A New Class of Non–Existence Examples for the Moral Hazard Problem*

Sofia Moroni and Jeroen Swinkels

April 10, 2013

Abstract

We provide a class of counter-examples to existence in a simple moral hazard problem in which the first-order approach is valid. These examples involve utility functions such as $\ln(w)$ in which utility diverges to negative infinity at a finite wealth level. In contrast to the Mirrlees example, unbounded likelihood ratios are not needed, and no sequence of contract can approach the full information first best. The examples expose a hidden (and not always correct) assumption in the existence proof of Jewitt, Kadan, and Swinkels (2008). We show that their result (and proof method) is valid under very mild conditions outside of this class of utility functions. In particular, if utility only goes to negative infinity when wealth does as well, then any utility function where the agent continues to dislike risk as wealth diverges to negative infinity will suffice.

1 Introduction

Mirrlees (1999) provides a classic example of a moral hazard problem with no optimal solution. In the example, which relies on an unbounded likelihood ratio, there is a sequence of contracts the cost of which converges to the full information first best, but no contract that exactly achieves the first best.

Jewitt, Kadan, and Swinkels (2008, henceforth JKS) provide a proof of existence when likelihood ratios are bounded. We point to an implicit, and not always correct, assumption in JKS by exhibiting a new class of examples in which there is non-existence of the optimal contract even if the likelihood ratio is bounded. The problem in these examples is that on the one hand, we know that an optimal contract must be of the form given by Hölmstrom (1979). But, in some settings, the principal can easily run out of room to provide adequate incentives by contracts of the required form. In particular, we show conditions under which any contract of the required form that gives the right utility provides inadequate incentives.

When utility diverges at a finite consumption level the existence problem is severe − for any given utility function and information structure one can find a specification of the agent's cost of effort and reservation utility such that an optimal contract does not exist. When utility diverges to $-\infty$ only as consumption does as well, we show a necessary and sufficient condition for the missing step in JKS to hold, and hence for existence. The condition is that, in a sense to be made

*We thank Larry Samuelson and Ohad Kadan for helpful comments.

precise, the agent remains risk averse as consumption diverges to $-\infty$. Under a technical regularity condition, the condition fails only if u has bounded slope, and so becomes essentially linear as consumption diverges. Hence, the existence problem is not severe for this case. For settings where the utility is bounded below, the original JKS proof goes through without problem.

As is intuitive given bounded likelihood ratios, we show that unlike in the Mirrlees example we cannot approach the first best. But, we show that one can approach the second best using contracts that bound utility strictly above $-\infty$,[1] and provide a characterization of both the limit to which these contracts converge, and the limit cost. The limit contract has an intuitive form. It can be thought of as beginning from a contract of the standard form that provides both too much utility and too little incentives, and then modifying it by providing very low utility on a small interval near the lowest signal, which both eliminates the extra utility the base contract provided and relaxes the incentive constraint so as to restore feasibility. Since the problem is too much utility and too little incentives, the "right" way to do this is to concentrate the modification closer and closer to the lowest signal (because it is here that the trade-off of providing extra incentives per unit of utility taken away is most favorable), and it is this that leads to non-existence.

2 Model

The model is standard. A risk neutral principal employs a risk averse agent. The agent's utility function over final wealth is $u : D \to \mathbb{R}$ where u is in C^2 with $u' > 0$ and $u'' < 0$. To focus attention, we assume $D \subseteq \mathbb{R}$ is an interval with upper bound ∞, and lower bound $\underline{d}$, where $\underline{d}$ may be finite or $-\infty$, and where $\underline{d}$ may or may not be in D. We assume $u(\underline{d}) \equiv \lim_{w \downarrow \underline{d}} u(w) = -\infty$, $\lim_{w \uparrow \infty} u(w) = \infty$, and $\lim_{w \uparrow \infty} u'(w) = 0$. Write also $u'(\underline{d})$ for $\lim_{w \downarrow \underline{d}} u'(w)$.[2]

The agent chooses an effort level $e \in [0, \bar{e}]$ which is unobservable to the principal. The effort cost is given by the function $c(e) \in C^2$ with $c'(e) > 0$ for $e > 0$ and $c''(e) > 0$. The agent's utility is additively separable and equal to

$$u(w) - c(e).$$

An outcome $x \in X = [0, 1]$ is realized according to $F(x|e) \in C^3$. The density of F is f, with $f(x|e) > 0$ for all x and e. As f is continuous, $f(x|e)$ is uniformly bounded from zero and so $f_e(x|e)/f(x|e)$ is also bounded. We assume that $\frac{\partial}{\partial x}\left(\frac{f_e(x|e)}{f(x|e)}\right)$ is strictly positive, so that the monotone likelihood ratio property (MLRP), holds.

The principal's gross benefit of e is given by

$$B(e) = \int x f(x|e)\, dx$$

The agent is compensated according to $\pi(x) : (0, 1] \to D$.[3] Let

$$U(\pi, e) \equiv \int u(\pi(x)) f(x|e) dx.$$

[1] These are cases for which JKS's existence proof is valid, and for which contracts have a simple characterization as truncated versions of standard Hölmstrom (1979) contracts.

[2] For the case $\underline{d} > -\infty$, it follows from $u(\underline{d}) = -\infty$ that $u'(\underline{d}) = \infty$. When $\underline{d} = -\infty$, it is possible that $u'(\underline{d}) > -\infty$. Assume for example that for $w < -1$, $u(w) = w + \frac{1}{w}$.

[3] We write things this way to allow for the possibility that at the zero measure point $x = 0$ the principal may wish to pay $\underline{d}$, where (recall) $\underline{d}$ may not be in D.Since $u(\underline{d}) = -\infty$, it cannot be that the principal wants to pay $\underline{d}$ on a positive measure set of signals in any contract satisfying the agent's individual rationality constraint.

The agent's net utility is given by $U(\pi, e) - c(e)$. Let

$$C(\pi, e) \equiv \int \pi(x) f(x|e) dx$$

be the expected cost to the principal of contract π and effort e.

The participation constraint is

$$U(\pi, e) - c(e) \geq u_0, \tag{IR}$$

where u_0 is the outside option. The incentive compatibility constraint is

$$e \in \arg\max_{\hat{e} \in [0, \bar{e}]} (U(\pi, \hat{e}) - c(\hat{e})).$$

Following Rogerson (1985), let the (doubly) relaxed incentive constraint be

$$U_e(\pi, e) \geq c'(e), \tag{IC}$$

and assume Convexity of the Distribution Function (CDFC) so that $F_{ee}(x|e) \geq 0$ for all x and e.[4] Under (CDFC) the First Order Approach is valid.

Write $\underline{b}$ for $\frac{1}{u'(\underline{d})}$. For any given $e > 0$ the relaxed cost minimization problem is

$$\begin{aligned} & \min_{\pi} C(\pi, e) \\ s.t. \quad & IC, IR. \end{aligned}$$

For any given e, $\mu > 0$, and $b \geq \underline{b}$, write

$$\delta(x|e) = \frac{f_e(x|e)}{f(x|e)} - \frac{f_e(0|e)}{f(0|e)}$$

and define the contract $\pi_{b,\mu,e}(\cdot)$ implicitly on $[0, 1]$ by

$$\frac{1}{u'(\pi_{b,\mu,e}(x))} = b + \mu \delta(x|e), \tag{1}$$

where we note that by MLRP, for $x > 0$, $\pi_{b,\mu,e}(x) \in D$, and where we follow the convention that $\pi_{\underline{b},\mu,e}(0) = \underline{d}$. Note that this is just a re-parametrization of the standard Hölmstrom (1979) (λ, μ)-contract, where in particular, $b = \lambda + \mu \frac{f_e(0|e)}{f(0|e)}$.

Lemma 1 *A necessary condition for optimality in the doubly relaxed problem is that the contract is of the form $\pi_{b,\mu,e}(\cdot)$ for some $b \geq \underline{b}$ and $\mu > 0$ where $U\left(\pi_{b,\mu,e}, e\right) = u_0 + c(e)$ and $U_e\left(\pi_{b,\mu,e}, e\right) = c'(e)$.*

Proof This is essentially standard (see Hölmstrom (1979)). But, given that one point of this paper is to argue that there is some lack of clarity in previous existence proofs, we provide an elementary proof for completeness. Let $\pi : (0, 1] \to D$ be an optimal contract. Choose disjoint Borel sets X_1, X_2, X_3 and X_4 of $(0, 1]$ with positive $F(\cdot|e)$-measure where $\frac{F_e(X_2|e)}{F(X_2|e)} > \frac{F_e(X_1|e)}{F(X_1|e)}$ and $\frac{F_e(X_4|e)}{F(X_4|e)} >$

[4] We could also impose the less stringent conditions from Jewitt (1988), except in the section in which we approach the second-best by constrained optimal contracts.

$\frac{F_e(X_3|e)}{F(X_3|e)}$. Consider raising utility by $\frac{z}{F(X_2|e)}$ on X_2, and lowering it by $\frac{z}{F(X_1|e)}$ on X_1. This leaves the expected utility of the agent unchanged, and changes his incentives at rate

$$\frac{F_e(X_2|e)}{F(X_2|e)} - \frac{F_e(X_1|e)}{F(X_1|e)},$$

while changing the expected cost to the principal at $z = 0$ at rate

$$\begin{aligned} & \frac{1}{F(X_2|e)} \int_{X_2} \frac{1}{u'(\pi(x))} f(x|e)\, dx - \frac{1}{F(X_1|e)} \int_{X_1} \frac{1}{u'(\pi(x))} f(x|e)\, dx \\ = \quad & E\left(\frac{1}{u'(\pi(x))}|X_2\right) - E\left(\frac{1}{u'(\pi(x))}|X_1\right). \end{aligned}$$

So, if

$$\frac{E\left(\frac{1}{u'(\pi(x))}|X_2\right) - E\left(\frac{1}{u'(\pi(x))}|X_1\right)}{\frac{F_e(X_2|e)}{F(X_2|e)} - \frac{F_e(X_1|e)}{F(X_1|e)}} < \frac{E\left(\frac{1}{u'(\pi(x))}|X_4\right) - E\left(\frac{1}{u'(\pi(x))}|X_3\right)}{\frac{F_e(X_4|e)}{F(X_4|e)} - \frac{F_e(X_3|e)}{F(X_3|e)}},$$

then the principal can improve on $\pi(\cdot)$ by increasing payments on X_2 and X_3, and decreasing them on X_1 and X_4 so as to restore the original incentives. It follows that a necessary condition for optimality is that the contract is of the form $\pi_{b,\mu,e}$. That $\mu > 0$ follows since otherwise $U_e(\pi, e) \leq 0$. That $b \geq \underline{b}$ follows since $\pi(x) \in D$ for all $x \in (0, 1]$, and hence $\frac{1}{u'(\pi(x))} > \underline{b}$. That IR binds follows from the standard thought experiment of removing a small constant from utility everywhere. That IC binds is also standard. If not, then pick disjoint intervals X_1 and X_2 where X_2 is strictly to the right of X_1. Lower payments on X_2 by z, and raise them on X_1 by z to restore IR. Since IC is slack, this contract is feasible for small z, and since payments on X_2 are strictly larger than on X_1 and the agent is risk averse, the principal saves money when z is small. ■

3 Non-existence: a class of simple examples

Let us turn to the new class of non-existence examples. In this section, we work with $u(w) = \ln w$. In the next section, we work with more general utility functions. Note in what follows that since we have assumed that $\frac{f_e}{f}$ is bounded, our class includes a set of cases for which Jewitt, Kadan, and Swinkels (2008) assert existence.

Lemma 2 *Let $u(w) = \ln w$. Then, $U_e(\pi_{0,1,e}, e)$ is finite, and for any $c(\cdot)$ with $U_e(\pi_{0,1,e}, e) < c'(e)$, there exists no optimal contract implementing e.*

The point of the proof is to show that $U_e(\pi_{0,1,e}, e)$ is in fact that strongest incentive that can be provided with a contract of the form required by Lemma 1.

Example 1 *Let*

$$f(x|e) = 1 + (e - 1)\left(x - \frac{1}{2}\right)$$

be the standard FGM linear copula. Then, it is straightforward that $\delta(x|1) = x$ and so

$$\begin{aligned} U_e(\pi_{0,1,e}, e) &= \int_0^1 (\ln \delta(x|1)) f_e(x|1)\, dx \\ &= \int_0^1 (\ln x)\left(x - \frac{1}{2}\right) dx = \frac{1}{4}. \end{aligned}$$

Hence, for any $c(\cdot)$ with $c'(1) > \frac{1}{4}$, there is no optimal contract implementing $e = 1$.

Proof of Lemma 2 That $U_e(\pi_{0,1,e}, e)$ is finite will be shown in Lemma 4. Because $u(w) = \ln w$, we have $\frac{1}{u'(w)} = w$, and so

$$U_e\left(\pi_{b,\mu,e}, e\right) = \int_0^1 \ln(b + \mu\delta(x|e)) f_e(x|e)\, dx,$$

from which

$$\frac{\partial}{\partial b} U_e\left(\pi_{b,\mu,e}, e\right) = \int_0^1 \frac{1}{b + \mu\delta(x|e)} f_e(x|e)\, dx,$$

where by $MLRP$, $\frac{1}{b+\mu\delta(x|e)}$ is positive and decreasing in x. Hence, since $f_e(\cdot|e)$ single crosses 0 from below, and so, since $\int_0^1 f_e(x|e) = 0$,

$$\frac{\partial}{\partial b} U_e\left(\pi_{b,\mu,e}\right) \leq 0.$$

But,

$$\frac{\partial}{\partial \mu} U_e\left(\pi_{b,\mu,e}\right) = \int_0^1 \frac{\delta(x|e)}{b + \mu\delta(x|e)} f_e(x|e)\, dx \geq 0,$$

with equality if and only if $b = 0$, since the fraction is constant in x for $b = 0$, and strictly increasing otherwise.

It follows that for each μ, $U_e\left(\pi_{b,\mu,e}\right)$ is maximized by taking $b = 0$, and that $U_e(\pi_{0,\mu,e})$ is independent of μ. Hence for any $c(\cdot)$ and e for which $c'(e) > U_e(\pi_{0,1,e})$, no contract of the form $\pi_{b,\mu,e}$ satisfies IC, and so by Lemma 1, there is in fact no optimal contract. ■

The difficulty for the proof in JKS is that it implicitly assumes that as $b \to \underline{b}$, so that utility at the worst outcome diverges to $-\infty$, expected utility and incentives diverge. As Example 1 illustrates, this can fail.

4 A necessary and sufficient condition for existence

We now present a necessary and sufficient condition for existence of the optimal contract for each specification of costs and the outside option. We will see that the condition always fails when $\underline{d} > -\infty$, and so the counterexample from the previous section is in fact very general. But, we will also see that when $\underline{d} = -\infty$, the condition is extremely mild, so that except in somewhat tortured examples, existence is guaranteed. We note also that the problem we identify in the proof by JKS is not an issue when $u(\underline{d}) > -\infty$. Hence, existence is essentially fine except in the case where utility goes to $-\infty$ at some finite pay level, but is in doubt in any such case.

4.1 Asymptotic risk aversion

Say that utility function u has *Asymptotic Risk Aversion (ARA)* if

$$\lim_{w \downarrow \underline{d}} \frac{-u(w)}{u'(w)} + w = -\infty.$$

When $\underline{d}$ is finite, then, since $\lim_{w \downarrow \underline{d}} \frac{-u(w)}{u'(w)} \geq 0$, ARA must fail. On the other hand, when $\underline{d} = -\infty$, then ARA is very weak. In particular, we have the following lemma which shows only if u' converges to a finite limit, so that the agent becomes effectively risk neutral as income tends to $-\infty$, can ARA fail.[5]

Lemma 3 *Assume that $\underline{d} = -\infty$. Assume also that $\frac{wu'(w)}{u(w)}$ and $\frac{-u''(w)}{u'(w)}$ have well defined limits as $w \to -\infty$. Then, sufficient for ARA is that $u'(\underline{d}) = -\infty$.*

Proof We want to prove that

$$\lim_{w \downarrow -\infty} -\frac{u(w)}{u'(w)} + w = -\infty. \tag{2}$$

The derivative of this expression is given by $\frac{u(w)u''(w)}{(u'(w))^2}$ and therefore we need

$$\lim_{w \downarrow -\infty} \int_w^0 \frac{u(w)u''(w)}{(u'(w))^2} = \infty.$$

For sufficiency, note that by the convergence test for integrals the previous expression holds if

$$\lim_{w \downarrow -\infty} \frac{\frac{u(w)u''(w)}{(u'(w))^2}}{-\frac{1}{w}} = \lim_{w \downarrow -\infty} \frac{u(w)}{u'(w)w} \underbrace{\frac{\frac{-u''(w)}{u'(w)}}{\frac{1}{w^2}}}_{=A} > 0. \tag{3}$$

If $\lim_{w \downarrow -\infty} \frac{u(w)}{u'(w)w} = 0$, then (2) holds. If $\lim_{w \downarrow -\infty} \frac{u(w)}{u'(w)w} > 0$, then (3) holds because A diverges. This observation also follows by the convergence test since the integral of the numerator of A diverges (given our assumption that $\lim_{w \downarrow -\infty} u' = \infty$ and given $\frac{u''}{u'} = \frac{\partial \ln(u')}{\partial w}$), while $\int_{-\infty}^{-1} \frac{1}{w^2} dw$ is finite. ■

The key implication of ARA is contained in the following lemma.

Lemma 4 *Fix $e > 0$ and $\mu > 0$. Then,*

$$\lim_{b \downarrow \underline{b}} U\left(\pi_{b,\mu,e}, e\right) = -\infty \tag{4}$$

if and only if u has ARA, and

$$\lim_{b \downarrow \underline{b}} U_e\left(\pi_{b,\mu,e}, e\right) = \infty \tag{5}$$

[5] When u' does converge to a finite limit, ARA may or may not be satsified. It fails when $u(\cdot)$ has the form $u(w) = w + \frac{1}{w}$ for $w < -2$, but is satisfied when $u(w) = w + \ln(-w)$ for $w < -2$.

if and only if u has ARA.[6,7]

Proof Fix $e > 0$ and $\mu > 0$. Since $F \in C^3$ it follows from our assumptions that $\frac{\partial}{\partial x}\left(\frac{f_e(x|e)}{f(x|e)}\right)$ is bounded away from 0 and ∞. The implication is that $h(\cdot|e)$ defined as the density of $\delta(x|e)$, is also bounded from below by some $\underline{h} > 0$, and above by some $\bar{h} < \infty$ on support $[0, \delta(1|e)]$. Let $\omega(t)$, defined by

$$\frac{1}{u'(\omega(t))} = t, \tag{6}$$

be the amount the agent is paid when $\frac{1}{u'} = t$. Then, changing variables,

$$\begin{aligned}\lim_{b\downarrow\underline{b}} U\left(\pi_{b,\mu,e}, e\right) &= \lim_{b\downarrow\underline{b}} \int_0^1 u(\pi_{b,\mu,e}) f(x|e)\, dx \\ &= \lim_{b\downarrow\underline{b}} \int_0^{\delta(1|e)} u(\omega(b+\mu z))\, h(z|e)\, dz,\end{aligned}$$

and,

$$\begin{aligned}\lim_{b\downarrow\underline{b}} U_e\left(\pi_{b,\mu,e}, e\right) &= \lim_{b\downarrow\underline{b}} \int_0^1 u(\pi_{b,\mu,e}) \frac{f_e(x|e)}{f(x|e)} f(x|e)\, dx \\ &= \lim_{b\downarrow\underline{b}} \int_0^{\delta(1|e)} u(\omega(b+\mu z)) \left(z + \frac{f_e(0|e)}{f(0|e)}\right) h(z|e)\, dz.\end{aligned}$$

Choose $\hat{z} > 0$ and $\hat{b} > \underline{b}$ such that $\hat{z} + \frac{f_e(0|e)}{f(0|e)} < 0$ and $u(\omega(\hat{b} + \mu\hat{z})) < 0$. Then, both

$$\lim_{b\downarrow\underline{b}} \int_{\hat{z}}^{\delta(1|e)} u(\omega(b+\mu z))\, h(z|e)\, dz$$

and

$$\lim_{b\downarrow\underline{b}} \int_{\hat{z}}^{\delta(1|e)} u(\omega(b+\mu z)) \left(z + \frac{f_e(0|e)}{f(0|e)}\right) h(z|e)\, dz$$

are well defined and finite by the dominated convergence theorem, since in particular, for $b \in [\underline{b}, \hat{b}]$, and $z \in [\hat{z}, \delta(1|e)]$

$$-\infty < u(\omega(\underline{b} + \mu\hat{z})) \leq u(\omega(b+\mu z)) \leq u(\omega(\hat{b} + \mu\delta(1|e))) < \infty$$

and since $\left(z + \frac{f_e(0|e)}{f(0|e)}\right)$ is bounded as well since $\delta(1|e) < \infty$.

But, on $[0, \hat{z}]$, h is bounded by $\underline{h}$ and $\bar{h}$, while

$$-\infty < \frac{f_e(0|e)}{f(0|e)} \bar{h}(z|e) \leq \left(z + \frac{f_e(0|e)}{f(0|e)}\right) h(z|e) \leq \left(\hat{z} + \frac{f_e(0|e)}{f(0|e)}\right) \underline{h} < 0.$$

[6] As will be seen in the proof, both limits are guaranteed to exist in the extended reals.

[7] Carlier and Dana (2004) also prove existence of an optimal contract in a principal agent context. They assume that both the principal and agent are strictly risk averse, and because they restrict utility functions to have full domain, our results do not immediately apply to their setting. But, for utility functions of full domain that fail ARA, we would again speculate that when $c'(\cdot)$ is sufficiently large, the principal will "run out of room" to provide sufficient incentives with contracts of the standard form.

It follows that the limits in (4) and (5) are well-defined, and that both (4) and (5) are equivalent to

$$\lim_{b \downarrow \underline{b}} \int_0^{\hat{z}} u\left(\omega\left(b+\mu z\right)\right) dz = -\infty$$

or, changing variables and discarding a constant, to

$$\lim_{b \downarrow \underline{b}} \int_b^{b+\mu\hat{z}} u\left(\omega\left(z\right)\right) dz = -\infty. \tag{7}$$

Integrating by parts,

$$\begin{aligned}
\int_b^{b+\mu\hat{z}} u\left(\omega\left(z\right)\right) dz &= u\left(\omega\left(z\right)\right) z\big|_b^{b+\mu\hat{z}} - \int_b^{b+\mu\hat{z}} u'\left(\omega\left(z\right)\right) \omega'\left(z\right) z dz \\
&= u(b+\mu\hat{z})(b+\mu\hat{z}) - u\left(\omega\left(b\right)\right) b - \int_b^{b+\mu\hat{z}} \omega'\left(z\right) dz \\
&= u(b+\mu\hat{z})\left(b+\mu\hat{z}\right) - u\left(\omega\left(b\right)\right) b - \omega(b+\mu\hat{z}) + \omega\left(b\right) \\
&= u(b+\mu\hat{z})\left(b+\mu\hat{z}\right) - \frac{u\left(\omega\left(b\right)\right)}{u'\left(\omega\left(b\right)\right)} - \omega(b+\mu\hat{z}) + \omega\left(b\right)
\end{aligned}$$

where both the simplification within the integral and the last equality use (6). But,

$$\lim_{b \downarrow \underline{b}} \left(u(b+\mu\hat{z})\left(b+\mu\hat{z}\right) - \omega(b+\mu\hat{z})\right) = \left(u(\underline{b}+\mu\hat{z})\left(\underline{b}+\mu\hat{z}\right) - \omega(\underline{b}+\mu\hat{z})\right)$$

and thus is finite and so (7) is equivalent to

$$\lim_{w \downarrow \underline{d}} \frac{-u\left(w\right)}{u'\left(w\right)} + w = -\infty,$$

as claimed. ■

Before we move on, we make two notes. First, an issue similar to what we have identified can also arise if $\lim_{w\uparrow\infty} u'\left(w\right) > 0$. Then, unless an analog to ARA holds, one can have a case where, holding fixed μ, utility does not diverge as we increase b towards the value where $b + \mu\frac{f_e(x|e)}{f(x|e)} = \frac{1}{\lim_{w\uparrow\infty} u'(w)}$. We rule this case out both because we find it less economically interesting, and because the technical point is already made in the case we examine. Second, Carlier and Dana (2005) do not, as far as we can tell, have the same problem as JKS. The key is that in their environment, the principal is also strictly risk averse with utility $v\left(\cdot\right)$, where in particular there is $\alpha \in (0,1)$, such that as $w \uparrow \infty$, $\frac{v(w)}{w^\alpha} \to 0$. This implies that as large amounts are transfered from the agent to the principal, there remains enough curvature in the system as to avoid the discontinuity issue here. In particular note that since $\frac{v(w)}{w^\alpha} \to 0$, $\lim_{w\uparrow\infty} v'(w) = 0$.

4.2 The Result and Proof

We begin with a small lemma. For the balance of the paper, assume that some $e > 0$ has been fixed.

Lemma 5 *Let u not have ARA. Then, there is a unique $\hat{\mu}$ such that if we let $\hat{\pi} = \pi_{\underline{b},\hat{\mu},e}$, then*

$$U(\hat{\pi}, e) = u_0 + c(e).$$

Proof This follows immediately from the intermediate value theorem, noting that by Lemma 4 $U\left(\pi_{\underline{b},\mu,e}, e\right)$ is finite and continuous in μ for all $\mu > 0$, and that for $x > 0$, $\pi_{\underline{b},\mu,e}(x)$ is increasing in μ with $\lim_{\mu\downarrow 0} \pi_{\underline{b},\mu,e}(x) = -\infty$ and $\lim_{\mu\uparrow 0} \pi_{\underline{b},\mu,e}(x) = \infty$ and so by the monotone convergence theorem, $\lim_{\mu\downarrow 0} U\left(\pi_{\underline{b},\mu,e}, e\right) = -\infty$ and $\lim_{\mu\uparrow\infty} U\left(\pi_{\underline{b},\mu,e}, e\right) = \infty$. ■

Proposition 2 *An optimal contact implementing e exists if either u has ARA or u does not have ARA, but $U_e(\hat{\pi}, e) \geq c'(e)$. An optimal contract does not exist if u does not have ARA and $U_e(\hat{\pi}, e) < c'(e)$*

Thus, when ARA holds, existence holds regardless of $c(\cdot)$ and u_0. But, when ARA fails, then existence will hold for some specifications of $c(\cdot)$ and u_0, but one can always find a specification of $c(\cdot)$ and u_0 where existence fails.

Proof of Necessity Assume $U_e(\hat{\pi}, e) < c'(e)$, and, given Lemma 1 let $(\tilde{b}, \tilde{\mu}) \neq (\underline{b}, \hat{\mu})$ be such that $\tilde{\pi} = \pi_{\tilde{b},\tilde{\mu},e}$ is optimal, where $U(\tilde{\pi}, e) = U(\hat{\pi}, e) = u_0 + c(e)$. Then, by Lemma 5, $\tilde{b} > \underline{b}$, and so, since $U(\tilde{\pi}, e) = U(\hat{\pi}, e)$, it must be that $\tilde{\mu} < \hat{\mu}$, and that $\tilde{\pi}$ single crosses $\hat{\pi}$ from above. But then, since

$$U(\hat{\pi}, e) - U(\tilde{\pi}, e) = \int_0^1 (u(\hat{\pi}) - u(\tilde{\pi})) f(x|e)\, dx = 0,$$

it follows from $MLRP$ that

$$U_e(\hat{\pi}, e) - U_e(\tilde{\pi}, e) = \int_0^1 (u(\hat{\pi}) - u(\tilde{\pi})) \frac{f_e(x|e)}{f(x|e)} f(x|e)\, dx > 0,$$

and so, as $\hat{\pi}$ fails IC, so does $\tilde{\pi}$. ■

Proof of Sufficiency The point here is simply that the JKS construction works once either (4) holds or $U_e(\hat{\pi}, e) \geq c'(e)$. In particular, assume (4). Then, for any given c, u_0, e, and $\mu \geq 0$, there is, by the intermediate value theorem, $b(\mu) > \underline{b}$ such that IR holds with equality at $\pi_{b(\mu),\mu,e}$. Because $b(\mu) > \underline{b}$, payoffs and incentives are continuous in b and μ at $(b(\mu), \mu)$, and so, as in JKS, $b(\cdot)$ is continuous and hence $U_e\left(\pi_{b(\mu),\mu,e}, e\right)$ is continuous in μ. But, $U_e\left(\pi_{b(0),0,e}, e\right) = 0$ and JKS argue that when μ is large enough, IC is slack.[8] But then, again by the intermediate value theorem, there is a μ where $U_e\left(\pi_{b(\mu),\mu,e}, e\right) = c'(e)$, and, as JKS show, this is sufficient to establish optimality. Similarly assume ARA fails, but $U_e(\hat{\pi}, e) \geq c'(e)$. Then, for $\mu < \hat{\mu}$, $b(\mu) > \underline{b}$ by definition of $\hat{\mu}$, and so on the domain $[0, \hat{\mu}]$, $b(\mu)$ and $U_e(\pi_{b(\mu),\mu,e}, e)$ are continuous in μ, and we are done as before by the intermediate value theorem. ■

[8] The argument is particularly simple here. Let x^* be such that $f_e(x^*|e) = 0$. Consider the contract π_z which pays z for $x \geq x^*$, and pays a constant $k(z)$ for $x < x^*$, where $k(z) > \underline{d}$ is chosen so that π_z satisfies IR with equality. Choose some $\hat{z}$ large enough that IC is slack at $\pi_{\hat{z}}$. Now, note that for μ sufficiently large, the contract $\pi_{b(\mu),\mu,e}$ will pay less than $\pi_{\hat{z}}(0) = k(\hat{z})$ at 0 (as $\mu \to \infty$, it must be that $b(\mu) \to \underline{b}$) but, since $b + \mu\delta(x^*|e)$ diverges, will for μ large enough, pays more than $\hat{z}$ at x^*. Thus, $\pi_{b(\mu),\mu,e}$ single crosses $\pi_{\hat{z}}$ from below, and so provides stronger incentives than it by Lemma 5 of JKS.

5 Characterizing near optimal contracts

Assume that there is no optimal contract. It still must be the case that over the set of feasible solutions implementing e in the (doubly relaxed) moral hazard problem there is an infimum C^* of expected costs. In this section, we characterize nearly optimal contracts. This gives us a tight expression for C^*, gives insight into how existence fails, and allows us to show that, in contrast to the Mirrlees counter-example, in this example, C^* is strictly higher than the full information first best.

Proposition 3 *Assume there is no optimal contract implementing e. Then, there is one (and only one) pair $\mu^* > 0$ and $\tau > 0$ such that if we let $\pi^* = \pi_{\underline{b},\mu^*,e}$, then*

$$U(\pi^*, e) = c(e) + u_0 + \tau$$

and

$$U_e(\pi^*, e) = c'(e) + \frac{f_e(0|e)}{f(0|e)}\tau.$$

Furthermore,

$$C^* = C(\pi^*, e) - \underline{b}\tau,$$

and $C^*(e)$ can be approximated by contracts each of which have a lower bound $d > \underline{d}$ on payments.

Note that it follows from JKS that π^* is the unique optimal solution to the auxiliary problem in which costs are $\hat{c}(\cdot)$ where $\hat{c}(e) = c(e) + \tau$, and $\hat{c}'(e) = c'(e) + \frac{f_e(0|e)}{f(0|e)}\tau$.

We begin with a Lemma.

Lemma 6 *Let π be any feasible contract. For $d > \underline{d}$, define $\pi_d(x) = \max(\pi(x), d)$. This is trivially also measurable. Then,*

$$\lim_{d \downarrow \underline{d}} \int u(\pi_d(x)) f(x|e)\, dx = \int u(\pi(x)) f(x|e)\, dx, \tag{8}$$

$$\lim_{d \downarrow \underline{d}} \int u(\pi_d(x)) f_e(x|e)\, dx = \int u(\pi(x)) f_e(x|e)\, dx \tag{9}$$

and

$$\lim_{d \downarrow \underline{d}} \int \pi_d(x) f(x|e)\, dx = \int \pi(x) f(x|e)\, dx. \tag{10}$$

Proof As $d \downarrow \underline{d}$, $u(\pi_d(x))$ and $\pi_d(x)$ decrease monotonically, and converge to $u(\pi(x))$ and $\pi(x)$. Hence, (8) and (10)follow by Lebesgue's monotone convergence theorem. To see (9), note that

$$\begin{aligned} & \left| \int u(\pi_d(x)) f_e(x|e) dx - \int u(\pi(x)) f_e(x|e)\, dx \right| \\ = & \left| \int (u(\pi_d(x)) - u(\pi_d(x))) \frac{f_e(x|e)}{f(x|e)} f(x|e)\, dx \right| \\ \leq & \; M \int (u(\pi_d(x)) - u(\pi_d(x))) f(x|e)\, dx \end{aligned}$$

which converges to 0 by (8). ■

Proof of Proposition 3 Define $\nu(\mu) = U\left(\pi_{\underline{b},\mu,e}, e\right) - c(e) - u_0$, and

$$K(\mu) = U_e\left(\pi_{\underline{b},\mu,e}, e\right) - c'(e) - \nu(\mu)\frac{f_e(0|e)}{f(0|e)}.$$

Any $\mu < \hat{\mu}(e)$ has $\nu(\mu) < 0$. By Lemma 5 and Proposition 2, $\nu(\hat{\mu}) = 0$ and $K(\hat{\mu}) < 0$. But, as $\mu \uparrow \infty$, $U\left(\pi_{\underline{b},\mu,e}, e\right)$, and hence $\nu(\mu)$ diverge to ∞, while $U_e\left(\pi_{\underline{b},\mu,e}, e\right)$ remains weakly positive. Hence, since $\frac{f_e(0|e)}{f(0|e)} < 0$, there is at least one $\mu^* > \hat{\mu}$ with $K(\mu^*) = 0$. By construction, μ^* and $\tau = \nu(\mu^*)$ satisfy the conditions of the proposition.

To see that μ^* is unique, note that

$$\begin{aligned}
\frac{\partial}{\partial\mu}K(\mu) &= \frac{\partial}{\partial\mu}\left(U_e\left(\pi_{\underline{b},\mu,e}, e\right) - \frac{f_e(0|e)}{f(0|e)}U\left(\pi_{\underline{b},\mu,e}, e\right)\right) \\
&= \frac{\partial}{\partial\mu}\left(\int_0^1 u\left(\omega\left(\underline{b} + \mu\delta(x|e)\right)\right)\delta(x|e)\, f(x|e)\, dx\right) \\
&= \int_0^1 u'\left(\omega\left(\underline{b} + \mu\delta(x|e)\right)\right)\omega'\left(\underline{b} + \mu\delta(x|e)\right)\delta^2(x|e)\, f(x|e)\, dx > 0.
\end{aligned}$$

Fix $\varepsilon > 0$, and define $\pi_{\varepsilon,d}$ by adding ε to $\pi^*(x)$ where $f_e(x|e) > 0$ and by replacing any payment below d by d. By Lemma 6, for any $\varepsilon > 0$, $d > \underline{d}$ can be chosen small enough that

$$\begin{aligned}
U(\pi_{\varepsilon,d}, e) &> u_0 + c(e) + \tau, \\
U_e(\pi_{\varepsilon,d}, e) &> c'(e) + \frac{f_e(0|e)}{f(0|e)}\tau, \text{ and} \\
C(\pi_{\varepsilon,d}, e) &< C(\pi^*, e) + \varepsilon.
\end{aligned}$$

For any $y \in (0,1]$, define the contract $\tilde{\pi}$ from $\pi_{\varepsilon,d}$ by reducing utility by $\frac{\tau}{F(y|e)}$ utils on $(0,y]$. Since $u(s) = u(d) - \frac{\tau}{F(y|e)}$ has solution $s > \underline{d}$, $\tilde{\pi}$ remains bounded, with $U(\tilde{\pi}, e) > u_0 + c(e)$ and $C(\tilde{\pi}, e) < C(\pi^*, e) + \varepsilon - \underline{b}\tau$. Since $U_e(\pi_{\varepsilon,d}, e) > c'(e) + \frac{f_e(0|e)}{f(0|e)}\tau$, for y small enough, $U_e(\tilde{\pi}, e) > c'(e)$. Thus, since $\varepsilon > 0$ was arbitrary, $C^* \leq C(\pi^*, e) - \tau\underline{b}$, and C^* can be approached by finite contracts.

Assume $C^* < C(\pi^*, e) - \tau\underline{b}$. Then, there is π with

$$\begin{aligned}
U(\pi, e) &\geq u_0 + c(e), \\
U_e(\pi, e) &\geq c'(e), \text{ and} \\
C(\pi, e) &= C(\pi^*, e) - \tau\underline{b} - \kappa
\end{aligned}$$

for some $\kappa > 0$. Fix $y > 0$ such that $\frac{1}{u'(\pi^*(y))} < \underline{b} + \frac{\kappa}{\tau}$ and for $z > 0$, define $\pi_z(x)$ from π^* by setting $u(\pi_z(x)) = u(\pi^*) + \frac{z}{F(y|e)}$ on $(0,y]$, so that $U(\pi_z, e) = u_0 + c(e) + \tau + z$. Note that

$$U_e(\pi_z, e) \geq c'(e) + \frac{f_e(0|e)}{f(0|e)}(\tau + z),$$

and

$$\frac{\partial}{\partial z}C(\pi_z, e)\,|_{z=0} \leq \frac{1}{u'(\pi^*(y))}.$$

Let $\alpha(z) = \frac{z}{\tau+z}$,and define $\tilde{\pi}_z$ implicitly on $(0,1]$ by

$$u(\tilde{\pi}_z(x)) = (1-\alpha(z))\, u(\pi_z(x)) + \alpha(z)\, u(\pi(x)). \tag{11}$$

Then,

$$\begin{aligned} U(\tilde{\pi}_z, e) &= (1-\alpha(z))\, U(\pi_z, e) + \alpha(z)\, U(\pi, e) \\ &\geq (1-\alpha(z))\,(c(e) + u_0 + \tau + z) + \alpha(z)\,(c(e) + u_0) \\ &= c(e) + u_0 + \tau, \end{aligned}$$

and

$$\begin{aligned} U_e(\hat{\pi}_z, e) &= (1-\alpha(z))\, U_e(\pi_z, e) + \alpha(z)\, U_e(\pi, e) \\ &\geq (1-\alpha(z))\left(c'(e) + \frac{f_e(0|e)}{f(0|e)}\tau + \frac{f_e(0|e)}{f(0|e)}z\right) + \alpha(z)\, c'(e) \\ &= c'(e) + \tau\frac{f_e(0|e)}{f(0|e)}, \end{aligned}$$

and so $\hat{\pi}_z$ is a feasible solution to the auxiliary problem.

Now by (11) ,

$$\frac{\partial \hat{\pi}_z(x)}{\partial z} = \frac{1}{u'(\hat{\pi}_z(x))}\left((1-\alpha(z))\frac{\partial u(\pi_z(x))}{\partial z} + \alpha'(z)\,(u(\pi(x)) - u(\pi_z(x)))\right),$$

and so

$$\left.\frac{\partial \hat{\pi}_z(x)}{\partial z}\right|_{z=0} = \frac{1}{u'(\pi^*(x))}\left(\frac{\partial u(\pi_z(x))}{\partial z} + \frac{1}{\tau}\,(u(\pi(x)) - u(\pi^*(x)))\right),$$

from which

$$\begin{aligned} \int \left.\frac{\partial \hat{\pi}_z(x)}{\partial z} f(x|e)\right|_{z=0} &= \int_0^y \frac{1}{u'(\pi^*(x))}\frac{1}{F(y|e)} f(x|e)\,dx + \frac{1}{\tau}\int \frac{1}{u'(\pi^*(x))}\,(u(\pi(x)) - u(\pi^*(x)))\, f(x|e)\,dx \\ &\leq \frac{1}{u'(\pi^*(y))} + \frac{1}{\tau}\int (\pi(x) - \pi^*(x))\, f(x|e)\,dx \\ &= \frac{1}{u'(\pi^*(y))} - \underline{b} - \frac{\kappa}{\tau} \\ &< 0 \end{aligned}$$

by choice of y, which contradicts that π^* is the optimal solution in the auxiliary problem. ■

In the Mirrlees example, non-existence is shown by exhibiting a sequence of contracts that implement e, but approach the cost $u^{-1}(u_0 + c(e))$ of a full information setting. Since $\mu^* > 0$, it follows immediately that in our setting the second best cost is strictly higher than the the first best. In particular, we have

$$C(\pi^*, e) > u^{-1}(u_0 + c(e) + \tau),$$

and so

$$C(\pi^*, e) - \underline{b}\tau > u^{-1}(u_0 + c(e) + \tau) - \underline{b}\tau \geq u^{-1}(u_0 + c(e)).$$

It also follows from Proposition 3 that if one lets b_d and μ_d be such that the optimal contract implementing e subject to paying no less than d pays the larger of d and the solution to $\frac{1}{u'(\pi(x))} = b_d + \mu_d \delta(x|e)$ then $(b_d, \mu_d) \to (\underline{b}, \mu_d)$.

References

[1] Carlier, G., and R. A. Dana, (2004), Existence and Monotonicity of Solutions to Moral Hazard Problems, **Journal of Mathematical Economics**, 41, 826-843.

[2] Holmström, Bengt, (1979); *Moral Hazard and Observability,* **Bell Journal of Economics,** 10, 74-91.

[3] Jewitt, Ian (1988); *Justifying the First Order Approach to Principal-Agent Problems,* **Econometrica**, 56, 1177-1190.

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

[5] Mirrlees, J., (1999); *The Theory of Moral Hazard and Unobservable Behavior: Part I,* **Review of Economic Studies,** 66, 3-21.