even moderately high output is sufficiently good news about effort to warrant the manager being paid – the threshold falls. Put differently, when the threshold is high, only high output levels are relevant. It does not matter that a low output level is less indicative of effort since, due to the high threshold, the manager receives zero anyway. When the output threshold is low across signal realizations, the manager is paid unless output is sufficiently low to be bad news about effort. When output is driven more by effort than luck, even moderately low output is sufficiently bad news about effort to warrant the manager not being paid – the threshold rises. The output thresholds in turn depends on the exogenous parameters of the model. For example, Appendix C shows that, if the participation constraint is non-binding and so the only goal of the contract is to provide incentives, an increase in the marginal cost of effort lowers the thresholds.

Another way to understand the precision effect is as follows. Concentrating incentives in states where output is driven more by effort than luck can yield the same overall incentives at a lower cost. Thus, the principal should provide more incentives in states where output is more precise. Whether "more incentives" corresponds to a higher or lower threshold depends on how high the threshold is to begin with. When the threshold is high, even if the manager works and achieves moderately good performance, he will not be paid; thus, reducing the threshold allows moderately good performance to be rewarded and thus increases incentives. However, if the threshold is low to begin with, the manager is paid even upon moderately poor performance. Raising the threshold means that moderately poor performance is no longer rewarded and thus increases incentives. In turn, increasing incentives in this state allows to lower incentives in other states when output is a less precise signal of effort, which either better insures the manager (lowering the cost of the contract if the IR binds) or diminishes his agency rent (if the IR does not bind).

Corollary 1 applies result (iv) of Proposition 1 to the case of a skew-normal output distribution and a non-binding participation constraint, to demonstrate how skewness affects the impact of the scale parameter on the threshold. In addition to location and scale parameters, the skew-normal distribution also has a "shape" parameter, which measures its skewness. Where the shape parameter is zero, the skew-normal distribution becomes a normal distribution.

Corollary 1 *Let the agent have a low reservation utility, $\overline{u} \leq u\left(\bar{W}\right) - C(\hat{e})$, so that the participation constraint is non-binding, and suppose that the output distribution is skew-normal. For a given signal s, the threshold q_s^* is increasing in the scale parameter if the signal is weakly bad news about effort ($\frac{\partial \phi_{\hat{e}}^s}{\partial e} \leq 0$) and the output distribution has nonnegative skewness.*

Intuitively, when the only purpose of the contract is to provide incentives, the manager is paid only if q and s are sufficiently good news about effort. When the signal s is (weakly) bad news, the likelihood ratio of output must be positive at the threshold q_s^*, to offset the bad news of the signal s. Since the likelihood ratio of output is zero at the peak, the threshold exceeds the peak due to MLRP. Moreover, for positively-skewed skew-normal distributions, the peak is above the location parameter so that $q_s^* > \max\left\{q_s^P, \xi_s\right\}$. Then, point (iv) of Proposition 1 implies that the threshold is higher for signals with a higher scale parameter.

Summing up the results of this section, a signal can affect the threshold even if it is not individually informative about effort. If the signal suggests that the entire output distribution has improved for reasons other than managerial effort, such as good industry performance, then all output levels are less indicative of effort and so the threshold rises. If the signal indicates that output is a more precise measure of effort, such as low industry volatility, then thresholds generally become less extreme – a previously high (low) threshold becomes lower (higher).

Before closing, we discuss three technical but important points. The first is to reiterate that none of the results in Proposition 1 are comparative statics, i.e. we are not changing underlying parameters such as the individual informativeness of the signal $\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s}$ or its location ξ_s. Such changes will, in general, change the Lagrange multipliers associated with the optimization program in equations (2)-(5) and thus have an ambiguous effect on the contract. Instead, our results are unique to an analysis of how the contract depends on the signal *realization*. We are holding constant the contracting setting, and showing how the contract depends on signal realizations that are associated with different parameters.

This difference contrasts our analysis with Chaigneau, Edmans, and Gottlieb (2018), who study how the optimal contract is affected by volatility, but not impact, location, or individual informativeness. Moreover, their volatility analysis is quite different. In their setting, there are no additional signals and the contract is written after volatility is realized, so volatility affects the contract that is offered. They show that a change in volatility ex ante affects the contract by altering the manager's incentives. In contrast, we study how incentives are allocated across signals – i.e. states of nature – that are associated with different output volatility. What matters instead is how the volatility associated with a given signal affects output informativeness. The contract is written before volatility (and other variables affected by the signal) is realized, and the contract is contingent on the signal – as is the case for performance-vesting equity.

Second, Proposition 1 contains "ceteris paribus" results where signal realizations s_i and s_j vary along only one dimension. For example, part (ii) shuts down the individual informativeness effect, considering signal realizations that affect the location parameter ξ_s but not individual informativeness $\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s}$. In reality, signal realizations may differ on multiple dimensions, and the location effect may counteract the individual informativeness effect. If industry performance is increasing in effort – for example, the manager's effort improves consumers' perception of the entire sector – then, one might think that high industry performance should be associated with a lower threshold, as it individually indicates high effort. However, this may be outweighed by the location effect: achieving a certain output is easier in an upswing. Thus, the threshold may be higher in an industry upswing, but Proposition 1 does not allow us to study this question analytically. Section 2.2 imposes more structure on the output distribution and allows us to analytically compare thresholds when signals differ across more than one dimension.

Third, even though the intuition behind the precision effect suggests that the effects of the impact and scale parameters should be in opposite directions, this is not always the case. For the impact parameter, what determines whether a threshold is "high" or "low" is how it compares with the peak of the distribution. A higher impact parameter corresponds to a steeper likelihood ratio of output: the ratio is more positive above the peak and more negative below the peak; at the peak it is zero and thus unaffected by the impact parameter.[6] Figure 1 illustrates. The solid line corresponds to $\zeta_{s_j} = 1$ and the dotted line to $\zeta_{s_i} = 2$. The left-hand side considers a positive cutoff for the likelihood ratio of output. Thus, the initial threshold $q^*_{s_j}$ is above the peak q^P_s, which is where the ratio is zero and the lines cross the x-axis. A higher impact parameter means that the new likelihood ratio of output at the initial threshold is even more positive, and so it equals the cutoff at a lower output level – thus the new threshold $q^*_{s_i}$ is lower. The right-hand side considers a negative cutoff, where the initial threshold $q^*_{s_j}$ is below the peak. A higher impact parameter corresponds to a lower likelihood ratio at $q^*_{s_j}$, and so the likelihood ratio equals the cutoff at a higher output level – thus the new threshold $q^*_{s_i}$ is higher. (We abuse language slightly by referring to "initial" and "new" parameters even though we are not conducting comparative statics but comparing parameters under different signal realizations.)

While the impact parameter changes the effect of effort on the output distribution, by construction it does not change the *equilibrium* output distribution. Recall that $h_s(\hat{e}) = \xi_s$: the location of the distribution is independent of ζ_s at the equilibrium effort level. Thus, the effect of the impact parameter depends on the threshold compared only with the distribution's peak, not its location.

In contrast, the scale parameter does change the equilibrium output distribution, spreading it out around the equilibrium location parameter – low output below the equilibrium location

[6]Recall that the likelihood ratio of output is given by equation (8). By definition, the (single) peak is such that $g' = 0$, and MLRP ensures that $g' < (>) 0$ below (above) the peak.

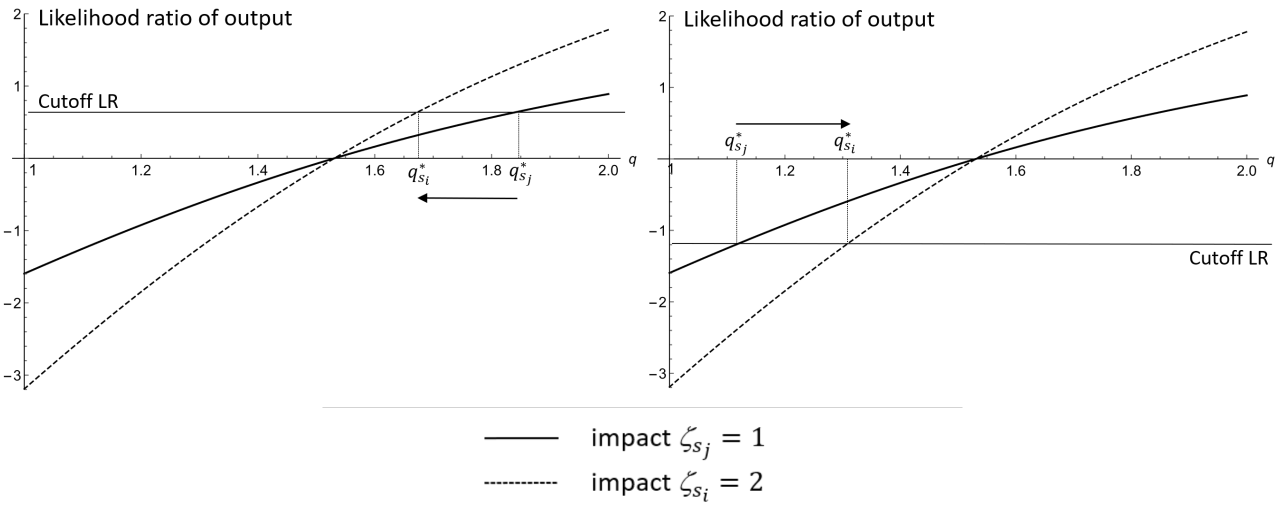

Figure 1: In both figures, the solid line is the likelihood ratio of output for a skew-normal distribution where location, scale, and impact parameters are all 1 and the shape parameter is 2. The peak of the distribution is $q_s^P \approx 1.531$. The dotted line is the likelihood ratio of output of the same distribution but with impact parameter 2.

parameter is a less negative signal of effort and high output is a less positive signal of effort. As a result, the effect of the scale parameter depends on two factors – the threshold compared to both the peak q_s^P and the equilibrium location parameter ξ_s as shown in part (iv). For symmetric distributions, the peak and location are the same, but for asymmetric distributions they are different. If the threshold is above (below) both the peak and location, a lower scale parameter corresponds to a higher (lower) likelihood ratio at the threshold. However, if the threshold is between these two points, the threshold could rise or fall. As a result, the scale parameter does not always have an opposite effect on the threshold to the impact parameter.

2.2 Option Repricing and Performance-Vesting

While Section 2.1 studied how a signal realization affects the threshold, we now study how it affects the slope of the contract above the threshold. So that the contract's slope can be defined in an unambiguous manner, we specialize the model to the case in which the optimal contract is linear above the threshold. As well as being theoretically clearer, this case is also practically applicable as many real-life contracts are piecewise linear. Indeed, if q is the stock price, the optimal contract is an option, and the slope above the threshold refers to the number of vesting options. Then, the model provides guidance on performance-based vesting – how the number of vesting options should depend on the performance measure s. Appendix D generalizes the analysis of the contract slope to the case of nonlinear contracts.

We first derive sufficient conditions under which the optimal contract is an option. Lemma 2 shows this is the case under the standard assumptions of log utility, normally-distributed output and limited liability. Formally, conditional on the signal s, let output q be normally distributed with mean $h_s(e)$ and standard deviation σ_s. The manager has log utility, $u(w) = \ln w$. Supplementary Appendix G provides a sufficient condition for the validity of the FOA under these assumptions[7] (the condition in Appendix B cannot be applied since log utility is unbounded).

From equation (1), the likelihood ratio under normal output is given by:

$$\begin{aligned} LR_s(q) := \frac{\frac{\partial f}{\partial e}(q, s|\hat{e})}{f(q, s|\hat{e})} &= \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \\ &= \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \frac{\frac{1}{\sigma\sqrt{2\pi}} \frac{2h'_s(\hat{e})(q-h_s(\hat{e}))}{2\sigma^2} \exp\left[-\frac{(q-h_s(\hat{e}))^2}{2\sigma^2}\right]}{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(q-h_s(\hat{e}))^2}{2\sigma^2}\right]} \\ &= \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \frac{\zeta_s}{\sigma_s^2}\left(q - \xi_s\right). \qquad (9) \end{aligned}$$

The $\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}}$ term is the standard individual informativeness effect. In general, the output inference effect $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$ comprises the precision and location effects; under the normal distribution (or any distribution with a linear likelihood ratio), $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$ can be cleanly decomposed into these two effects, allowing us to see the parameters that drive them. The precision effect is given by the first component of the second term, $\frac{\zeta_s}{\sigma_s^2}$. The signal s increases the precision of output as a measure of effort through increasing the impact of effort on output ζ_s or reducing the volatility of output σ_s. Here, since the normal distribution is symmetric, these parameters do always have opposite effects. The location effect is given by the final component, ξ_s. The signal affects expected output ξ_s and thus changes the location of the output distribution.

Lemma 2 below shows that the optimal contract gives the manager n_s^* options with strike price q_s^*.

Lemma 2 *(Optimal contract, log utility and normal output): The optimal contract under log*

[7] The condition for the validity of the FOA in the case without an additional signal is remarkably simple: it is $C''(e) \geq \frac{\bar{e}}{\sigma^2}$ for all $e \in [0, \bar{e}]$. For example, with a quadratic effort cost, $C(e) = \alpha e + \frac{\beta}{2}e^2$, and the condition is $\beta \geq \frac{\bar{e}}{\sigma^2}$.

utility and normally-distributed output consists of $n_s^* \geq 0$ *options with a strike price of* q_s^**:*

$$w(q) = n_s^* \max\{q - q_s^*, 0\}. \tag{10}$$

Given limited liability, the minimum payment is zero; given MLRP, this minimum payment will be made for all outputs below a threshold. Above the threshold, the payment is positive and determined so that the manager's marginal utility is the inverse of a linear transformation of the likelihood ratio (see Lemma 1). With log utility, marginal utility is the inverse of the payment, and so the payment equals a linear transformation of the likelihood ratio. With normally-distributed output, the likelihood ratio is linear in output, and so the payment is linear in output. Overall, the payment is zero below a threshold and linear in output above the threshold. This corresponds to an option contract, where the strike price is the threshold.

To our knowledge, Lemma 2 and our condition for the validity of the FOA in this setting provide the first sufficient conditions for the optimality of options with a risk-averse manager.[8] More generally, the linearity of the contract above the threshold, and thus the optimality of the option contract, holds not only for the normal distribution but for any distribution that has a linear likelihood ratio (for example, the gamma distribution).

Having derived the optimal contract in closed form, Proposition 2 studies how the signal affects each dimension of the contract.

Proposition 2 *(Effect of signal on vesting and strike price):*

(i) The number of options received ex post by the manager n_s^* *is proportional to* $\frac{\zeta_s}{\sigma_s^2}$*.*

(ii) The strike price q_s^* *is given by:*

$$q_s^* = \xi_s + \frac{\sigma_s^2}{\zeta_s}\left(K - \frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s}\right), \tag{11}$$

where $K \equiv \frac{\bar{W} - \lambda}{\mu}$*.*

[8] Jewitt, Kadan, and Swinkels (2008) show that the contract is "option-like" with risk aversion and agent limited liability, in that incentives are zero for low output and positive for high output, but do not identify conditions under which the increasing portion of the contract is linear. Hemmer, Kim, and Verrecchia (1999) identify a linear likelihood ratio and log utility as leading to the contract having a linear portion, but do not combine them with limited liability to obtain an option contract. In addition, Jewitt, Kadan, and Swinkels (2008) assume the Rogerson (1985) conditions to guarantee the validity of the FOA, but these conditions do not hold under the normal distribution; Hemmer, Kim, and Verrecchia (1999) assume the Jewitt (1988) conditions but they do not hold under limited liability. We derive a general condition for the validity of the FOA that holds in the setting of limited liability, normal output and log utility. The standard model justifying options under moral hazard is Innes (1990), which requires the agent to be risk-neutral. In addition, under risk neutrality, the manager is the residual claimant for $q \geq q^*$, so that the number of options is fixed at 1 and does not depend on the signal realization. Under risk aversion, this need not be the case.

The intuition for the number of options is as follows. As in any principal-agent model, pay is increasing in the likelihood ratio. The number of options represents pay-performance sensitivity, and is thus increasing in the sensitivity of the likelihood ratio to output, $\frac{dLR_s(q)}{dq} = \frac{\zeta_s}{\sigma_s^2}$. As is standard, the strike price is the level of output that, if not reached, it is sufficiently likely that the agent shirked that it is optimal to pay him zero. In this setting, the strike price can be derived in closed form.

Note that Proposition 2 is different from Proposition 1 – part (ii) is not simply Proposition 1 applied to the case of log utility and normally-distributed output. Proposition 1 compared signal realizations that differ only across a single dimension, e.g. either the impact parameter ζ_s or the scale parameter σ_s, but not both. Moving to this setting allows Proposition 2 to compare contracts analytically when signals differ across multiple dimensions. For example, part (i) allows signal realizations to differ in both ζ_s and σ_s. In an economic expansion, not only might effort have a greater effect on output ζ_s (e.g. if effort has a multiplicative effect on firm value), but also volatility σ_s might be higher or lower. The ability to compare contracts when signals differ across multiple dimensions will be important for the applications later in this section.[9]

We now discuss the two parts of Proposition 2 in turn.

2.2.1 Performance-based vesting

Part (i) of Proposition 2 studies how a signal realization affects the number of options given to the manager. A signal has value if it affects any component of the likelihood ratio in equation (9): $\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s}$ (the individual informativeness effect), $\frac{\zeta_s}{\sigma_s^2}$ (the precision effect), or ξ_s (the location effect). The existence of such a signal will, in general, alter the Lagrange multiplier μ and thus scale up or down the number of options $n_s^* = \mu \frac{\zeta_s}{\sigma_s^2}$ received across all signal realizations. However, part (i) shows that the number of options received will depend on the actual signal realization only via the precision effect and not via the location or individual informativeness effects.

The intuition is as follows. The number of vesting options represents pay-performance

[9]While we allow signal realizations to differ across multiple dimensions, it is important to reiterate that we are holding the distribution of signals fixed. Proposition 2 would not arise under a comparative statics analysis that compares contracts under different signal distributions, where the principal observes the parameters of the distribution before writing the contract. Here, the contract is set before the signal is observed, and the agent does not know the signal at the time he makes his effort decision. This is what allows the principal to efficiently concentrate incentives (by giving more options) in states of the world where output is more informative. If the agent already knew the state when he took his effort decision, to implement $\hat{e}$ the principal would need to provide the same incentives in all states: see Edmans and Gabaix (2011).

sensitivity ("PPS"). Pay should be more sensitive to performance, i.e. the number of vesting options should be higher, upon signals where output is a more precise measure of effort. This arises if either effort has a greater effect on output (ζ_s is higher) or output is less volatile (σ_s is lower). Proposition 2 derives PPS in closed form and shows how it depends on the *ratio* of impact ζ_s to (the square of) scale σ_s. We can thus understand precisely how PPS varies with parameters of the output distribution under various signal realizations.

To our knowledge, part (i) is the first theoretical justification of why performance-based vesting may be optimal in a standard moral hazard model where the agent only makes an effort decision. The only other justification of performance-based vesting of which we are aware is Marinovic and Varas (2019), where the agent also takes a manipulation action, and the role of performance-based vesting is to deter manipulation. In that model, performance-based vesting is always optimal. Given the "performance-based vesting" terminology, one might think that it will similarly always be optimal in an effort-only model, as long as the performance measure is informative about effort. However, this is not the case, and we provide a framework to understand under what conditions performance-vesting is optimal.

Specifically, it might seem that a signal that is individually indicative of effort (i.e. increases $\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}}$) should lead to more vesting; indeed, current performance-vesting practices award more equity after beating performance thresholds. However, Proposition 2 shows that positive signals of effort should increase the level of pay for all output realizations (reduce the strike price) rather than the sensitivity of pay to output (increase the number of vesting options). Similarly, one might think that a signal that indicates that high output is due to luck (i.e. the output distribution has shifted to the right) should lead to less vesting – the intuition behind relative performance evaluation. Indeed, Bettis et al. (2018) find that 48% of firms that use performance-based vesting have at least one performance measure that is calculated relative to peers. However, Proposition 2 shows that the equilibrium location parameter ξ_s (which is affected by peer performance) changes the strike price, not the number of vesting options. In sum, vesting should not depend on signals that do not affect output precision.

2.2.2 Strike price

Part (ii) of Proposition 2 turns to the second dimension of the contract, the strike price. Unlike the number of vesting options, which depends only on the precision effect, the strike price is driven by all three effects. The intuition is as follows. The individual informativeness effect matters because it captures what a signal individually conveys about effort, regardless of the output realization. The strike price affects the payment for all outputs above the strike price. Thus, what the signal individually conveys about effort affects the strike price – signals that are more indicative of effort are associated with lower strike prices. The location effect matters for a similar reason – if the signal indicates that the output distribution has shifted to the left, a given output is more indicative of effort, and so the strike price should be lower.

As in Section 2.1, the precision effect can be either positive or negative. Equation (11) shows that the effect of precision $\frac{\zeta_s}{\sigma_s^2}$ on the strike price q_s^* depends on $K \lesseqgtr \frac{\partial \phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}$, which is equivalent to $q_s^* \lesseqgtr \xi_s$. If $q_s^* > \xi_s$, i.e. strike prices are high across signal realizations, the manager is only paid if output is high. If output is a more precise signal of effort ($\frac{\zeta_s}{\sigma_s^2}$ is higher), even moderately high outputs are positive signals of effort – the strike price falls. One case in which $K > \frac{\partial \phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}$ (and thus $q_s^* > \xi_s$) is when the manager is above his reservation utility ($\lambda = 0$ yields $K \equiv \frac{\bar{W}-\lambda}{\mu} > 0$) and the signal is individually uninformative ($\frac{\partial \phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s} = 0$). If $q_s^* < \xi_s$, i.e. strike prices are low across signal realizations, then greater precision $\frac{\zeta_s}{\sigma_s^2}$ increases the strike price due to the converse argument. In both cases, greater output precision leads to the strike price being closer to the mean of the stock price distribution.

Note that equation (11) shows that the individual informativeness effect is scaled by the inverse of signal precision, $\frac{\sigma_s^2}{\zeta_s}$, and so the individual informativeness effect interacts with the precision effect. The intuition is as follows. A signal that is individually bad news about effort leads to a higher strike price. If output is an imprecise measure of effort, then the strike price needs to be raised a lot, otherwise it would likely be exceeded purely by luck.

Since the strike price is driven by the location and precision effects, as well as the individual informativeness effect, it may be lower even if the signal individually indicates low effort. Consider two signal realizations, L and H, such that $\frac{\partial \phi_{\hat{e}}^L/\partial e}{\phi_{\hat{e}}^L} < \frac{\partial \phi_{\hat{e}}^H/\partial e}{\phi_{\hat{e}}^H}$: L is individually worse news about effort than H. Despite this, the strike price may be lower under L ($q_L^* < q_H^*$). This can happen for two reasons. Consider the following simple cases.

One case is $\xi_L < \xi_H$ and $\frac{\zeta_L}{\sigma_L^2} = \frac{\zeta_H}{\sigma_H^2}$, so the signal has a location effect but not a precision effect. Here, any given output q is better news about effort under L than H, since $\frac{\frac{\partial f}{\partial e}(q|\hat{e},L)}{f(q|\hat{e},L)} > \frac{\frac{\partial f}{\partial e}(q|\hat{e},H)}{f(q|\hat{e},H)}$ for any q. Equation (11) shows that the strike price is optimally lower under L if the location effect (the difference between ξ_L and ξ_H) outweighs the individual informativeness effect (the difference between $\frac{\partial \phi_{\hat{e}}^L/\partial e}{\phi_{\hat{e}}^L}$ and $\frac{\partial \phi_{\hat{e}}^H/\partial e}{\phi_{\hat{e}}^H}$). For example, let s be the number of industry competitors. Few competitors ($s = H$) are individually a better signal of effort than many competitors ($s = L$), because they suggest that the manager has driven out competition or deterred entry. This consideration may be outweighed by a second effect – a given level of output is a more positive signal of effort the more competitors there are, since competition causes a downward shift in the location of the output distribution. Equation (11) shows that the strike price may optimally be lower when there are more competitors, despite $s = L$ individually indicating low effort, as this consideration may be outweighed by the location effect.

A second case is $\frac{\zeta_H}{\sigma_H^2} < \frac{\zeta_L}{\sigma_L^2}$, $\xi_H = \xi_L = \xi$, so the signal has a precision effect but not a location effect. Since $\frac{\zeta_H}{\sigma_H^2} < \frac{\zeta_L}{\sigma_L^2}$, output q is more informative about effort under L than H which means that q_L^* should be closer than q_H^* to ξ, ceteris paribus. Since $\frac{\partial \phi_{\hat{e}}^L / \partial e}{\phi_{\hat{e}}^L} < \frac{\partial \phi_{\hat{e}}^H / \partial e}{\phi_{\hat{e}}^H}$, the individual informativeness effect means that $q_L^* > q_H^*$, ceteris paribus. Thus, if both q_L^* and q_H^* are below ξ, the precision and individual informativeness effects work in the same direction and so we unambiguously have $q_L^* > q_H^*$. We thus focus on the interesting case where both q_L^* and q_H^* are above ξ. We have $q_L^* < q_H^*$ if the precision effect (the difference between $\frac{\zeta_H}{\sigma_H^2}$ and $\frac{\zeta_L}{\sigma_L^2}$) is sufficiently large to outweigh the individual informativeness effect (the difference between $\frac{\partial \phi_{\hat{e}}^H / \partial e}{\phi_{\hat{e}}^H}$ and $\frac{\partial \phi_{\hat{e}}^L / \partial e}{\phi_{\hat{e}}^L}$).

For example, let $s = H$ be high volatility (e.g. of the stock price, cash flow, or profit) and $s = L$ be low volatility. If the manager's effort involves finding and pursuing risky investment opportunities, then $s = L$ is individually bad news about effort. In addition, it also means that output is a precise measure of managerial effort. Thus, high output combined with low volatility can indicate effort more than high output combined with high volatility. Even though low volatility individually indicates low effort, in combination with high output it indicates high effort, and so can be associated with a lower strike price ($q_L^* < q_H^*$).

Summing up the results of these two sub-sections, a signal realization should be associated with more vesting options if it is associated with a higher optimal sensitivity of pay to output, due to output being a more precise measure of effort. It should be associated with a lower strike price if it increases the optimal level of pay. This will be the case if the signal indicates high effort, that the location of the equilibrium output distribution has shifted to the left, that output precision has increased and strike prices are high, or that precision has decreased and strike prices are low.

2.2.3 Applications

We now apply the results of parts (i) and (ii) of Proposition 2 to three types of signal.

Economic conditions

First, let s be a signal of economic conditions. Economic conditions are outside the manager's control and thus individually uninformative about effort ($\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} = 0$), but may have location and precision effects.

Economic conditions will affect vesting if and only if they affect either ζ_s or σ_s. Starting with the former (ζ_s), if good economic conditions increase the manager's impact on output ζ_s (e.g. because effort has a multiplicative effect on firm value), vesting should be increasing in economic conditions. If instead bad economic conditions increase the manager's impact, e.g. if the stakes are higher in bad times, vesting should be decreasing in economic conditions. Moving to the latter (σ_s), if risk σ_s is lower (higher) in good economic conditions, then output is a more (less) precise signal of effort and so vesting should be higher (lower).

That vesting should depend on economic conditions suggests that it should be affected by luck. This contrasts with current performance-vesting practices which assume that vesting should depend on performance measures within the manager's control – and if individually uninformative measures are used, this is to filter out external factors that affect the location of the output distribution (the intuition behind relative performance evaluation). However, Proposition 2 shows that the location effect should only affect the strike price, not vesting.

Economic conditions will affect the strike price if they affect ξ_s, ζ_s, or σ_s. The strike price should depend on the location effect. If the firm's business is pro-cyclical, then good economic conditions are associated with a high location parameter ξ_s, which increases the strike price. The opposite is true if the business is counter-cyclical. In addition, the strike price should depend on the precision effect. If good economic conditions increase output precision $\frac{\zeta_s}{\sigma_s^2}$ and the manager is above his reservation utility, then the strike price is decreasing in economic conditions. The opposite is true if good economic conditions reduce output precision.

We now present an example that illustrates the effect of economic conditions on the contract. We consider a signal s that affects only the location and scale parameters, but not the impact parameter, and where good economic conditions are associated with lower volatility. As a result, the number of vesting options is higher in an expansion, because lower volatility means greater output precision. The strike price is affected by both the location and precision effects. On the one hand, an expansion is associated with a higher location parameter which tends to increase the strike price. On the other hand, it is also associated with a lower scale parameter which tends to decrease the strike price. In our example, the latter effect dominates, so the strike price is higher in a recession – the opposite of the relative performance evaluation effect.

Example 1 *Suppose that the manager has log utility, outside wealth* $\bar{W} = 2$*, and reservation utility* $\bar{u} = 0$*. The cost of effort is characterized by* $C(\hat{e}) = 0.5$ *and* $C'(\hat{e}) = 0.1$*. The signal* s *is the state of the economy, which can be a recession (*$\tilde{s} = s_R$*, which occurs with probability* 0.25*) or expansion (*$\tilde{s} = s_E$*). The signal is individually uninformative about effort: (*$\frac{\partial \phi_{\hat{e}}^s}{\partial e} = 0$*). At equilibrium effort* $\hat{e}$*, output is normally distributed with location and scale parameters* $\xi_{s_R} = 10$ *and* $\sigma_{s_R} = 1.7$ *in a recession, and* $\xi_{s_E} = 11$ *and* $\sigma_{s_E} = 1.0$ *in an expansion. For both signals,*

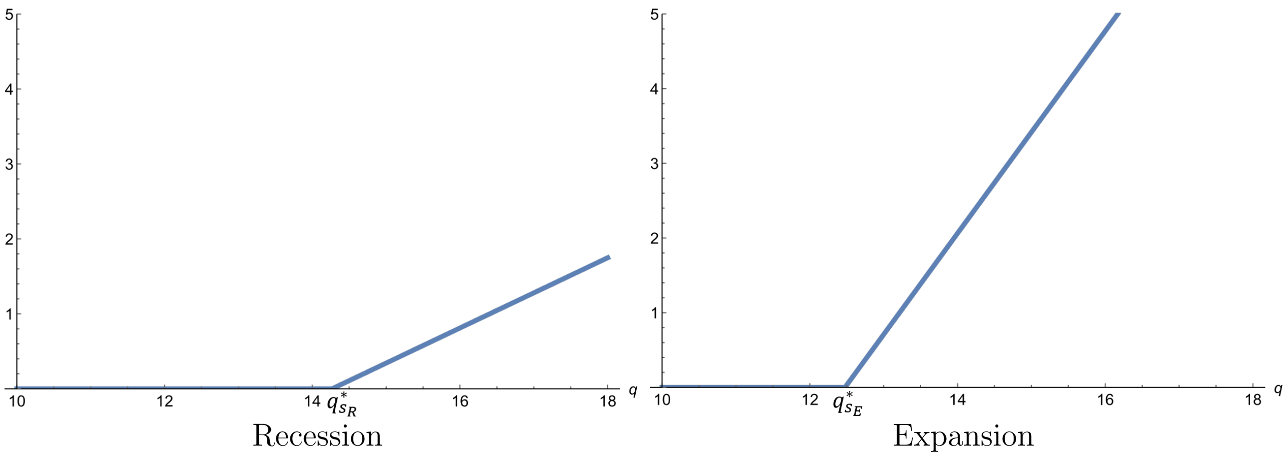

Figure 2: The payoff of the manager as a function of output in a recession ($s = s_R$) on the left, and in an expansion ($s = s_E$) on the right, in Example 1.

the impact of effort on output is the same: $\zeta_s = 1$.

Since $\ln \bar{W} - C(\hat{e}) \geq \bar{u}$, *the participation constraint is nonbinding. We find a strike price of* $q^*_{s_R} \approx 14.27$ *in a recession, and* $q^*_{s_E} \approx 12.48$ *in an expansion. Indeed, we have* $q^*_s = \xi_s + \frac{\sigma_s^2}{\zeta_s}\frac{\bar{W}}{\mu}$*: in a recession, a lower equilibrium location parameter* ξ_s *decreases the strike price, whereas a higher scale parameter* σ_s *increases it. The latter effect dominates so that the strike price is higher in a recession. Moreover, because output is a more precise measure of effort in an expansion, the number of vesting options is higher in an expansion. The state-contingent contract is depicted in Figure 2.*

Accounting performance

Now let s be an accounting performance measure, such as profits or cash flows, which Bettis et al. (2018) show to be common vesting determinants. The main difference with the previous section is that, unlike economic conditions, accounting performance is individually informative about effort. In addition, the settings in which s affects the impact and scale parameters will be different when s is accounting performance rather than economic conditions.

There are two cases to consider, and we discuss the effects on both vesting and the strike price together. First, suppose that the ratio $\frac{\zeta_s}{\sigma_s^2}$ is decreasing in profits. This may be the case for a start-up, where the baseline scenario is low profits and a low stock price – high profits increase stock price volatility because investors speculate over whether the high profits are sustainable. Starting with vesting, it is decreasing in profits because they are associated with greater volatility. (While profits also individually indicate effort, this is irrelevant for vesting.)

Moving to the strike price, it is affected by profits in two ways. The individual informativeness effect means that the strike price will be decreasing in profits; the precision effect can go in either direction depending on whether strike prices are high or low. In addition to its own effect, the precision effect also dampens the individual informativeness effect. For low profits, output volatility is low, and therefore the strike price does not need to increase much to punish the manager.

Second, suppose that the ratio $\frac{\zeta_s}{\sigma_s^2}$ is increasing in profits. This may arise due to the well-known "leverage effect" of Black (1976) and Christie (1982), where higher profitability reduces a firm's leverage and thus equity volatility. It may also be true for a mature firm, where the baseline scenario is high profits and a high stock price. High profits imply "business as usual", where the stock price is less volatile and thus more informative about effort. Low profits likely mean that the business was disrupted, for example by new entrants, so the stock price is more volatile and less informative about effort. In addition, if effort has a multiplicative effect on firm value, the impact parameter ζ_s is higher in a more profitable firm. In this case, vesting is increasing in profits. As in the first case, the strike price tends to be decreasing in profits due to the individual informativeness effect, and the precision effect can go in either direction. Now, the precision effect dampens the response to high profits. For high profits, output volatility is low, and therefore the strike price does not need to decrease much to reward the manager.

In addition to depending on firm characteristics, the impact of accounting performance on the strike price may also depend on the manager's tenure. Accounting performance realized early in the manager's tenure will mostly be attributable to exogenous factors, and it will have an effect on the stock price to the extent that it is partly unexpected. For example, announcing unexpectedly high profits will increase the location of the stock price distribution, and therefore the strike price. By contrast, accounting performance realized later in the manager's tenure will typically be informative about his effort, and the individual informativeness effect is such that high accounting performance tends to reduce the strike price.

Sustainability performance

Finally, let s be a measure of sustainability performance, such as employee satisfaction, customer satisfaction, or carbon emissions. Practitioners argue that sustainability measures should be incorporated into CEO contracts; for example, this is advocated by the 2020 European Commission study on sustainable corporate governance. While one reason for doing so is if the firm's objective is broader than just shareholder value, our model suggests that it should be incorporated even if the objective function is shareholder value alone. For example, Bettis et al. (2018) find that the vesting of some equity grants depends on customer satisfaction, which is often used as a sustainability metric but unlikely to be a separate objective from shareholder value. It may seem that the impact of sustainability is automatic because high sustainability is individually informative about managerial effort. However, since sustainability can also change location and precision, the effects are nuanced.

There is evidence that increases in sustainability are associated with lower stock price volatility (Hoepner et al. 2020). If this association is not exclusively driven by a firm fixed effect, then vesting should be increasing in sustainability, because the stock price q is more informative about managerial effort upon high sustainable performance (higher $\frac{\zeta_s}{\sigma_s^2}$). However, vesting should only be increasing in sustainability if it affects output precision, not if it indicates managerial effort.

In addition, sustainability performance could affect the location parameter. In the general equilibrium model of Pastor et al. (2020), more sustainable firms have higher stock prices if customers' demand for sustainability is unexpectedly high. This suggests that, in this case, higher sustainability performance should be associated with a higher location parameter ξ_s, which increases the strike price. Intuitively, this avoids rewarding or punishing executives for changes in the stock price due to unexpected changes in customer tastes.

3 Implications

This section summarizes the normative implications of our model for compensation design, and compares them to current practice and existing empirical studies. The first three implications concern the level of the threshold and apply to any threshold-based contract, such as dismissal contracts, bonuses, and performance-vesting stock or options. The fourth concerns the number of vesting options and applies to performance-vesting options.

1. All else equal, thresholds should fall upon signals that are individually good news about effort.

This implication stems from part (i) of Proposition 1. A signal that is good news about effort should increase the payment to the manager, regardless of his current payment level, unless limited liability binds. Proposition 1 achieves this by reducing the output threshold; this increases the payment as long as output exceeds the threshold, but has no effect if output is below (i.e. limited liability binds). While it seems intuitive that pay should rise with signals that individually indicate effort, this is not always the case in reality. Rather than decreasing the threshold, additional performance signals typically enter the contract as a separate bonus component as shown in Figure 3 below (adapted from Edmans, Gabaix, and Jenter (2017)). This differs from Proposition 1 in several ways. First, the signal has no effect on pay if it is

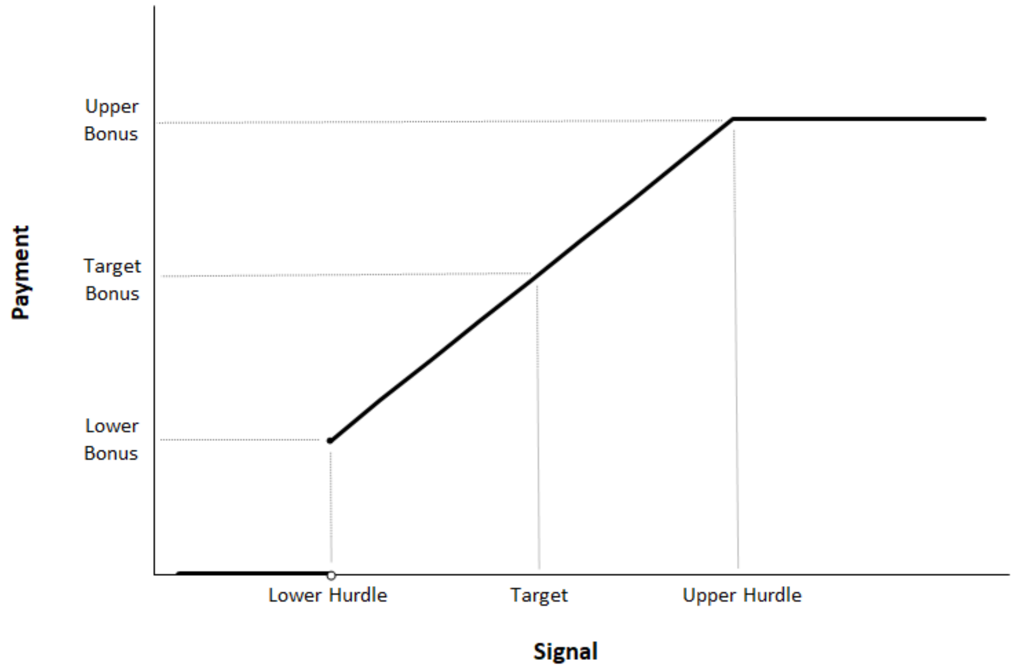

Figure 3: Example of a bonus payment used in practice as a function of the signal s.

below the lower hurdle or above the higher hurdle. In our model, as long as limited liability does not bind, pay should always be higher if the signal individually indicates effort. Second, pay jumps discontinuously once the signal reaches the lower hurdle, but Proposition 1 features no such discontinuities. Third, the effect of the signal on pay does not depend on whether output is above the threshold, as it should.

2. All else equal, thresholds should fall upon signals that indicate the output distribution has shifted to the left.

This implication stems from part (ii) of Proposition 1, and indeed often occurs in reality. Bettis et al. (2018) find that 48% of performance-vesting equity grants calculate at least one performance measure on a relative rather than an absolute basis. This effectively reduces the threshold when peer performance is weak, as this indicates that the output distribution has shifted to the left. In addition to this implication being implemented on a contractual basis through vesting being based on relative performance measures, it is sometimes implemented on a discretionary basis. For example, the BP board lowered the cash flow target in the bonus of CEO Bob Dudley in 2015, because the Deepwater Horizon disaster caused a leftward shift in the distribution of cash flows.

3. All else equal, thresholds become less extreme (high thresholds fall and low thresholds rise) upon signals that indicate that output is a precise measure of effort.

This implication stems from parts (iii) and (iv) of Proposition 1. Overall, signals associated with lower σ (such as low market, industry or firm volatility) and/or higher ζ (such as low regulation, high product market fluidity, high industry competition, or high industry disruption)[10] increase the precision of output as a signal of effort and thus lead to less extreme thresholds: low thresholds rise and high thresholds fall. Intuitively, if output is more informative about effort, then less extreme output levels are needed to determine whether the manager should be paid. However, we are not aware of cases in which thresholds are altered due to changes in the precision of output as a measure of effort. In reality, while practitioners recognize that signals should affect contracts if they are individually informative or indicate shifts in the output distribution, we are not aware of cases in which signals affect contracts due to indicating changes in output precision.

Corollary 1, which applies part (iv) of Proposition 1 to the case of skewed distributions, suggests that if output is positively skewed, thresholds should be higher upon signal realizations that indicate high volatility. In reality, stock price distributions are typically positively skewed (see, e.g., Albuquerque (2012)), and Bettis et al. (2010, Table 8) show that stock price thresholds in performance-vesting equity are increasing in stock return volatility, even though they do not depend on most other firm-level variables. However, their result studies the impact of volatility ex ante, i.e. before the contract has been signed; Corollary 1 focuses on how the threshold should vary ex post with realized volatility.

Applying points 1-3 to option contracts suggests that it is sometimes optimal to reduce the option strike price. Liljeblom, Pasternack, and Rosenberg (2011) study the determinants of the option strike price ex ante, i.e., when the contract is set, in Finland where there are no tax and accounting considerations that favor at-the-money options as in the US. However, they do not study how the strike price is reset ex post depending on the signal realization. Brenner, Sundaram, and Yermack (2000) find empirically that repricing nearly always involves a lowering of the strike price, and follows poor stock price performance (both absolute and industry-adjusted). Chance, Kumar, and Todd (2000) also find that repricing follows poor stock returns, but that these are not due to market or industry conditions. Our model suggests that

[10]See RegData's Industry Regulation Index for an example of an industry regulation index, and Hoberg, Phillips, and Prabhala (2013) for an example of a measure of product market fluidity. Merger waves are a potential measure of industry disruption; Harford (2005) finds that they are driven by economic, regulatory, and technological shocks.

a reduction in the strike price should *generally* be prompted by positive, rather than negative, signals of effort, suggesting that such practices are suboptimal.[11] However, the examples in Section 2.2.2 (of more industry competitors and low risk) provide conditions under which repricing after poor performance may be optimal, contrary to concerns that it is universally inefficient because it rewards failure (e.g. Bebchuk and Fried (2004)).

4. The number of vesting options should increase if and only if a signal indicates that output is a precise measure of effort.

This implication stems from part (i) of Proposition 2. While practitioners recognize that certain signals should lead to adjustments to the level of pay (either through separate bonuses or changes in thresholds), we are not aware of cases in which signals affect pay-performance sensitivity due to indicating changes in output precision.

This result is related to studies on the relation between firm risk and PPS (e.g. Garen (1994), Aggarwal and Samwick (1999)). Those studies estimate this relation cross-sectionally between firms, whereas our prediction is within firm but across states of the world. Aggarwal and Samwick (1999) argue that a negative relation between PPS and stock return volatility is a "key prediction" of the principal-agent model of executive compensation. However, Prendergast's (2002) review of the evidence finds a mixed relationship, and points out that this may arise because risk can be endogenous either to the agent's effort or the job to which he is assigned. Our model yields a related prediction which is not affected by this endogeneity concern: even if the manager could change the overall level of risk, the realized level of risk will still depend on the state of the world. For a given manager, PPS should be higher (via additional vesting) in states where the firm's stock is less volatile because the stock price is a better signal of effort.

Moreover, part (i) of Proposition 2 is an "if and only if" result – it suggests that the number of vesting options should not change with signals that only affect individual informativeness or the location parameter. In practice, vesting sometimes depends on measures of accounting performance, even though they may not clearly affect the impact or scale parameters. Furthermore, combining points 1-4 above helps us understand how signals that affect all three of individual informativeness, location, and precision should be used. It suggests that the common practice of making pay depend on a performance measure only through performance-based vesting may not actually be optimal. This is efficient if the performance measure only has a precision effect, but most measures also have individual informativeness or location effects;

[11] Acharya, John, and Sundaram (2000) also study the repricing of options theoretically. In their model, repricing is not undertaken to make use of additional informative signals, but instead to maintain effort incentives when options fall out of the money.

thus, they should also affect the strike price.

The conventional justification of performance-based vesting is that it rewards the manager for good performance. However, its primary effect is to change the slope of the contract, which is only optimal if the signal realization affects either impact or scale (or both). If the principal wishes to reward the manager for good performance, either because the signal is individually indicative of effort, or because the signal shows that a given output was harder to achieve (the location effect), this should be done by lowering the strike price. Changing the number of vesting options fails to change the payment below the strike price. Thus, the normative implication of this section is that performance-based options should typically involve the strike price, rather than the number of vesting options, depending on performance.

4 Conclusion

This paper has studied how signal realizations should affect contracts, thus providing guidance on practical contract design. In a general optimal contracting model with limited liability, the optimal contract involves a threshold output level. The signal realization affects the threshold in two main ways. The first is the individual informativeness effect – signals that individually indicate high effort should be associated with lower thresholds, all else equal. The second is the output inference effect – signals affect the information that the principal infers about effort from the output level. Due to this second effect, the impact of a signal realization on the threshold is ambiguous; signals that are individually indicative of effort may end up being associated with higher thresholds. The output inference effect in turn can be decomposed into two components. The location effect arises if the signal indicates that the location of the output distribution has shifted to the left (right), in which case the threshold decreases (increases). The precision effect occurs if the signal indicates that output is riskier and/or more impacted by effort. Greater output precision (higher impact and/or lower risk) leads to high thresholds falling and low thresholds rising.

We then study how the signal realization affects the slope of the contract above the threshold. To do so, we apply our model to study option contracts, which are linear above the threshold (strike price), allowing us to define the slope in an unambiguous way. This application has implications for performance-vesting options. In practice, performance signals only affect the slope (i.e. number of vesting options), but our analysis suggests that they should instead affect the strike price. Only a limited set of signals – those that affect output precision – should also affect the number of vesting options.

Our model features a single effort decision by the manager and a single component of output, and studies how to incorporate an additional signal into a contract. In future research, it would be interesting to study a multi-tasking model where there are multiple actions and multiple components of output. For example, in Holmström and Tirole (1993), the manager takes both a short-term and a long-term action, and the principal cares about both short-term earnings and liquidation value. How to incorporate an additional signal, that is informative about one or both actions, into the contract is an open question. A second potential extension would be to explore in more detail how the signal affects optimal risk-sharing. While the focus of this paper is on how signals affect contracts through what they imply about effort, it does contain a channel through which the signal affects risk-sharing – the precision effect shows that, when output is a less precise indicator of effort, pay is less sensitive to output. Future research could allow the manager's wealth or utility function to vary with the signal, thus leading to a richer analysis on how signals affect risk-sharing.

References

[1] Acharya, Viral V., Kose John, and Rangarajan K. Sundaram (2000): "On the optimality of resetting executive stock options." *Journal of Financial Economics* 57, 65–101.

[2] Aggarwal, Rajesh K., Andrew A. Samwick (1999): "The other side of the trade-off: the impact of risk on executive compensation." *Journal of Political Economy* 107, 65–105.

[3] Albuquerque, Rui (2012): "Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns." *Review of Financial Studies* 25, 1630–1673.

[4] Allen, Franklin and Douglas Gale (1992): "Measurement distortion and missing contingencies in optimal contracts." *Economic Theory* 2, 1–26.

[5] Amershi, Amin H. and John S. Hughes (1989): "Multiple signals, statistical sufficiency, and Pareto orderings of best agency contracts." *RAND Journal of Economics* 20, 102–112.

[6] Axelson, Ulf and Sandeep Baliga (2009): "Liquidity and manipulation of executive compensation schemes." *Review of Financial Studies* 22, 3907–3939.

[7] Bebchuk, Lucian Arye and Jesse M. Fried (2004): *Pay without performance: The unfulfilled promise of executive compensation*. Harvard University Press, Cambridge.

[8] Bettis, J. Carr, John Bizjak, Jeffrey Coles, and Swaminathan Kalpathy (2010): "Stock and option grants with performance-based vesting provisions." *Review of Financial Studies* 23, 3849–3888.

[9] Bettis, J. Carr, John Bizjak, Jeffrey Coles, and Swaminathan Kalpathy (2018): "Performance-vesting provisions in executive compensation." *Journal of Accounting and Economics* 66, 194–221.

[10] Black, Fischer (1976): "Studies of stock price volatility changes." *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association*, 177–181.

[11] Brenner, Menachem, Rangarajan K. Sundaram, and David Yermack (2000): "Altering the terms of executive stock options." *Journal of Financial Economics* 57, 103–128.

[12] Chaigneau, Pierre, Alex Edmans and Daniel Gottlieb (2018): "Does improved information improve incentives?" *Journal of Financial Economics* 130, 291–307.

[13] Chaigneau, Pierre, Alex Edmans and Daniel Gottlieb (2019): "The informativeness principle without the first-order approach." *Games and Economic Behavior* 113, 743–755.

[14] Chance, Don M., Raman Kumar, and Rebecca B. Todd (2000): "The 'repricing' of executive stock options." *Journal of Financial Economics* 57, 129–154.

[15] Christie, Andrew A. (1982): "The stochastic behavior of common stock variances: Value, leverage and interest rate effects." *Journal of Financial Economics* 4, 407–432.

[16] Dittmann, Ingolf, Ernst Maug, and Oliver G. Spalt (2013): "Indexing executive compensation contracts." *Review of Financial Studies* 26, 3182–3224.

[17] Edmans, Alex and Xavier Gabaix (2011): "Tractibility in incentive contracting." *Review of Financial Studies* 24, 2865–2894.

[18] Edmans, Alex, Xavier Gabaix and Dirk Jenter (2017): "Executive compensation: A survey of theory and evidence." *Handbook of the Economics of Corporate Governance* (Benjamin E. Hermalin and Michael S. Weisbach, eds.) Elsevier/North-Holland: New York.

[19] Gale, Douglas and Martin Hellwig (1985): "Incentive-compatible debt contracts: The one-period problem." *Review of Economic Studies* 52, 647–663.

[20] Garen, John E. (1994): "Executive compensation and principal–agent theory." *Journal of Political Economy* 102, 1175–1199.

[21] Gjesdal, Froystein (1982): "Information and incentives: the agency information problem." *Review of Economic Studies* 49, 373–390.

[22] Gopalan, Radhakrishnan, Todd T. Milbourn and Fenghua Song (2010): "Strategic flexibility and the optimality of pay for sector performance." *Review of Financial Studies* 23, 2060–2098.

[23] Grossman, Sanford J., and Oliver D. Hart (1983): "An analysis of the principal-agent problem." *Econometrica* 51, 7–45.

[24] Hartman-Glaser, Barney and Benjamin Hébert (2020): "The insurance is the lemon: Failing to index contracts." *Journal of Finance* 75, 463–506.

[25] Hemmer, Thomas (2004): "Lessons lost in linearity: A critical assessment of the general usefulness of LEN models in compensation research." *Journal of Management Accounting Research* 16, 149–162.

[26] Harford, Jarrad (2005): "What drives merger waves?" *Journal of Financial Economics* 77, 529–560.

[27] Hoberg, Gerard, Gordon Phillips, and Nangpurnanand Prabhala (2013): "Product market threats, payouts, and financial flexibility." *Journal of Finance* 69, 293–324.

[28] Hoepner, Andreas G. F., Ioannis Oikonomou, Zacharias Sautner, Laura T. Starks, and Xiaoyan Zhou (2020): "ESG shareholder engagement and downside risk." Working Paper, University College Dublin.

[29] Hoffmann, Florian and Sebastian Pfeil (2010): "Reward for luck in a dynamic agency model." *Review of Financial Studies* 23, 3329–3345.

[30] Holmström, Bengt (1979): "Moral hazard and observability." *Bell Journal of Economics* 10, 74–91.

[31] Holmström, Bengt and Jean Tirole (1993): "Market liquidity and performance monitoring." *Journal of Political Economy* 101, 678–709.

[32] Holmström, Bengt (2017): "Pay for performance and beyond." *American Economic Review* 107, 1753–1777

[33] Innes, Robert D. (1990): "Limited liability and incentive contracting with ex-ante action choices." *Journal of Economic Theory* 52, 45–67.

[34] Jewitt, Ian (1988): "Justifying the first-order approach to principal–agent problems." *Econometrica* 56, 1177–1190.

[35] Jewitt, Ian, Ohad Kadan and Jeroen M. Swinkels (2008): "Moral hazard with bounded payments." *Journal of Economic Theory* 143, 59–82.

[36] Johnson, Shane A. and Yisong S. Tian (2000): "Indexed executive stock options." *Journal of Financial Economics* 57, 35–64.

[37] Kadan, Ohad and Jeroen M. Swinkels (2008): "Stocks or options? Moral hazard, firm viability, and the design of compensation contracts." *Review of Financial Studies* 21, 451–482.

[38] Kim, Son Ku (1995): "Efficiency of an information system in an agency model." *Econometrica* 63, 89–102.

[39] Liljeblom, Eva, Daniel Pasternack, and Matts Rosenberg (2011): "What determines stock option contract design?" *Journal of Financial Economics* 102, 293–316.

[40] Marinovic, Ivan and Felipe Varas (2019): "CEO horizon, optimal pay duration, and the escalation of short-termism." *Journal of Finance* 74, 2011–2053.

[41] Murphy, Kevin J. (2013): "Executive compensation: Where we are, and how we got there." *Handbook of the Economics of Finance Volume 2A: Corporate Finance*, edited by George M. Constantinides, Milton Harris, and Rene M. Stulz, 211–356. Elsevier, North-Holland.

[42] Oyer, Paul (2004): "Why do firms use incentives that have no incentive effects?" *Journal of Finance* 59, 1619–1649.

[43] Pastor, Lubos, Robert F. Stambaugh and Lucian A. Taylor (2020): "Sustainable investing in equilibrium." *Journal of Financial Economics*, forthcoming.

[44] Prendergast, Canice (2002): "The tenuous trade-off between risk and incentives." *Journal of Political Economy* 110, 1071–1102.

[45] Rogerson, William P. (1985): "The first order approach to principal–agent problems." *Econometrica* 53, 1357–1368.

[46] Townsend, Robert M. (1979): "Optimal contracts and competitive markets with costly state verification." *Journal of Economic Theory* 21, 265–293.

A Proofs

Proof of Lemma 1: For now we ignore the LL (5). By Lemma 4, when (31) holds, we can replace the IC in (4) by the first-order condition ("FOC"):

$$\sum_s \left[\frac{\partial \phi^s_{\hat{e}}}{\partial e} \int_0^{+\infty} u\left(\bar{W} + w_s\left(q\right)\right) f(q|\hat{e}, s) dq + \phi^s_{\hat{e}} \int_0^{+\infty} u\left(\bar{W} + w_s\left(q\right)\right) \frac{\partial f}{\partial e}(q|\hat{e}, s) dq \right] = C'(\hat{e}). \tag{12}$$

The FOC with respect to $w_s\left(q\right)$ in the program in (2), (3), and (4) is:

$$\phi^s_{\hat{e}} f(q|\hat{e}, s) - \lambda \phi^s_{\hat{e}} u'(\bar{W} + w_s\left(q\right)) f(q|\hat{e}, s) - \mu u'(\bar{W} + w_s\left(q\right)) \left[\frac{\partial \phi^s_{\hat{e}}}{\partial e} f(q|\hat{e}, s) + \phi^s_{\hat{e}} \frac{\partial f}{\partial e}(q|\hat{e}, s) \right] = 0$$

$$\Leftrightarrow \quad \frac{1}{u'(\bar{W} + w_s\left(q\right))} = \lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}} / \partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right]. \tag{13}$$

Due to LL, we have $\underline{m}(q) = \bar{W}$ and $\overline{m}(q) = \infty$, using the notation in Jewitt, Kadan, and Swinkels (2008). Using the FOC in (13), the reasoning in Proposition 1 of Jewitt, Kadan, and Swinkels (2008) applies for any given signal realization s, so that the optimal contract for a given s is defined implicitly by:

$$\frac{1}{u'(\bar{W} + w_s\left(q\right))} = \begin{cases} \lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}} / \partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] & \text{if } \lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}} / \partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \geq \frac{1}{u'(\bar{W})}, \\ \frac{1}{u'(\bar{W})} & \text{if } \lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}} / \partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] < \frac{1}{u'(\bar{W})}, \end{cases} \tag{14}$$

with $\lambda \geq 0$ and $\mu > 0$. This can be rewritten as equation (6). ■

Proof of Proposition 1: From equation (14), we have $w_s(q) > 0$ if and only if:

$$\lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}} / \partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] > \frac{1}{u'(\bar{W})}.$$

If the support of the output distribution is $[0, \infty)$ and if, for a given s,

$$\lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}} / \partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(0|\hat{e}, s)}{f(0|\hat{e}, s)} \right] > \frac{1}{u'(\bar{W})}, \tag{15}$$

then $q^*_s = 0$. Otherwise (if the support is $[0, \infty)$ and equation (15) does not hold, or if the support is instead $(-\infty, \infty)$), since $\frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)}$ is increasing in q by MLRP, for $u'(\bar{W}) > 0$ the threshold q_s^* is implicitly defined by:

$$\begin{aligned} \lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q_s^* | \hat{e}, s)}{f(q_s^* | \hat{e}, s)} \right] &= \frac{1}{u'(\bar{W})} \\ \Leftrightarrow \qquad \frac{\frac{\partial f}{\partial e}(q_s^* | \hat{e}, s)}{f(q_s^* | \hat{e}, s)} &= \frac{1}{\mu} \left(\frac{1}{u'(\bar{W})} - \lambda \right) - \frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} \end{aligned} \tag{16}$$

where the right-hand side ("RHS") is independent of q, and $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$ is increasing in q by MLRP.

Using equation (8) that defines the likelihood ratio of output, we have:

$$\frac{d}{dq} \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} = \frac{\zeta_s}{\sigma_s^2} G(q), \tag{17}$$

where

$$G(q) := -\frac{g''\left(\frac{q-\xi_s}{\sigma_s}\right) g\left(\frac{q-\xi_s}{\sigma_s}\right) - \left(g'\left(\frac{q-\xi_s}{\sigma_s}\right)\right)^2}{\left(g\left(\frac{q-\xi_s}{\sigma_s}\right)\right)^2}. \tag{18}$$

Due to MLRP, equations (17) and (18) imply that $G(q) > 0 \ \forall q$. From equation (8), since $\zeta_s > 0$, $\sigma_s > 0$, and $g(\cdot) > 0$, a distribution with location and scale parameters that satisfies MLRP is such that $g'(\cdot) > 0$ if q is lower than a threshold, and $g'(\cdot) < 0$ if q is higher than this threshold, i.e., the probability density function ("PDF") g is single peaked – the output corresponding to the peak is denoted by q_s^P.

For part (i), there are four cases to consider:

- If the support of the output distribution is $[0, \infty)$ and (15) holds for s_i and s_j, then $q_{s_i}^* = q_{s_j}^* = 0$.
- If the support is $[0, \infty)$ and (15) holds for s_i but not s_j, then $q_{s_i}^* = 0 < q_{s_j}^*$.
- If the support is $[0, \infty)$, it is impossible for (15) to hold for s_j but not s_i since $LR_{s_i}(q) > LR_{s_j}(q) \, \forall q$.
- If the support is $[0, \infty)$ and (15) holds for neither s_i nor for s_j, or if the support is instead $(-\infty, \infty)$, $q_{s_i}^*$ and $q_{s_j}^*$ are both described by equation (16). Since $\mu > 0$ and $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$ is increasing in q by MLRP, it follows from (16) that $\frac{\partial \phi_{\hat{e}}^{s_i}/\partial e}{\phi_{\hat{e}}^{s_i}} > \frac{\partial \phi_{\hat{e}}^{s_j}/\partial e}{\phi_{\hat{e}}^{s_j}}$ is associated with $q_{s_i}^* < q_{s_j}^*$, all else equal across signals.

For part (ii), there are the same four cases to consider. The first three cases are exactly the same as part (i). For the fourth case, $q^*_{s_i}$ and $q^*_{s_j}$ are both described by (16). For $\xi_{s_i} < \xi_{s_j}$, all else equal across signals we have $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s_i)}{f(q|\hat{e},s_i)} > \frac{\frac{\partial f}{\partial e}(q|\hat{e},s_j)}{f(q|\hat{e},s_j)}$ for any q. Then by MLRP and $\mu > 0$, it follows from (16) that $q^*_{s_i} < q^*_{s_j}$, all else equal across signals.

For part (iii), using equation (8) we have:

$$\frac{\partial \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}}{\partial \zeta_s} = \underbrace{-\frac{1}{\sigma_s}}_{\leq 0} \underbrace{\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)}}_{\geq 0 \text{ for } q \leq q^P_s} . \tag{19}$$

Changes in ζ_s do not change the peak of the distribution (such that $f'(q|\hat{e},s) = 0$), which is the same for s_i and s_j, and denoted by q^P_s. There are four cases to consider:

- If the support is $[0,\infty)$ and (15) holds for s_i and s_j, then $q^*_{s_i} = q^*_{s_j} = 0$.
- If the support is $[0,\infty)$ and (15) holds for s_i but not s_j, then $q^*_{s_i} = 0 < q^*_{s_j}$. Given MLRP and equation (19), this is possible for $q^*_s > q^P_s$ but not for $q^*_s < q^P_s$ $(s \in \{s_i, s_j\})$.
- If the support is $[0,\infty)$ and (15) holds for s_j but not s_i, then $q^*_{s_j} = 0 < q^*_{s_i}$. Given MLRP and (19), this is possible for $q^*_s < q^P_s$ but not for $q^*_s > q^P_s$ $(s \in \{s_i, s_j\})$.
- If the support is $[0,\infty)$ and (15) does not hold, or if the support is instead $(-\infty,\infty)$, $q^*_{s_i}$ and $q^*_{s_j}$ are both described by (16). Again, since $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$ is increasing in q by MLRP, a higher likelihood ratio of output implies a lower threshold q^*_s, as implicitly defined in (16). For $q^*_s \leq q^P_s$, the RHS of (19) is negative at $q = q^*_s$, so that a higher ζ_s is associated with a higher threshold q^*_s. For $q^*_s \geq q^P_s$, the RHS of (19) is positive at $q = q^*_s$, so that a higher ζ_s is associated with a lower threshold q^*_s.

For part (iv), using (8) we have:

$$\frac{\partial \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}}{\partial \sigma_s} = \frac{\zeta_s}{\sigma_s^2} \underbrace{\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)}}_{> 0 \text{ for } q < q^P_s} + \frac{\zeta_s}{\sigma_s} \underbrace{G(q)}_{\geq 0} \underbrace{\left(\frac{\xi_s - q}{\sigma_s^2}\right)}_{> 0 \text{ for } q < \xi_s} .$$

We also have $\zeta_s > 0$ and $\sigma_s > 0$, so that:

$$\text{sign}\left(\frac{\partial \frac{\frac{\partial f}{\partial e}(q_s^*|\hat{e},s)}{f(q_s^*|\hat{e},s)}}{\partial \sigma_s}\right)\begin{cases} > 0 & \text{if } q_s^* < q_s^P \text{ and } q_s^* < \xi_s \\ < 0 & \text{if } q_s^* > q_s^P \text{ and } q_s^* > \xi_s \end{cases}. \tag{20}$$

There are four cases to consider:

- If the support is $[0, \infty)$ and (15) holds for s_i and s_j, then $q_{s_i}^* = q_{s_j}^* = 0$.
- If the support is $[0, \infty)$ and (15) holds for s_i but not s_j, then $q_{s_i}^* = 0 < q_{s_j}^*$. Given MLRP and equation (20), this is possible for $q_s^* < q_s^P$ and $q_s^* < \xi_s$ but not for $q_s^* > q_s^P$ and $q_s^* > \xi_s$ ($s \in \{s_i, s_j\}$).
- If the support is $[0, \infty)$ and (15) holds for s_j but not s_i, then $q_{s_j}^* = 0 < q_{s_i}^*$. Given MLRP and (20), this is possible for $q_s^* > q_s^P$ and $q_s^* > \xi_s$ but not for $q_s^* < q_s^P$ and $q_s^* < \xi_s$ ($s \in \{s_i, s_j\}$).
- If the support is $[0, \infty)$ and (15) does not hold, or if the support is instead $(-\infty, \infty)$, $q_{s_i}^*$ and $q_{s_j}^*$ are both described by equation (16). Since $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$ is increasing in q by MLRP, a higher likelihood ratio of output implies a lower threshold q_s^*, as implicitly defined in (16); the relation between σ_s and the threshold is then given by (20). ■

Proof of Corollary 1: The threshold q_s^* is implicitly defined in (16). With $\mu > 0$, $u' > 0$, a non-binding IR ($\lambda = 0$) and a signal realization that is weakly bad news about effort ($\partial \phi_{\hat{e}}^s / \partial e \leq 0$), the threshold is such that:

$$\frac{\frac{\partial f}{\partial e}(q_s^*|\hat{e}, s)}{f(q_s^*|\hat{e}, s)} = \frac{1}{\mu}\frac{1}{u'(\bar{W})} - \frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} > 0$$

Since $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$ is increasing in q by MLRP, an increase in the likelihood ratio of output implies a lower threshold q_s^*. Using equation (20), we know that:

$$\text{sign}\left(\frac{dq_s^*}{d\sigma_s}\right)\begin{cases} \leq 0 & \text{if } q_s^* \leq q_s^P \text{ and } q_s^* \leq \xi_s \\ \geq 0 & \text{if } q_s^* \geq q_s^P \text{ and } q_s^* \geq \xi_s \end{cases} \tag{21}$$

For a distribution with location and scale parameters, $\lambda = 0$, and $\partial \phi_{\hat{e}}^s / \partial e \leq 0$, equation

(16) rewrites as:

$$-\frac{\zeta_s}{\sigma_s}\frac{g'\left(\frac{q_s^*-\xi_s}{\sigma_s}\right)}{g\left(\frac{q_s^*-\xi_s}{\sigma_s}\right)}=\frac{1}{\mu}\frac{1}{u'(\bar{W})}-\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}>0 \tag{22}$$

The peak of the output distribution, q_s^P, is defined implicitly by:

$$\frac{1}{\sigma_s^2}g'\left(\frac{q_s^P-\xi_s}{\sigma_s}\right)=0. \tag{23}$$

For a single-peaked distribution, equations (22) and (23) and MLRP imply that $q_s^* > q_s^P$. In this case, we know from (21) that q_s^* is increasing in σ_s if $q_s^* \geq \xi_s$.

Under a skew-normal distribution for q with location parameter $h_s(e)$, scale parameter σ_s, and shape parameter α_s, the likelihood ratio of output is:

$$\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}=\frac{1}{\sigma_s}\left(\frac{q-\xi_s}{\sigma_s}-\alpha_s\frac{\varphi\left(\alpha_s\frac{q-\xi_s}{\sigma_s}\right)}{\Phi\left(\alpha_s\frac{q-\xi_s}{\sigma_s}\right)}\right), \tag{24}$$

In addition, we know from (8) and (23) that the likelihood ratio of output is zero at the peak q_s^P. Using equation (24) with $\alpha_s \geq 0$, this implies that $q_s^P \geq \xi_s$, with equality for $\alpha_s = 0$. In sum, with a skew-normal distribution, we have $q_s^P \geq \xi_s$ if and only if the distribution has nonnegative skewness. In this case, $q_s^* > q_s^P$ implies $q_s^* > \xi_s$. Thus, q_s^* is increasing in σ_s. ■

Proof of Lemma 2: We start by characterizing the optimal contract that induces effort $\hat{e}$. For effort e, we have $q \sim \mathcal{N}(h_s(e), \sigma_s^2)$. Let φ_s be the PDF of the normal distribution with mean zero and standard deviation σ_s.

Letting $W_s(q) = w_s(q) + \bar{W}$ to simplify notation, the IC is:

$$\hat{e} \in \arg\max_e \sum_s \phi_e^s \int \ln\left(W_s(q)\right)\varphi_s\left(q - h_s(e)\right)dq - C(e).$$

The IR and LL are, respectively:

$$\sum_s \phi_{\hat{e}}^s \int \ln\left(W_s(q)\right)\varphi_s\left(q - h_s(\hat{e})\right)dq - C(\hat{e}) \geq \overline{u},$$

and

$$W_s(q) \geq \bar{W} \;\forall q, s.$$

To simplify notation, we will work with the manager's indirect utility, $u_s(q) := \ln W_s(q)$, so that $W_s(q) = \exp[u_s(q)]$. This step is without loss of generality. The next step, which is not without loss of generality, is to replace the IC by its FOC:

$$\sum_s \frac{\partial \phi^s_{\hat{e}}}{\partial e} \int u_s(q) \varphi_s(q-\xi_s)\, dq + \sum_s \phi^s_{\hat{e}} \int u_s(q) \zeta_s \frac{q-\xi_s}{\sigma_s^2} \varphi_s(q-\xi_s)\, dq - C'(\hat{e}) \begin{cases} \geq 0 & \text{if } \hat{e} = \bar{e} \\ = 0 & \text{if } \hat{e} \in (0, \bar{e}) \end{cases} .$$

This FOC rewrites as:

$$\sum_s \int u_s(q) \varphi_s(q-\xi_s) \left[\frac{\partial \phi^s_{\hat{e}}}{\partial e} + \phi^s_{\hat{e}} \zeta_s \frac{q-\xi_s}{\sigma_s^2} \right] dq - C'(\hat{e}) \begin{cases} \geq 0 & \text{if } \hat{e} = \bar{e} \\ = 0 & \text{if } \hat{e} \in (0, \bar{e}) \end{cases} . \tag{25}$$

Since replacing the IC by its FOC is not always valid, after solving the firm's relaxed program, we will need to verify that its solution satisfies the IC. The principal's *relaxed program* is:

$$\max_{u_s(\cdot)} \sum_s \phi^s_{\hat{e}} \int (q - \exp[u_s(q)]) \varphi_s(q-\xi_s)\, dq$$

subject to (25),

$$\sum_s \phi^s_{\hat{e}} \int u_s(q) \varphi_s(q-\xi_s)\, dq - C(\hat{e}) \geq \overline{u}, \tag{26}$$

and

$$u_s(q) \geq \ln \bar{W} \ \forall q, s.$$

Lemma 3 solves this optimization problem.

Lemma 3 *The optimal contract that implements $\hat{e} > 0$ in the relaxed program satisfies:*

$$W_s(q) = \begin{cases} \bar{W} & \text{for } q \leq \xi_s + \frac{\sigma_s^2}{\zeta_s} \left[\frac{\bar{W}-\lambda}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} \right] \\ \lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q-\xi_s}{\sigma_s^2} \right] & \text{for } q > \xi_s + \frac{\sigma_s^2}{\zeta_s} \left[\frac{\bar{W}-\lambda}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} \right] \end{cases},$$

where $\lambda \geq 0$ and $\mu > 0$.

Proof. The relaxed program maximizes a strictly concave function subject to linear constraints, so the FOC below, the complementary slackness conditions and the constraints are necessary and sufficient. Pointwise optimization gives:

$$\begin{aligned} &- \exp[u_s(q)] \phi^s_{\hat{e}} \varphi_s(q-\xi_s) + \mu \left[\frac{\partial \phi^s_{\hat{e}}}{\partial e} + \phi^s_{\hat{e}} \zeta_s \frac{q-\xi_s}{\sigma_s^2} \right] \varphi_s(q-\xi_s) \\ &+ \lambda \phi^s_{\hat{e}} \varphi_s(q-\xi_s) + \lambda_{LL}(q,s) = 0, \end{aligned}$$

where $\lambda_{LL}(q,s)$ are the multipliers associated with the LL. Letting $\tilde{\lambda}_{LL}(q,s) \equiv \frac{\lambda_{LL}(q,s)}{\phi^s_{\hat{e}}\varphi_s(q-\xi_s)} \geq 0$, we can rewrite the FOC as:

$$W_s(q) = \lambda + \tilde{\lambda}_{LL}(q,s) + \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q-\xi_s}{\sigma^2_s} \right]. \tag{27}$$

There are two cases to consider. First, if $\lambda = 0$ in the optimal contract, it can be verified that the following solves the necessary and sufficient optimality conditions:

$$W_s(q) = \begin{cases} \bar{W} & \text{for } q \leq \xi_s + \frac{\sigma^2_s}{\zeta_s}\left(\frac{\bar{W}}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} \right) \\ \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q-\xi_s}{\sigma^2_s} \right] & \text{for } q > \xi_s + \frac{\sigma^2_s}{\zeta_s}\left(\frac{\bar{W}}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} \right) \end{cases}, \tag{28}$$

where μ is chosen so that the IC holds (it can be shown that such $\mu > 0$ exists and is unique). To see this, note that when the LL binds, we have $W_s(q) = \bar{W}$ and $\tilde{\lambda}_{LL}(q,s) \geq 0$. Then, the FOC becomes:

$$\tilde{\lambda}_{LL}(q,s) = \bar{W} - \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q-\xi_s}{\sigma^2_s} \right] - \lambda = \bar{W} - \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q-\xi_s}{\sigma^2_s} \right] \geq 0,$$

which is positive because $q \leq \xi_s + \frac{\sigma^2_s}{\zeta_s}\left(\frac{\bar{W}}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} \right)$. When the LL does not bind ($W_s(q) > \bar{W}$), we have $\tilde{\lambda}_{LL}(q,s) = 0$, so that the FOC becomes:

$$W_s(q) = \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q-\xi_s}{\sigma^2_s} \right],$$

where μ is chosen so that the IC holds. If the resulting contract satisfies the IR in (26), then indeed $\lambda = 0$ at the optimal contract, which is described in (28). This establishes that an option with state-contingent strike price $\xi_s + \frac{\sigma^2_s}{\zeta_s}\left(\frac{\bar{W}}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} \right)$ and slope $n^*_s := \mu \frac{\zeta_s}{\sigma^2_s}$ solves the relaxed program when we are implementing $\hat{e} > 0$.

Now suppose that the resulting contract does not satisfy IR in (26). Then we have $\lambda > 0$. It can be verified that the following solves the necessary and sufficient optimality conditions:

$$W_s(q) = \begin{cases} \bar{W} & \text{for } q \leq \xi_s + \frac{\sigma^2_s}{\zeta_s}\left(\frac{\bar{W}-\lambda}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} \right) \\ \lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q-\xi_s}{\sigma^2_s} \right] & \text{for } q > \xi_s + \frac{\sigma^2_s}{\zeta_s}\left(\frac{\bar{W}-\lambda}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} \right) \end{cases},$$

where λ and μ are chosen so that the IR and IC hold. To see this, note that when the LL binds, we have $W_s(q) = \bar{W}$ and $\tilde{\lambda}_{LL}(q,s) \geq 0$. Then, the FOC becomes:

$$\tilde{\lambda}_{LL}(q,s) = \bar{W} - \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q - \xi_s}{\sigma_s^2} \right] - \lambda \geq 0,$$

which is positive because $q \leq \xi_s + \frac{\sigma_s^2}{\zeta_s}\left(\frac{\bar{W}-\lambda}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}}\right)$. When the LL does not bind ($W_s(q) > \bar{W}$), we have $\tilde{\lambda}_{LL}(q,s) = 0$, so that the FOC becomes:

$$W_s(q) = \lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q - \xi_s}{\sigma_s^2} \right].$$

■

This establishes that an option with state-contingent strike price $q_s^* = \xi_s + \frac{\sigma_s^2}{\zeta_s}\left(\frac{\bar{W}-\lambda}{\mu} - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}}\right)$ and slope $n_s^* := \mu \frac{\zeta_s}{\sigma_s^2}$ solves the relaxed program when we are implementing $\hat{e} > 0$. Let $K := \frac{\bar{W}-\lambda}{\mu}$, which is independent from q and s. ■

Proof of Proposition 2: We rely on the proof of Lemma 2.

For part (i), the slope of the option contract is $n_s^* := \mu \frac{\zeta_s}{\sigma_s^2}$.

For part (ii), we can write the optimal contract as:

$$w_s(q) = \max \left\{ \lambda + \mu \left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \zeta_s \frac{q - \xi_s}{\sigma_s^2} \right] - \bar{W}, 0 \right\}.$$

The strike price q_s^* is such that $w_s(q) > 0$ if and only if $q \geq q_s^*$, as derived in the proof of Lemma 2. With $K = \frac{\bar{W}-\lambda}{\mu}$, we have $q_s^* = \xi_s + \frac{\sigma_s^2}{\zeta_s}\left(K - \frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}}\right)$. ■

B The First-Order Approach With Limited Liability

Prior literature has developed conditions to justify the FOA, such as Rogerson (1985) and Jewitt (1988). However, using the FOA in models with limited liability is not straightforward.

First, some conditions rely on the optimal contract derived in the absence of contracting constraints (Jewitt (1988), Kim and Jung (2015)), and are thus not applicable under limited liability. Second, Rogerson's (1985) condition on the convexity of the cumulative distribution function, which is generalized to the multi-signal case by Sinclair-Desgagné (1994), does not rely on the contract and thus can be used with contracting constraints (e.g., Kadan and Swinkels (2008), Jewitt, Kadan, and Swinkels (2008)). However, it imposes a strong restriction on the probability distribution of the performance measure, which is not satisfied by probability distributions commonly used to model stock returns, such as the normal, gamma, and lognormal (Hemmer (2004)). Third, an alternative approach, used by Dittmann and Maug (2007) in calibration and Kirkegaard (2017), first derives the optimal contract assuming the FOA is valid, and then imposes conditions based on this contract that guarantee the validity of the FOA. Since the optimal contract generally cannot be characterized in closed form, it is hard to obtain analytical conditions for the validity of the FOA.

We present a new condition for the validity of the FOA under limited liability for utility functions that are bounded from above, which is the case for many common utility functions, such as constant absolute risk aversion and constant relative risk aversion ("CRRA") with a relative risk aversion γ larger than 1. The condition uses limited liability to obtain a lower bound on the manager's utility which, combined with the upper bound, allows us to specify conditions that rule out profitable non-local deviations. Crucially, these conditions do not depend on the (endogenous) contract, nor do they require the strong convexity condition in Rogerson (1985).

Let K_e^+ and K_e^- denote the integral of the positive and negative parts of the second derivative of the joint distribution $f(q,s|e)$ with respect to effort:

$$K_e^+ \;:=\; \sum_s \int_{\underline{q}}^{+\infty} \max\left\{\frac{\partial^2 f}{\partial e^2}(q,s|e),\ 0\right\} dq, \tag{29}$$

$$K_e^- \;:=\; \sum_s \int_{\underline{q}}^{+\infty} \min\left\{\frac{\partial^2 f}{\partial e^2}(q,s|e),\ 0\right\} dq. \tag{30}$$

Lemma 4, proven in Supplementary Appendix E, provides a sufficient condition for the FOA to be valid.

Lemma 4 *(First-Order Approach): Suppose that*

$$K_e^- u(\bar{W}) + K_e^+ \lim_{c \nearrow \infty} u(c) < C''(e) \tag{31}$$

for all $e \in (0, \bar{e})$. Then, the FOA is valid.

The intuition for the condition in Lemma 4 is as follows. For the FOA to be valid, we must ensure that the manager's objective (expected utility of payments minus effort costs) is a concave function of effort. In general, the effect of effort on the expected utility of payments may be sufficiently locally convex that it offsets the impact of a strictly convex effort cost. We therefore need to provide an upper bound on the convexity of the (endogenous) payments.

With limited liability, $u(\bar{W})$ is a lower bound on the manager's utility, and by monotonicity, $\lim_{c\nearrow\infty} u(c)$ is an upper bound on the manager's utility. Lemma 4 uses these bounds to limit the convexity in the manager's expected utility.

We close by showing how researchers can apply the condition in Lemma 4 to special cases. Example 2 below considers a quadratic effort cost.

Example 2 *Let $\bar{W} = 10$, $u(w) = \frac{w^{-2}}{-2}$ (i.e., the manager has CRRA utility $\gamma = 3$), the cost of effort is $C(e) = \alpha e + \frac{\beta}{2}e^2$, with $\alpha \geq 0$ and $\beta > 0$, the set of possible effort levels is $e \in [0, 10]$, output follows a normal distribution with mean e and standard deviation 2, $S = 1$, and the manager is protected by limited liability. Then we have: $\int_{\underline{q}}^{\infty} \min\left\{\frac{\partial^2 f}{\partial e^2}(q|e,s), 0\right\} dq \approx -0.121$ for all e, $u(\bar{W}) = -\frac{1}{200}$, $C''(e) = \beta$, and the condition in Lemma 4 is simply $\beta > \frac{0.121}{200}$.*

Equation (31) can be simplified in special cases. Since $K_e^- < 0$, it is easier to satisfy condition (31) when the manager's outside wealth $\bar{W}$ is higher. When $\lim_{c\nearrow+\infty} u(c) = 0$, as with constant absolute risk aversion and CRRA with $\gamma > 1$, (31) becomes

$$K_e^- u(\bar{W}) < C''(e)$$

for all $e \in (0, \bar{e})$.

Lemma 4 can also be applied to settings in which there is no additional signal. In that case, we have:

$$K_e^+ = \int_{\underline{q}}^{\infty} \max\left\{\frac{\partial^2 f}{\partial e^2}(q|e),\ 0\right\} dq \ \text{ and } \ K_e^- = \int_{\underline{q}}^{\infty} \min\left\{\frac{\partial^2 f}{\partial e^2}(q|e),\ 0\right\} dq.$$

C Determinants of the Threshold

Without loss of generality, let an increase in c parametrize an increase in the marginal cost of effort that leaves the equilibrium cost of effort $C(\hat{e})$ unchanged. Suppose that $\overline{u} \leq u\left(\bar{W}\right) - C(\hat{e})$, i.e., the IR is nonbinding and $\lambda = 0$. Assuming that the FOA holds, plugging the optimal contract (6) into the IC (4) yields:

$$\sum_s \left[\int_{\underline{q}}^{q_s^*} u\left(\bar{W}\right) \frac{\partial f}{\partial e}(q,s|\hat{e}) dq + \int_{q_s^*}^{\infty} u\left(u'^{-1}\left(1 \Big/ \mu \frac{\frac{\partial f}{\partial e}(q,s|\hat{e})}{f(q,s|\hat{e})}\right)\right) \frac{\partial f}{\partial e}(q,s|\hat{e}) dq\right] = C'(\hat{e}). \tag{32}$$

From equation (16) that describes the thresholds, with $\lambda = 0$:

$$LR_s(q_s^*|\hat{e}) \;=\; \frac{\frac{\partial f}{\partial e}(q_s^*, s|\hat{e})}{f(q_s^*, s|\hat{e})} = \frac{1}{\mu}\frac{1}{u'(\bar{W})} > 0. \tag{33}$$

This implies that $\frac{\partial f}{\partial e}(q, s|\hat{e}) > 0 \;\forall q > q_s^*$ due to MLRP, so that the derivative of the left-hand side ("LHS") of the IC in equation (32) with respect to μ is positive (recall that, by construction, $u'^{-1}\left(1/\mu LR_s(q_s^*|\hat{e})\right) = \bar{W}$). To maintain incentive compatibility following an increase in c that raises the RHS of (32), the Lagrange multiplier μ must therefore rise for the LHS of (32) to increase correspondingly. In turn, using MLRP, $u' > 0$, and equation (33), a higher μ implies a lower q_s^* for any s. In sum, when IR does not bind, a higher c results in a lower q_s^* for any s. Since the likelihood ratio is positive at the threshold (see (33)), an increase in c brings the threshold q_s^* closer to the point where the likelihood ratio $LR_s(q|\hat{e})$ is zero. Intuitively, a higher marginal cost of effort requires the principal to offer higher incentives, which she ensures by lowering the threshold.

D Performance Signals and Pay-Performance Sensitivity

This section studies how the signal realization affects pay-performance sensitivity ("PPS") above the threshold pay-performance, generalizing Section 2.2 beyond piecewise linear contracts. We return to the main model of Section 1 and no longer require log utility or normally-distributed output. The only additional assumption that we impose is CRRA utility. This is because the curvature of the utility function plays an important role in the slope of the contract, and CRRA utility allows us to capture this curvature with a single parameter, γ. Moreover, CRRA is widely used for executive pay, in particular for calibration (see, e.g., Dittmann and Maug (2007) and references cited therein).

We define PPS as $\frac{w_s(q)-w_s(q_0)}{q-q_0}$ for $q > q_0 > \max\{q_s^P, \xi_s, q_s^*\}$ for $s \in \{s_i, s_j\}$. It represents the slope of the contract between any two outputs q and q_0 where the payment is strictly positive (since q_0 is above the threshold). Proposition 3, proven in Supplementary Appendix F, studies how PPS depends on the signal realization.

Proposition 3 *(Effect of signal on pay-performance sensitivity) Suppose that the FOA holds, and consider outputs q and q_0 such that $q > q_0 > \max\{q_s^P, \xi_s, q_s^*\}$ for $s \in \{s_i, s_j\}$. All else equal across signals:*

(i) If signal realization s_i is individually more indicative of effort than s_j ($\frac{\partial\phi_{\hat{e}}^{s_i}/\partial e}{\phi_{\hat{e}}^{s_i}} > \frac{\partial\phi_{\hat{e}}^{s_j}/\partial e}{\phi_{\hat{e}}^{s_j}}$), PPS is higher (lower) under s_i if $\gamma < (>)1$.

(ii) If signal realization s_i is associated with a higher equilibrium location parameter than s_j ($\xi_i > \xi_j$), PPS is higher (lower) under s_i if the likelihood ratio of output is weakly concave and $\gamma \geq 1$ (weakly convex and $\gamma \leq 1$).

(iii) If signal realization s_i is associated with a higher impact parameter than s_j ($\zeta_i > \zeta_j$), PPS is higher under s_i if $\gamma \leq 1$.

(iv) If signal realization s_i is associated with a higher scale parameter than s_j ($\sigma_i > \sigma_j$), PPS is lower under s_i than s_j if the likelihood ratio of output is weakly convex and $\gamma \leq 1$.

To understand how a signal realization affects PPS, consider the local PPS at output q and signal s for $w_s(q) > 0$:

$$w_s'(q) = \underbrace{\frac{\mu}{\gamma} \frac{\partial}{\partial q} \left\{ \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right\}}_{\text{A}} \left(\underbrace{\lambda + \mu \left[\frac{\partial\phi_{\hat{e}}^{s}/\partial e}{\phi_{\hat{e}}^{s}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right]}_{\text{B}} \right)^{\frac{1}{\gamma}-1} . \tag{34}$$

A signal can affect PPS by affecting term A or B in equation (34). Term A, which is positive by MLRP, is the slope of the likelihood ratio of output. A signal may affect this slope in two ways. The first is a location effect: the likelihood ratio of output under s_i is a rightward slope of the likelihood ratio under signal s_j. When the likelihood ratio is concave (convex) in output, the location effect increases (decreases) PPS at a given output level. The second is a precision effect: the likelihood ratio is steeper under s_i than s_j, i.e. higher for high output and lower for low output. The precision effect increases PPS regardless of the curvature of the likelihood ratio – it concentrates incentives in states of the world (signal realizations) where output is more informative about effort. (Note that, unlike in the main paper, here the location and precision effects refer to the impact on PPS rather than the threshold.)

Term B, which is positive (see equation (37) in Supplementary Appendix F), is a linear transformation of the likelihood ratio. It captures the *contract curvature effect*: if the signal realization affects the payment $w_s(q)$, this in turn affects PPS if the payment above the threshold is nonlinear in term B. When $\gamma < 1$ ($\gamma > 1$), the payment above the threshold is convex (concave) in term B. Intuitively, γ affects not only risk aversion but also prudence – downside risk aversion, which depends on the third derivative of the utility function. When $\gamma < 1$, the effect of prudence (which favors convex contracts as they protect the agent from downside risk) dominates the effect of risk aversion (which favors concave contracts), so that the optimal contract is convex in a linear transformation of the likelihood ratio, i.e. term B.[12] When prudence is high, it is efficient to concentrate rewards on very positive outcomes rather than moderately positive outcomes. If the payment above the threshold is convex (concave) in term B, then if a signal realization increases the payment at a given output, PPS increases (decreases). When $\gamma = 1$ (as in Section 2.2), the contract is linear in term B, so changes to the payment do not affect PPS.

The likelihood ratio of the signal affects only term B; the likelihood ratio of output affects both terms. Thus, in part (i) of Proposition 3, which concerns the likelihood ratio of the signal, only the contract curvature effect applies. Signals that are individually indicative of effort increase pay; if $\gamma < 1$, the contract is convex in term B, and so this increase in pay also increases PPS. The effect is reversed if $\gamma > 1$.

In part (ii), which concerns the likelihood ratio of output, both the location and contract curvature effects apply. Starting with the former, if $\xi_i > \xi_j$, the output distribution under s_i is shifted to the right. If the likelihood ratio is concave in output, a higher location of the output distribution means a greater slope of the likelihood ratio at any given output level, which increases PPS. Moving to the latter, at any given output, the payment is lower under s_i than s_j. When $\gamma \geq 1$ ($\gamma \leq 1$), the payment is concave (convex) in a linear transformation of the likelihood ratio, which means a higher (lower) PPS under s_i.

In parts (iii) and (iv), both the precision and contract curvature effects apply. In part (iii), if $\zeta_i > \zeta_j$, output is more informative about effort under s_i and so PPS is higher (the precision effect). In addition, above the peak of the distribution, a larger impact parameter leads to a higher likelihood ratio and therefore a higher payment, which weakly increases PPS if and only if $\gamma \leq 1$ (the contract curvature effect). Thus, if $\gamma \leq 1$, both effects go in the same direction and so PPS is unambiguously higher. In part (iv), if $\sigma_i > \sigma_j$, output is less informative about effort under s_i and so PPS is lower (the precision effect). In addition, above the peak and location parameter of the distribution, a more volatile output leads to a lower likelihood ratio and therefore a lower payment, which weakly decreases PPS if and only if $\gamma \leq 1$ (the contract curvature effect).[13]

[12]For a formal result, see Proposition 1 and Claim 4 in Chaigneau, Sahuguet, and Sinclair-Desgagné (2017).

[13]There is a subtle distinction between the informativeness effects for the impact and scale parameters, for the same reason that the impact and scale parameters do not always have opposite effects on the threshold in Proposition 1. The impact parameter affects the slope of the likelihood ratio but not the equilibrium output distribution. In contrast, the scale parameter affects the slope of the likelihood ratio and also spreads out the equilibrium output distribution. Thus, a higher scale parameter results in a lower likelihood ratio for high output levels, which further diminishes the sensitivity of the likelihood ratio to output (and thus PPS) if the likelihood ratio of output is convex – reinforcing the effects described in the main text. This explains why part (iv) also requires the likelihood ratio of output to be weakly convex whereas part (iii) does not.

References

[1] Chaigneau, Pierre, Nicolas Sahuguet, and Bernard Sinclair-Desgagné (2017): "Prudence and the convexity of compensation contracts." *Economics Letters* 157, 14–16.

[2] Dittmann, Ingolf and Ernst Maug (2007): "Lower salaries and no options? On the optimal structure of executive pay." *Journal of Finance* 62, 303–343.

[3] Hemmer, Thomas, Oliver Kim and Robert E. Verrecchia (1999): "Introducing convexity into optimal compensation contracts." *Journal of Accounting and Economics* 28, 307–328.

[4] Innes, Robert D. (1993): "Debt, futures and options: Optimal price-linked financial contracts under moral hazard and limited liability." *International Economic Review* 34, 271–295.

[5] Jung, Jin Yong and Son Ku Kim (2015): "Information space conditions for the first-order approach in agency problems." *Journal of Economic Theory* 160, 243–279.

[6] Kirkegaard, René (2017): "A unifying approach to incentive compatibility in moral hazard problems. " *Theoretical Economics* 12, 25–51.

[7] Matthews, Steven A. (2001): "Renegotiating moral hazard contracts under limited liability and monotonicity." *Journal of Economic Theory* 97, 1–29.

[8] Sinclair-Desgagné, Bernard (1994): "The first-order approach to multi-signal principal-agent problems." *Econometrica* 62, 459–466.

Supplementary Appendix for
"How Should Performance Signals Affect Contracts?"

E Proof of Lemma 4

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

$$\max_e \mathbb{E}\left[u\left(\bar{W}+w_s(q)\right)\right]-C(e) \;\Leftrightarrow\; \max_e \sum_s \int_{\underline{q}}^{+\infty} u\left(\bar{W}+w_s(q)\right) f(q,s|e)dq - C(e)$$

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

$$\sum_s \int_{\underline{q}}^{+\infty} u\left(\bar{W}+w_s(q)\right) \frac{\partial^2 f(q,s|e)}{\partial e^2} dq < C''(e) \quad \forall e \in (0,\bar{e}). \tag{35}$$

Assume that u is bounded from above, with $\lim_{w\to\infty} u(w) \equiv u^+$. In addition, with limited liability, the minimum payment is $w_s(q)=0$; with an increasing utility function, this implies that the minimum value of u is $u(\bar{W})$. Therefore, for any $\{q,s\}$:

$$u\left(\bar{W}+w_s(q)\right) \in \left[u(\bar{W}), u^+\right]$$

Using notations K_e^+ and K_e^- defined in equations (29) and (30), the expression on the LHS of equation (35) can then be rewritten as:

$$\sum_s \int_{\underline{q}}^{+\infty} u\left(\bar{W}+w_s(q)\right) \min\left\{\frac{\partial^2 f(q,s|e)}{\partial e^2},0\right\}dq + \sum_s \int_{\underline{q}}^{+\infty} u\left(\bar{W}+w_s(q)\right) \max\left\{\frac{\partial^2 f(q,s|e)}{\partial e^2},0\right\}dq \tag{36}$$

As established above, we have $u\left(\bar{W}+w_s(q)\right) \geq u(\bar{W})$ for any q,s, and $u\left(\bar{W}+w_s(q)\right) \leq u^+$ for any q,s. Therefore, for any q,s such that $\frac{\partial^2 f(q,s|e)}{\partial e^2} \leq 0$ we have $u\left(\bar{W}+w_s(q)\right)\frac{\partial^2 f(q,s|e)}{\partial e^2} \leq u(\bar{W})\frac{\partial^2 f(q,s|e)}{\partial e^2}$; and for any q,s such that $\frac{\partial^2 f(q,s|e)}{\partial e^2} \geq 0$ we have $u\left(\bar{W}+w_s(q)\right)\frac{\partial^2 f(q,s|e)}{\partial e^2} \leq u^+\frac{\partial^2 f(q,s|e)}{\partial e^2}$. Integrating and summing over q and s, this implies that expression (36) is less than $K_e^- u(\bar{W}) + K_e^+ u^+$, which completes the proof. ■

F Proof of Proposition 3

From equation (8), since $\zeta_s > 0$, $\sigma_s > 0$, and $g(\cdot) > 0$, a distribution with location and scale parameters that satisfies MLRP is such that $g'(\cdot) > 0$ if q is lower than a threshold, and $g'(\cdot) < 0$ if q is higher than this threshold, i.e., the PDF g is single peaked – the output corresponding to the peak is denoted by q_s^P. For symmetric distributions, the single peak of the distribution, which is such that $g'\left(\frac{q_s^P - \xi_s}{\sigma_s}\right) = 0$, is at $q_s^P = \xi_s$ for $s \in \{s_i, s_j\}$. With CRRA utility, characterized by $u'(w) = w^{-\gamma}$ and $u'^{-1}(w) = w^{-\frac{1}{\gamma}}$, provided that the FOA holds we have:

$$w_s(q) = \begin{cases} \left(\lambda + \mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}} - \bar{W} & \text{if } \lambda + \mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right] \geq \frac{1}{u'(\bar{W})} \\ 0 & \text{if } \lambda + \mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right] < \frac{1}{u'(\bar{W})} \end{cases}. \tag{37}$$

For $\gamma > 1$, we use the condition for the FOA in Lemma 4. For $\gamma \leq 1$, we derive a condition for FOA to hold in this setting. The FOA holds if:

$$\sum_s \int_{\underline{q}}^{+\infty} u\left(\bar{W} + w_s(q)\right) \frac{\partial^2 f(q,s|e)}{\partial e^2} dq < C''(e) \quad \forall e \in (0, \bar{e}), \tag{38}$$

where $u(w) = \frac{w^{1-\gamma}}{1-\gamma}$ if $\gamma < 1$ and $\ln w$ if $\gamma = 1$, and $w_s\left(q\right)$ is defined by equation (37).

Part (i): Suppose that signals s_i and s_j differ only in their individual informativeness: $\frac{\partial\phi_{\hat{e}}^{s_i}/\partial e}{\phi_{\hat{e}}^{s_i}} > \frac{\partial\phi_{\hat{e}}^{s_j}/\partial e}{\phi_{\hat{e}}^{s_j}}$ (s_i and s_j are associated with the same output distribution). For notational convenience, let $\tilde{\phi}_s \equiv \frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}$ and $w_s(q) \equiv W_q(\tilde{\phi}_s)$, which is a continuous function of $\tilde{\phi}_s$. For a given q, we can write:

$$w_{s_i}(q) - w_{s_j}(q) = W_q(\tilde{\phi}_i) - W_q(\tilde{\phi}_j) = \int_{\tilde{\phi}_j}^{\tilde{\phi}_i} \frac{\partial W_q(\tilde{\phi})}{\partial\tilde{\phi}} d\tilde{\phi}.$$

Holding all else constant including Lagrange multipliers (we are comparing two signal realizations, i.e. we do not change parameters of the contracting environment), for given q and s such that $w_s(q) > 0$:

$$\frac{\partial W_q(\tilde{\phi})}{\partial \tilde{\phi}} = \frac{\partial}{\partial \tilde{\phi}} \left\{ \left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma}} - \bar{W} \right\}$$

$$= \underbrace{\frac{\mu}{\gamma}}_{> 0} \underbrace{\left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma} - 1}}_{\geq 0 \text{ since } w_s(q) \geq 0}, \tag{39}$$

For given q and q_0, we have

$$\frac{w_{s_i}(q) - w_{s_i}(q_0)}{q - q_0} \geq \frac{w_{s_j}(q) - w_{s_j}(q_0)}{q - q_0} \quad \Leftrightarrow \quad w_{s_i}(q) - w_{s_j}(q) - \big(w_{s_i}(q_0) - w_{s_j}(q_0)\big) \geq 0. \tag{40}$$

Thus, for given q and q_0 such that $q > q_0$ and $w_s(q_0) > 0$, we have:

$$\begin{aligned} & w_{s_i}(q) - w_{s_j}(q) - \big(w_{s_i}(q_0) - w_{s_j}(q_0)\big) \\ = \; & \frac{\mu}{\gamma} \int_{\tilde{\phi}_j}^{\tilde{\phi}_i} \left(\left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma} - 1} - \left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q_0|\hat{e}, s)}{f(q_0|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma} - 1} \right) d\tilde{\phi}. \end{aligned}$$

We have: $q > q_0$ which implies $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)} > \frac{\frac{\partial f}{\partial e}(q_0|\hat{e},s)}{f(q_0|\hat{e},s)}$ by MLRP, so that, for a given $\tilde{\phi}_s$:

$$\begin{aligned} \lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] &> \lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q_0|\hat{e}, s)}{f(q_0|\hat{e}, s)} \right] \\ \left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma} - 1} &> \left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q_0|\hat{e}, s)}{f(q_0|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma} - 1} \quad \text{iff } \gamma < 1. \end{aligned} \tag{41}$$

If $\gamma < 1$, the condition in (41) holds, and so (40) also holds. If $\gamma < 1$, (41) does not hold, and so (40) does not either. If $\gamma = 1$, the equation in (41) is satisfied as an equality so that the expression on the right in (40) is equal to zero.

Part (ii): Suppose that signals s_i and s_j differ only in their equilibrium location parameter, with $\xi_i > \xi_j$. Let $w_s(q) \equiv W_q(\xi_s)$, which is a continuous function of ξ_s, since $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$ is by assumption continuously differentiable in the equilibrium location parameter ξ_s. For a given q, we have:

$$w_{s_i}(q) - w_{s_j}(q) = W_q(\xi_i) - W_q(\xi_j) = \int_{\xi_j}^{\xi_i} \frac{\partial W_q(\xi)}{\partial \xi} d\xi.$$

Holding all else constant including Lagrange multipliers, for given q and s such that $w_s(q) > 0$, we have:

$$\begin{aligned}
\frac{\partial W_q(\xi)}{\partial \xi} &= \frac{\partial}{\partial \xi}\left\{\left(\lambda + \mu\left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}} - \bar{W}\right\} \\
&= \underbrace{\frac{\mu}{\gamma}}_{>0}\ \frac{\partial}{\partial \xi}\left\{\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right\}\underbrace{\left(\lambda + \mu\left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1}}_{\geq 0 \text{ since } w_s(q) \geq 0}, \qquad (42)
\end{aligned}$$

where:

$$\begin{aligned}
\frac{\partial}{\partial \xi}\left\{\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right\} &= \frac{\partial}{\partial \xi}\left\{-\frac{\zeta_s}{\sigma_s}\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)}\right\} \\
&= -\frac{\zeta_s}{\sigma_s^2}\underbrace{G(q)}_{\geq 0}, \qquad (43)
\end{aligned}$$

where G is defined in equation (18). For given q and q_0, we have:

$$\frac{w_{s_i}(q) - w_{s_i}(q_0)}{q - q_0} \geq \frac{w_{s_j}(q) - w_{s_j}(q_0)}{q - q_0} \quad \Leftrightarrow \quad w_{s_i}(q) - w_{s_j}(q) - \big(w_{s_i}(q_0) - w_{s_j}(q_0)\big) \geq 0. \qquad (44)$$

Thus, for given q and q_0 such that $q > q_0$ and $w_s(q_0) > 0$, we have:

$$\begin{aligned}
& w_{s_i}(q) - w_{s_j}(q) - \big(w_{s_i}(q_0) - w_{s_j}(q_0)\big) \\
= \ & \frac{\mu}{\gamma}\frac{\zeta_s}{\sigma_s^2}\int_{\xi_j}^{\xi_i}\Bigg(-G(q)\left(\lambda + \mu\left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1} \\
& + G(q_0)\left(\lambda + \mu\left[\frac{\partial \phi^s_{\hat{e}}/\partial e}{\phi^s_{\hat{e}}} + \frac{\frac{\partial f}{\partial e}(q_0|\hat{e},s)}{f(q_0|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1}\Bigg)d\xi. \qquad (45)
\end{aligned}$$

From equation (17) and the definition of $G(q)$ in (18), the likelihood ratio of output is weakly concave in q if and only if $G'(q) \leq 0$ in which case $0 \leq G(q) \leq G(q_0)$ since $q > q_0$. In addition, $q > q_0$ implies $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)} > \frac{\frac{\partial f}{\partial e}(q_0|\hat{e},s)}{f(q_0|\hat{e},s)}$ by MLRP, so that, for a weakly concave likelihood ratio and

$\gamma \geq 1$:

$$\begin{aligned}
\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right] &> \lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q_0|\hat{e},s)}{f(q_0|\hat{e},s)}\right] \\
\left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1} &\leq \left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q_0|\hat{e},s)}{f(q_0|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1} \\
G(q)\left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1} &\leq G(q_0)\left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q_0|\hat{e},s)}{f(q_0|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1} \quad (46)
\end{aligned}$$

We conclude that if the likelihood ratio is nonconvex and $\gamma \geq 1$, then (46) holds and, using equation (45), (44) holds too. Symmetrically, if the likelihood ratio is weakly convex (so that $G'(q) \geq 0$) and $\gamma \leq 1$, then the inequality in (46) is reversed, so that it is reversed in condition (44) too. Finally, if the likelihood ratio is linear (so that $G'(q) = 0$) and $\gamma = 1$, then (46) holds as an equality, and so does condition (44).

Part (iii): Suppose that signals s_i and s_j differ only in their impact parameter, with $\zeta_i > \zeta_j$. Let $w_s(q) \equiv W_q(\zeta_s)$, which is a continuous function of ζ_s, since $\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}$ is by assumption continuously differentiable in the parameter ζ_s. For a given q, we have:

$$w_{s_i}(q) - w_{s_j}(q) = W_q(\zeta_i) - W_q(\zeta_j) = \int_{\zeta_j}^{\zeta_i} \frac{\partial W_q(\zeta)}{\partial \zeta} d\zeta. \quad (47)$$

As above, holding all else constant including Lagrange multipliers, for given q and s such that $w_s(q) > 0$:

$$\begin{aligned}
\frac{\partial W_q(\zeta)}{\partial\zeta} &= \frac{\partial}{\partial\zeta}\left\{\left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}} - \bar{W}\right\} \\
&= \frac{\mu}{\gamma}\frac{\partial}{\partial\zeta}\left\{\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right\}\underbrace{\left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1}}_{>0 \text{ since } w_s(q)>0}, \quad (48)
\end{aligned}$$

where

$$\frac{\partial}{\partial\zeta}\left\{\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right\} = \frac{\partial}{\partial\zeta}\left\{-\frac{\zeta_s}{\sigma_s}\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)}\right\} = -\underbrace{\frac{1}{\sigma_s}}_{<0}\underbrace{\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)}}_{<0 \text{ for } q>q_s^P}. \quad (49)$$

Thus, at output q, we have:

$$\int_{\zeta_j}^{\zeta_i} \frac{\partial W_q(\zeta)}{\partial \zeta} d\zeta = \int_{\zeta_j}^{\zeta_i} \frac{\mu}{\gamma} \frac{1}{\sigma_s} \underbrace{\left(-\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)} \right)}_{>0 \text{ for } q > q_s^P} \underbrace{\left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma}-1}}_{>0 \text{ since } w_s(q)>0} d\zeta.$$

In sum, with $q > q_0 > \max\{q_s^P, q_s^*, \xi_s\}$, we have:

$$\begin{aligned} & w_{s_i}(q) - w_{s_j}(q) - \left[w_{s_i}(q_0) - w_{s_j}(q_0) \right] = W_q(\zeta_i) - W_q(\zeta_j) - [W_{q_0}(\zeta_i) - W_{q_0}(\zeta_j)] \\ = \; & \frac{\mu}{\gamma} \frac{1}{\sigma_s} \int_{\zeta_j}^{\zeta_i} \left(\left(-\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)} \right) \left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma}-1} \right. \\ & \left. - \left(-\frac{g'\left(\frac{q_0-\xi_s}{\sigma_s}\right)}{g\left(\frac{q_0-\xi_s}{\sigma_s}\right)} \right) \left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q_0|\hat{e}, s)}{f(q_0|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma}-1} \right) d\zeta. \end{aligned} \tag{50}$$

For $q > \max\{q_s^P, q_s^*, \xi_s\}$, both $-\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)}$ and $\left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma}-1}$ are positive and weakly increasing in q (by MLRP) if $\gamma \leq 1$. Therefore, if $\gamma \leq 1$, expression (50) is positive. Using (44), this means that, with $\zeta_i > \zeta_j$, the PPS measure $\frac{w_s(q)-w_s(q_0)}{q-q_0}$ is higher under s_i than under s_j.

Part (iv): Suppose that signals s_i and s_j differ only in their scale parameter, with $\sigma_i > \sigma_j$. Let $w_s(q) \equiv W_q(\sigma_s)$, which is a continuous function of σ_s, since $\frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)}$ is by assumption continuously differentiable in the scale parameter σ_s. For a given q, we have:

$$w_{s_i}(q) - w_{s_j}(q) = W_q(\sigma_i) - W_q(\sigma_j) = \int_{\sigma_j}^{\sigma_i} \frac{\partial W_q(\sigma)}{\partial \sigma} d\sigma. \tag{51}$$

Holding all else constant including Lagrange multipliers, for given q and s such that $w_s(q) > 0$:

$$\begin{aligned} \frac{\partial W_q(\sigma)}{\partial \sigma} &= \frac{\partial}{\partial \sigma} \left\{ \left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma}} - \bar{W} \right\} \\ &= \underbrace{\frac{\mu}{\gamma}}_{>0} \frac{\partial}{\partial \sigma} \left\{ \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right\} \underbrace{\left(\lambda + \mu \left[\frac{\partial \phi_{\hat{e}}^s / \partial e}{\phi_{\hat{e}}^s} + \frac{\frac{\partial f}{\partial e}(q|\hat{e}, s)}{f(q|\hat{e}, s)} \right] \right)^{\frac{1}{\gamma}-1}}_{>0 \text{ since } w_s(q)>0}, \end{aligned}$$

where:

$$\frac{\partial}{\partial\sigma}\left\{\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right\} = \frac{\partial}{\partial\sigma}\left\{-\frac{\zeta_s}{\sigma_s}\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)}\right\}$$
$$= \underbrace{\frac{\zeta_s}{\sigma_s^2}}_{>0}\underbrace{\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)}}_{<0\text{ for }q>q_s^P} + \underbrace{\frac{\zeta_s}{\sigma_s}}_{>0}\underbrace{G(q)}_{\geq 0}\underbrace{\frac{\xi_s-q}{\sigma_s^2}}_{<0\text{ for }q>\xi_s}. \quad (52)$$

So, at output q:

$$\int_{\sigma_j}^{\sigma_i}\frac{\partial W_q(\sigma)}{\partial\sigma}d\sigma = \int_{\sigma_j}^{\sigma_i}\frac{\mu}{\gamma}\frac{\zeta_s}{\sigma_s^2}\left(\underbrace{\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)} + G(q)\frac{\xi_s-q}{\sigma_s}}_{<0\text{ for }q>\max\{\xi_s,q_s^P\}}\right)\underbrace{\left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1}}_{>0\text{ since }w_s(q)>0}d\sigma$$

In sum, with $q > q_0 > \max\{q_s^P, q_s^*, \xi_s\}$:

$$\begin{aligned} w_{s_i}(q) - w_{s_j}(q) - \left(w_{s_i}(q_0) - w_{s_j}(q_0)\right) &= W_q(\sigma_i) - W_q(\sigma_j) - (W_{q_0}(\sigma_i) - W_{q_0}(\sigma_j)) \\ &= \frac{\mu}{\gamma}\frac{\zeta_s}{\sigma_s^2}\int_{\sigma_j}^{\sigma_i}\left(\left(\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)} + G(q)\frac{\xi_s-q}{\sigma_s}\right)\left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right.\right. \\ &\quad \left.\left. - \left(\frac{g'\left(\frac{q_0-\xi_s}{\sigma_s}\right)}{g\left(\frac{q_0-\xi_s}{\sigma_s}\right)} + G(q)\frac{\xi_s-q_0}{\sigma_s}\right)\left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q_0|\hat{e},s)}{f(q_0|\hat{e},s)}\right]\right)\right) \end{aligned}$$

For $q > \max\{q_s^P, q_s^*, \xi_s\}$, $\left(\frac{g'\left(\frac{q-\xi_s}{\sigma_s}\right)}{g\left(\frac{q-\xi_s}{\sigma_s}\right)} + G(q)\frac{\xi_s-q}{\sigma_s}\right)$ is negative and weakly decreasing in q if the likelihood ratio of output is weakly convex (by MLRP and since then $G'(q) \geq 0$, as per point (ii) above), while $\left(\lambda+\mu\left[\frac{\partial\phi_{\hat{e}}^s/\partial e}{\phi_{\hat{e}}^s}+\frac{\frac{\partial f}{\partial e}(q|\hat{e},s)}{f(q|\hat{e},s)}\right]\right)^{\frac{1}{\gamma}-1}$ is positive and weakly increasing in q (by MLRP) if $\gamma \leq 1$. Therefore, if $\gamma \leq 1$ and the likelihood ratio of output is weakly convex, expression (53) is negative. Using inequality (44), this means that, with $\sigma_i > \sigma_j$, the PPS measure $\frac{w_s(q)-w_s(q_0)}{q-q_0}$ is lower under s_i than under s_j if the likelihood ratio of output is weakly convex and $\gamma \leq 1$. ■

G The First-Order Approach with Limited Liability, Normally-Distributed Output, and Log Utility

This Appendix provides sufficient conditions for the FOA in the setting considered in Section 2.2, with limited liability, normally-distributed output, and log utility. We first derive the optimal contract and provide a sufficient condition for the FOA without an additional signal. Given effort $e \in [0, \bar{e}]$, output is determined by

$$q = e + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

Proposition 4 *Suppose $C''(e) \geq \frac{\bar{e}}{\sigma^2}$ for all $e \in [0, \bar{e}]$. Let $\{w^*(\cdot), e^*\}$ be the optimal contract and the effort it implements. Then, there exist $\lambda > 0$ and $q^* \leq e^* + \frac{\sigma^2}{\lambda}\bar{W}$ such that*

$$w^*(q) = \frac{\lambda}{\sigma^2} \cdot \max\{q - q^*,\ 0\}.$$

Moreover, $q^* = e^* + \frac{\sigma^2}{\lambda}\bar{W}$ if the IR does not bind.

For example, with a quadratic effort cost, $C(e) = \alpha e + \frac{\beta}{2}e^2$, for $\alpha > 0$ and $\beta > 0$, we have $C''(e) = \beta$ for all e, and the sufficient condition for the FOA is simply $\beta \geq \frac{\bar{e}}{\sigma^2}$.

Proof of Proposition 4:

As usual, let φ denote the PDF of the standard normal distribution. Let $W(q) \equiv w(q) + \bar{W}$ denote the manager's consumption. His IC, IR, and LL are, respectively:

$$\begin{aligned}
& e \in \arg\max_{\hat{e} \in [0, \bar{e}]} \int \ln\left(W(q)\right) \frac{1}{\sigma}\varphi\left(\frac{q - \hat{e}}{\sigma}\right) dq - C(\hat{e}), \\
& \int \ln\left[W\left(q\right)\right] \frac{1}{\sigma}\varphi\left(\frac{q - e}{\sigma}\right) dq - C(e) \geq 0, \\
& W\left(q\right) \geq \bar{W}\ \forall q.
\end{aligned}$$

To simplify notation, we will work with the manager's indirect utility, $u(q) \equiv \ln W\left(q\right)$, so that $W(q) = \exp\left[u(q)\right]$. This step is without loss of generality. The next step, which in general is not without loss of generality, is to replace the IC by its FOC:

$$\int u(q)\left(\frac{q-e}{\sigma^3}\right)\varphi\left(\frac{q-e}{\sigma}\right)dq - C'(e)\begin{cases}\geq 0 & \text{if } \; e=\bar{e}\\ =0 & \text{if } \; e\in(0,\bar{e})\\ \leq 0 & \text{if } \; e=0\end{cases}, \tag{54}$$

where we used the fact that $\varphi'(q) = -x\varphi(q)$, so that $\frac{d}{de}\left[\varphi\left(\frac{q-e}{\sigma}\right)\right] = \frac{q-e}{\sigma^2}\cdot\varphi\left(\frac{q-e}{\sigma}\right)$. Since replacing the IC by its FOC is not always valid, after solving the firm's relaxed program, we will need to verify that its solution satisfies the IC.

Writing in terms of the manager's indirect utility, the IR becomes

$$\int u(q)\frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right)dq - C(e) \geq 0. \tag{55}$$

It is convenient to multiply both sides of LL by $\frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right) > 0$, rewriting it as:

$$\frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right)u(q) \geq \frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right)\ln\bar{W}\ \forall q. \tag{56}$$

The principal's *relaxed program* is:

$$\max_{u(\cdot),e}\int\{q-\exp[u(q)]\}\frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right)dq,$$

subject to (54), (55) and (56).

As in Grossman and Hart (1983), we break down this program in two parts. First, we consider the solution of the relaxed program holding each effort $e\in[0,\bar{e}]$ fixed:

$$\min_{u(\cdot)}\int\exp[u(q)]\frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right)dq,$$

subject to (54), (55) and (56).

The optimal contract to implement the lowest effort ($e^* = 0$) pays a fixed wage. The utility given to the manager is set at the lowest level that still satisfies both LL and IR: $u(q) = \max\{\ln\bar{W},\ C(0)\}$ for all q. To see this, notice that a constant utility $u(q) = u^*$ always satisfies (54):

$$\int u^*\frac{q}{\sigma^3}\varphi\left(\frac{q}{\sigma}\right)dq - C'(0) = \frac{u^*}{\sigma^3}\times\int q\varphi\left(\frac{q}{\sigma}\right)dq - C'(0) = -C'(0) \leq 0.$$

Lemma 5 obtains the solution of the relaxed program for $e^* > 0$.

Lemma 5 *The optimal contract that implements $e^* > 0$ in the relaxed program is:*

$$w(q) = \frac{\lambda}{\sigma^2} \cdot \max\{q - q^*,\ 0\},$$

where $q^* \leq e^* + \frac{\sigma^2}{\lambda}\bar{W}$ (with equality if the IR does not bind).

Proof. The (infinite-dimensional) Lagrangian gives the following FOC:

$$-\exp[u(q)]\frac{1}{\sigma}\varphi\left(\frac{q-e^*}{\sigma}\right)+\lambda\left(\frac{q-e^*}{\sigma^3}\right)\varphi\left(\frac{q-e^*}{\sigma}\right)+\mu_{IR}\frac{1}{\sigma}\varphi\left(\frac{q-e^*}{\sigma}\right)+\mu_{LL}(q)\frac{1}{\sigma}\varphi\left(\frac{q-e^*}{\sigma}\right)=0,$$

where λ is the multiplier associated with (54), and μ_{LL} and μ_{IR} are the multipliers associated with (55) and (56). Since the program corresponds to the minimization of a strictly convex function subject to linear constraints, the FOC above, along with the standard complementary slackness conditions and the constraints, are sufficient for an optimum. Substitute $\exp[u(q)] = W(q)$ and simplify the FOC above to obtain:

$$W(q) = \lambda\frac{q-e^*}{\sigma^2} + \mu_{IR} + \mu_{LL}(q).$$

Suppose first that the IR does not bind, so that $\mu_{IR} = 0$. Then, the FOC becomes

$$W(q) = \lambda\frac{q-e^*}{\sigma^2} + \mu_{LL}(q).$$

For $W(q) > \bar{W}$, complementary slackness gives $\mu_{LL}(q) = 0$, so that:

$$W(q) = \lambda \times \frac{q-e^*}{\sigma^2},$$

which exceeds $\bar{W}$ if and only if

$$\lambda \times \frac{q-e^*}{\sigma^2} > \bar{W} \iff q > e^* + \frac{\sigma^2\bar{W}}{\lambda} \equiv q^*.$$

For $W(q) = \bar{W}$, the FOC becomes:

$$\bar{W} = \lambda\frac{q-e^*}{\sigma^2} + \mu_{LL}(q) \therefore \mu_{LL}(q) = \bar{W} - \lambda\frac{q-e^*}{\sigma^2},$$

so that $\mu_{LL}(q) \geq 0$ if and only if

$$\bar{W} \geq \lambda \times \frac{q - e^*}{\sigma^2} \iff q \leq q^*.$$

Therefore, the optimal contract is

$$W(q) = \max\left\{\frac{\lambda\,(q - e^*)}{\sigma^2},\ \bar{W}\right\} = \begin{cases} \frac{\lambda(q-e^*)}{\sigma^2} & \text{if } q \geq q^* \\ \bar{W} & \text{if } q \leq q^* \end{cases}.$$

Writing in terms of the firm's payments, we have

$$w(q) = W(q) - \bar{W} = \frac{\lambda}{\sigma^2}\max\{q - q^*,\ 0\},$$

where the last equality uses the definition of q^*. The firm gives the manager an option with strike price $q^* = e^* + \frac{\sigma^2}{\lambda}\bar{W} > e^*$ and a slope $\frac{\lambda}{\sigma^2}$ chosen so that (54) holds (which can be shown to exist and be unique).

Next, suppose that the IR binds so that $\mu_{IR} \geq 0$. Then, for $W(q) > \bar{W}$, we must have

$$W(q) = \lambda\frac{q - e^*}{\sigma^2} + \mu_{IR},$$

so that

$$W(q) > \bar{W} \iff \mu_{IR} > \bar{W} - \lambda\frac{q - e^*}{\sigma^2}.$$

For $W(q) = \bar{W}$, we have:

$$\bar{W} = \lambda\frac{q - e^*}{\sigma^2} + \mu_{IR} + \mu_{LL}(q),$$

so that $\mu_{LL}(q) \geq 0$ if and only if

$$\mu_{LL}(q) = \bar{W} - \lambda\frac{q - e^*}{\sigma^2} - \mu_{IR} \geq 0$$

$$\iff \bar{W} - \lambda\frac{q - e^*}{\sigma^2} \geq \mu_{IR}.$$

Define the strike price q^* as the solution to

$$\bar{W} - \lambda\frac{q^* - e^*}{\sigma^2} = \mu_{IR},$$

that is,

$$q^* \equiv e^* + \frac{\sigma^2}{\lambda}\left(\bar{W} - \mu_{IR}\right) \le e^* + \frac{\sigma^2}{\lambda}\bar{W}.$$

Combining both conditions, we obtain

$$W(q) = \begin{cases} \frac{\lambda}{\sigma^2}(q - q^*) + \bar{W} & \text{if } q \ge q^* \\ \bar{W} & \text{if } q \le q^* \end{cases},$$

which again corresponds to an option with strike price q^* and slope $\frac{\lambda}{\sigma^2}$. Here, λ and q^* are chosen so that both (54) and (55) hold with equality. ■

Lemma 6 gives an upper bound on λ:

Lemma 6 *Suppose $e^* > 0$ is the effort that solves the firm's relaxed program. Then the optimal contract is*

$$w(q) = \max\left\{\frac{\lambda}{\sigma^2}(q - q^*),\ 0\right\},$$

where $0 < \lambda < \sqrt{2\pi}\sigma e^*$ and $q^* \le e^* + \frac{\sigma^2}{\lambda}\bar{W}$.

Proof. From Lemma 5, we need to show that $\lambda \le \sqrt{2\pi}\sigma e^*$. Recall that the optimal contract that implements effort $e > 0$ is the option:

$$w(q) = \max\left\{\frac{\lambda}{\sigma^2}(q - q^*),\ 0\right\},$$

where $q^* \le \frac{\sigma^2}{\lambda}\bar{W} + e$. Since the firm's net profits $q - w(q)$ are increasing in the strike price q^* (holding constant all other variables, including effort), its profits are bounded above by the profits from offering the option with the highest strike price ($\bar{q} = \frac{\sigma^2}{\lambda}\bar{W} + e \ge q^*$), which equal

$$e - \left[\frac{\lambda}{\sigma^2}\int_{\frac{\sigma^2}{\lambda}\bar{W}+e}^{\infty}\left(q - e - \frac{\sigma^2}{\lambda}\bar{W}\right)\frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right)dq\right].$$

Let $z \equiv q - e - \frac{\sigma^2}{\lambda}\bar{W}$, so that $q = z + e + \frac{\sigma^2}{\lambda}\bar{W}$. Note that $q \ge \frac{\sigma^2}{\lambda}\bar{W} + e$ if and only if $z \ge 0$. Thus, we can rewrite this expression as

$$e - \left[\frac{\lambda}{\sigma^2}\int_0^{\infty} z\frac{1}{\sigma}\varphi\left(\frac{z + \frac{\sigma^2}{\lambda}\bar{W}}{\sigma}\right)dz\right].$$

Moreover, since $\varphi(z)$ is decreasing in z for $z > 0$, it follows that

$$\varphi\left(\frac{z+\frac{\bar{W}}{\lambda}}{\sigma}\right) < \varphi\left(\frac{z}{\sigma}\right) \quad \forall z > 0.$$

Thus, the firm's profits are strictly less than

$$e - \frac{\lambda}{\sigma^2}\int_0^\infty \frac{z}{\sigma}\varphi\left(\frac{z}{\sigma}\right)dz. \tag{57}$$

Apply the following change of variables $y = \frac{z}{\sigma}$ (so that $z = \sigma y$, $dz = \sigma dy$) to write

$$\int_0^\infty \frac{z}{\sigma}\varphi\left(\frac{z}{\sigma}\right)dz = \sigma\int_0^\infty y\varphi\left(y\right)dy.$$

Integrating by parts gives

$$\int_0^\infty y\varphi(y)dy = \left[-\varphi(y)\right]_0^\infty = \varphi(0) = \frac{1}{\sqrt{2\pi}}.$$

Substituting into (57), the firm's profits are strictly less than

$$e - \frac{1}{\sqrt{2\pi}}\cdot\frac{\lambda}{\sigma}.$$

Since the firm can always obtain a profit of zero by paying zero wages and implementing zero effort, we must have

$$e - \frac{1}{\sqrt{2\pi}}\cdot\frac{\lambda}{\sigma} > 0 \iff \lambda < \sqrt{2\pi}\sigma e.$$

■

Lemma 7 provides an additional upper bound:

Lemma 7 *For any $q^* \in \mathbb{R}$, $e \in [0, \bar{e}]$, $\sigma > 0$ and $\lambda > 0$, we have*

$$\int \ln\left[\bar{W} + \frac{\lambda}{\sigma^2}\cdot\max\left\{(q-q^*)\,,\ 0\right\}\right]\left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right]\frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right)dq$$

$$\leq \int \left[\bar{W} + \frac{\lambda}{\sigma^2}\cdot\max\left\{(q-q^*)\,,\ 0\right\}\right]\left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right]\frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right)dq.$$

Proof. For notational simplicity, let $y \equiv \frac{q^{*}-e}{\sigma}$, apply the change of variables $z \equiv \frac{q-e}{\sigma}$, and let

$$g(z) \equiv \bar{W} + \frac{\lambda}{\sigma} \cdot \max\{z-y,\ 0\} - \ln\left[\bar{W} + \frac{\lambda}{\sigma} \cdot \max\{z-y,\ 0\}\right].$$

Then, the inequality in Lemma 7 can be written as

$$\int_{-\infty}^{\infty} g(z)\left(z^{2}-1\right)\varphi(z)\,dz \geq 0.$$

We claim that $g(\cdot)$ is non-decreasing. To see this, note that, for $z \leq y$, $g(z) = \bar{W} - \ln \bar{W}$ (which is constant in z). For $z > y$, we have

$$g'(z) = \frac{\lambda}{\sigma}\left(\frac{\bar{W} - 1 + \frac{\lambda}{\sigma}(z-y)}{\bar{W} + \frac{\lambda}{\sigma}(z-y)}\right),$$

which is positive for all $z > y$ since $\bar{W} \geq 1$. Since g is non-decreasing, we have $g(q) \geq g(-q)$ for $q \geq 0$ and $\frac{d}{dq}[g(q) - g(-q)] \geq 0$. Note that, applying the change of variables $\tilde{z} = -z$ and using the symmetry of $(z^{2}-1)\varphi(z)$ around zero, we have:

$$\int_{-\infty}^{0} g(z)\left(z^{2}-1\right)\varphi(z)\,dz = -\int_{0}^{\infty} g(-z)\left(z^{2}-1\right)\varphi(z)\,dz. \tag{58}$$

Therefore,

$$\begin{aligned}
\int g(z)(z^{2}-1)\varphi(z)\,dz &= \int_{-\infty}^{0} g(z)(z^{2}-1)\varphi(z)\,dz + \int_{0}^{\infty} g(z)(z^{2}-1)\varphi(z)\,dz \\
&= -\int_{0}^{\infty} g(-z)(z^{2}-1)\varphi(z)\,dz + \int_{0}^{\infty} g(z)(z^{2}-1)\varphi(z)\,dz \\
&= \int_{0}^{\infty} [g(z) - g(-z)](z^{2}-1)\varphi(z)\,dz \\
&= \int_{0}^{1} [g(z) - g(-z)](z^{2}-1)\varphi(z)\,dz + \int_{1}^{\infty} [g(z) - g(-z)](z^{2}-1)\varphi(z)\,dz \\
&\geq \int_{0}^{1} [g(1) - g(-1)](z^{2}-1)\varphi(z)\,dz + \int_{1}^{\infty} [g(1) - g(-1)](z^{2}-1)\varphi(z)\,dz \\
&= [g(1) - g(-1)]\int_{0}^{\infty} (z^{2}-1)\varphi(z)\,dz = 0,
\end{aligned}$$

where the first line opens the integral between positive and negative values of z, the second line substitutes (58), the third line combines the terms from the two integrals, and the fourth line opens the integral between $z \leq 1$ and $z \geq 1$. The fifth line is the crucial step, which uses the following two facts: (i) $z^{2} > (<) 1$ for $z > (<) 1$, and (ii) $g(z) - g(-z)$ is non-decreasing for all z. Therefore, substituting $g(z) - g(-z)$ by its upper bound where the term inside the integral is negative, and by its lower bound where it is positive, lowers the value of the integrand. The sixth line combines terms and uses the fact that

$$\int_0^\infty \left(z^2 - 1\right) \varphi\left(z\right) dz = \left[-z\varphi(z)\right]_0^\infty = 0.$$

■

Lemma 8 shows that the solution of the relaxed program also solves the firm's program if the effort cost is sufficiently convex, i.e. the FOA is valid.

Lemma 8 *Suppose $C''(e) \geq \frac{\bar{e}}{\sigma}$ for all $e \in [0, \bar{e}]$. Then, the solution of the firm's program coincides with the solution of the relaxed program.*

Proof. The manager's utility from choosing effort e is:

$$U(e;\ q^*, \lambda) \equiv \int \ln\left[\bar{W} + \frac{\lambda}{\sigma^2} \cdot \max\left\{(q - q^*),\ 0\right\}\right] \frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right) dq - C(e).$$

We know from previous results that $0 < \lambda < \sqrt{2\pi}\sigma e^*$. The FOA is valid if

$$\frac{\partial^2 U}{\partial e^2}(e;\ q^*, \lambda) \leq 0$$

for all $e \in [0, \bar{e}]$, all $q^* \in \mathbb{R}$, and $\lambda \in (0, \sqrt{2\pi}\sigma\bar{e})$. Differentiating gives

$$\begin{aligned} \frac{\partial^2 U}{\partial e^2} &= \int \ln\left[\bar{W} + \frac{\lambda}{\sigma^2} \cdot \max\left\{(q - q^*),\ 0\right\}\right] \frac{1}{\sigma}\frac{d^2}{de}\left[\varphi\left(\frac{q-e}{\sigma}\right)\right] dq - C''(e) \\ &= \frac{1}{\sigma^2}\int \ln\left[\bar{W} + \frac{\lambda}{\sigma^2} \cdot \max\left\{(q - q^*),\ 0\right\}\right] \frac{1}{\sigma}\left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right] \varphi\left(\frac{q-e}{\sigma}\right) dq - C''(e) \end{aligned}, \tag{59}$$

where the second line uses the fact that $\frac{d^2}{de^2}\left[\varphi\left(\frac{q-e}{\sigma}\right)\right] = \frac{1}{\sigma^2}\left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right]\varphi\left(\frac{q-e}{\sigma}\right)$. Note that

$$\begin{aligned} & \int \ln\left[\bar{W} + \frac{\lambda}{\sigma^2} \cdot \max\left\{(q - q^*),\ 0\right\}\right] \left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right] \frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right) dq \\ \leq\ & \int \left[\bar{W} + \frac{\lambda}{\sigma^2} \cdot \max\left\{(q - q^*),\ 0\right\}\right] \left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right] \frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right) dq \\ =\ & \frac{\lambda}{\sigma^2} \cdot \int \left[\max\left\{(q - q^*),\ 0\right\}\right] \left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right] \frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right) dq \\ =\ & \frac{\lambda}{\sigma^2} \cdot \int_{q^*}^\infty \left(q - q^*\right) \left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right] \frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right) dq, \end{aligned}$$

where the inequality uses the result from Lemma 7, the third line follows from $\int \left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right] \frac{1}{\sigma}\varphi\left(\frac{q-e}{\sigma}\right) dq = 0$ (a standard normal variable has variance 1), and the fourth line opens the max operator. Substituting in the expression from (59), we obtain the following sufficient condition for the validity of the FOA:

$$\frac{\lambda}{\sigma^4} \cdot \int_{q^*}^{\infty} (q - q^*) \left[\left(\frac{q-e}{\sigma} \right)^2 - 1 \right] \frac{1}{\sigma} \varphi \left(\frac{q-e}{\sigma} \right) dq \le C''(e) \tag{60}$$

for all $e \in [0, \bar{e}]$, $q^* \in \mathbb{R}$, and $\lambda \in (0, \sqrt{2\pi}\sigma\bar{e})$.

Let $\xi(q^*) \equiv \int_{q^*}^{\infty} (q - q^*) \left[\left(\frac{q-e}{\sigma} \right)^2 - 1 \right] \frac{1}{\sigma} \varphi \left(\frac{q-e}{\sigma} \right) dq$. We claim that $\xi'(q^*) \left\{ \begin{matrix} > \\ < \end{matrix} \right\} 0 \iff q^* \left\{ \begin{matrix} < \\ > \end{matrix} \right\} e$. Differentiating yields:

$$\xi'(q^*) = - \int_{q^*}^{\infty} \left[\left(\frac{q-e}{\sigma} \right)^2 - 1 \right] \frac{1}{\sigma} \varphi \left(\frac{q-e}{\sigma} \right) dq. \tag{61}$$

Note that

$$\frac{d}{dq} \left[- \left(\frac{q-e}{\sigma} \right) \varphi \left(\frac{q-e}{\sigma} \right) \right] = - \frac{1}{\sigma} \varphi \left(\frac{q-e}{\sigma} \right) - \left(\frac{q-e}{\sigma} \right) \frac{1}{\sigma} \varphi' \left(\frac{q-e}{\sigma} \right) = \left[\left(\frac{q-e}{\sigma} \right)^2 - 1 \right] \frac{1}{\sigma} \varphi \left(\frac{q-e}{\sigma} \right),$$

where the last equality uses the fact that $\varphi'(q) = -q\varphi(q)$. Therefore,

$$\int \left[\left(\frac{q-e}{\sigma} \right)^2 - 1 \right] \frac{1}{\sigma} \varphi \left(\frac{q-e}{\sigma} \right) dq = - \left(\frac{q-e}{\sigma} \right) \varphi \left(\frac{q-e}{\sigma} \right).$$

Substituting back into (61) gives:

$$\xi'(q^*) = - \left(\frac{q^*-e}{\sigma} \right) \varphi \left(\frac{q^*-e}{\sigma} \right) \left\{ \begin{matrix} > \\ < \end{matrix} \right\} 0 \iff q^* \left\{ \begin{matrix} < \\ > \end{matrix} \right\} e.$$

Therefore, $\xi(\cdot)$ is maximized at $q^* = e$, so that, by condition (60), it suffices to show that

$$\frac{\lambda}{\sigma^4} \cdot \xi(e) \le C''(e). \tag{62}$$

Evaluating ξ at e gives:

$$\xi(e) = \int_{e}^{\infty} (q - e) \left[\left(\frac{q-e}{\sigma} \right)^2 - 1 \right] \frac{1}{\sigma} \varphi \left(\frac{q-e}{\sigma} \right) dq.$$

Performing the change of variables $z \equiv \frac{q-e}{\sigma}$, we obtain

$$\xi(e) = \int_e^\infty \left(\frac{q-e}{\sigma}\right)\left[\left(\frac{q-e}{\sigma}\right)^2 - 1\right]\varphi\left(\frac{q-e}{\sigma}\right)dq = \sigma\int_0^\infty z\left(z^2-1\right)\varphi(z)\,dz. \qquad (63)$$

Integrating by parts gives

$$\int z\left(z^2-1\right)\varphi(z)\,dz = -z^2\varphi(z) + \int z\varphi(z)\,dz,$$

where we let $(z^2-1)\varphi(z)\,dz = dv$ so that $v = -z\varphi(z)$, and we let $u = z$, so that $du = dz$. Therefore

$$\int_0^\infty z\left(z^2-1\right)\varphi(z)\,dz = \int_0^\infty z\varphi(z)\,dz.$$

Using $\frac{d}{dz}[-\varphi(z)] = z\varphi(z)$, we have

$$\int_0^\infty z\left(z^2-1\right)\varphi(z)\,dz = [-\varphi(z)]_0^{+\infty} = \varphi(0) = \frac{1}{\sqrt{2\pi}}.$$

Substituting into (63), yields

$$\xi(e) = \frac{\sigma}{\sqrt{2\pi}}.$$

Substituting into (62), we obtain the following sufficient condition:

$$\frac{\lambda}{\sqrt{2\pi}\sigma^3} \le C''(e),$$

which is true for all $e \in [0, \bar{e}]$ and all $\lambda \in (0, \sqrt{2\pi}\sigma\bar{e})$ if and only if

$$C''(e) \ge \frac{\bar{e}}{\sigma^2} \quad \forall e \in [0, \bar{e}].$$

■

Proposition 5 provides a sufficient condition for the FOA with an additional performance signal, for a subset of signal distributions.

Proposition 5 *We consider the same setting as in Proposition 2, and a signal distribution such that: (i) $h_s''(e) \le 0$ for all s; (ii) ϕ_e^s linear in e for all s; (iii) $h_{s_1}(e) \le h_{s_2}(e)$, $h_{s_1}'(e) \le h_{s_2}'(e)$, and $\sigma_{s_1} \ge \sigma_{s_2}$ for any s_1, s_2 with $\frac{d\phi_e^{s_1}}{de} > 0 > \frac{d\phi_e^{s_2}}{de}$ and any $e \in [0, \bar{e}]$. Then the FOA is valid if $C''(e) \ge \sum_s \phi_{e^*}^s h_s(\bar{e}) \frac{\sum_s \frac{\phi_e^s}{\sigma_s^3}(h_s'(e))^2}{\sum_s \frac{\phi_{e^*}^s}{\sigma_s}}$ for all $e \in [0, \bar{e}]$.*

Proof of Proposition 5: Let φ denote the PDF of the standard normal distribution. Let $W_s(q) := \bar{W} + w_s(q)$ denote the manager's consumption (i.e., the manager's initial wealth $\bar{W}$ plus his pay). His IC, IR, and LL are, respectively:

$$
\begin{aligned}
& e \in \arg\max_{\hat{e}\in[0,\bar{e}]} \sum_s \phi_e^s \int \ln\left(W_s(q)\right) \frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(\hat{e})}{\sigma_s}\right) dq - C(\hat{e}), \\
& \sum_s \phi_e^s \int \ln\left(W_s(q)\right) \frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right) dq - C(e) \geq 0, \\
& W_s(q) \geq \bar{W} \ \forall q, s.
\end{aligned}
$$

To simplify notation, we will work with the manager's indirect utility, $u_s(q) := \ln\left(W_s(q)\right)$, so that $W_s(q) = \exp[u_s(q)]$. This step is without loss of generality. The next step, which in general is not without loss of generality, is to replace the IC by its FOC:

$$
\sum_s \phi_e^s \int u_s(q) \frac{q-h_s(e)}{\sigma_s^3}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right) dq - C'(e) \begin{cases} \geq 0 & \text{if } e = \bar{e} \\ = 0 & \text{if } e \in (0,\bar{e}) \\ \leq 0 & \text{if } e = 0 \end{cases}. \tag{64}
$$

Since replacing the IC by its FOC is not always valid, after solving the firm's relaxed program, we will need to verify that its solution satisfies the IC. It is convenient to multiply both sides of LL by $\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)\phi_e^s > 0$, rewriting it as:

$$
\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)\phi_e^s u_s(q) \geq \frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)\phi_e^s \ln \bar{W} \ \forall q, s. \tag{65}
$$

The principal's *relaxed program* is:

$$
\max_{\{u_s(q)\}_{q,s},e} \sum_s \phi_e^s \int \left\{q - \exp\left[u_s(q)\right]\right\} \frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right) dq
$$

subject to (64), (65), and

$$
\sum_s \phi_e^s \int u_s(q) \frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right) dq - C(e) \geq 0. \tag{66}
$$

As in Grossman and Hart (1983), we break down this program in two parts. First, we consider the solution of the relaxed program holding each effort $e \in [0, \bar{e}]$ fixed:

$$\min_{u(\cdot)} \sum_s \phi_e^s \int \exp\left[u_s(q)\right] \frac{1}{\sigma_s} \varphi\left(\frac{q - h_s(e)}{\sigma_s}\right) dq$$

subject to (64), (65), and (66).

The optimal contract to implement the lowest effort ($e^* = 0$) pays a fixed wage. The utility given to the manager is set at the lowest level that still satisfies both the LL and IR: $u_s(q) = \max\{\ln \bar{W},\ C(0)\}$ for all q, s.

Lemma 9 obtains the solution of the relaxed program for $e^* > 0$.

Lemma 9 *The optimal contract that implements $e^* > 0$ in the relaxed program is:*

$$w_s(q) = \frac{\lambda}{\sigma_s^2} \cdot \max\left\{q - q_s^*, 0\right\},$$

where $q_s^* \le \sigma_s^2 \frac{\bar{W}}{\lambda} + h_s(e^*)$ (with equality if the IR does not bind).

Proof. The (infinite-dimensional) Lagrangian associated with this program is:

$$\begin{aligned}
&\sum_s \phi_e^s \int \exp\left[u_s(q)\right] \frac{1}{\sigma_s} \varphi\left(\frac{q - h_s(e^*)}{\sigma_s}\right) dq \\
&+\lambda \left[\sum_s \phi_e^s \int u_s(q) \frac{q - h_s(e^*)}{\sigma_s^3} \varphi\left(\frac{q - h_s(e^*)}{\sigma_s}\right) dq - C'(e^*)\right] \\
&+\mu_{IR} \left[\sum_s \phi_e^s \int u_s(q) \frac{1}{\sigma_s} \varphi\left(\frac{q - h_s(e^*)}{\sigma_s}\right) dq - C(e^*)\right] \\
&+\mu_{LL}(q,s) \frac{1}{\sigma_s} \varphi\left(\frac{q - h_s(e^*)}{\sigma_s}\right) \phi_e^s u_s(q).
\end{aligned}$$

The FOC is:

$$-\exp\left[u_s(q)\right] \frac{1}{\sigma_s} \varphi\left(\frac{q - h_s(e^*)}{\sigma_s}\right) \phi_e^s + \lambda \frac{q - h_s(e^*)}{\sigma_s^3} \varphi\left(\frac{q - h_s(e^*)}{\sigma_s}\right) \phi_e^s$$

$$+\mu_{IR} \frac{1}{\sigma_s} \varphi\left(\frac{q - h_s(e^*)}{\sigma_s}\right) \phi_e^s + \mu_{LL}(q,s) \frac{1}{\sigma_s} \varphi\left(\frac{q - h_s(e^*)}{\sigma_s}\right) \phi_e^s = 0,$$

where λ is the multiplier associated with (64), and μ_{LL} and μ_{IR} are the multipliers associated with (65) and (66). Since the program corresponds to the minimization of a strictly convex function subject to linear constraints, the FOC above, along with the standard complementary slackness conditions and the constraints, are sufficient for an optimum. Substitute $\exp\left[u_s\left(q\right)\right] = W_s(q)$ and simplify the FOC above to obtain:

$$W_s(q) = \lambda \frac{q - h_s\left(e^*\right)}{\sigma_s^2} + \mu_{IR} + \mu_{LL}(q, s).$$

By complementary slackness, we must have $\mu_{IR} \geq 0$ (with $\mu_{IR} = 0$ if IR does not bind). Similarly, $\mu_{LL}(q) \geq 0$ with equality if $W_s(q) > \bar{W}$. Thus, for $W_s(q) > \bar{W}$, we must have

$$W_s(q) = \lambda \frac{q - h_s\left(e^*\right)}{\sigma_s^2} + \mu_{IR} > \bar{W},$$

which can be rearranged as

$$q > \sigma_s^2 \frac{\bar{W} - \mu_{IR}}{\lambda} + h_s\left(e^*\right) =: q_s^*.$$

For $W_s(q) = \bar{W}$, we must have

$$\mu_{LL}(q, s) = \bar{W} - \lambda \frac{q - h_s\left(e^*\right)}{\sigma_s^2} - \mu_{IR} \geq 0 \iff q \leq q_s^*.$$

Combining both conditions, we obtain

$$W_s(q) = \max\left\{\lambda \frac{q - h_s\left(e^*\right)}{\sigma_s^2} + \mu_{IR}, \bar{W}\right\} = \bar{W} + \frac{\lambda}{\sigma_s^2} \cdot \max\left\{q - q_s^*, 0\right\}.$$

Thus,

$$w_s(q) = \frac{\lambda}{\sigma_s^2} \cdot \max\left\{q - q_s^*, 0\right\}.$$

Finally, since $\mu_{IR} \geq 0$,

$$q_s^* = \sigma_s^2 \frac{\bar{W} - \mu_{IR}}{\lambda} + h_s\left(e^*\right) \leq h_s\left(e^*\right) + \sigma_s^2 \frac{\bar{W}}{\lambda},$$

with equality if IR does not bind (in which case, we have $\mu_{IR} = 0$). ■

Lemma 10 gives an upper bound on λ:

Lemma 10 *Suppose $e^* > 0$ is the effort that solves the firm's relaxed program. Then the* *optimal contract is*

$$w_s(q) = \max\left\{\frac{\lambda}{\sigma_s^2}\left(q - q_s^*\right),\ 0\right\},$$

where $0 < \lambda < \frac{\sqrt{2\pi}\sum_s \phi_{e^*}^s h_s(e^*)}{\sum_s \frac{\phi_{e^*}^s}{\sigma_s^2}}$ *and* $q_s^* \leq h_s(e^*) + \sigma_s^2 \frac{\bar{W}}{\lambda}$.

Proof. From Lemma 10, we need to show that $\lambda \leq \frac{\sqrt{2\pi}\sum_s \phi_{e^*}^s h_s(e^*)}{\sum_s \frac{\phi_{e^*}^s}{\sigma_s^2}}$. Recall that the optimal contract that implements effort $e^* > 0$ is the option:

$$w_s(q) = \max\left\{\frac{\lambda}{\sigma_s^2}\left(q - q_s^*\right),\ 0\right\},$$

where $q_s^* \leq h_s(e^*) + \sigma_s^2 \frac{\bar{W}}{\lambda}$. Since the firm's net profits $q - w(q)$ are increasing in the strike price q_s^* (holding constant all other variables, including effort), its profits are bounded above by the profits from offering the option with the highest strike price for each signal s ($\bar{q}_s = h_s(e^*) + \sigma_s^2 \frac{\bar{W}}{\lambda} \geq q_s^*$), which equal

$$\sum_s \phi_{e^*}^s h_s(e^*) - \sum_s \phi_{e^*}^s \left[\frac{\lambda}{\sigma_s^2}\int_{h_s(e^*)+\sigma_s^2\frac{\bar{W}}{\lambda}}^{\infty}\left(q - h_s(e^*) - \sigma_s^2\frac{\bar{W}}{\lambda}\right)\frac{1}{\sigma_s}\varphi\left(\frac{q - h_s(e^*)}{\sigma_s}\right)dq\right].$$

For each s, let $z \equiv q - h_s(e^*) - \sigma_s^2 \frac{\bar{W}}{\lambda}$, so that $q = z + h_s(e^*) + \sigma_s^2 \frac{\bar{W}}{\lambda}$. Note that $q \geq h_s(e^*) + \sigma_s^2 \frac{\bar{W}}{\lambda}$ if and only if $z \geq 0$. Thus, we can rewrite this expression as

$$\sum_s \phi_{e^*}^s h_s(e^*) - \sum_s \phi_{e^*}^s \left[\frac{\lambda}{\sigma_s^2}\int_0^{\infty} z\frac{1}{\sigma_s}\varphi\left(\frac{z + \frac{\sigma_s^2}{\lambda}\bar{W}}{\sigma_s}\right)dz\right].$$

Moreover, since $\varphi(z)$ is decreasing in z for $z > 0$, it follows that, for any s,

$$\varphi\left(\frac{z + \sigma_s^2\frac{\bar{W}}{\lambda}}{\sigma_s}\right) < \varphi\left(\frac{z}{\sigma_s}\right) \quad \forall z > 0.$$

Thus, the firm's profits are strictly less than

$$\sum_s \phi_{e^*}^s h_s(e^*) - \sum_s \phi_{e^*}^s \frac{\lambda}{\sigma_s^2}\int_0^{\infty}\frac{z}{\sigma_s}\varphi\left(\frac{z}{\sigma_s}\right)dz. \tag{67}$$

Apply the following change of variables $y = \frac{z}{\sigma_s}$ (so that $z = \sigma_s y$, $dz = \sigma_s dy$) to write

$$\int_0^\infty \frac{z}{\sigma_s}\varphi\left(\frac{z}{\sigma_s}\right)dz = \sigma_s \int_0^\infty y\varphi\left(y\right)dy.$$

Integrating by parts gives

$$\int_0^\infty y\varphi(y)dy = [-\varphi(y)]_0^\infty = \varphi(0) = \frac{1}{\sqrt{2\pi}}.$$

Substituting into (67), the firm's profits are strictly less than

$$\sum_s \phi_{e^*}^s h_s\left(e^*\right) - \sum_s \phi_{e^*}^s \frac{1}{\sqrt{2\pi}} \cdot \frac{\lambda}{\sigma_s}.$$

Since the firm can always obtain a profit of zero by paying zero wages and implementing zero effort, we must have

$$\sum_s \phi_{e^*}^s h_s\left(e^*\right) - \frac{\lambda}{\sqrt{2\pi}}\sum_s \frac{\phi_{e^*}^s}{\sigma_s} > 0 \iff \lambda < \frac{\sqrt{2\pi}\sum_s \phi_{e^*}^s h_s\left(e^*\right)}{\sum_s \frac{\phi_{e^*}^s}{\sigma_s}}.$$

■

Lemma 11 provides an additional upper bound:

Lemma 11 *For any $q_s^* \in \mathbb{R}\,\forall s$, $e \in [0,\bar{e}]$, $e^* \in [0,\bar{e}]$, $\sigma_s > 0\,\forall s$, and $\lambda > 0$, we have*

$$\begin{aligned} &\sum_s \phi_e^s \int \ln\left[\bar{W} + \frac{\lambda}{\sigma_s^2}\cdot\max\left\{q - q_s^*,\ 0\right\}\right]\left(h_s'(e)\right)^2\left[\frac{\left(q - h_s\left(e\right)\right)^2}{\sigma_s^4} - \frac{1}{\sigma_s^2}\right]\frac{1}{\sigma_s}\varphi\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)dq \\ \leq\ &\sum_s \phi_e^s \int \left[\bar{W} + \frac{\lambda}{\sigma_s^2}\cdot\max\left\{q - q_s^*,\ 0\right\}\right]\left(h_s'(e)\right)^2\left[\frac{\left(q - h_s\left(e\right)\right)^2}{\sigma_s^4} - \frac{1}{\sigma_s^2}\right]\frac{1}{\sigma_s}\varphi\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)dq. \end{aligned}$$

Proof. For notational simplicity, for each s, let $y_s := \frac{q_s^* - h_s(e)}{\sigma_s}$, apply the change of variables $z_s := \frac{q - h_s(e)}{\sigma_s}$, and let

$$g_s(z) := \bar{W} + \frac{\lambda}{\sigma_s}\cdot\max\left\{z - y_s,\ 0\right\} - \ln\left[\bar{W} + \frac{\lambda}{\sigma_s}\cdot\max\left\{z - y_s,\ 0\right\}\right].$$

Then, the inequality in Lemma 11 can be written as

$$\sum_s \frac{\phi_e^s}{\sigma_s^3}(h_s'(e))^2 \int_{-\infty}^{\infty} g_s(z)\left(z^2-1\right)\varphi(z)\,dz \geq 0. \tag{68}$$

The terms ϕ_e^s, σ_s, and $(h_s'(e))^2$ are positive, so it remains to prove that this integral is positive. We claim that, for each s, $g_s(\cdot)$ is non-decreasing. To see this, notice that, for $z_s \leq y_s$, $g_s(z) = \bar{W} - \ln \bar{W}$ (which is constant in z_s). For $z_s > y_s$, we have

$$g_s'(z) = \frac{\lambda}{\sigma_s}\left(\frac{\bar{W} - 1 + \frac{\lambda}{\sigma_s}(z-y)}{\bar{W} + \frac{\lambda}{\sigma_s}(z-y)}\right),$$

which is positive for all $z_s > y_s$ since $\bar{W} \geq 1$. Since g is non-decreasing, we have $g_s(q) \geq g_s(-q)$ for $q \geq 0$ and $\frac{d}{dq}[g_s(q) - g_s(-q)] \geq 0$. Note that, applying the change of variables $\tilde{z} = -z$ and using the symmetry of $(z^2-1)\varphi(z)$ around zero, we have:

$$\int_{-\infty}^{0} g_s(z)\left(z^2-1\right)\varphi(z)\,dz = -\int_{0}^{\infty} g_s(-z)\left(z^2-1\right)\varphi(z)\,dz. \tag{69}$$

Therefore,

$$\begin{aligned}
\int g_s(z)\,(z^2-1)\,\varphi(z)\,dz &= \int_{-\infty}^{0} g_s(z)\,(z^2-1)\,\varphi(z)\,dz + \int_0^\infty g_s(z)\,(z^2-1)\,\varphi(z)\,dz \\
&= -\int_0^\infty g_s(-z)\,(z^2-1)\,\varphi(z)\,dz + \int_0^\infty g_s(z)\,(z^2-1)\,\varphi(z)\,dz \\
&= \int_0^\infty [g_s(z) - g_s(-z)]\,(z^2-1)\,\varphi(z)\,dz \\
&= \int_0^1 [g_s(z) - g_s(-z)]\,(z^2-1)\,\varphi(z)\,dz + \int_1^\infty [g_s(z) - g_s(-z)]\,(z^2-1)\,\varphi(z)\,dz \\
&\geq \int_0^1 [g_s(1) - g_s(-1)]\,(z^2-1)\,\varphi(z)\,dz + \int_1^\infty [g_s(1) - g_s(-1)]\,(z^2-1)\,\varphi(z)\,dz \\
&= [g_s(1) - g_s(-1)]\int_0^\infty (z^2-1)\,\varphi(z)\,dz = 0,
\end{aligned}$$

where the first line opens the integral between positive and negative values of z, the second line substitutes (69), the third line combines the terms from the two integrals, and the fourth line opens the integral between $z \leq 1$ and $z \geq 1$. The fifth line is the crucial step, which uses the following two facts: (i) $z^2 > (<)1$ for $z > (<)1$, and (ii) $g_s(z) - g_s(-z)$ is non-decreasing for all z. Therefore, substituting $g_s(z) - g_s(-z)$ by its upper bound where the term inside the integral is negative, and by its lower bound where it is positive, lowers the value of the integrand. The sixth line combines terms and uses the fact that

$$\int_0^\infty \left(z^2-1\right)\varphi(z)\,dz = [-z\varphi(z)]_0^\infty = 0.$$

■

Lemma 12 shows that the solution of the relaxed program also solves the firm's program if the effort cost is sufficiently convex, i.e. the FOA is valid.

Lemma 12 *Suppose* $C''(e) \geq \sum_s \phi^s_{e^*} h_s(\bar{e}) \frac{\sum_s \frac{\phi^s_e}{\sigma^3_s}(h'_s(e))^2}{\sum_s \frac{\phi^s_{e^*}}{\sigma_s}}$ *for all* $e \in [0, \bar{e}]$. *Then, the solution of the firm's program coincides with the solution of the relaxed program.*

Proof. The manager's utility from choosing effort e is:

$$U(e; \{q^*_s\}, \lambda) := \sum_s \phi^s_e \int \ln \left[\bar{W} + \frac{\lambda}{\sigma^2_s} \cdot \max\{q - q^*_s,\ 0\} \right] \frac{1}{\sigma_s} \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) dq - C(e).$$

We know from previous results that $0 < \lambda < \frac{\sqrt{2\pi} \sum_s \phi^s_{e^*} h_s(\bar{e})}{\sum_s \frac{\phi^s_{e^*}}{\sigma_s}}$. The FOA is valid if

$$\frac{\partial^2 U}{\partial e^2}(e; \{q^*_s\}, \lambda) \leq 0$$

for all $e \in [0, \bar{e}]$, all $q^*_s \in \mathbb{R}$, and $\lambda \in \left(0, \frac{\sqrt{2\pi} \sum_s \phi^s_{e^*} h_s(\bar{e})}{\sum_s \frac{\phi^s_{e^*}}{\sigma_s}} \right)$. Differentiating gives

$$\begin{aligned}
\frac{\partial^2 U}{\partial e^2} &= \textstyle\sum_s \int \ln \left[\bar{W} + \frac{\lambda}{\sigma^2_s} \cdot \max\{q - q^*_s,\ 0\} \right] \frac{1}{\sigma_s} \frac{d^2}{de^2} \left[\phi^s_e \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) \right] dq - C''(e) \\
&= \textstyle\sum_s \int \ln \left[\bar{W} + \frac{\lambda}{\sigma^2_s} \cdot \max\{q - q^*_s,\ 0\} \right] \frac{1}{\sigma_s} \frac{d}{de} \left[\frac{d\phi^s_e}{de} \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) + \phi^s_e \frac{d}{de} \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) \right] dq - C''(e) \\
&= \textstyle\sum_s \int \ln \left[\bar{W} + \frac{\lambda}{\sigma^2_s} \cdot \max\{q - q^*_s,\ 0\} \right] \frac{1}{\sigma_s} \left[\frac{d^2\phi^s_e}{de^2} \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) + 2 \frac{d\phi^s_e}{de} \frac{d}{de} \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) \right. \\
&\quad \left. + \phi^s_e \frac{d^2}{de^2} \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) \right] dq - C''(e) \\
&= \textstyle\sum_s \frac{d^2\phi^s_e}{de^2} \int \ln \left[\bar{W} + \frac{\lambda}{\sigma^2_s} \cdot \max\{q - q^*_s,\ 0\} \right] \frac{1}{\sigma_s} \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) dq \\
&\quad + 2 \textstyle\sum_s \frac{d\phi^s_e}{de} \frac{h'_s(e)}{\sigma^2_s} \int \ln \left[\bar{W} + \frac{\lambda}{\sigma^2_s} \cdot \max\{q - q^*_s,\ 0\} \right] \frac{q - h_s(e)}{\sigma_s} \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) dq \\
&\quad + \textstyle\sum_s \frac{\phi^s_e}{\sigma^2_s} \int \ln \left[\bar{W} + \frac{\lambda}{\sigma^2_s} \cdot \max\{q - q^*_s,\ 0\} \right] \frac{1}{\sigma_s} \\
&\quad \times \left[(h'_s(e))^2 \left[\frac{(q - h_s(e))^2}{\sigma^2_s} - 1 \right] + h''_s(e)(q - h_s(e)) \right] \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) dq - C''(e)
\end{aligned}, \tag{70}$$

where the last equality uses the fact that

$$\frac{d^2}{de^2} \left[\varphi \left(\frac{q - h_s(e)}{\sigma_s} \right) \right] = \frac{1}{\sigma^2_s} \left[(h'_s(e))^2 \left(\frac{(q - h_s(e))^2}{\sigma^2_s} - 1 \right) + h''_s(e)(q - h_s(e)) \right] \varphi \left(\frac{q - h_s(e)}{\sigma_s} \right).$$

First, with ϕ_e^s linear in e (assumption (ii) in Proposition 5), $\frac{d^2\phi_e^s}{de^2} = 0\,\forall s$, so that the first term on the RHS of (70) is zero.

Second, the second term on the RHS of (70) can be rewritten as:

$$\begin{aligned}
& 2\sum_s \frac{d\phi_e^s}{de}\frac{h_s'(e)}{\sigma_s^2}\int \ln\left[\bar{W} + \frac{\lambda}{\sigma_s^2}\cdot\max\{q - q_s^*,\ 0\}\right]\frac{q - h_s(e)}{\sigma_s}\varphi\left(\frac{q - h_s(e)}{\sigma_s}\right)dq \\
= \; & 2\sum_s \frac{d\phi_e^s}{de}\frac{h_s'(e)}{\sigma_s^2}\left[\int_{-\infty}^{q_s^*}\ln\left(\bar{W}\right)\frac{q - h_s(e)}{\sigma_s}\varphi\left(\frac{q - h_s(e)}{\sigma_s}\right)dq\right. \\
& \left. + \int_{q_s^*}^{\infty}\ln\left(\bar{W} + \frac{\lambda}{\sigma_s^2}(q - q_s^*)\right)\frac{q - h_s(e)}{\sigma_s}\varphi\left(\frac{q - h_s(e)}{\sigma_s}\right)dq\right],
\end{aligned} \tag{71}$$

where $q_s^* = \sigma_s^2\frac{\bar{W} - \mu_{IR}}{\lambda} + h_s(e^*)$. For a given e, letting $\zeta_s := \frac{q - h_s(e)}{\sigma_s}$ and $\zeta_s^* := \frac{q_s^* - h_s(e)}{\sigma_s}$, we have:

$$\begin{aligned}
& \int \ln\left[\bar{W} + \frac{\lambda}{\sigma_s^2}\cdot\max\{q - q_s^*,\ 0\}\right]\frac{q - h_s(e)}{\sigma_s}\frac{1}{\sigma_s}\varphi\left(\frac{q - h_s(e)}{\sigma_s}\right)dq \\
= \; & \int \ln\left[\bar{W} + \frac{\lambda}{\sigma_s}\cdot\max\{\zeta - \zeta_s^*,\ 0\}\right]\zeta\varphi(\zeta)\,d\zeta \\
= \; & \int_{-\infty}^{\zeta_s^*}\ln\left(\bar{W}\right)\zeta\varphi(\zeta)\,d\zeta + \int_{\zeta_s^*}^{\infty}\ln\left[\bar{W} + \frac{\lambda}{\sigma_s}(\zeta - \zeta_s^*)\right]\zeta\varphi(\zeta)\,d\zeta \geq 0,
\end{aligned} \tag{72}$$

where the inequality follows from $\bar{W} \geq 1$ and the symmetry of the normal distribution. This shows that, in equation (71), the term in brackets is increasing in $h_s(e)$ and in $h_s'(e)$, and decreasing in σ_s, all else equal. Note that, as $\sum_s \phi_e^s = 1\,\forall e$, we have $\sum_s \frac{d\phi_e^s}{de} = 0$, which implies

$$\sum_{s|\frac{d\phi_e^s}{de}>0}\frac{d\phi_e^s}{de} = -\sum_{s|\frac{d\phi_e^s}{de}<0}\frac{d\phi_e^s}{de}.$$

In sum, with assumption (iii), the expression in (71) is negative.

Third, we now show that the third term on the RHS of (70) is negative. With $\phi_e^s \geq 0$ and $\sigma_s > 0$ for all s, with $h_s''(e) \leq 0$ for all s (assumption (i)), and with equation (72), we have:

$$\sum_s \frac{\phi_e^s}{\sigma_s}h_s''(e)\int \ln\left[\bar{W} + \frac{\lambda}{\sigma_s^2}\cdot\max\{q - q_s^*,\ 0\}\right]\frac{q - h_s(e)}{\sigma_s}\frac{1}{\sigma_s}\varphi\left(\frac{q - h_s(e)}{\sigma_s}\right)dq \leq 0. \tag{73}$$

Moreover:

$$\begin{aligned}
& \textstyle\sum_s \frac{\phi_e^s}{\sigma_s^2} \int \ln\left[\bar{W} + \frac{\lambda}{\sigma_s^2}\cdot \max\left\{q - q_s^*,\ 0\right\}\right]\left[(h_s'(e))^2\left[\frac{(q-h_s(e))^2}{\sigma_s^2} - 1\right] + h_s''(e)\left(q - h_s\left(e\right)\right)\right]\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)dq \\
= & \textstyle\sum_s \frac{\phi_e^s}{\sigma_s^2} \int \ln\left[\bar{W} + \frac{\lambda}{\sigma_s^2}\cdot \max\left\{q - q_s^*,\ 0\right\}\right](h_s'(e))^2\left[\frac{(q-h_s(e))^2}{\sigma_s^2} - 1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)dq \\
& + \textstyle\sum_s \frac{\phi_e^s}{\sigma_s} h_s''(e) \int \ln\left[\bar{W} + \frac{\lambda}{\sigma_s^2}\cdot \max\left\{q - q_s^*,\ 0\right\}\right]\frac{q-h_s(e)}{\sigma_s}\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)dq \\
\leq & \textstyle\sum_s \frac{\phi_e^s}{\sigma_s^2}(h_s'(e))^2 \int \left[\bar{W} + \frac{\lambda}{\sigma_s^2}\cdot \max\left\{q - q_s^*,\ 0\right\}\right]\left[\frac{(q-h_s(e))^2}{\sigma_s^2} - 1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)dq \\
= & \textstyle\sum_s \lambda\frac{\phi_e^s}{\sigma_s^4}(h_s'(e))^2 \cdot \int \max\left\{q - q_s^*,\ 0\right\}\left[\frac{(q-h_s(e))^2}{\sigma_s^2} - 1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)dq \\
= & \textstyle\sum_s \lambda\frac{\phi_e^s}{\sigma_s^4}(h_s'(e))^2 \cdot \int_{q_s^*}^{\infty}\left(q - q_s^*\right)\left[\frac{(q-h_s(e))^2}{\sigma_s^2} - 1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)dq,
\end{aligned}$$

where the first equality separates the sum into two components, the inequality that follows uses the result from Lemma 11 and equation (73), the next equality follows from

$$\int\left[\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)^2 - 1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)dq = 0,$$

(a standard normal variable has variance 1), and the last equality opens the max operator. Substituting in the expression from (70), we obtain the following sufficient condition for the validity of the FOA:

$$\lambda\sum_s \frac{\phi_e^s}{\sigma_s^4}\left(h_s'\left(e\right)\right)^2 \cdot \int_{q_s^*}^{\infty}\left(q - q_s^*\right)\left[\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)^2 - 1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)dq \leq C''(e) \tag{74}$$

for all $e \in [0, \bar{e}]$, $q_s^* \in \mathbb{R}$, and $\lambda \in \left(0, \frac{\sqrt{2\pi}\sum_s \phi_{e^*}^s h_s(\bar{e})}{\sum_s \frac{\phi_{e^*}^s}{\sigma_s}}\right)$. Let

$$\xi_s(q_s^*) := \int_{q_s^*}^{\infty}\left(q - q_s^*\right)\left[\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)^2 - 1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)dq.$$

We claim that

$$\xi_s'(q_s^*)\left\{\begin{array}{c} > \\ < \end{array}\right\}0 \iff q_s^*\left\{\begin{array}{c} < \\ > \end{array}\right\}h_s\left(e\right). \tag{75}$$

Differentiating yields:

$$\xi_s'(q_s^*) = -\int_{q_s^*}^{\infty}\left[\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)^2 - 1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q - h_s\left(e\right)}{\sigma_s}\right)dq. \tag{76}$$

Note that

$$\frac{d}{dq}\left[-\left(\frac{q-h_s(e)}{\sigma_s}\right)\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)\right] = -\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)-\left(\frac{q-h_s(e)}{\sigma_s}\right)\frac{1}{\sigma_s}\varphi'\left(\frac{q-h_s(e)}{\sigma_s}\right)$$
$$= \left[\left(\frac{q-h_s(e)}{\sigma_s}\right)^2-1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right),$$

where the last equality uses the fact that $\varphi'(q) = -q\varphi(q)$. Therefore,

$$\int\left[\left(\frac{q-h_s(e)}{\sigma_s}\right)^2-1\right]\frac{1}{\sigma_s}\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)dq = -\left(\frac{q-h_s(e)}{\sigma_s}\right)\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right).$$

Substituting back into (76) gives:

$$\xi_s'(q_s^*) = -\left(\frac{q_s^*-h_s(e)}{\sigma_s}\right)\varphi\left(\frac{q_s^*-h_s(e)}{\sigma_s}\right)\left\{\begin{array}{c}>\\<\end{array}\right\}0 \iff q_s^*\left\{\begin{array}{c}<\\>\end{array}\right\}h_s(e).$$

Therefore, $\xi_s(\cdot)$ is maximized at $q_s^* = h_s(e)$, so that, by condition (74), it suffices to show that

$$\lambda\sum_s\frac{\phi_e^s}{\sigma_s^4}\left(h_s'(e)\right)^2\cdot\xi_s(h_s(e)) \le C''(e), \tag{77}$$

for all $e \in [0,\bar{e}]$ and $\lambda \in \left(0, \frac{\sqrt{2\pi}\sum_s \phi_{e^*}^s h_s(\bar{e})}{\sum_s \frac{\phi_{e^*}^s}{\sigma_s}}\right)$. Evaluating ξ_s at $h_s(e)$ gives:

$$\xi_s(h_s(e)) = \int_{h_s(e)}^{\infty}\frac{q-h_s(e)}{\sigma_s}\left[\left(\frac{q-h_s(e)}{\sigma_s}\right)^2-1\right]\varphi\left(\frac{q-h_s(e)}{\sigma_s}\right)dq.$$

Performing the change of variables $z_s \equiv \frac{q-h_s(e)}{\sigma_s}$, we obtain

$$\xi_s(h_s(e)) = \sigma_s\int_0^{\infty} z\left(z^2-1\right)\varphi(z)\,dz. \tag{78}$$

Integrating by parts gives

$$\int z\left(z^2-1\right)\varphi(z)\,dz = -z^2\varphi(z)+\int z\varphi(z)\,dz,$$

where we let $(z^2-1)\,\varphi(z)\,dz = dv$ so that $v = -z\varphi(z)$, and we let $u = z$, so that $du = dz$.

Therefore

$$\int_0^\infty z\left(z^2-1\right)\varphi\left(z\right)dz=\int_0^\infty z\varphi\left(z\right)dz.$$

Using $\frac{d}{dz}\left[-\varphi(z)\right]=z\varphi\left(z\right)$, we have

$$\int_0^\infty z\left(z^2-1\right)\varphi\left(z\right)dz=\left[-\varphi(z)\right]_0^{+\infty}=\varphi(0)=\frac{1}{\sqrt{2\pi}}.$$

Substituting into (78), yields

$$\xi_s(h_s(e))=\frac{\sigma_s}{\sqrt{2\pi}}.$$

Substituting into (77), we obtain the following sufficient condition:

$$\frac{\lambda}{\sqrt{2\pi}}\sum_s\frac{\phi_e^s}{\sigma_s^3}\left(h_s'\left(e\right)\right)^2\leq C''(e),$$

which is true for all $e\in[0,\bar{e}]$ and all $\lambda\in\left(0,\frac{\sqrt{2\pi}\sum_s\phi_{e^*}^s h_s(\bar{e})}{\sum_s\frac{\phi_{e^*}^s}{\sigma_s}}\right)$ if

$$\sum_s\phi_{e^*}^s h_s\left(\bar{e}\right)\frac{\sum_s\frac{\phi_e^s}{\sigma_s^3}\left(h_s'\left(e\right)\right)^2}{\sum_s\frac{\phi_{e^*}^s}{\sigma_s}}\leq C''(e)\ \ \forall e\in[0,\bar{e}].$$

■