The No-Upward-Crossing Condition, Comparative Statics, and the Moral-Hazard Problem

Hector Chade* and Jeroen Swinkels†

September 2019

Abstract

We define and explore the No-Upward-Crossing (NUC), a condition satisfied by every parameterized family of distributions commonly used in economic applications. Under smoothness assumptions, NUC is equivalent to log-supermodularity of the negative of the derivative of the distribution with respect to the parameter. It is characterized by a natural monotone comparative static, and is central in establishing quasi-concavity in a family of decision problems. As an application, we revisit the first-order approach to the moral hazard problem. NUC simplifies the relevant conditions for the validity of the first-order approach and gives them an economic interpretation. We provide extensive analysis of sufficient conditions for the first-order approach for exponential families.

Keywords. Log-supermodularity, Quasi-concavity, Moral Hazard, First-Order Approach.

*Arizona State University, `hector.chade@asu.edu`

†Northwestern University, `j-swinkels@northwestern.edu`

We are grateful to John Conlon, Ian Jewitt, Ohad Kadan, a Co-Editor, and three anonymous referees for very helpful comments and suggestions that significantly improved the paper.

1 Introduction

In this paper, we introduce, motivate, and show the usefulness of a condition on a parameterized family of distributions. We term the condition "No-Upward-Crossing," or *NUC*, a choice of terminology that we will justify shortly. *NUC* is satisfied for every distribution that we are aware of that is commonly used in economic applications. Indeed, *NUC* holds if and only if an intuitive comparative static holds, and hence, while one can construct examples where *NUC* fails, such examples are necessarily somewhat artificial. *NUC* simplifies the analysis of some important economic problems, including the question of when the first-order condition is sufficient for a global optimum in a variety of problems, and when some natural comparative statics results hold. *NUC* also allows for economic interpretations of otherwise hard-to-interpret technical conditions.

To be concrete, consider the family of distributions $\{F(\cdot|a)\}_{a\in A}$ on the reals, parameterized by $a \in A \subset \mathbb{R}$, where we speak of the argument x as the "outcome" and often refer to a as the "effort." We assume that F satisfies strict first-order stochastic dominance (*FOSD*). That is, increases in effort strictly decrease the probability of an outcome below any given interior output.

To see the motivation and definition of *NUC*, consider two ordered pairs of effort, $a_l < a'_l$ and $a_h < a'_h$ where $a_l \leq a_h$ and $a'_l \leq a'_h$ (and where in the interesting case, at least one of the later two inequalities is strict). For example, a might be the amount of exercise a subject gets, where a_l is "completely sedentary" and a'_l is "occasionally goes for a stroll," while a_h is "walks regularly" and a'_h is "jogs on a regular basis." Let x be the number of miles the subject is able to cover on foot in a particular thirty-minute period.

Now, fix a threshold t, and compare the probabilities of various types coming in above or below the threshold. Consider first a low threshold t, say one mile. Then, both the walker and the jogger will almost surely exceed t, and so $F(t|a_h) - F(t|a'_h)$ will be small. But, there is likely to be a significant increase in the probability that the occasional stroller versus the sedentary subject exceeds the threshold, so that $F(t|a_l) - F(t|a'_l)$ is larger. On the other hand, if we take the threshold t to be three miles, then the opposite will hold, as covering three miles in thirty minutes is probably almost impossible for either of the unfit types, but more likely to discriminate between the two fitter types. Driven by this intuition, the condition we impose, *NUC*, is simply that the ratio of these two probability differences rises with the threshold. That is, *NUC holds if*

$$\frac{F(t|a_h) - F(t|a'_h)}{F(t|a_l) - F(t|a'_l)}$$

is increasing in t.

When F is twice-continuously differentiable, *NUC* holds *if and only if* $-F_a$ (which is positive by *FOSD*) is log-supermodular (*lsm*) in output and effort. Since $-F_a$ is a measure of the marginal return to effort, this characterization says that if extra effort has diminishing marginal returns at given x, then in proportionate terms these returns fall more slowly above x. If extra effort has increasing marginal returns at x, then in proportionate terms they rise even faster above x.

The condition that $-F_a$ is *lsm* is in turn equivalent to the condition that for each real number τ, and for each effort a, $F_{aa}(\cdot|a) - \tau F_a(\cdot|a)$ *is never first strictly negative and then strictly positive.* This version of *NUC* is useful in many applications, and the property is the genesis of the name.[1]

We next turn to sufficient conditions for *NUC*. These are useful in practice, both to check *NUC* and to tie *NUC* to well-known classes of distributions, and also help to build our intuition. We provide a set of such conditions, interpret them economically, and show that exponential families satisfy the most stringent such condition. Along the way, we show that *NUC* is automatic for any distribution that satisfies the monotone likelihood ratio property (*MLRP*) and is totally positive of order three (TP_3).[2]

In information economics it is so standard as to be barely commented on to assume *MLRP*, which is that the density or probability distribution f associated with F is *lsm*. This is not because *MLRP* is without loss of generality. Rather, *MLRP* is invoked because it both has a clear economic motivation and simplifies the analysis. We think of *NUC* in the same way. As for *MLRP*, *NUC* is not without loss (although we argue the loss is pretty mild). But, as for *MLRP*, we will show that *NUC* has a clear economic motivation and usefully simplifies both analysis and interpretation. Indeed, *NUC* is *MLRP*'s fraternal twin. To see the family relationship, note that, when F is differentiable, *MLRP* is the condition that F_x is *lsm*, and hence each condition is a log-supermodularity condition on one of the derivatives of F.

One central reason for our belief in the naturalness of *NUC* is that it holds if and only if an intuitive comparative static holds, one somewhat related to a statistical thought experiment discussed by Jewitt, Kadan, and Swinkels (2008). Consider a university department with junior faculty of varying ability who face a tenure standard that depends on whether their research output exceeds some given threshold t, where t is set exogenously, for example, by the central administration. The department gets a negative payoff when low-ability faculty are tenured, and a positive payoff when their ability is above some threshold. It can offer the junior faculty more or less aid in their research–mentoring, research assistants, equipment–with research output being stochastically distributed according to $F(\cdot|\alpha(a, \delta))$, where α is an increasing function of ability, a, and the amount of aid received, δ. The department faces the trade-off that more aid makes it more likely that high-ability faculty make the tenure threshold, which the department likes, but also makes it more likely that low-ability faculty make the hurdle, which the department dislikes.

The comparative static that we want is that *if the tenure standard is raised, then the optimal amount of research aid to offer does not go down.* This seems to us very intuitive: When the

[1] In the moral hazard context, a weaker version of this condition appears in one result in Jung and Kim (2015a) as sufficient to justify the first-order approach. We discuss this paper further below.

[2] TP_3 has appeared in the literature on the first-order approach to the moral hazard problem, most prominently in Jewitt (1988), p.1182 (where one can find a definition and discussion), and also in Jung and Kim (2015a).

tenure threshold is low, the high-ability are likely to exceed the threshold without help, and so giving aid predominantly helps the low-ability. Hence, the optimal amount of aid is low. When the threshold is higher, the low-ability are unlikely to exceed the threshold even with substantial aid, but aid may well lift the high-ability above the threshold. Thus, the optimal level of aid will be higher. We show that this comparative static–that optimal aid rises with the tenure standard–will hold for all relevant settings if and only if *NUC* holds.

We also explore when *NUC* might fail.[3] At a mathematical level, such examples are easy to construct–indeed, we provide a recipe for doing so. Exploring examples suggests that failing *NUC* while satisfying *MLRP* is hard. We show that *NUC* fails most naturally in a situation where F is a mixture of two distributions, where each satisfies *NUC*, but the mixture does not.[4]

A main application of *NUC* is the following one. Imagine that one receives utility $v(x)$ from outcome x, but that effort comes at some utility cost $c(a)$, so that one wishes to maximize $U(a) = \int v(x)dF(x|a) - c(a)$ by choice of a.[5] The central question is under what conditions is U strictly quasi-concave, so that a solution to $U_a(a) = 0$ is also a global maximizer of U.[6]

NUC very much simplifies the analysis of this problem. We show that where $U_a = 0$,

$$U_{aa}(a) = -\int v'(x)\left(F_{aa}(x|a) - \tau F_a(x|a)\right)dx,$$

where $\tau = c_{aa}(a)/c_a(a)$. By one of the equivalent characterizations of *NUC*, $F_{aa}(\cdot|a) - \tau F_a(\cdot|a)$ does not go from negative to positive. Hence, if v is increasing and concave, then v' puts more weight on $F_{aa}(\cdot|a) - \tau F_a(\cdot|a)$ where it is positive and less where it is negative. But then, a sufficient condition for U to be strictly quasi-concave is that $\int(F_{aa}(x|a) - \tau F_a(x|a))dx > 0$, which by integration by parts is equivalent to

$$\frac{\left(\mathbb{E}[x|a]\right)_{aa}}{\left(\mathbb{E}[x|a]\right)_a} < \frac{c_{aa}(a)}{c_a(a)}.$$

Hence, a *single* integral involving $F_{aa} - \tau F_a$ needs to be checked, and the relevant inequality has the simple and clean economic interpretation that expected output is (in proportionate terms) less convex in effort than c. Obviously, if F_{aa} is positive and c convex in a, then the result is immediate.[7] A convenient implication of *NUC* is that an almost equally simple argument yields U_{aa} negative without convexity requirements in F and c.

If v is not concave, then we show that it is enough to find a strictly increasing and differentiable

[3]We are very grateful to two referees who helped us in thinking about the intuition for when *NUC* might fail.

[4]This is related to the fact that the mixture of two distributions each satisfying *MLRP* need not itself do so.

[5]In this paper, an integral $\int$ without delimiters is understood to be over the entire range of relevant values.

[6]We will actually analyze a slightly more general setup where there is an unknown state θ distributed according to Γ, and thus the decision maker chooses an action a (independent of θ) that maximizes $\int\int v(x,\theta)dF(x|a,\theta)d\Gamma(\theta) - c(a)$. The conditions we are about to describe must then apply pointwise for each θ.

[7]In the moral hazard context, Rogerson (1985) showed that F convex in a validates the first-order approach.

function q such that v'/q' is decreasing, and such that the expectation of q is less convex than c. This generalization turns out to be very useful when we turn to the moral hazard problem.

One way to satisfy these assumptions is to assume expected output is concave in effort and expected costs are convex. We view this as unnecessarily restrictive. To see why, note first that there is no reason why the economically natural way of writing down such a problem will lead to a convex c. For instance, one might think about effort expended on a given day on writing a paper, the cost per minute of effort will initially decrease as one "gets into" the problem and then eventually strongly increase, yielding a c which is initially concave and then convex. One could also imagine a setting where F does not lead to a concave expected outcome, but one is willing to assume sufficient convexity in c as to overcome this.

A highly relevant application of the results just described is the question of the validity of the first-order approach (*FOA*) in the classical moral hazard problem (Holmstrom (1979), Mirrlees (1975)) in which a risk-averse agent chooses effort but a principal can see–and reward–only a noisy signal of that effort.[8] Implementing a specific effort by the agent requires the design of an optimal contract that deters *all* possible deviations, a decidedly intractable problem. The *FOA* focuses on the relaxed problem that considers only local deviations in effort. The question is when a solution to this problem satisfies all of the omitted constraints. But that is exactly the problem considered above, where the extra interest comes because v is itself endogenous.

We first show how a version of the central result of Jewitt (1988) falls out as an immediate corollary to our analysis. In particular, mimicking Jewitt, under the right conditions on the curvature of the agent's utility function and likelihood ratio, the agent's utility from income, v, is concave in output for the optimal contract solving the relaxed problem for any given effort. But then, from above, the agent's expected utility is a quasi-concave function of effort as long as expected output is less convex in effort than c, and so the *FOA* is validated.

Except for the fact that we incorporate the curvature of c, our goal here is not to generalize Jewitt (1988).[9] Rather, it is to show how *NUC* simplifies both the application and interpretation of his central result. First, for each a, rather than a continuum of integrals (of the form $\int_{\underline{x}}^{x} F_{aa}(s|a)ds$ for each x), only a single integral must be checked. Second, as discussed above, this integral has the correct sign if and only if expected output is less convex in effort than c, and

[8] If the talent of the agent is unknown, as in the standard two-period career-concern model of Dewatripont, Jewitt, and Tirole (1999) without explicit contracts, then the setting described in footnote 6 applies.

[9] Brown, Chiang, Ghosh, and Wolfstetter (1986) provide some limiting curvature and complementarity conditions on the agent's utility for wage and effort (not necessarily additively separable). Simultaneous work by Jung and Kim (2015b) also looks at the curvature of c, but follows a different approach, which relies on a double-crossing property between the agent's utility for income and disutility of effort. Note that an alternative to including the curvature of c is to linearize it by rescaling effort, thus folding any curvature of c into F. But, while this is conceptually straightforward, inverting the relevant cost function may be intractable and the resulting conditions hard to check and interpret. Forcing c to be linear also makes it essentially impossible to build a model where the expected cost to the principal of inducing effort is continuous at zero effort, since effort zero can be implemented with a flat contract, while, since the marginal cost of effort is positive at zero, implementing any positive effort requires imposing a strictly positive amount of risk on the agent.

hence the condition takes on a simple economic interpretation.

We then turn to a version of a central result of Jung and Kim (2015a), who focus on the distribution of the likelihood ratio rather than on the distribution of the outcome itself. Indeed, *fix* any given action $\hat{a}$, and consider the likelihood ratio function $\hat{l}$ evaluated at that $\hat{a}$. If the expectation of the function $\hat{l}$ is concave in effort (that is, $\hat{l}$ continues to be evaluated at $\hat{a}$, but the expectation is taken with respect to $f(\cdot|a)$ as a varies), then we show that as an immediate consequence of the construction involving v'/q', one can drop Jewitt's condition on the shape of the likelihood ratio.[10] Further, the concavity condition on the expected likelihood ratio is weaker than that on the expected outcome. Here again, our role is not to generalize Jung and Kim (2015a), but to show how *NUC* clarifies the analysis.

We are not the first to make the observation that F_{aa} may be well-behaved, and that this can simplify checking the Jewitt conditions or the Jung and Kim variations to them. Indeed, the related idea that it is useful that F_{aa} crosses zero appropriately has already appeared in Jung and Kim (2015a), as one of the conditions that help to justify the *FOA* (see their Proposition 7 and Lemma 2), by showing that the sufficient conditions in Jewitt (1988) hold. Less directly, the observation is at the heart of the simplification that Jewitt (1988) (Corollary 1) makes when discussing exponential families (which satisfy *NUC*). Our main contribution is to explore the considerable generality with which *NUC* holds, and to explore and understand its foundations and implications. Also, our analysis shows that, exploiting *NUC*, one can show *directly* that the *FOA* is valid, without the need to show that (the continuum of integrals in) Jewitt's conditions hold. This has pedagogical value, since our proof of the validity of *FOA* is only slightly more difficult than that of Rogerson (1985) but without the restrictive convexity of F.

The final part of out paper is devoted to a deeper exploration of the exponential families. In particular, we examine the question of when the expectation of the likelihood ratio is indeed less convex in effort than c. Since exponential families subsume many of the most common distributions used in applications, and have a number of other desirable properties for the modeler, a complete off-the-shelf result on the *FOA* in this setting is of considerable practical use.

The paper proceeds as follows. Section 2 presents the model. Section 3 defines *NUC*, characterizes it for twice-continuously differentiable functions, and presents several sufficient conditions for *NUC*. Section 4 discusses the relationship between *NUC* and comparative statics. Section 5 shows how *NUC* aids in the analysis of the quasi-concavity of the objective function of an optimization problem. Section 6 discusses when *NUC* fails. Section 7 applies those results to the validity of the *FOA* in the moral-hazard problem. Finally, Section 8 discusses exponential families. Proofs omitted from the main text are in the Appendix.

[10] This condition has a slightly less economically intuitive interpretation, but extends the analysis beyond a concave likelihood function.

2 The Setting

Let A be a subset of $\mathbb{R}$ and let X be an interval of $\mathbb{R}$ with infimum $\underline{x}$ and supremum $\bar{x}$ in the extended reals. We will often speak of $a \in A$ as the effort taken by an agent and $x \in X$ as the outcome, reflecting that one central use of our ideas will be in principal-agent settings. The outcome x has cumulative distribution (cdf) conditional on a, $F(\cdot|a) : X \to [0, 1]$, with, as appropriate, density or probability distribution function $f(\cdot|a)$.[11] We assume that $F_a(x|\cdot) < 0$ for all interior x, so that first-order stochastic dominance (*FOSD*) strictly holds. Regarding f, we will occasionaly also assume the stronger condition that f is strictly log-supermodular (*lsm*) in a and x (or equivalently, satisfies strict *MLRP*), so that $f(\cdot|a_h)/f(\cdot|a_l)$ is strictly increasing in x when $a_h > a_l$. When f is differentiable in a, this is equivalent to $l(\cdot|a) \equiv f_a(\cdot|a)/f(\cdot|a)$ being strictly increasing in x. We willl be very explicit when we impose *MLRP*.[12]

3 The No-Upward-Crossing Condition

Say that F satisfies *No Upward Crossing* (*NUC*) if for all $\{a_l, a'_l, a_h, a'_h\}$ with $a_l < a'_l$, $a_h < a'_h$, $a_l \leq a_h$, and $a'_l \leq a'_h$,

$$\frac{F(x|a'_h) - F(x|a_h)}{F(x|a'_l) - F(x|a_l)} \tag{1}$$

is increasing in x (recall that we use increasing in the weak sense). We will justify the choice of nomenclature below.

In words, F satisfies *NUC* if the ratio by which an increase in action from a_h to a'_h versus an increase in action from a_l to a'_l affects F goes up with the outcome x. That is, changes in effort between two lower effort levels matter relatively more to F at low outcomes, while changes in effort between two higher effort levels matter relatively more at high outcomes.

3.1 *NUC* for Continuous Distributions

Under some smoothness assumptions on F, *NUC* has a simple characterization.

Proposition 1 *Let A and X be intervals, and F be $\mathcal{C}^2$. Then the following are equivalent:*

(i) NUC,

[11] When f is a density, we will freely impose high-order differentiability assumptions on f and F in the interest of simplicity and clarity, although many of the results of the paper rely on less. We also assume that the functions f_a, f_{aa}, F_a, and F_{aa} are integrable, and take for granted the validity of interchanging differentiation and integration, which can be justified under mild conditions (see Chade and Swinkels (2016)).

[12] We use increasing, decreasing, convex, concave, etc., in the weak sense, adding 'strictly' when appropriate. A twice continuously differentiable real-valued function g with domain on a rectangle of the plane is supermodular if $\partial^2 g/\partial x \partial y \geq 0$, and it is log-supermodular if $\log g$ is supermodular. We use the notation $g \geq 0$, etc., to mean that $g(x) \geq 0$ for all x. For simplicity, we will often omit the argument of functions when it causes no confusion. We will also use $=_s$ to mean "has strictly the same sign as."

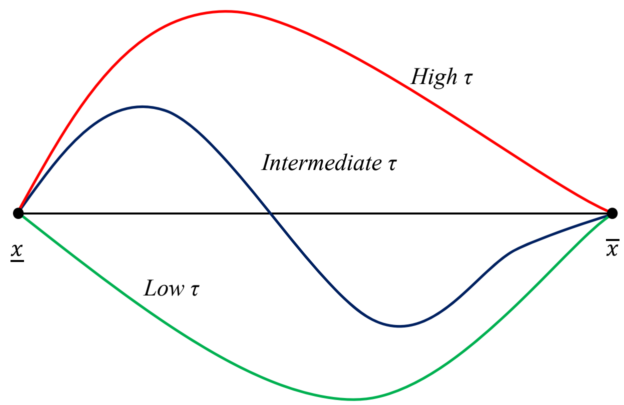

Figure 1: ***NUC***. The figure depicts the behavior of the function $z(\cdot, \tau, a)$ for three different levels of τ: depending on τ it can be always negative, always positive, or first positive and then negative. In all cases, it starts and ends at zero.

(ii) $-F_a$ *is* log-*supermodular in* a *and* x*, and,*

(iii) for each $a \in A$*, and* $\tau \in \mathbb{R}$*,* $F_{aa}(\cdot|a) - \tau F_a(\cdot|a)$ *never crosses zero from below on the interior of* X*.*

The proof is in Appendix A.1. To show that (i) and (ii) are equivalent one expresses the ratio in (1) as a ratio of integrals of $-F_a$, and then shows that the derivative of this ratio with respect to x has the sign of the difference of the expectation of $f_a(x|\cdot)/F_a(x|\cdot)$ over $[a_h, a'_h]$ and $[a_l, a'_l]$ with respect to the (artificial) density on a given by

$$\xi(a) = \frac{-F_a(x|a)}{\int(-F_a(x|s))ds}.$$

Since $a'_h \geq a'_l$ and $a_h \geq a_l$ (with at least one inequality strict except in the trivial case) this difference is always positive if and only if $f_a(x|\cdot)/F_a(x|\cdot)$ is increasing, which is equivalent to $-F_a$ *lsm*.

Condition (iii) is very useful in applications, and is the progenitor of the term *NUC*. Under *NUC*, $z(\cdot, \tau, a) \equiv F_{aa}(\cdot|a) - \tau F_a(\cdot|a)$ has only three possible sign patterns: for any given a, and depending on τ, it can be everywhere positive, everywhere negative, or first positive and then negative (see Figure 1). It has, in particular, no upward crossing on the interior of X.

Condition (ii) helps in our intuitive understanding of *NUC*. Note that $-F_a$ is the amount by which extra effort raises the probability of an outcome above x, and thus $-F_{aa}/(-F_a)$ measures the proportionate change in the benefit of extra effort as effort is increased. By (ii), *NUC* is equivalent to this proportional change being more favorable at higher outcomes. That is, at points x where $F_{aa} > 0$, so that extra effort has diminishing marginal returns, $-F_{aa}/(-F_a)$ is becoming less negative: the diminishing returns are smaller at higher outputs. Similarly if $F_{aa} < 0$, so that there are increasing marginal returns to effort at x, then the increasing returns are yet larger at higher x.[13]

Using Proposition 1, we see that *MLRP* and *NUC* are in the same spirit. In particular, *MLRP* is the condition that F_{ax}/F_x is increasing in x, while *NUC* is the condition that F_{aa}/F_a is increasing in x. Thus, *MLRP* asks that F_x is log-supermodular in a and x, while, given *FOSD*, *NUC* asks for the same condition on $-F_a$. Each condition fails in specific examples (see below), but imposes useful regularity on the problem. Indeed, in the location families $F(x|a) = Q(x-a)$, we have $F_x(x|a) = Q'(x-a) = -F_a(x|a)$, and so $MLRP$ and NUC reduce to the same condition.

Remark 1 *As suggested by the example in the Introduction, our intuition for the monotonicity of the expression in* (1) *is strongest when $a_l' < a_h$. It can be seen from the proof of Proposition 1 (see in particular* (19)*) that even if one weakened the definition of NUC to only consider cases where $a_l' < a_h$, NUC would remain equivalent to $-F_a$ lsm. Hence, for $F \in \mathcal{C}^2$, the two definitions agree.*

3.2 Three Sufficient Conditions for *NUC*

In this section, we provide three increasingly strong conditions that imply *NUC*. We continue to assume that A and X are intervals, and take F to be $\mathcal{C}^5$. These conditions help to check *NUC* in examples, and to build intuition.

Proposition 2 *Assume strict MLRP and let F be $\mathcal{C}^5$. Then,*

$$(\log f)_{ax} \ lsm \Longrightarrow f^2 (\log f)_{ax} \ lsm \Longrightarrow F^2 (\log F)_{ax} \ lsm \Longrightarrow NUC.$$[14]

The proof is in Appendix A.2.[15] Note that $(\log f)_{ax} > 0$ is equivalent to the condition that f is strictly *lsm* (strict $MLRP$). Hence, the conditions on f in this proposition can be interpreted as repeated applications of *lsm*.

The condition $(\log f)_{ax}$ *lsm* is most stark when we consider *exponential families*. In Section 8, we show that for such families, $(\log(\log f)_{ax})_{ax} = 0$ (and indeed, that only the exponential families satisfy this condition). Leading examples of exponential families are the Exponential, Poisson, Gamma, Normal, and Beta distributions, and truncations thereof.

To interpret the condition $(\log f)_{ax}$ *lsm*, fix a, and for ε small, note that the likelihood ratio of a versus $a-\varepsilon$ given x is (recall that $l = f_a/f$ is the likelihood ratio in differential form)

$$\frac{f(x|a)}{f(x|a-\varepsilon)} \cong 1 + \varepsilon l(x|a).$$

[13] Since log-supermodularity is robust to an increasing transformation of x or a, so is *NUC*. The proof of this assertion is immediate and thus omitted.

[14] Because f is strictly *lsm*, each of the relevant objects is strictly positive on $(\underline{x}, \bar{x})$.

[15] We are very grateful to a referee who helped us towards a simpler proof of the second implication.

So, l is steep around x if and only if changes in the outcome around x provide significantly different information about a versus $a-\varepsilon$. Take $x'' > x'$. Then, since $(\log f)_{ax} = l_x$, the condition $(\log f)_{ax}$ *lsm* is equivalent to the condition that as a increases, l becomes relatively steeper at x'' versus x'. That is, as a goes up, changes in the outcome near x'' become more informative about a versus $a - \varepsilon$ relative to changes in the outcome near x'.

This is intuitive. If a is low, then *any* high outcome may be largely a matter of (good) luck, with the relative probability of these outcomes not depending much on small differences in effort. This leads to a relatively flat l at high outcomes. Conversely, when a is high, it is low outcomes that are largely a matter of (bad) luck, leading to a relatively flat l at low outcomes.

Finally, we connect *NUC* to TP_3, a condition explored by Jewitt (1988) (p. 1182).

Lemma 1 *Assume MLRP and let F be $\mathcal{C}^5$. Then,*

$$f^2 (\log f)_{ax} \text{ strictly lsm} \Longrightarrow TP_3 \Longrightarrow f^2 (\log f)_{ax} \text{ lsm}$$

Hence, by Proposition 2, *NUC* is considerably more permissive than TP_3. See Appendix A.3.

4 *NUC*, Threshold Tests, and Comparative Statics

In this section, we provide a first application of *NUC*. This application is economically relevant in its own right, and also illuminates an intuitive foundation for *NUC*. We begin with a simplified setting, and show that an intuitive comparative static holds *if and only if NUC* is in force. We then consider a more elaborate and economically realistic version of the problem, and show that strict *NUC* remains the right condition to imply the desired comparative static. Together, these results substantially strengthen our belief that *NUC* is a natural condition to impose.

A university department which has junior faculty of ability $\theta \in \{\ell, h\}$, where the probability of type ℓ is p, and the probability of type h is $1-p$. The department can offer research support $\delta \in \{0, 1\}$ to its junior faculty. Research output with ability θ and support δ is distributed according to $F(\cdot|\alpha(\theta,\delta))$, where $\alpha(\theta, 0) < \alpha(\theta, 1)$ for $\theta \in \{\ell, h\}$, and $\alpha(\ell, \delta) < \alpha(h, \delta)$ for $\delta \in \{0, 1\}$. That is, output is stochastically increased both by ability and by the level of research support. Tenure is granted if and only if output exceeds a threshold t. The department gets utility -1 from tenuring type ℓ, utility 1 from tenuring type h, and utility 0 from not tenuring, for an expected payoff of

$$\pi(t, \delta) = -p\,(1 - F(t|\alpha(\ell, \delta))) + (1-p)\,(1 - F(t|\alpha(h, \delta)))\,. \tag{2}$$

Say that preferences are *monotone* if for all pairs (p, α), the function $\pi(t, 1) - \pi(t, 0)$ never crosses zero from above at some interior t. That is, the department never wants to support faculty facing an easy tenure hurdle, but not support them when they face a harder one. As discussed in the introduction, we find monotonicity quite intuitive.

Proposition 3 *Preferences are monotone if and only if F satisfies NUC.*

Proof Using (2) at $\delta = 0$ and $\delta = 1$, and rearranging, at any interior t,

$$\pi(t,1) - \pi(t,0) =_s -\frac{p}{1-p} + \frac{F(t|\alpha(h,1)) - F(t|\alpha(h,0))}{F(t|\alpha(\ell,1)) - F(t|\alpha(\ell,0))}, \tag{3}$$

using *FOSD*. The result follows from the definition of *NUC*, since the range of $\frac{p}{1-p}$ is $[0,\infty)$. □

Remark 2 *This setting is equivalent to one in which an observer is trying to guess a state based only on the information of whether output exceeds threshold t, and is choosing which of two information environments ($\delta = 0$ or $\delta = 1$) they prefer for any given t. This is similar to a setting considered by Jewitt, Kadan, and Swinkels (2008). In this interpretation, it is intuitive to let the observer also guess according to her prior, simply ignoring whether or not the threshold was exceeded. In Appendix A.4, we show that if F is $\mathcal{C}^2$ and satisfies a log-concavity condition, then any failure of NUC allows one to choose α and p such that (a) the observer uses her information (and so optimally guesses state h if and only if the threshold is exceeded), and (b) a failure of monotonicity occurs.*

Remark 3 *This sort of situation is ubiquitous in any setting where what is reported to the decision maker is a coarsening of a raw "score". For example, in many areas restaurant goers can only see whether a given restaurant earned an "A" or "B" placard, corporate lenders can only see which of a small number of possible ratings a borrowing firm received, a professional school may provide only a coarse report of student performance, and Amazon provides customers only some of the information it uses to certify the vendors using its platform. Our condition corresponds to a situation where if the threshold score for a high rating goes up, the users of the information do not switch their preferences from an easier inspection/grading/certification system to a harsher one.*

To see that *NUC* is really the "right" condition for this sort of problem, let us elaborate our base setting. Let θ have arbitrary distribution Γ (atomic or otherwise), let $\delta \in [0,1]$, and let $v(\theta)$ be the value to the department of tenuring a faculty member of ability θ, where we assume v single crosses zero from below (we do not need the natural but stronger condition that v is increasing). Let output for the faculty member be distributed as $F(\cdot|\alpha(\theta,\delta))$, where α_θ and α_δ are strictly positive. For given t and δ, the payoff to the department is thus

$$\pi(t,\delta) = \int v(\theta)(1 - F(t|\alpha(\theta,\delta)))d\Gamma(\theta).$$

This problem embeds the problem originally considered, and so *NUC* remains necessary for monotonicity. In its strict form, it also remains sufficient:

Proposition 4 *Fix (p, α), and assume that F is $\mathcal{C}^2$ and satisfies NUC strictly (i.e., that $-F_a$ is strictly ℓsm). Then, for every pair $\delta_\ell < \delta_h$, $\pi(\cdot, \delta_h) - \pi(\cdot, \delta_\ell)$ has the strict single-crossing property, and hence the optimal choice of δ is increasing in t.*[16]

Proof It is enough to show that if $\pi(t, \delta_h) - \pi(t, \delta_\ell)$ is zero, then $(\pi(t, \delta_h) - \pi(t, \delta_\ell))_t > 0$. But

$$\pi(t, \delta_h) - \pi(t, \delta_\ell) = \int y(\theta, t) d\Gamma(\theta), \tag{4}$$

where $y(\theta, t) = v(\theta)(F(t|\alpha(\theta, \delta_\ell)) - F(t|\alpha(\theta, \delta_h)))$, and hence,

$$(\pi(t, \delta_h) - \pi(t, \delta_\ell))_t = \int y_t(\theta, t) d\Gamma(\theta) = \int \frac{y_t(\theta, t)}{y(\theta, t)} y(\theta, t) d\Gamma(\theta).$$

Since v single-crosses zero from below, and since $F(t|\alpha(\cdot, \delta_\ell)) - F(t|\alpha(\cdot, \delta_h)) > 0$, it follows that $y(\cdot, t)$ is first strictly negative and then strictly positive. Hence, since $\int y(\theta, t) d\Gamma(\theta) = 0$, it is enough, using an inequality in Beesack (1957), to show that $y_t(\cdot, t)/y(\cdot, t)$ is strictly increasing.[17] This is established in Appendix A.5, where the proof hinges on the fact that

$$\frac{y_t(\theta, t)}{y(\theta, t)} = \frac{f(t|\alpha(\theta, \delta_\ell)) - f(t|\alpha(\theta, \delta_h))}{F(t|\alpha(\theta, \delta_\ell)) - F(t|\alpha(\theta, \delta_h))},$$

which can be expressed as an expectation of $f_a(t|\cdot)/F_a(t|\cdot)$ with respect to the density ξ on a, where recall that $\xi(a) = -F_a(t|a)/\int -F_a(t|s)ds$, and where we condition on $a \in [\alpha(\theta, \delta_l), \alpha(\theta, \delta_h)]$. Since α is increasing in θ, the result follows since the conditional expectations of an increasing function over an interval increases in the endpoints of that interval. But, by Proposition 1, $f_a(\cdot|a)/F_a(\cdot|a)$ is strictly increasing since F satisfies *NUC* strictly. $\square$

5 *NUC* and Quasi-Concave Expectations

NUC is of particular use when one is considering maximizing the expectation of a function with respect to a parameterized distribution. To be concrete, let an agent have ability $\theta \in \Theta \subseteq \mathbb{R}$ that she views as coming from prior Γ, let the relationship between effort, ability, and output be given by $F(\cdot|a, \theta)$, and let the payoff to the agent of output x and ability θ be $v(x, \theta)$. Let the cost of effort a to the agent be $c(a)$, with c strictly increasing and $\mathcal{C}^2$. Then, the agent maximizes

$$U(a) = \int_\Theta \left(\int_X v(x, \theta) f(x|a, \theta) dx \right) d\Gamma(\theta) - c(a).$$

[16] In particular, if $t_h > t_l$, then the smallest optimal δ at t_h is larger than the largest optimal δ at t_l.

[17] The version of Beesack's inequality we use states that if G is a measure, h is a function that strictly single crosses zero from below, and q is a strictly increasing function, then $\int hdG = 0 \Longrightarrow \int qhdG > 0$.

If Γ is degenerate and v is independent of θ, then this is the problem faced by an agent in a standard moral hazard problem with contract v (in utils), a topic on which we will have more to say in Section 7. If Γ is non-degenerate, then this is a key building block for a career-concerns model, where v, which does not depend on θ, is the market's estimate of the value of the agent given output x and the market's conjectured effort level by the agent.[18]

In any such application, analysis of the problem via the first-order condition is very convenient. But, to do so, one needs to know that the first-order condition characterizes the global optimum. In the following proposition, we use *NUC* to provide such a result.

Proposition 5 *Assume that, for each θ, $F(\cdot|\cdot,\theta)$ satisfies NUC, where $F(\cdot|\cdot,\theta)$ has a $\mathcal{C}^2$ density $f(\cdot|\cdot,\theta)$, and that there is a differentiable function q of x and θ that is strictly increasing in x and such that, for each θ, v_x/q_x is decreasing in x and*

$$\frac{(\mathbb{E}[q|a])_{aa}}{(\mathbb{E}[q|a])_{a}} < \frac{c_{aa}(a)}{c_a(a)}. \tag{5}$$

Then, U is strictly quasi-concave, and hence the first-order condition characterizes the optimal choice of a.

Proof We have

$$U_a(a) = \int_\Theta \int_X v(x,\theta) f_a(x|a,\theta)\,dx\,d\Gamma(\theta) - c_a(a),$$

and

$$U_{aa}(a) = \int_\Theta \int_X v(x,\theta) f_{aa}(x|a,\theta)\,dx\,d\Gamma(\theta) - c_{aa}(a).$$

When $U_a = 0$,

$$c_{aa}(a) = \frac{c_{aa}(a)}{c_a(a)} c_a(a) = \int_\Theta \int_X \frac{c_{aa}(a)}{c_a(a)} v(x,\theta) f_a(x|a,\theta)\,dx\,d\Gamma(\theta),$$

and so, substituting and rearranging , we have

$$U_{aa}(a) = \int_\Theta \int_X v(x,\theta) \left(f_{aa}(x|a,\theta) - \frac{c_{aa}(a)}{c_a(a)} f_a(x|a,\theta) \right) dx\,d\Gamma(\theta).$$

It is thus enough that the inner integral is negative for each θ. Integrating the inner integral by parts, it suffices that for each θ

$$\begin{aligned} 0 &< \int_X v_x(x,\theta) \left(F_{aa}(x|a,\theta) - \frac{c_{aa}(a)}{c_a(a)} F_a(x|a,\theta) \right) dx \\ &= \int_X \frac{v_x(x,\theta)}{q_x(x,\theta)} q_x(x,\theta) \left(F_{aa}(x|a,\theta) - \frac{c_{aa}(a)}{c_a(a)} F_a(x|a,\theta) \right) dx. \end{aligned}$$

[18] See Dewatripont, Jewitt, and Tirole (1999) for details on the two-period career-concerns model.

Since $q_x(\cdot,\theta) > 0$ for all θ, it follows by *NUC* that $q_x\left(F_{aa} - (c_{aa}/c_a)\, F_a\right)$ is never first strictly negative and then strictly positive. Hence, since v_x/q_x is positive and decreasing in x by assumption, we can apply another inequality of Beesack (1957).[19] It is in particular sufficient that for each θ

$$\int_X q_x(x,\theta)\left(F_{aa}(x|a,\theta) - \frac{c_{aa}(a)}{c_a(a)}F_a(x|a,\theta)\right)dx > 0, \tag{6}$$

or, equivalently (integrating by parts and rearranging), that

$$\frac{(\mathbb{E}[q|a])_{aa}}{(\mathbb{E}[q|a])_a} < \frac{c_{aa}(a)}{c_a(a)},$$

for all θ, and we are done. □

In decision problems under uncertainty where, for example, v is the utility function for income, it is commonly assumed that v is concave in x, and then $q(x,\theta) = x$ suffices. Then (5) asks simply that for each type of the agent, c is at least as convex as the expected value of the outcome.[20] This holds if expected outcome is concave and costs are convex (with one strictly so), but can also easily hold in applications where expected outcome is not concave or costs are not convex. An attractive feature of Proposition 5 is that both *NUC* and (5) have clean interpretations.

As we will see presently (see in particular Corollary 2), the generality offered by q is very useful in applications. To see where q comes from, note that it may be that either v is not concave in x, or expected output is insufficiently concave. But, it may also be that by a change of variables, the requisite properties hold. This is the role of q, which, for any given θ, corresponds to simply stretching the x axis according to $z = q(x,\theta)$. Denote the inverse of q with respect to x by φ. If under this relabelling, $\hat{v}(z,\theta) = v(\varphi(z,\theta),\theta)$ is concave in z for each θ, while the expectation of $z = q(x,\theta)$ is more concave than c for each θ, then we are done as well.[21]

6 When does *NUC* Fail?

In this section, we explore settings where *NUC* fails. We begin with a continuous distribution, and then provide a discrete example that exposes intuition.

[19] This version of Beesack's inequality states that if G is a measure, h is a function that strictly single crosses zero from below (above), and r is an increasing (decreasing) positive function, then $\int hdG > 0 \Longrightarrow \int rhdG > 0$.

[20] In the two-period career-concerns model v is a composition of functions whose curvature is much more difficult to pin down. When v will be concave in this case is well beyond the scope of this paper.

[21] If v is independent of θ, then the decision maker's expected utility can be written as $\int v(x)(\int f(x|a,\theta)d\Gamma(\theta))dx$, and so one might hope to work simply with the density $\int f(x|a,\theta)d\Gamma(\theta)$. But, as the next section illustrates, *NUC* need not be inherited by a mixture of distributions that satisfy *NUC*, and so it may be convenient to retain the θ-structure here as well.

For an example where NUC fails but strict $MLRP$ holds, let $a \in [0, 0.48]$, $x \in [0,1]$, and

$$f(x|a) = \left(\frac{1}{6} - 2\left(\frac{1}{2} - x\right)^2\right)\frac{a^2}{2} + 4\left(x - \frac{1}{2}\right)a + 1,$$

which is quadratic in a with coefficients that depend on x. It can be checked that f satisfies strict $MLRP$. But, f_{aa} has sign pattern $-/+/-$, and so $F_{aa} = \int_0^x f_{aa}$ will be first strictly negative and then strictly positive. In fact, this example contains a recipe for constructing examples where NUC fails: the key step is to appropriately craft the coefficient of a^2.

To see a discrete example, consider a student who chooses an effort level in $\{0,1,2\}$, with probability and cumulative distributions over grades given effort given by

$$f = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1-x & x \\ 1-y & y-w & w \\ 1/2 & 1/2 & 0 \end{bmatrix} \end{array} \text{ and } F = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & 1 & x \\ 1 & y & w \\ 1 & 1/2 & 0 \end{bmatrix} \end{array}.$$

where rows indicate effort and columns indicate grades. Then, strict $MLRP$ is equivalent to $(1/2)/(1/2) > (1-y)/(y-w)$ and $(y-w)/w > (1-x)/x$ or equivalently, $w < 2y - 1$ and $w < yx$, and thus requires $y > 1/2$. Similarly, NUC holds if $(1/2 - y)/(-w) > (y-1)/(w-x)$ or equivalently, using that by strict $FOSD$, $w - x < 0$, $w < (2y-1)x$. Thus, strict $MLRP$ holds but NUC fails when

$$x \in (0,1)\,,\ y \in (1/2, 1)\,, \text{ and } w \in ((2y-1)\,x, yx) \cap [0, 2y-1).$$

In Figure 2, y is fixed in $(1/2, 1)$, with the shaded areas representing pairs (x, w) where $MLRP$ holds, with the lighter area being where NUC also holds.[22]

In these examples, increasing effort from 0 to 1 moves substantial weight from B to A, while increasing effort from 1 to 2 moves relatively more weight from C to B. More generally, NUC fails when, starting from a low effort level, incremental effort has an effect more on high outcomes, while starting from a high effort level, incremental effort has an effect more on low outcomes.[23]

[22] In this example, NUC implies $MLRP$. An example that satisfies NUC and $FOSD$, but fails $MLRP$ is

$$f = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 3/4 & 1/8 & 1/8 \end{bmatrix} \end{array} \text{ and } F = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & 1 & 1/2 \\ 1 & 3/4 & 1/4 \\ 1 & 1/4 & 1/8 \end{bmatrix} \end{array},$$

where $MLRP$ fails because $f(B|2)/f(C|2) < f(B|1)/f(C|1)$.

[23] If (x, y, w) is uniform on $[0,1]^3$, then the probability of NUC given $MLRP$ is 87%, where we had to explore several low-dimensional parameterizations of this three by three example before we found one where NUC ever failed while $MLRP$ held.

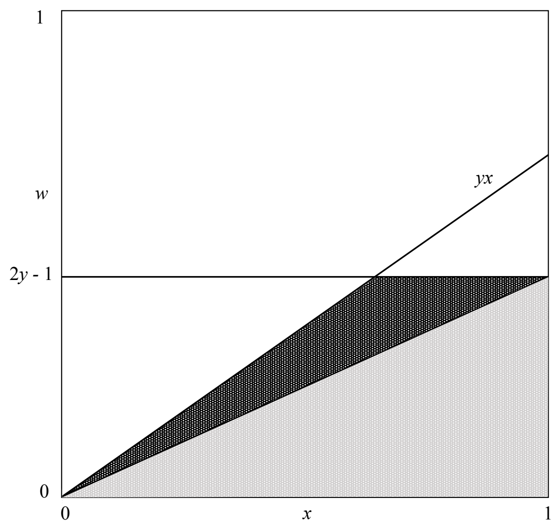

Figure 2: Failure of $\boldsymbol{NUC}$.

If $x = 1/2$, $y = 3/4$, and $w = 3/8$, then

$$f = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 3/8 & 3/8 \\ 1/2 & 1/2 & 0 \end{bmatrix} \end{array} \text{ and } F = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & 1 & 1/2 \\ 1 & 3/4 & 3/8 \\ 1 & 1/2 & 0 \end{bmatrix} \end{array}.$$

For a professor with known grading standards, this situation is quite odd–turning in some work should primarily turn C's into B's, while turning in all the work instead of some of the work should be relatively more important in turning B's into A's.

To see how such a failure of NUC might still occur, imagine f reflects a professor with unknown type, which can be soft (S) or harsh (H), with $\mathbb{P}[S]/\mathbb{P}[H] = 3$, and with

$$f_S = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 1/2 & 1/6 \\ 2/3 & 1/3 & 0 \end{bmatrix} \end{array} \text{ and } f_H = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \end{array},$$

with associated F_S and F_H. Each distribution (weakly) satisfies $MLRP$ and NUC, but as we saw above the mixture of f_S and f_H, given by f, fails NUC. Indeed, changes in effort from 0 to 1 only affect the grade given by the soft professor, who is quite likely to change a B to an A in response. In contrast, the change in effort from 1 to 2, because it affects the grade given by the harsh professor, has a relatively larger effect on B versus C, leading NUC to fail.[24]

7 *NUC* and The First Order Approach

An important context in which NUC is very helpful is in verifying the validity of the first-order approach in the standard moral hazard problem. In this section, we remind the reader of the basic issue, and then see how NUC and Proposition 5 simplify the analysis.

A risk-neutral principal hires a strictly risk-averse agent whose effort $a \in A$ is unobservable to the principal. The principal sees a signal $x \in X$ which is distributed according to $F(\cdot|a)$, where f is strictly *lsm* (strict *MLRP*). The agent's utility for income is u, his cost of effort c, both of which are assumed to be $\mathcal{C}^2$, and his wage as a function of output w. For any contract w and effort a, the agent's expected utility is $\int u(w(x))f(x|a)dx - c(a)$, while the principal's expected profit is $\int(x - w(x))f(x|a)dx$. The agent has an outside option that yields utility u_0. The principal's problem is to choose a contract w and recommend an effort a to maximize expected profits subject to incentive compatibility and participation. Formally, the principal's problem is

$$\begin{aligned} \max_{w,a} \quad & \int (x - w(x))f(x|a)dx \\ \text{s.t.} \quad & \int u(w(x))f(x|a)dx - c(a) \geq u_0 \\ & a \in \arg\max_{a' \in A} \int u(w(x))f(x|a')dx - c(a'). \end{aligned} \tag{7}$$

As is standard (Holmstrom (1979), Mirrlees (1975)), the FOA begins by considering the relaxed problem in which (7) is replaced by the agent's first order condition

$$\int u(w(x))f_a(x|a)dx = c_a(a) \tag{8}$$

for which an optimal solution will be of the form (recall that $l \equiv f_a/f$)

$$\frac{1}{u'(w(x))} = \lambda + \mu l(x|a), \tag{9}$$

where $\lambda > 0$ is the Lagrange multiplier associated with the participation constraint, and $\mu > 0$ the one associated with (8) (see Jewitt (1988) for the proof that μ is strictly positive). The question the FOA addresses is when we can conclude that the solution to this relaxed problem satisfies (7) as well. But this will be true as long as $\int u(w(x))f(x|a)dx - c\,(a)$ is quasi-concave in a.[25] Hence,

[24] This is mathematically the same point as the fact that the convex combination of two distributions each satisfying $MLRP$ need not satisfy $MLRP$.

[25] As in Proposition 5, we will derive conditions for strict quasiconcavity, but there are versions of all the results in this section where the appropriate inequality is a weak one.

from Proposition 5, the key question is to understand the behavior of of $u(w(\cdot))$, where one can exploit the structure inherent in (9) to put structure on $u(w(\cdot))$.

As in Jewitt (1988), let ρ carry $1/u'$ to u.[26] Then, from (9), if the action being implemented is $\hat{a}$, then $u(w(x)) = \rho(\lambda + \mu l(x|\hat{a}))$, and so

$$(u(w(x)))_x = \rho'(\lambda + \mu l(x|\hat{a}))\mu l_x(x|\hat{a}), \tag{10}$$

which is strictly positive since $l_x > 0$ by strict *MLRP*. We can thus use Proposition 5 in one of two ways. The first is a direct variation of Jewitt (1988), Theorem 1.

Corollary 1 *Assume f is $\mathcal{C}^2$, F satisfies NUC and MLRP, ρ is concave, l is concave in x for each a, and*

$$\frac{(\mathbb{E}[x|a])_{aa}}{(\mathbb{E}[x|a])_a} < \frac{c_{aa}(a)}{c_a(a)}. \tag{11}$$

Then, the FOA *is valid.*

Proof Immediate from Proposition 5, taking q to be the identity function, and noting that v, given by $v(x) = u(w(x))$ is concave under the stated conditions on ρ and l. $\square$

Jewitt (1988), Corollary 1, observes that for exponential families (defined in Section 8) with $l(\cdot|a)$ concave, it is enough to check that expected output is concave in effort. The force of our result is that under *NUC*, this basic insight of exponential families holds *regardless* of the distribution. Under *NUC*, there is only *one* integral to check (recall that (11) is equivalent to (6) with $q_x = 1$), instead of a continuum of such expectations in Jewitt (1988). In addition, the relevant integral has a simple economic interpretation.

Proposition 5 also yields the following corollary, which, modulo the use of the disutility of effort, is similar to a result in Jung and Kim (2015a) (see their Proposition 7 and Lemma 2).

Corollary 2 *Assume f is $\mathcal{C}^2$, F satisfies NUC and MLRP, ρ is concave, and that for all a and $\hat{a}$,*

$$\frac{(\mathbb{E}[\hat{l}|a])_{aa}}{(\mathbb{E}[\hat{l}|a])_a} < \frac{c_{aa}(a)}{c_a(a)}, \tag{12}$$

where $\hat{l} = l(\cdot|\hat{a})$. Then, the FOA is valid.

Proof Immediate from Proposition 5 and from (10), taking $q = l(\cdot|\hat{a})$, $v(x) = u(w(x))$, and noting that $v'/l_x(\cdot|\hat{a}) = \mu\rho'$ is decreasing, since ρ is concave by assumption. $\square$

As mentioned in the introduction, we stress that both corollaries follow effortless from Proposition 5, which in turn follows easily from a *single* integration by parts plus the application of an integral inequality to sign a *single* integral.

[26] That is, define $\rho(\cdot)$ by $\rho(z) = u\left([u']^{-1}(1/z)\right)$.

We will see in the next section that for exponential families, (12) is easy to check. Also, removing the concavity condition on $l(\cdot|\hat{a})$ is especially useful for some exponential families, where $\mathbb{E}[l(\cdot|\hat{a})|a]$ is quite tractable; see Section 8, Example 1.

Remark 4 *If $l(\cdot|a)$ is concave in x for each a, and if NUC holds, then (12) is weaker than* (11).

See Appendix A.6 for a direct proof, and Jung and Kim (2015a), Proposition 8, for an alternative argument. Thus, Corollary 2 generalizes Jewitt (1988), Theorem 1 in three directions. First, it incorporates the curvature of c. Second, it allows for examples in which l is not concave. Third, even when l is concave, the integral condition typically becomes strictly weaker, as shown by the remark above. For examples where l is convex but the integral condition holds, and where l is concave but the distinction between (11) and (12) has real bite, see Section 8.

So far we have explored the implications of Proposition 5 for the moral hazard problem, but there is also an interesting implication of the condition $(\log(\log f)_{ax})_{ax} \geq 0$ which as Proposition 2 shows, under strict *MLRP* is a sufficient condition for *NUC*. Pick (a, λ, μ) and $(\hat{a}, \hat{\lambda}, \hat{\mu})$ with $\hat{a} > a$. Let $\phi(x) = \lambda + \mu l(x|a)$ and $\hat{\phi}(x) = \hat{\lambda} + \hat{\mu} l(x|\hat{a})$ be the corresponding contracts considered as functions from x to $1/u'$. Since ρ is strictly increasing, the associated monetary contracts $w = \rho(\phi)$ and $\hat{w} = \rho(\hat{\phi})$ have the same crossing properties as ϕ and $\hat{\phi}$. Now, $\hat{\phi}'(x)/\phi'(x) = \hat{\mu} l_x(x|\hat{a})/\mu l_x(x|a)$, and thus

$$\left(\frac{\hat{\phi}'(x)}{\phi'(x)}\right)_x =_s \left(\frac{l_{xx}}{l_x}(x|\hat{a}) - \frac{l_{xx}}{l_x}(x|a)\right) =_s \left(\frac{l_{xx}}{l_x}\right)_a = (\log(\log f)_{ax})_{ax},$$

where we remind the reader that "$=_s$" means "has strictly the same sign as," and where we note that by the last equality and the premise, $(l_{xx}/l_x)_a$ is everywhere positive. So, $(\log(\log f)_{ax})_{ax} \geq 0$ holds if and only if the ratio of the slope of the higher effort contract to the slope of the lower effort contract increases in x, so that the higher effort contract is "more convex" than the lower effort contract. An implication is that (except if they are the same contract), $\hat{\phi}$ can cross ϕ at most twice, and if it does so, it does so first from above and then from below.[27] For the exponential families, $(\log(\log f)_{ax})_{ax} = 0$, and hence contracts either coincide everywhere, cross exactly once, or do not cross at all, something that is potentially useful to the modeler.

We have followed the standard way to study the validity of the *FOA*, which imposes conditions on the conditional distribution of the outcome x (and also on u and c). The central point of Jung and Kim (2015a), however, is that one can instead focus on the distribution of the likelihood ratio l, which leads to conditions for the validity of the *FOA* that do not require strict *MLRP* on l, and that apply also to the multidimensional-signal case.

For completeness, let us see how Proposition 5 implies a key part of their analysis. To do so,

[27] If $\left(\log(\log f)_{ax}\right)_{ax} < 0$ then $\hat{\phi}$ and ϕ cross at most twice, with $\hat{\phi}$ crossing ϕ first from below, then from above.

fix $\hat{a}$, and for each $\zeta \in \mathbb{R}$, define $X(\zeta, \hat{a}) = \{x | l\,(x|\hat{a}) \leq \zeta\}$. Let

$$G(\zeta|a, \hat{a}) \equiv \int_{X(\zeta,\hat{a})} f(s|a)ds$$

be the probability, given effort a, that output satisfies $l(x|\hat{a}) \leq \zeta$, and let $g(\cdot|a, \hat{a})$ be the associated density. Let $r(\zeta) \equiv \rho(\lambda + \mu\zeta)$ for each ζ, and write the agent's expected utility given effort a as $\int r(\zeta)g(\zeta|a, \hat{a})d\zeta - c(a)$. We then have the following variation on Proposition 7 and Lemma 2 of Jung and Kim (2015a), which unlike Corollary 2 does not assume that l is increasing in x.

Proposition 6 *Assume that $G(\cdot|a, \hat{a})$ satisfies FOSD in a for each $\hat{a}$, and $G(\cdot|\cdot, \hat{a})$ is $\mathcal{C}^2$ and satisfies NUC, for each $\hat{a}$. Assume also that ρ is concave, and that for all a and $\hat{a}$,*

$$\frac{\left(\mathbb{E}_{G(\cdot|a,\hat{a})}[\zeta|a]\right)_{aa}}{\left(\mathbb{E}_{G(\cdot|a,\hat{a})}[\zeta|a]\right)_{a}} < \frac{c_{aa}(a)}{c_a(a)}.$$

Then, the FOA is valid.

Proof Immediate from Proposition 5, with ζ taking the role of x, $G(\cdot|\cdot, \hat{a})$ taking the role of F, and r taking the role of v, taking q as the identity and noting that $r' = \mu\rho'$ is decreasing, since ρ is concave by assumption. $\square$

The degree to which this result is useful depends on the degree to which one can check that $G(x|\cdot, \hat{a})$ satisfies *FOSD* in a for each x and $\hat{a}$, and that $G(\cdot|\cdot, \hat{a})$ satisfies *NUC*, for each $\hat{a}$. Note in particular that once one has abandoned *MLRP*, the sets $X(\zeta, \hat{a})$ are in principle arbitrary, and so it is not clear what primitives on F are required even for *FOSD*, let alone *NUC*. See Jung and Kim (2015a) Section 4.3 for some positive examples.

As an alternative, one could tackle the multidimensional-signal case directly, as in Jewitt (1988) and as in the general analysis in Conlon (2009). The difficulty in extending our results to the multidimensional case is to come up with the analog of the crossing condition that characterizes *NUC* in the one-dimensional case. One case in which *NUC* remains useful is the two-signal case when signals are independent (this case was also analyzed by Jewitt (1988)). In this case, it is easy to use strict *MLRP*, ρ concave, and *NUC* on the distribution of each signal to justify the *FOA*. To see this, let y be a second signal with support on an interval with infimum $\underline{y}$ and supremum $\overline{y}$, and with parameterized distribution P and density p. Assume that F and P are $\mathcal{C}^2$ and that f and p satisfy strict *MLRP*. Denote by ℓ the likelihood ratio of y, that is, $\ell \equiv p_a/p$, and for any $\hat{a}$ the principal wants to implement, let $\hat{\ell} \equiv \ell(\cdot|\hat{a})$. Then we have the following result, whose proof is in Appendix A.7:

Proposition 7 *Assume that x and y are independent signals with $\mathcal{C}^2$ densities f and p that satisfy strict MLRP, and with distributions F and P that satisfy NUC. Assume also that ρ is* *concave, and that for all a and $\hat{a}$, $\hat{l}$ and $\hat{\ell}$ satisfy (12). Then the FOA is valid.*

Here again, *NUC* simplifies the analysis and leads to interpretable sufficient conditions.

8 The Exponential Families

In this section, we explore the exponential families. As mentioned, the Exponential, Poisson, Gamma, Normal, Beta, and their truncations are exponential families. Such truncations are important when applying some of the results of this section to the *FOA*, since they bound the likelihood ratio and thus rule out a standard nonexistence issue. Motivated by our previous results, we focus on the behavior of $\mathbb{E}[x|a]$ and $\mathbb{E}[l(\cdot|\hat{a})|a]$.

Recall that a family of densities $\{f(\cdot|a)\}_{a\in A}$ is a (one-parameter) exponential family if it can be expressed as

$$f(x|a) = m(a)n(x)e^{H(a)j(x)} \tag{13}$$

where $n \geq 0$ and $m(a) = 1/\int n(x)e^{H(a)j(x)}dx$.[28] Let H be analytic, set $h \equiv H'$, and assume that $h > 0$. Note that

$$l(x|a) = h(a)j(x) + \frac{m'(a)}{m(a)} \tag{14}$$

and so $l_x(x|a) = h(a)j'(x)$. Hence f satisfies *MLRP* if and only if j is increasing, as we henceforth assume. Finally, note that $l_{ax}(x|a) = h'(a)j'(x)$, and so

$$(\log(\log f)_{ax})_{ax} = \left(\frac{l_{ax}}{l_x}\right)_x = \left(\frac{h'}{h}\right)_x = 0,$$

as claimed following Proposition 2.

With this in hand, let us return to the moral hazard problem, and in particular, to the question of when the relevant integral condition holds for exponential families. Then, by (14), the inequality (12) reduces simply to

$$\frac{(\mathbb{E}[j|a])_{aa}}{(\mathbb{E}[j|a])_a} < \frac{c_{aa}(a)}{c_a(a)}, \tag{15}$$

where the problem is more tractable because we can throw away the multiplicatively separable factor $h(\hat{a}) > 0$. By Corollary 1, if j is concave, then it also suffices to check

$$\frac{(\mathbb{E}[x|a])_{aa}}{(\mathbb{E}[x|a])_a} < \frac{c_{aa}(a)}{c_a(a)}, \tag{16}$$

as observed by Jewitt (1988), which may be simpler in some settings. Examples include the

[28] It is implicit in this definition that n, H, and j are chosen such that $n(x)e^{H(a)j(x)}$ is integrable for all a in $(\underline{a}, \bar{a})$. It follows (e.g., Lehmann and Romano (2005), Theorem 2.7.1) that if H is analytic on $(\underline{a}, \bar{a})$, then so is m, and that f has finite moments of all orders.

Exponential and Poisson distributions.

There are two reasons to want to go further. First, once j is non-concave, it is not enough to check concavity of output. Second, even when j is concave, concavity of output is more than we need, and may unnecessarily exclude cases of interest. In the following example, j is concave, but it is both easier to check concavity of $\mathbb{E}[j|x]$ than $\mathbb{E}[x|a]$, and critical to do so.

Example 1 *Consider*

$$f(x|a) = a^b x^{a^b-1} = a^b e^{(a^b-1)\log x}$$

for $x \in [0,1]$, $a > 0$, and $b > 0$. As an exponential family, NUC holds. It is then a matter of simple calculation that, since $j = \log x$, $\mathbb{E}[j|a] = -1/a^b$, which is concave in a for any $b > 0$, while $\mathbb{E}[x|a] = a^b/(a^b+1)$, which fails to be concave for any $b > 1$.[29]

Our next result illuminates what is needed to satisfy (15). To simplify notation, write σ for the standard deviation of j and $\gamma = \mathbb{E}[(j(x)-\bar{j})^3|a]/\sigma^3$ for the skewness of j.

Proposition 8 *Let F be an exponential family. Then for each a, (15) holds if and only if*

$$\frac{h'(a)}{h(a)} + h(a)\sigma\gamma < \frac{c_{aa}(a)}{c_a(a)}. \tag{17}$$

This holds if c is strictly relatively more convex than H (since $h'/h = H''/H'$) and j is negatively skewed. Sufficient for j to be negatively skewed is that H is positive and j'/n is decreasing.

The proof is in Appendix A.8, and its main step shows that $(\mathbb{E}[j])_{aa}/(\mathbb{E}[j])_a = (h'(a)/h(a)) + h(a)\sigma\gamma$. The result implies that, for exponential families, the *FOA* is valid if c is strictly relatively more convex than H, $H \geq 0$, and j'/n is decreasing in x. Given Proposition 8, to construct simple examples where j need not be concave but where the integral condition holds, let j be an arbitrary increasing and differentiable function and take $n = j'$ and H positive and concave.

Remark 5 *The exponential families satisfy* $(\log f)_{ax}$ *lsm weakly. We can, however, modify this family and construct tractable F for which* $(\log f)_{ax}$ *lsm is strict. Say that F is a* blended exponential family *if it can be written in the form $f(x|a) = m(a)n(x)e^{j(x)H(a)+\hat{j}(x)\hat{H}(a)}$, where j, $\hat{j}$, H, and $\hat{H}$ are increasing. Then, $l(x|a) = j(x)h(a) + \hat{j}(x)\hat{h}(a) + \frac{m'(a)}{m(a)}$ and* $(\log f)_{ax}(x|a) = l_x(x|a) = j'(x)h(a) + \hat{j}'(x)\hat{h}(a) > 0$*, and so MLRP is satisfied. Further,*

$$(\log(\log f)_{ax})_a(x|a) = \frac{j'(x)h'(a) + \hat{j}'(x)\hat{h}'(a)}{j'(x)h(a) + \hat{j}'(x)\hat{h}(a)},$$

[29] In this example the likelihood ratio is unbounded below. This is easy to modify by assuming that $x \in [\eta, 1+\eta]$ with $\eta > 0$, so that $f(x|a) = (a^b x^{a^b-1})/((1+\eta)^{a^b} - \eta^{a^b})$. One can verify that for each $b > 1$, if one takes η sufficiently small, the expectation of $j(x)$ is concave in a but the expectation of x is not.

and it is now possible to construct distributions with a rich set of behaviors for $(\log(\log f)_{ax})_{ax}$.[30] *Indeed, take* $H(a) = a^2/2$ *and* $\hat{H}(a) = -(1-a)^2/2$. *Then,* $h(a) = a$, $\hat{h}(a) = 1-a$, $l_x(x|a) = j'(x)a + \hat{\jmath}'(x)(1-a)$, *and* $(\log(\log f)_{ax})_a(x|a) = (j'(x) - \hat{\jmath}'(x))/aj'(x) + (1-a)\hat{\jmath}'(x)$. *By suitable choice of* $\hat{\jmath}(x)$ *and* $j(x)$, *we can make* $(\log(\log f)_{ax})_{ax}$ *strictly positive. For example, if* $j''/j' > \hat{\jmath}''/\hat{\jmath}'$, *then* $(\log(\log f)_{ax})_{ax} > 0$, *and thus NUC holds.*

9 Concluding Remarks

We have introduced, motivated, and illustrated the economic relevance of a log-supermodularity condition, which we call *NUC*, on a parameterized family of distributions. We provided a characterization for *NUC* under differentiability assumptions, as well as several sufficient conditions that are easier to check in some settings. We showed that *NUC* has a strong intuitive foundation in terms of a natural monotonicity property in a statistical decision problem.

We showed that *NUC* is useful in the analysis of some interesting economic problems. It is especially relevant when characterizing a global optimum using first-order conditions in some problems under uncertainty. Two such problems are the principal-agent problem with moral hazard, where a technical hurdle is to justify the *FOA*, and the career-concerns problem, where the agent's first-order condition becomes an equilibrium condition under rational expectations.

We also explored the limitations of *NUC*, and provided examples where *NUC* fails. Although *NUC* can fail, we contend that the instances where this happens are somewhat artificial, lending further credibility to *NUC* as a natural condition.

In the last part of the paper, we provided a thorough analysis of the validity of the *FOA* in the moral-hazard problem using *NUC*, and related the results with those in the analysis in Jewitt (1988) and Jung and Kim (2015a). In particular, we explored the usefulness of *NUC* in the case of exponential families, and illustrated its tractability in commonly used examples of this class. It is our hope on a forward-going basis that, as for *MLRP*, *NUC* will turn out to simplify and clarify analysis in a broad set of problems.

A Appendix: Omitted Proofs

A.1 Proof of Proposition 1

Lemma 2 *Let* $\chi : \mathbb{R}^2 \to \mathbb{R}$ *be strictly positive and* $\mathcal{C}^2$. *Then* χ *is lsm if and only if for each* $\tau \in \mathbb{R}$, *and each* a, $-\chi_a(\cdot, a) + \tau\chi(\cdot, a)$ *is never first strictly negative and then strictly positive.*

[30] One can actually go further by letting $f(x|a) = m(a)n(x)e^{\int j(x,s)H(a,s)dK(s)}$ for some distribution K, and run through the same derivation.

Proof Sufficiency follows since

$$-\chi_a(x,a)+\tau\chi(x,a)=_s \tau-\frac{\chi_a(x,a)}{\chi(x,a)}.$$

Since χ is *lsm*, χ_a/χ is increasing in x, so once $-\chi_a(x,a)+\tau\chi(x,a)$ is negative, it remains so.

For necessity, assume χ is not *lsm*; then there are $x',x''\in(\underline{x},\bar{x})$ with $x''>x'$, such that

$$\frac{\chi_a}{\chi}(x'',a)<\frac{\chi_a}{\chi}(x',a),$$

and so for any $\tau\in(\frac{\chi_a}{\chi}(x'',a),\frac{\chi_a}{\chi}(x',a))$, $-\chi_a(x',a)+\tau\chi(x',a)<0<-\chi_a(x'',a)+\tau\chi(x'',a)$, contradicting the premise. □

Proof of Proposition 1 That (ii) is equivalent to (iii) is immediate by applying Lemma 2 to $\chi(x,a)=-F_a\left(x|a\right)$, which is positive under our maintained assumption of *FOSD*.

To see that (i) and (ii) are equivalent, let

$$R\left(x\right)\equiv\frac{F\left(x|a_h'\right)-F\left(x|a_h\right)}{F\left(x|a_l'\right)-F\left(x|a_l\right)}=\frac{-\int_{a_h}^{a_h'}F_a\left(x|s\right)ds}{-\int_{a_l}^{a_l'}F_a\left(x|s\right)ds},$$

and so,

$$\begin{aligned}
R'\left(x\right) &=_s \left(\log R\right)_x\\
&=\frac{\int_{a_h}^{a_h'}f_a\left(x|s\right)ds}{\int_{a_h}^{a_h'}F_a\left(x|s\right)ds}-\frac{\int_{a_l}^{a_l'}f_a\left(x|s\right)ds}{\int_{a_l}^{a_l'}F_a\left(x|s\right)ds}\\
&=\int_{a_h}^{a_h'}\frac{f_a}{F_a}\left(x|s\right)\frac{-F_a\left(x|s\right)}{\int_{a_h}^{a_h'}\left(-F_a\left(x|s\right)\right)ds}ds-\int_{a_l}^{a_l'}\frac{f_a}{F_a}\left(x|s\right)\frac{-F_a\left(x|s\right)}{\int_{a_l}^{a_l'}\left(-F_a\left(x|s\right)\right)ds}ds \qquad (18)\\
&=\mathbb{E}_\xi\left(\frac{f_a}{F_a}\left(x|a\right)|a\in\left[a_h,a_h'\right]\right)-\mathbb{E}_\xi\left(\frac{f_a}{F_a}\left(x|a\right)|a\in\left[a_l,a_l'\right]\right).
\end{aligned}$$

Assume that $-F_a$ is lsm in (x,a). Then, f_a/F_a increases in a, and (18) implies that $R'\geq 0$, since the expectation of an increasing function increases in either bound of the conditioning set.

Assume that $R'\geq 0$. Then, for any interior x, any $a''>a'$, and for each $\varepsilon>0$, it follows from (18) that

$$\mathbb{E}_\xi\left(\frac{f_a\left(x|a\right)}{F_a\left(x|a\right)}|a\in\left[a'',a''+\varepsilon\right]\right)-\mathbb{E}_\xi\left(\frac{f_a\left(x|a\right)}{F_a\left(x|a\right)}|a\in\left[a',a'+\varepsilon\right]\right)\geq 0, \qquad (19)$$

and so, taking $\varepsilon\to 0$,

$$\frac{f_a\left(x|a''\right)}{F_a\left(x|a''\right)}\geq\frac{f_a\left(x|a'\right)}{F_a\left(x|a'\right)}.$$

Hence, $-F_a$ is lsm in (x,a). □

A.2 Proof of Proposition 2

The first implication is trivial since f is lsm and the product of lsm functions is lsm.

$f^2 (\log f)_{ax}$ $lsm \Longrightarrow F^2 (\log F)_{ax}$ lsm: Note that $F^2 (\log F)_{ax} = f_a F - F_a f$. Now

$$\frac{(f_a F - F_a f)_a}{f_a F - F_a f} = \frac{f_{aa} F + f_a F_a - F_{aa} f - F_a f_a}{f_a F - F_a f} = \frac{\frac{f_{aa}}{f} - \frac{F_{aa}}{F}}{\frac{f_a}{f} - \frac{F_a}{F}} \equiv \eta(x|a),$$

and so, to establish that $f_a F - F_a f$ is lsm, we need to show that $\eta_x \geq 0$. But,

$$\begin{aligned}\frac{f_{aa}}{f}(x|a) - \frac{F_{aa}}{F}(x|a) &= \int_{\underline{x}}^{x} \left(\frac{f_{aa}}{f}(x|a) - \frac{f_{aa}}{f}(z|a) \right) \frac{f(z|a)}{F(x|a)} dz = \int_{\underline{x}}^{x} \int_{z}^{x} \left(\frac{f_{aa}}{f}(s|a) \right)_x ds \frac{f(z|a)}{F(x|a)} dz \\ &= \int_{\underline{x}}^{x} \left(\frac{f_{aa}}{f}(s|a) \right)_x \left(\int_{\underline{x}}^{s} \frac{f(z|a)}{F(x|a)} dz \right) ds = \int_{\underline{x}}^{x} \left(\frac{f_{aa}}{f}(s|a) \right)_x \frac{F(s|a)}{F(x|a)} ds,\end{aligned}$$

where the second equallity uses the Fundamental Theorem of Calculus and the third equality exchanges the order of integration over the domain $\underline{x} \leq z \leq s \leq x$. Similarly,

$$\frac{f_a}{f}(x|a) - \frac{F_a}{F}(x|a) = \int_{\underline{x}}^{x} \left(\frac{f_a}{f}(s|a) \right)_x \frac{F(s|a)}{F(x|a)} ds,$$

and so

$$\begin{aligned}\eta &= \frac{\int_{\underline{x}}^{x} \left(\frac{f_{aa}}{f}(s|a) \right)_x F(s|a) ds}{\int_{\underline{x}}^{x} \left(\frac{f_a}{f}(s|a) \right)_x F(s|a) ds} \\ &= \int_{\underline{x}}^{x} \frac{\left(\frac{f_{aa}}{f}(s|a) \right)_x}{\left(\frac{f_a}{f}(s|a) \right)_x} \frac{\left(\frac{f_a}{f}(s|a) \right)_x F(s|a)}{\int_{\underline{x}}^{x} \left(\frac{f_a}{f}(\tau|a) \right)_x F(\tau|a) d\tau} ds \\ &= \int_{\underline{x}}^{x} \beta(s) \frac{\psi(s)}{\Psi(x)} ds,\end{aligned}$$

where

$$\beta(s) = \frac{\left(\frac{f_{aa}}{f}(s|a) \right)_x}{\left(\frac{f_a}{f}(s|a) \right)_x}, \; \psi(s) = \frac{\left(\frac{f_a}{f}(s|a) \right)_x F(s|a)}{\int \left(\frac{f_a}{f}(\tau|a) \right)_x F(\tau|a) d\tau},$$

and Ψ is the cumulative distribution function of ψ.

We would then be done if β is increasing, since the conditional expectation of an increasing function over an interval increases in the endpoints of that interval. But,

$$\beta = \frac{\left(\frac{f_{aa}}{f}\right)_x}{\left(\frac{f_a}{f}\right)_x} = \frac{f_{aax}f - f_{aa}f_x}{f_{ax}f - f_a f_x} = \frac{(f_{ax}f - f_a f_x)_a}{f_{ax}f - f_a f_x}, \tag{20}$$

which is increasing in x, since $f_{ax}f - f_a f_x = f^2(log f)_{ax}$ is lsm by assumption.

$F^2(\log F)_{ax}$ $lsm \implies NUC$: Fix a, and at any $x \in (\underline{x}, \bar{x}]$, differentiate the identity $F_a = FF_a/F$ by a to get

$$F_{aa} = F_a \frac{F_a}{F} + F\left(\frac{F_a}{F}\right)_a,$$

and so

$$\nu \equiv F_{aa} - \tau F_a = F\left(\left(\frac{F_a}{F}\right)^2 + \left(\frac{F_a}{F}\right)_a - \tau\frac{F_a}{F}\right),$$

where we think of ν as a function purely of x. Thus

$$\nu' = f\left(\left(\frac{F_a}{F}\right)^2 + \left(\frac{F_a}{F}\right)_a - \tau\frac{F_a}{F}\right) + F\left(2\left(\frac{F_a}{F}\right)\left(\frac{F_a}{F}\right)_x + \left(\frac{F_a}{F}\right)_{ax} - \tau\left(\frac{F_a}{F}\right)_x\right)$$

or

$$\nu'(x) = \frac{f}{F}\nu(x) + F\left(\frac{F_a}{F}\right)_x (x|a)(r(x) - \tau) \tag{21}$$

where

$$r = 2\left(\frac{F_a}{F}\right) + \frac{\left(\frac{F_a}{F}\right)_{ax}}{\left(\frac{F_a}{F}\right)_x},$$

and where it is standard that lsm is preserved by integration, so f is lsm implies that F is lsm and thus $(F_a/F)_x > 0$. Note also that

$$r = \left(\log\left(F^2\left(\frac{F_a}{F}\right)_x\right)\right)_a = \left(\log\left(F_{ax}F - F_a F_x\right)\right)_a, \tag{22}$$

and hence r is increasing in x if and only if $(F_{ax}F - F_a F_x) = F^2(\log F)_{ax}$ is lsm.

Assume that F fails NUC, so that there is $x' < x''$ such that $\nu(x') < 0 < \nu(x'')$. Then, there must be $\tilde{x} \in (x', x'')$ such that $\nu(\tilde{x}) = 0$, and $\nu'(\tilde{x}) \geq 0$, and so from (21), $r(\tilde{x}) \geq \tau$. Since r is increasing it also follows from (21) that for all $x \in (\tilde{x}, \bar{x})$, if $\nu > 0$ then $\nu' > 0$. Thus, since $\nu(x'') > 0$, ν is strictly increasing after x'', and so $\nu(\bar{x}) > \nu(x'') > 0$. But, $\nu(\bar{x}) = F_{aa}(\bar{x}) - \tau F_a(\bar{x}) = 0$, contradicting that F fails NUC. $\square$

A.3 Proof of Lemma 1

Let

$$d = \begin{vmatrix} f & f_a & f_{a^2} \\ f_x & f_{ax} & f_{a^2x} \\ f_{x^2} & f_{ax^2} & f_{a^2x^2} \end{vmatrix}.$$

Given *MLRP*, Karlin (1955), Theorem 2 shows that necessary for TP_3 is that $d \geq 0$ for each x and a, and sufficient is that $d > 0$ for each x and a (see Karlin (1955), pp $289-290$ for a discusion of the case $d = 0$). Lemma 1 then follows from the observation that

$$\begin{aligned}
&\left(\log\left(f^2\left(\log f\right)_{ax}\right)\right)_{ax} = \left(\log\left(f_{ax}f - f_af_x\right)\right)_{ax} = \left(\frac{\left(f_{ax}f - f_af_x\right)_a}{f_{ax}f - f_af_x}\right)_x = \left(\frac{f_{a^2x}f - f_{a^2}f_x}{f_{ax}f - f_af_x}\right)_x \\
&=_s \left(f_{a^2x^2}f - f_{a^2}f_{x^2}\right)\left(f_{ax}f - f_af_x\right) - \left(f_{a^2x}f - f_{a^2}f_x\right)\left(f_{ax^2}f - f_af_{x^2}\right) \\
&= f\left(-f_af_xf_{a^2x^2} + f_xf_{a^2}f_{ax^2} + f_af_{x^2}f_{a^2x} - f_{ax}f_{a^2}f_{x^2} + ff_{a^2x^2}f_{ax} - ff_{ax^2}f_{a^2x}\right) \\
&=_s f\left(f_{ax}f_{a^2x^2} - f_{a^2x}f_{ax^2}\right) - f_a\left(f_xf_{a^2x^2} - f_{x^2}f_{a^2x}\right) + f_{a^2}\left(f_xf_{ax^2} - f_{x^2}f_{ax}\right) \\
&= d.
\end{aligned}$$

A.4 Proof of Remark 2

Let us now formalize and prove Remark 2.

Proposition 9 *Assume that A is an interval, that F is $\mathcal{C}^2$, and that $F(t|\cdot)$ and $1 - F(t|\cdot)$ are strictly log-concave for all t interior. Then if F fails NUC, then there exists α, p, and $\hat{t}$ such that $\pi(\cdot,1) - \pi(\cdot,0)$ crosses zero from above at $\hat{t}$, and such that on a neighborhood of $\hat{t}$,* $\max\{\pi(\cdot,1), \pi(\cdot,0)\} > \max\{0, 1-2p\}$.

Noting that $\max\{0, 1-2p\}$ is the observer's payoff from guessing using her prior, this means that near $\hat{t}$, the observer is strictly better off to use the results of the test rather than act according to the prior, where just to the left of $\hat{t}$, the observer strictly prefers $\delta = 1$, while just to the right of $\hat{t}$, the observer strictly prefers $\delta = 0$.

Proof If *NUC* fails, then by Proposition 1, Part (ii), there exists $\hat{t}$, a_l, and $a_h > a_l$ such that

$$\frac{f_a}{F_a}(\hat{t}|a_l) > \frac{f_a}{F_a}\left(\hat{t}|a_h\right),$$

and so there is $\delta > 0$ such that

$$\mathbb{E}_\xi\left(\frac{f_a(\hat{t}|a)}{F_a(\hat{t}|a)}|a \in [a_l, a_l + \delta]\right) - \mathbb{E}_\xi\left(\frac{f_a(\hat{t}|a)}{F_a(\hat{t}|a)}|a \in [a_h, a_h + \delta]\right) > 0.$$

Take $\alpha(l,0) = a_l$, $\alpha(l,1) = a_l + \delta$, $\alpha(h,0) = a_h$, and $\alpha(h,1) = a_h + \delta$, and choose p such that

$$\frac{p}{1-p} = \frac{F\left(\hat{t}|a_h + \delta\right) - F\left(\hat{t}|a_h\right)}{F\left(\hat{t}|a_l + \delta\right) - F\left(\hat{t}|a_l\right)}.$$

Then, by (3) and (18), $\pi(t,1) - \pi(t,0)$ strictly crosses zero from above at $\hat{t}$. It remains to show that $\pi(\hat{t},1) = \pi(\hat{t},0) > \max\{0, 1-2p\}$.

Let us show first that $\pi(\hat{t},0) > 1-2p$, or equivalently, that

$$-p\left(1 - F(\hat{t}|a_l)\right) + (1-p)\left(1 - F(\hat{t}|a_h)\right) > 1 - 2p.$$

This is equivalent to

$$\frac{F(\hat{t}|a_h)}{F(\hat{t}|a_l)} < \frac{p}{1-p} = \frac{F\left(\hat{t}|a_h + \delta\right) - F\left(\hat{t}|a_h\right)}{F\left(\hat{t}|a_l + \delta\right) - F\left(\hat{t}|a_l\right)},$$

where the equality uses our choice of p. Rearranging the end expressions (recalling that the denominator on the *rhs* is negative), we arrive at

$$\frac{F(\hat{t}|a_h)}{F(\hat{t}|a_l)} > \frac{F\left(\hat{t}|a_h + \delta\right)}{F\left(\hat{t}|a_l + \delta\right)},$$

which holds for all $a_h > a_l$ and $\delta > 0$ since

$$\frac{F(\hat{t}|a_h + \tau)}{F(\hat{t}|a_l + \tau)} = \exp \int_{a_l}^{a_h} \left(\frac{\partial}{\partial a} \log F\left(\hat{t}|a + \tau\right) \right) da,$$

which is strictly decreasing in τ since $F\left(\hat{t}|\cdot\right)$ is strictly log-concave. The proof that $\pi(\hat{t},1) > 0$ is analogous, using that $1 - F(\hat{t}|\cdot)$ is strictly log-concave. $\square$

A.5 Proof that $y_t(\cdot,t)/y(\cdot,t)$ is Increasing

$y(\theta,t) = v(\theta)(F(t|\alpha(\theta,\delta_\ell)) - F(t|\alpha(\theta,\delta_h)))$

Note that

$$
\begin{aligned}
\frac{y_t(\theta,t)}{y(\theta,t)} &= \frac{f(t|\alpha(\theta,\delta_\ell)) - f(t|\alpha(\theta,\delta_h)}{F(t|\alpha(\theta,\delta_l)) - F(t|\alpha(\theta,\delta_h))} \\
&= \frac{1}{F(t|\alpha(\theta,\delta_l)) - F(t|\alpha(\theta,\delta_h))} \int_{\alpha(\theta,\delta_l)}^{\alpha(\theta,\delta_h)} (-f_a(t|a))da \\
&= \frac{1}{F(t|\alpha(\theta,\delta_l)) - F(t|\alpha(\theta,\delta_h))} \int_{\alpha(\theta,\delta_l)}^{\alpha(\theta,\delta_h)} \frac{f_a(t|a)}{F_a(t|a)} \left(-F_a(t|a)\right) da \\
&= \frac{\int -F_a(t|s)ds}{F(t|\alpha(\theta,\delta_l)) - F(t|\alpha(\theta,\delta_h))} \int_{\alpha(\theta,\delta_l)}^{\alpha(\theta,\delta_h)} \frac{f_a(t|a)}{F_a(t|a)} \frac{-F_a(t|a)}{\int -F_a(t|s)ds} da \\
&= \mathbb{E}_\xi \left(\frac{f_a(t|a)}{F_a(t|a)} | a \in [\alpha(\theta,\delta_l), \alpha(\theta,\delta_h)] \right)
\end{aligned}
$$

Since $f_a(t|\cdot)/F_a(t|\cdot)$ is increasing, and since α is increasing in θ, the result follows. □

A.6 Proof of Remark 4

It is enough to show that when $l\,(\cdot|a)$ is concave, then $(\mathbb{E}[\hat{l}|a])_{aa}/(\mathbb{E}[\hat{l}|a])_a \leq (\mathbb{E}[x|a])_{aa}/(\mathbb{E}[x|a])_a$ for all a. Consider the densities $F_a\hat{l}_x/\int F_a\hat{l}_x dx$ and $F_a/\int F_a dx$ and note that, since $\hat{l}\,(\cdot|a)$ is concave, the first density is likelihood ratio dominated by the second (i.e., the ratio of the first density to the second is decreasing in x). Thus,

$$
\frac{(\mathbb{E}[\hat{l}|a])_{aa}}{(\mathbb{E}[\hat{l}|a])_a} = \frac{-\int \hat{l}_x F_{aa} dx}{-\int \hat{l}_x F_a dx} = \int \frac{F_{aa}}{F_a} \frac{-F_a \hat{l}_x}{\int -F_a \hat{l}_x dx} dx \leq \int \frac{F_{aa}}{F_a} \frac{-F_a}{\int -F_a dx} dx = \frac{(\mathbb{E}[x|a])_{aa}}{(\mathbb{E}[x|a])_a},
$$

where the first and last equalities follow by integration by parts, and the inequality because F_{aa}/F_a is increasing in x by *NUC* (see Proposition 1 (*ii*)) and by likelihood ratio dominance.[31]

A.7 Proof of Proposition 7

Let $v(x,y) \equiv u(w(x,y))$ for all (x,y), where w is the contract conditioned on the realization of the two signals. The agent's problem is

$$
\max_a \int\int v(x,y) f(x|a) p(y|a) dx dy - c(a). \tag{23}
$$

The first-order condition is

$$
\int \left(\int v(x,y) f_a(x|a) dx \right) p(y|a) dy + \int \left(\int v(x,y) p_a(y|a) dy \right) f(x|a) dx - c_a(a) = 0.
$$

[31] Note that if F_{aa}/F_a is not a constant (which can only happen in the trivial case $F_{aa} \equiv 0$) and if $\hat{l}$ is strictly concave, then this inequality is strict.

which is equal to (by integration by parts)

$$\int \left(\int v_x(x,y)(-F_a(x|a))dx \right) p(y|a)dy + \int \left(\int v_y(x,y)(-P_a(y|a))dy \right) f(x|a)dx - c_a(a) = 0.$$

The second derivative can be written as follows:

$$\begin{aligned}&\int \left(\int v_x(x,y)(-F_{aa}(x|a))dx \right) p(y|a)dy + \int \left(\int v_y(x,y)(-P_{aa}(y|a))dy \right) f(x|a)dx \\ &+ 2 \int \int v_{xy}(x,y) F_a(x|a) P_a(y|a) dxdy - c_{aa}(a),\end{aligned}$$

where the last integral follows by integrating two terms by parts.

Using the first-order condition, we obtain

$$\begin{aligned}&- \int \left(\int v_x(x,y) \left(F_{aa} - \frac{c_{aa}(a)}{c_a(a)} F_a(x|a) \right) dx \right) p(y|a)dy \\ &- \int \left(\int v_y(x,y) \left(P_{aa}(y|a) - \frac{c_{aa}(a)}{c_a(a)} P_a(y|a) \right) dy \right) f(x|a)dx \\ &+ 2 \int \int v_{xy}(x,y) F_a(x|a) P_a(y|a) dxdy.\end{aligned}$$

We will show that this expression is strictly negative under the premises. In particular, note that by Holmstrom (1979), $v(x,y) = \rho(\lambda + \mu\hat{l} + \mu\hat{\ell})$. Thus, $v_x = \rho'\mu\hat{l}_x$ and $v_y = \rho'\mu\hat{\ell}_y$, which are both strictly positive under strict *MLRP*. Also, $v_{xy} = \rho''\mu^2\hat{l}_x\hat{\ell}_y$, which is negative if ρ is concave. It follows that the last term is negative by *FOSD*. Regarding the first term, it is strictly negative since $\rho' > 0$, $\mu > 0$, and

$$\frac{(\mathbb{E}[\hat{l}|a])_{aa}}{(\mathbb{E}[\hat{l}|a])_a} < \frac{c_{aa}(a)}{c_a(a)},$$

and similarly for the second term. Hence, the *FOA* is valid, and we are done. □

A.8 Proof of Proposition 8

We begin with two lemmas. To simplify notation, we set $k \equiv n(x)e^{j(x)H(a)}$.

Lemma 3 *If F is an exponential family, then*

$$\lim_{x \to \bar{x}} F_a(x|a)j(x) = \lim_{x \to \underline{x}} F_a(x|a)j(x) = 0.$$

Proof Consider first the case $x \to \bar{x}$. Note that

$$F = 1 - m(a) \int_x^{\bar{x}} k ds$$

and so

$$F_a = -m(a)\left(h(a)\int_x^{\bar{x}} j(s)kds + \frac{m'(a)}{m(a)}\int_x^{\bar{x}} kds\right).$$

But, since $m(a) = 1/\int kdx$,

$$\frac{m'(a)}{m(a)} = -\frac{h(a)\int jkds}{\int kds} = -h(a)\bar{j}, \tag{24}$$

where $\bar{j} = \mathbb{E}[j(x)|a]$, and hence

$$F_a = -h(a)m(a)\int_x^{\bar{x}} (j(s)-\bar{j})kds$$

and

$$j(x)F_a = -h(a)m(a)\int_x^{\bar{x}} j(x)(j(s)-\bar{j})kds.$$

But, by $MLRP$, for x sufficiently large, $j(s)-\bar{j} > 0$ for all $s > x$, and so for x sufficiently large,

$$|j(x)F_a(x|a)| \leq |h(a)m(a)|\left|\int_x^{\bar{x}} j(s)(j(s)-\bar{j})kds\right|.$$

But, since $\sigma^2 = m(a)\int j(s)(j(s)-\bar{j})kds$ is finite (see Footnote 28), it must be that

$$\lim_{x\to\bar{x}}\int_x^{\bar{x}} j(s)(j(s)-\bar{j})kds = 0,$$

and so, since $|h(a)m(a)|$ is a constant independent of x, we have $\lim_{x\to\bar{x}} j(x)F_a(x|a) = 0$. The case $x \to \underline{x}$ is similar. $\square$

Lemma 4 *Let F be an exponential family. Then $(\sigma^2)_a = h(a)\sigma^3\gamma$.*

Proof We have

$$\sigma^2 = \frac{\int j^2kdx}{\int kdx} - \frac{(\int jkdx)^2}{(\int kdx)^2},$$

and hence

$$
\begin{aligned}
\left(\sigma^2\right)_a &= \left(\frac{\int j^2 k dx}{\int k dx} - \frac{\left(\int j k dx\right)^2}{\left(\int k dx\right)^2}\right)_a \\
&= h(a)\left(\frac{\int j^3 k dx}{\int k dx} - \frac{\int j^2 k dx \int j k dx}{\left(\int k dx\right)^2} - 2\frac{\int j k dx}{\int k dx}\left(\frac{\int j^2 k dx}{\int k dx} - \frac{\left(\int j k dx\right)^2}{\left(\int k dx\right)^2}\right)\right) \\
&= h(a)\left(\mathbb{E}[j^3|a] - 3\bar{j}\mathbb{E}[j^2|a] + 2\bar{j}^3\right) \\
&= h(a)\mathbb{E}[(j-\bar{j})^3|a] \\
&= h(a)\sigma^3\gamma,
\end{aligned}
$$

where the fourth equality is standard from the third central moment of a distribution. □

This in hand, note that

$$
(\mathbb{E}[j])_a = \int j f_a dx = \int j \frac{f_a}{f} f dx = \int j\left(hj + \frac{m'}{m}\right) f dx = h(a)\int j(j-\bar{j}) f dx = h\left(a\right)\sigma^2
$$

where the third equality uses (14) and the fourth (24).

Differentiating again and using Lemma 4, we obtain

$$
(\mathbb{E}[j])_{aa} = h'(a)\sigma^2 + h(a)(\sigma^2)_a = h'(a)\sigma^2 + h^2(a)\sigma^3\gamma.
$$

Hence, $(\mathbb{E}[j])_{aa}/(\mathbb{E}[j])_a = (h'(a)/h(a)) + h(a)\sigma\gamma$, and thus (15) holds if and only if (17) holds.

To prove the final assertion, note that we desire that

$$
-\gamma =_s -\mathbb{E}[(j-\bar{j})^3|a] = -\int \left(j^2 - 2\bar{j}j\right)\left(j-\bar{j}\right) f dx \geq 0.
$$

Integrating by parts yields

$$
\begin{aligned}
-\gamma &=_s -\left(j^2 - 2\bar{j}j\right)\int_{\underline{x}}^{x}\left(j(s)-\bar{j}\right) f(s) ds\Big|_{\underline{x}}^{\bar{x}} + 2\int\left(j-\bar{j}\right) j' \int_{\underline{x}}^{x}\left(j(s)-\bar{j}\right) f(s) ds dx \\
&= 2\int\left(j-\bar{j}\right) j' \int_{\underline{x}}^{x}\left(j(s)-\bar{j}\right) f(s) ds dx,
\end{aligned}
$$

where the integrand in the last expression has sign pattern $+/-$ since $\int_{\underline{x}}^{x}\left(j(s)-\bar{j}\right) f(s) ds \leq 0$, and since $j' > 0$. By a standard integral inequality (see Beesack (1957)), it would thus be enough that j'/f is decreasing and

$$
\int (j-\bar{j}) f \int_{\underline{x}}^{x} (j(s)-\bar{j}) f(s) ds dx \geq 0. \tag{25}
$$

To see that (25) holds, note that since

$$(j(x) - \bar{j})f(x) = \frac{\partial}{\partial x} \int_{\underline{x}}^{x} (j(s) - \bar{j})f(s)ds,$$

the left hand side of (25) is equal to

$$\frac{1}{2} \left(\int_{\underline{x}}^{x} (j(s) - \bar{j})\, f(s)ds \right)^2 \Bigg|_{\underline{x}}^{\bar{x}} = 0.$$

To see that j'/f is decreasing, note that

$$\left(\frac{j'}{f}\right)_x =_s \frac{j''}{j'} - \frac{f_x}{f} = \frac{j''}{j'} - \frac{n'}{n} - j'(x)H(a) \leq \left(\log \frac{j'}{n}\right)_x =_s \left(\frac{j'}{n}\right)_x \leq 0$$

where the two sign equalities use $j' > 0$, the second equality uses (13), the first inequality follows since $H(a) \geq 0$, and the second inequality follows by assumption. $\square$

References

Beesack, P. R. (1957): "A Note on an Integral Inequality," *Proceedings of the Americal Mathematical Society*, 8(5), 875–879.

Brown, M., S. Chiang, S. Ghosh, and E. Wolfstetter (1986): "A new class of sufficient conditions for the first-order approach to the principal-agent problem," *Economics Letters*, 21, 1–6.

Chade, H., and J. Swinkels (2016): "Disentangling Moral Hazard and Adverse Selection," Arizona State University Working Paper.

Conlon, J. (2009): "Two New Conditions Supporting the First-Order Approach to Multisignal Principal–Agent Problems," *Econometrica*, 77, 249–278.

Dewatripont, M., I. Jewitt, and J. Tirole (1999): "The economics of career concerns, part I: Comparing information structures," *The Review of Economic Studies*, 66, 183–198.

Holmstrom, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 74–91.

Jewitt, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, pp. 1177–1190.

Jewitt, I., O. Kadan, and J. Swinkels (2008): "Moral Hazard with Bounded Payments," *Journal of Economic Theory*, 143, 59–82.

Jung, J. Y., and S. K. Kim (2015a): "Information space conditions for the first-order approach in agency problems," *Journal of Economic Theory*, 160, 243–279.

——— (2015b): "Justifying the First-Order Approach to the Principal-Agent Problem without Concavity," Seoul National University Working Paper.

Karlin, S. (1955): "Pólya Type Distributions, II," Technical Report 1, Office of Naval Research, NR 042–995.

Lehmann, E., and J. Romano (2005): *Testing Statistical Hypothesis*. Springer, third edn.

Mirrlees, J. (1975): "On Moral Hazard and the Theory of Unobservable Behavior," Nuffield College.

Rogerson, W. (1985): "The First-Order Approach to Principal-Agents Problems," *Econometrica*, 53, 1357–1367.