Games and Economic Behavior 86 (2014) 1–11

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Note

Wealth effects and agency costs [☆]

Hector Chade [a,*], Virginia N. Vera de Serio [b]

[a] *Arizona State University, Department of Economics, Tempe, AZ 85287-3806, United States*
[b] *Universidad Nacional de Cuyo, Facultad de Ciencias Económicas, I.C.B., Mendoza, 5500 Argentina*

ARTICLE INFO

Article history:
Received 8 January 2011
Available online 18 March 2014

JEL classification:
D86

Keywords:
Moral hazard
Principal–agent model
Contracts
Wealth effects

ABSTRACT

We analyze how the agent's initial wealth affects the principal's expected profits in the standard principal–agent model with moral hazard.
We show that if the principal prefers a poorer agent for all specifications of action sets, probability distributions, and disutility of effort, then the agent's utility of income *must* exhibit a coefficient of absolute prudence less than three times the coefficient of absolute risk aversion for all levels of income, thus strengthening the sufficiency result of Thiele and Wambach (1999). Also, we prove that there is no condition on the agent's utility of income alone that will make the principal prefer *richer* agents. Moreover, we show that, for an interesting class of problems, the principal prefers a relatively poorer agent if agent's wealth is sufficiently *large*. Finally, we discuss how alternative ways of modeling the agent's outside option affects the principal's preferences for agent's wealth.

Published by Elsevier Inc.

1. Introduction

The principal–agent problem with moral hazard is one of the cornerstones of the theory of incentives. In its standard formulation, a risk neutral principal hires a risk averse agent to perform a task, and their relationship is regulated by a contract based on a signal that depends on the agent's unobservable action.[1] In many applications, it is realistic to assume that the principal faces a pool of agents who differ in their wealth. A natural question then is: Does the principal prefer to hire a poorer or a richer agent?

The difficulty in answering this question lies in the way in which agent's wealth impacts the principal's problem. First, an increase in wealth increases the value of the agent's outside option, making it harder for the principal to induce the agent to accept the contract. Second, an increase in wealth affects the agent's attitudes towards risk and thus how costly it is for the principal to induce the agent to bear risk. Third, and more subtly, an increase in agent's wealth increases the risk of the contract that implements any given action. Although the first effect has an unambiguous impact on the principal's expected cost of implementing any action, the last two effects combined have an unclear impact. In a nice paper, Thiele and Wambach (1999) (TW henceforth) proved that if the agent's utility function is additively separable in income and effort, and her utility of income exhibits a coefficient of absolute prudence that is less than three times its coefficient of absolute risk aversion, then the principal prefers a poorer agent. Since many common utility functions satisfy TW's sufficient condition, their result yields an interesting class of problems in which a clear answer to the above question obtains.

[☆] We are grateful to an associate editor and two anonymous referees whose excellent suggestions significantly improved the paper. We also thank Eddie Schlee for helpful comments.

[*] Corresponding author.

[1] The obvious references are Holmstrom (1979) and Grossman and Hart (1983). For a recent contribution to the development of the principal–agent framework, see Jewitt et al. (2008).

http://dx.doi.org/10.1016/j.geb.2014.02.012
0899-8256/Published by Elsevier Inc.

This note continues the analysis of agency costs driven by wealth effects. We provide answers to the following questions: Can a weaker condition than TW's suffice? Is there an analogous condition under which the principal prefers a richer agent instead? Does the principal prefer a poorer agent in 'extreme' cases, such as when the disutility of effort becomes small or the agent's wealth becomes large?

We first show that TW's condition is *tight*: i.e., if the principal prefers a poorer agent across all principal–agent problems (i.e., for all action sets, probability distribution of the observable outcome, disutility of effort, and wealth levels), then the agent's utility of income *must* satisfy TW's condition. Indeed, if their condition fails for some level of income, then one can construct a robust principal–agent problem where the principal prefers a richer agent. Second, we show that given a principal–agent problem, the principal *always* prefers a poorer agent if the task involved entails a (suitably) small disutility of effort. This result implies that there is no analogous condition to TW's under which the principal prefers a *richer* agent. Third, we show that in an important class of problems, the principal *always* prefers a poorer agent if wealth is large enough. Finally, we discuss how alternative assumptions on the agent's reservation utility affect the principal's preference for a poorer or a wealthier agent.

These results are useful in a variety of applications (see TW for other illustrations). For instance, consider the shareholders/CEO application of the principal–agent model. The above results suggest that it might be a bad idea for a firm to hire a very rich CEO who will be costly to motivate. This important consideration is absent in the executive compensation literature that uses the Holmstrom and Milgrom (1987) model, where wealth effects are irrelevant. Also, knowing the principal's preferences for agent's wealth can serve as a building block in matching models where principals are also heterogeneous along some dimension. Finally, wealth effects are also crucial for understanding dynamic moral hazard problems (e.g., Chiappori et al., 1994; Park, 2004; Spear and Wang, 2005). In all these settings, it is helpful to have simple conditions on primitives that yield an unambiguous impact of agent's wealth on principal's profit. Given how few comparative statics properties are available on the principal–agent problem, we view our results as a useful step in this direction.

As mentioned, this paper complements the results of Thiele and Wambach (1999), and also those of a recent paper by Kadan and Swinkels (2013) that, as an application of their analysis without the first-order approach, generalizes TW's result and provides some results on the case with a final wealth or a final transfer constraint.

Section 2 describes the model. Section 3 contains the main results. Section 4 concludes. The main proofs are in Appendix A, and the rest in the (Online) Appendix B.

2. The model and TW's result

2.1. The model

The set-up is the standard principal–agent problem with moral hazard (e.g., Grossman and Hart, 1983). A principal hires an agent to perform a certain task, but since her effort is unobservable, the contract is based on a stochastic output that depends on her effort. The only difference with the standard model is that we assume that the agent has an observable 'initial wealth,' a positive scalar denoted by θ.

The principal is risk neutral and maximizes expected profits defined as the difference between expected output and expected compensation paid to the agent. The agent is risk averse, with utility function for income-action pairs (I, a) given by $V(I+\theta) - \psi(a)$, where $V : (I_\ell, \infty) \to \mathbb{R}$, $I_\ell \geqslant -\infty$, is three times continuously differentiable, strictly increasing, and strictly concave; i.e., $V'(\cdot) > 0$, and $V''(\cdot) < 0$. Also, $\lim_{I \to I_\ell} V(I) = -\infty$. In turn, $\psi(\cdot)$ is nonnegative for all actions a, and it is strictly increasing in a.

Let $\bar{I}$ be the constant income level the agent could obtain with certainty elsewhere if she did not work for the principal. Then her reservation utility is $V(\bar{I}+\theta)$.[2]

We denote by $R(\cdot) = -V''(\cdot)/V'(\cdot)$ and $P(\cdot) = -V'''(\cdot)/V''(\cdot)$ the coefficients of absolute risk aversion and prudence, respectively, associated with $V(\cdot)$.

Let A be the set of feasible actions (e.g., effort levels) available to the agent. We focus on the two most oft-used cases in applications, namely, A is either a finite set $a_1 < a_2 < \cdots < a_m$, or an interval $[0, \bar{a}]$. Wlog, the lowest action in each case is costless for the agent (i.e., $\psi(a_1) = 0$ and $\psi(0) = 0$).

The observable output is denoted by q, and it assumes values in $Q = \{q_1, \ldots, q_n\}$, where wlog we assume that $q_1 < q_2 < \cdots < q_n$. The probability of observing q_i, $i = 1, \ldots, n$, when the agent's action is a is denoted by $\pi_i(a)$, and it is positive for all i and a. We denote by $\pi(a)$ the vector $(\pi_1(a), \pi_2(a), \ldots, \pi_n(a))$.

When $A = [0, \bar{a}]$, we further assume that $\psi(\cdot)$ and $\pi_i(\cdot)$ are twice continuously differentiable in a (three times in one result in Section 3.3), and that $\psi(\cdot)$ is strictly convex in a, i.e., $\psi''(\cdot) > 0$ for every action a, with $\psi'(0) = 0$.

Since the agent's action is unobservable, the principal offers a compensation contract $(I_1, I_2, \ldots, I_n)$ contingent on output and recommends an action a to the agent. Let $B(a) = \sum_{i=1}^{n} \pi_i(a) q_i$ be the expected value of output given action a, and let $C(a, \theta)$ be the minimum cost for the principal of implementing action a if the agent's wealth is θ. As in Grossman and Hart (1983), one can split the analysis of the problem in two steps: first, for each action a, find the contract that minimizes the expected cost to the principal and obtain $C(a, \theta)$; second, find the action that maximizes $B(a) - C(a, \theta)$.

[2] We discuss in Section 3.4 alternative assumptions about the agent's reservation utility.

This completes the description of the model. Notice that in terms of primitives, we can succinctly denote a principal–agent problem by $(V(\cdot), \psi(\cdot), \pi(\cdot), A, Q, \bar{I}, \theta)$.

2.2. The cost minimization problem

The function $C(a, \theta)$ solves:

$$C(a,\theta) = \min_{I_1,\dots,I_n} \sum_{i=1}^{n} \pi_i(a) I_i$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \pi_i(a) V(I_i + \theta) - \psi(a) \geqslant V(\bar{I} + \theta) \tag{1}$$

$$a \in \arg\max_{a' \in A} \sum_{i=1}^{n} \pi_i\big(a'\big) V(I_i + \theta) - \psi\big(a'\big), \tag{2}$$

where (1) is the participation constraint and (2) is the incentive constraint.

If the action set is finite, then (2) consists of a finite number of incentive constraints $\sum_{i=1}^n \pi_i(a) V(I_i+\theta) - \psi(a) \geqslant \sum_{i=1}^n \pi_i(a') V(I_i+\theta) - \psi(a')$, for all a'. If the action set is an interval, then we replace (2) by the first-order condition of the agent's problem $\sum_{i=1}^n \pi_i'(a) V(I_i+\theta) - \psi'(a) = 0$, and we assume that $\pi(\cdot)$ satisfies the monotone likelihood ratio property (MLRP) and the convexity of the distribution function condition (CDFC), so that the 'first-order approach' is valid (Rogerson, 1985).[3]

An equivalent formulation of the problem with the contract written in utility units is as follows: $\min_{v_1,\dots,v_n} \sum_{i=1}^n \pi_i(a) h(v_i) - \theta$ subject to $\sum_{i=1}^n \pi_i(a) v_i - \psi(a) \geqslant V(\bar{I}+\theta)$ and $a \in \arg\max_{a' \in A} \sum_{i=1}^n \pi_i(a') v_i - \psi(a')$, where $h(\cdot)$ denotes the inverse function of $V(\cdot)$, i.e., $h(\cdot) = V^{-1}(\cdot)$, and $v_i = V(I_i + \theta)$. We use both formulations interchangeably.

The utility formulation consists in minimizing a strictly convex function subject to linear constraints. When the constraint set is nonempty, a solution exists (Grossman and Hart, 1983); hence it is unique. Moreover, it is characterized by the Kuhn–Tucker conditions (convex objective and linear constraints imply that no additional regularity condition is needed besides feasibility). And if it is empty for some a, its cost is set to infinity. To avoid repeating this proviso in each case below, we will assume the constraint set is nonempty for every action. (Under MLRP and CDFC, a sufficient condition when $A = [0, \bar{a}]$ is that, for each a, $\pi_i'(a) \neq 0$ for some i, and when A is finite, that for each k, $\pi_i(a_k) \neq \pi_i(a_{k-1})$ for some i. Strict MLRP for all a imply them.)

2.3. A remark on differentiability

Below we will repeatedly differentiate the cost function $C(a, \theta)$, and sometimes also the optimal contract that implements a. This can be justified as follows (the proofs of these assertions are in Section 3 of Appendix B). If all we are interested in is the differentiability of the cost function with respect to θ, then one can show that Corollary 4 in Milgrom and Segal (2002) delivers the result.[4] Under MLRP and CDFC, a simple adaptation of Lemma 2 in Jewitt et al. (2008) allows us to show that the optimal contract and the cost function are continuously differentiable. Moreover, an application of the Implicit Function Theorem reveals that they are twice continuously differentiable if both $\pi(\cdot)$ and $\psi(\cdot)$ are three times continuously differentiable.

2.4. TW's result and intuition

The behavior of $C(a, \theta)$ as θ changes plays a fundamental role in the analysis. An application of the Envelope Theorem yields (see TW Proposition 1)

$$\begin{aligned} \frac{\partial C(a,\theta)}{\partial \theta} &= \left(\sum_{i=1}^{n} \pi_i(a) \frac{1}{V'(I_i+\theta)} \right) V'(\bar{I}+\theta) - 1 \\ &= V'\big(h(\bar{v})\big) \left(\sum_{i=1}^{n} \pi_i(a) \frac{1}{V'(h(v_i))} - \frac{1}{V'(h(\bar{v}))} \right) \\ &= g(\bar{v})^{-1} \left(\sum_{i=1}^{n} \pi_i(a) g(v_i) - g(\bar{v}) \right), \end{aligned} \tag{3}$$

[3] All we need is the validity of the first-order approach. We use MLRP and CDFC for simplicity.

[4] One can apply that corollary after transforming the variables from v_i to $z_i = v_i - \bar{v}$, so that θ appears only in the objective function, and showing that one can restrict attention to a compact feasible set. This requires a non-trivial proof of the continuity of the optimal contract.

where we have set for notational simplicity $\bar{v} = V(\bar{I} + \theta)$ and $g(\cdot) = 1/V'(h(\cdot))$, which is the derivative of $h(\cdot)$, the inverse function of $V(\cdot)$.

From (3), the cost of implementing an action is increasing (decreasing) in agent's wealth if and only if $\sum_{i=1}^{n} \pi_i(a)g(v_i)$ is bigger (smaller) than $g(\bar{v})$.

Thiele and Wambach (1999) provided the following condition on $V(\cdot)$ for the principal to prefer poorer agents for all choices of the other primitives of the model. We include an alternative simple proof of their result that relies on Jensen's inequality.

Proposition 1 *(Sufficiency, TW). If $V(\cdot)$ satisfies $P(I+\theta) \leqslant 3R(I+\theta)$ for all $I+\theta$, then the principal's cost of implementing any action higher than the lowest one is an increasing function of the agent's wealth θ. As a result, the principal's expected profit is a decreasing function of the agent's wealth θ.*

Proof. Simple algebra shows that $P(I+\theta) \leqslant 3R(I+\theta)$ for all $I+\theta$ if and only if $g(\cdot)$ is convex in v (see Amir and Czupryna, 2004). Therefore, for any solution of the cost minimization problem where a is not the lowest action

$$\sum_{i=1}^{n} \pi_i(a)g(v_i) \geqslant g\left(\sum_{i=1}^{n} \pi_i(a)v_i\right) = g\big(\bar{v} + \psi(a)\big) > g(\bar{v}), \tag{4}$$

where the first inequality follows from Jensen's inequality, the equality from the binding participation constraint (1), and the last inequality from $\psi(a) > 0$. Thus, $\sum_{i=1}^{n} \pi_i(a)g(v_i) > g(\bar{v})$, which implies that $\partial C(a,\theta)/\partial\theta > 0$ by (3). As $B(a) - C(a,\theta)$ is decreasing in θ for every a, so is $\max_{a \in A} B(a) - C(a,\theta)$. □

As we stated in the Introduction, many common utility functions satisfy this condition, and hence the result holds for an interesting class of principal–agent problems.

A direct explanation of TW's result follows from the Envelope Theorem. Suppose an increase in the agent's wealth increases her reservation utility by dv utils, i.e., wealth raises by $g(\bar{v})\,dv$. One feasible choice for the principal is to increase the agent's utility v_i by dv for each i, which leaves the incentive and participation constraints unaffected. The expected cost of such a choice is $\sum_i \pi_i(a)g(v_i)\,dv$ for the principal. When $g(\cdot)$ is convex in v, then $\sum_i \pi_i(a)g(v_i)\,dv > g(\bar{v})\,dv$. By the Envelope Theorem, the same holds if the principal responds optimally to the agent's increase in wealth, and then (3) shows that the increase in wealth increases the principal's cost of implementing action a.[5]

A less direct but much more intuitive economic explanation is based on the 'gaps' used in the proof of Proposition 1. That is, the difference between $\sum_i \pi_i(a)g(v_i)$ and $g(\bar{v})$ can be decomposed into two wedges: the first one is the gap between

$$g\left(\sum_{i=1}^{n} \pi_i(a)v_i\right) = g\big(\bar{v} + \psi(a)\big) \quad \text{and} \quad g(\bar{v}), \tag{5}$$

which is clearly positive. It captures the intuition that a wealthier agent needs to be paid more in order to induce her to accept the job offered by the principal, which entails a disutility of effort $\psi(a)$.[6] The second is the gap between

$$\sum_{i=1}^{n} \pi_i(a)g(v_i) \quad \text{and} \quad g\left(\sum_{i=1}^{n} \pi_i(a)v_i\right), \tag{6}$$

which can be positive or negative depending on the function $g(\cdot)$, and whose intuition is more subtle. To see this, notice that (3) is the change in the mean of the contract $(I_1, \ldots, I_n)$ when θ increases, and this is a priori ambiguous. The variance of the contract, however, increases in θ (see Appendix B) due to the incentive constraints. This does not mean that the principal's cost increases in θ, for the agent's risk aversion is also changing. For example, if the agent exhibits increasing absolute risk aversion, then it is clear that the principal's cost increases in θ: the agent bears additional risk and she is more risk averse when her wealth goes up, which implies that her average wage increases and hence so does the principal's cost. If the agent instead exhibits decreasing absolute risk aversion, then a trade off ensues (the contract entails higher risk, but she is more capable of bearing it). If risk aversion does not decrease too fast (e.g., $g(\cdot)$ is convex in v), then the variance effect dominates and the principal's cost goes up.

A nice way to see these effects is to start from a riskless case and assume that the contract makes the agent bear a small risk ε, with mean zero and variance σ_ε^2. Using the standard approximation for risk aversion in the small, her compensation

[5] Along the same lines, in the utility formulation of the problem it is easy to see that an increase in θ by $d\theta$ has a direct impact on the principal's cost equal to $-d\theta$, and an indirect one via the participation constraint, since the reservation utility increases by $g(\bar{v})^{-1}\,d\theta$ utils, and each util has a 'shadow cost' for the principal given by $\sum_i \pi_i(a)g(v_i)$ (the value of the Lagrange multiplier of the participation constraint). Adding both effects yields $(-1 + g(\bar{v})^{-1}(\sum_i \pi_i(a)g(v_i)))\,d\theta$, which is (3).

[6] Note in passing that this is the only gap present in the first-best case in which the action is observable, and thus the principal always prefers a poorer agent in that case.

must be increased by $dI = 0.5R(\bar{v})\sigma_\varepsilon^2$ to keep her incentives constant. Hence, the change in utility is $dv = g(\bar{v})^{-1}dI$, which has mean zero and variance $\sigma_v^2 = g(\bar{v})^{-2}\sigma_\varepsilon^2$. This yields

$$dI = 0.5R(\bar{v})g(\bar{v})^2\sigma_v^2.$$

The Envelope Theorem argument above reveals that the first-order effect on σ_v^2 of a change in wealth is negligible. Thus, dI increases in $\bar{v}$, and hence the principal prefers a poorer agent, if and only if $R(\cdot)g(\cdot)^2$ increases in $\bar{v}$, and differentiation shows that this holds if and only if TW's condition is satisfied. When risk aversion increases in wealth, then $R(\cdot)g(\cdot)^2$ trivially increases since both terms are increasing, but there is trade-off when $R(\cdot)$ is decreasing, as it can offset the increase in $g(\cdot)$ if it falls too fast.

3. Main results

3.1. Tightness of TW's condition

The proof of Proposition 1 reveals that there is some unused *slack* in the participation constraint, namely, $\sum_{i=1}^n \pi_i(a)v_i = \bar{v} + \psi(a)$ implies $\sum_{i=1}^n \pi_i(a)v_i - \bar{v} > 0$ since $\psi(a) > 0$ for all actions above the lowest one. That is, convexity of $g(\cdot)$ does not appear to be a tight condition. This begs the question of whether a *weaker* condition on $V(\cdot)$ suffices. The next result shows that the answer is negative: $V(\cdot)$ must satisfy TW's condition if the principal prefers a richer agent for *all* choices of the other primitives of the principal–agent problem. That is, the condition is indeed *necessary* in a precise sense.[7]

Proposition 2 *(Tightness).* (i) *If the principal's cost of implementing an action higher than the lowest one is increasing in the agent's wealth θ for all choices of $(\psi(\cdot), \pi(\cdot), A, Q, \bar{I}, \theta)$, then $V(\cdot)$ satisfies TW's condition.*

(ii) *If $V(\cdot)$ does not satisfy TW's condition, then the principal prefers a richer agent in some principal–agent problem.*

The proof is in Appendix A, but we sketch the main idea of part (i) here.[8] Suppose that TW's condition fails, i.e., that for some $\bar{v}$ we have $g''(\bar{v}) < 0$. By continuity, $g''(v) < 0$ for all v in some neighborhood of $\bar{v}$. It seems that we could then adjust the other primitives so that for some action a with a small disutility $\psi(a)$, the riskiness of the optimal $(v_1, \ldots, v_n)$ would be small enough to yield $\sum_{i=1}^n \pi_i(a)g(v_i) < g(\bar{v})$ (i.e., $\partial C(a,\theta)/\partial\theta < 0$ and thus the principal prefers a richer agent). The challenge with this argument is to make the gap in (6) negative and large while simultaneously keeping the gap in (5) positive but small enough. It is not a priori clear that this is feasible, since a small gap in (5) can be associated with a contract with v_i very close to $\bar{v}$ for all i, and thus the gap in (6) might end up being smaller in size than the gap in (5).

We produce, however, a principal–agent problem with a continuum of actions and a 'sufficiently convex' $\psi(\cdot)$ such that, for a given action a, it keeps $\psi(a)$ small – to ensure that the gap between $g(\bar{v} + \psi(a))$ and $g(\bar{v})$ is small – and $\psi'(a)$, which drives the riskiness of the contract, large enough – to ensure that the wedge between $\sum_{i=1}^n \pi_i(a)g(v_i)$ and $g(\bar{v} + \psi(a))$ is larger than the previous gap. Intuitively, $\psi(\cdot)$ is very elastic up to a, so the participation constraint can be met with a small increase in the agent's pay, thus keeping the gap in (5) small. At a and beyond, however, it is highly inelastic, and hence the principal must increase the agent's incentives sharply if he wants to implement any such action, i.e., increase the riskiness in the v_i's, thereby making the gap in (6) large.

As a formal hint for why this works, consider the first two terms of the Taylor's expansion of $g(\cdot)$ around $\bar{v}$. Multiplying by $\pi_i(a)$ and summing over i yield

$$\sum_{i=1}^n \pi_i(a)g(v_i) - g(\bar{v}) \approx g'(\bar{v})\sum_{i=1}^n \pi_i(a)(v_i - \bar{v}) + \frac{1}{2}g''(\bar{v})\sum_{i=1}^n \pi_i(a)(v_i - \bar{v})^2. \tag{7}$$

Now, $\sum_{i=1}^n \pi_i(a)(v_i - \bar{v}) = \psi(a)$ by (1), while $\sum_{i=1}^n \pi_i(a)(v_i - \bar{v})^2$ also depends positively on $\psi'(a)$. So if $\psi(\cdot)$ is 'sufficiently convex' as specified above, the right side is negative (recall $g''(\bar{v}) < 0$) and thus $C(a,\cdot)$ is *strictly decreasing* in θ.[9] Then part (ii) completes the specification of the primitives of the problem so that the principal will indeed find it *optimal* to implement such an action a. Thus, we *cannot* weaken TW's condition if we want a parsimonious condition (imposed *solely* on the utility of income) that yields a preference for poorer agents for *all* principal–agent problems.

3.2. Wealth effects for small disutility of effort

Suppose we restrict attention to contracting situations in which the task involved entails a small disutility of effort for the agent, as it would be the case in applications where the agent is hired to perform some minor task. Will the principal prefer a poorer agent?

---

[7] This does not contradict a recent paper by Chiu (2010) that shows that risk aversion alone is sufficient for the principal to prefer a poorer agent if *all* wages in the contract are *greater* than $\bar{I}$, since this requires restrictions on other primitives besides the agent's utility of income.

[8] We are grateful to an anonymous referee for providing us with this more intuitive proof.

[9] Obviously, to make all this precise the proof in Appendix A shows that the right side is negative when instead of an approximation we use an exact second-order Taylor expansion.

The answer to this question will also shed light on the following important one: Could we find a condition analogous to TW's under which the principal always prefers a *richer* agent? Such a condition would be useful in applications in the same way as TW's condition is. As we shall see below, the answer is negative.

Fix a problem $(V(\cdot), \psi(\cdot), \pi(\cdot), A, Q, \bar{I}, \theta)$ *without* imposing TW's condition. As in Grossman and Hart (1983) (Section 5), we parameterize the agent's disutility of effort as $\tilde{\psi}(a) = \eta\psi(a)$, with $\eta \geqslant 0$. Also, assume that the solution of the cost minimization problem is continuously differentiable in η, and denote the cost function by $C(a, \theta, \eta)$.

Note that if the disutility of effort were zero, then the optimal contract would pay a flat wage equal to the reservation wage; i.e., if $\eta = 0$, then $v_i = \bar{v}$ for all i. Hence, $C(a, \theta, 0) = \bar{I}$ implies that $\partial C(a, \theta, \eta)/\partial\theta$ vanishes when evaluated at $\eta = 0$. We will show that this derivative is positive for $\eta > 0$ small enough; that is, the principal's cost of implementing an action is increasing in agent's wealth when her disutility of effort is small. As a result, the principal prefers a *poorer* agent in this case.

Proposition 3 *(Small disutility of effort).* (i) *For any action a above the lowest one, there is an $\eta_a > 0$ such that the cost of implementing action a strictly increases in agent's wealth for all $0 < \eta < \eta_a$.*

(ii) *The principal prefers a poorer agent if her disutility of effort is sufficiently small.*

The proof is in Appendix A. For an intuition of part (i), consider again the approximation of $\sum_{i=1}^{n} \pi_i(a)g(v_i) - g(\bar{v})$ given by (7). Both terms on the right side of (7) go to zero as η approaches zero, but the second term vanishes at a faster rate since it depends on η^2. Thus, for η small enough, the first term on the right side dominates and the left side becomes positive. Unlike the counterexample of Proposition 2, which uses a sufficiently convex disutility of effort function, now the convexity of $\tilde{\psi}(\cdot)$ reduces as η goes to zero. As a result, the gap between $g(\bar{v} + \tilde{\psi}(a))$ and $g(\bar{v})$ becomes larger relative to the gap between $g(\bar{v} + \tilde{\psi}(a))$ and $\sum_{i=1}^{n} \pi_i(a)g(v_i)$, and the result ensues. In turn, part (ii) shows that the bound in (i) for *each* action can be made uniform for *all* actions. This implies that $C(a, \cdot)$ is increasing in θ for all a when η is sufficiently small, and thus $\max_a B(a) - C(a, \theta)$ decreases in θ, i.e., the principal prefers a poorer agent.

The most important implication of Proposition 3 is that *there is no condition imposed solely on $V(\cdot)$ such that the principal's cost of implementing an action is decreasing in agent's wealth for all* $(\psi(\cdot), \pi(\cdot), A, Q, \bar{I}, \theta)$. The proof is simple. If such a condition existed, then the principal would prefer a richer agent when the disutility of effort has the functional form assumed in Proposition 3, i.e., parameterized by η. But the principal always prefers poorer agents when $\eta > 0$ is close to zero, contradiction.

For simplicity, we have focused on a specific parametrization of the disutility of effort that is multiplicatively separable in η and a, which allowed us to find a uniform bound on η and it was sufficient for the implication mentioned above. Two comments are in order. First, instead of finding a uniform bound on η, one could ask whether the principal prefers a poorer agent if the *equilibrium* disutility of effort is small enough. If the optimal contract is twice continuously differentiable, then the answer is affirmative in the continuum of actions case (see Section 6 of Appendix B). Second, a careful inspection of the proof of Proposition 3 reveals that it goes through if we assume a more general $\psi(\cdot, \cdot)$ that is twice continuously differentiable in (a, η), strictly increasing in η for any a above the lowest action, with $\psi(a, 0) = 0$, $\psi_a(a, 0) = 0$, and $\psi_\eta(a, 0) > 0$.

3.3. Rich agents and agency costs

Suppose the pool of agents from which the principal draws the one he hires consists of fairly rich individuals. For instance, the principal could be a firm seeking to hire a CEO. In this case, when would the principal prefer a relatively poorer agent from that pool?

In this section we fix a principal–agent problem $(V(\cdot), \psi(\cdot), \pi(\cdot), A, Q, \bar{I}, \theta)$ and ask whether the principal prefers a poorer agent when wealth is *sufficiently large* (without imposing TW's condition). We provide sufficient conditions on the primitives under which an affirmative answer to this question obtains.

Addressing this question turns out to be technically complex (unlike η in the previous section, now θ affects both $\bar{v}$ and v_i, which complicates the limiting argument). We do, however, derive a result that holds in an important class of principal–agent problems that subsumes cases commonly used in applications. We leave it as an open problem the generalization of the result beyond the class of principal–agent problems considered.

We assume in this section that $V(\cdot)$ is unbounded above. Using the definition of $g(\cdot)$, it is easy to verify that $-g''(\cdot)/g'(\cdot) = (P(h(\cdot)) - 3R(h(\cdot)))/V'(h(\cdot))$.

In the next results we will make use of the following conditions:

(a) There is a threshold $\tilde{v}$ such that either $g(\cdot)$ is convex in v when $v \in (\tilde{v}, \infty)$, or $g(\cdot)$ is concave in v when $v \in (\tilde{v}, \infty)$ and $\lim_{v\to\infty} -g''(v)/g'(v) = 0$.
(b) For any $a \in A$ there is an optimal $(v_1, v_2, \ldots, v_n)$ with $|v_i - \bar{v}| \leqslant K_a$ for all i, where $K_a > 0$ is independent of $\bar{v}$.
(c) $\sup_{a\in A} K_a < \infty$.
(d) The principal's optimal action is bounded away from the lowest action for all θ.

Proposition 4 *(Rich agents).* (i) *Assume* (a) *and* (b)*. Then, for any action a above the lowest one, there is a threshold $\theta_a < \infty$ such that the cost of implementing action a strictly increases in agent's wealth for all $\theta > \theta_a$. Furthermore, if in addition condition* (c) *holds, then given any $\tilde{a}$ above the lowest action, there is a $\theta^* < \infty$ such that the cost of implementing action a strictly increases in agent's wealth for all $a \geqslant \tilde{a}$ and all $\theta > \theta^*$.*

(ii) *Assume* (a)–(c)*. If A is finite, or if $A = [0, \bar{a}]$ and* (d) *holds, then the principal prefers a poorer agent when agent's wealth is sufficiently large.*

Although the proof is long and technical (see Appendix A), we can again use (7) to informally and readily convey the idea underlying part (i). Rewrite it as

$$\sum_{i=1}^{n} \pi_i(a) g(v_i) - g(\bar{v}) \approx g'(\bar{v}) \left(\psi(a) + \frac{1}{2} \frac{g''(\bar{v})}{g'(\bar{v})} \sum_{i=1}^{n} \pi_i(a)(v_i - \bar{v})^2 \right).$$

Conditions (a) and (b) ensure that the second term inside the parenthesis on the right side goes to a nonnegative number as θ (and thus $\bar{v}$) go to infinity.[10] Hence, for wealth levels large enough, the left side is positive, and the result follows. In terms of the gaps mentioned before, as wealth grows large the riskiness of the optimal contract remains bounded while its mean increases; hence, at some point the gap between $\sum_{i=1}^{n} \pi_i(a) g(v_i)$ and $g(\bar{v} + \psi(a))$ becomes smaller relative to that between $g(\bar{v} + \psi(a))$ and $g(\bar{v})$.

Proposition 4 provides conditions for the principal to favor a relatively poorer agent when the pool of agents consists of wealthy ones. To be sure, its usefulness hinges on the plausibility of conditions (a)–(d). It is easy to show that (a) is satisfied by many utility functions in the HARA class that are commonly used in applications. But (b)–(d) are more delicate, as they refer to properties of the optimum. Since little is known about the optimal contract's *functional form* in the principal–agent model with moral hazard, it is unclear when they would hold from primitives. We show that *all* these conditions are met in the canonical case with two outcomes and either a finite or a continuum of actions if MLRP and CDFC hold that is commonly used in economic and financial applications (see Section 8 in Appendix B for the proof).

3.4. Remarks on the agent's outside option

Both TW and this paper assume that the agent's outside option consists of a constant wage $\bar{I}$ that the agent collects if she does not work for the principal.[11] A natural interpretation is that the outside option is *retirement*, since if she worked for another principal then she might face a contract that either entails some disutility of effort or variability in wages or both at the other job. Notice, however, that if at the alternative job the participation constraint has the retirement option as the reservation utility *and* it is binding, then the agent's reservation utility in our contracting problem is indeed the retirement utility, even though the agent need not retire but take the other job if she rejects the contract. This is the interpretation that is implicit in our analysis.

This argument relies crucially on a binding participation constraint at the other job, which need not be the case if, say, there is a lower bound on wages at the alternative job, or if the compensation scheme is not tailored to the agent's characteristics (e.g., as in a 'one-size-fits-all' commission contracts for salespeople at some large firms). Although a thorough analysis of this issue involving multiple principals is beyond the scope of this paper and left for future research, there are a couple of interesting cases for which we can provide clear-cut results. They illustrate the importance of how the agent's reservation utility is specified (the proofs are in Section 7 of Appendix B).

Consider first a scenario where at the alternative job the agent's action is contractible and there is a lower bound $m \geqslant 0$ on wages that is binding, so that the agent obtains $V(m + \theta) - \psi(\hat{a}) > V(\bar{I} + \theta)$ at her outside option, where $\hat{a}$ is the action implemented at the alternative job. We show that for any action $a < \hat{a}$, if $g(\cdot)$ is *concave* in v, then the principal prefers a *richer* agent. Thus, if the principal optimally implements an action smaller than $\hat{a}$, then his profits will be increasing in θ. This result does *not* obtain when the reservation utility is the utility of the retirement outside option.

Assume now a scenario where the alternative job compensation is risky and the agent receives $\sum_j \hat{\pi}_j(\hat{a}) V(\bar{I}_j + \theta) - \psi(\hat{a}) > V(\bar{I} + \theta)$, where $\bar{I}_j$ does not depend on θ for all j (e.g., as in the 'one-size-fits-all' example above) and the participation constraint is slack.[12] We show that for any $a \geqslant \hat{a}$, if $V(\cdot)$ exhibits decreasing absolute risk aversion and $g(\cdot)$ is convex in v, then the principal prefers a poorer agent. Thus, if the principal optimally implements an action bigger than $\hat{a}$, then his profits will be decreasing in θ. That is, TW's result extends but under more *stringent* conditions.

[10] In a nutshell, condition (a) guarantees that either $g(\cdot)$ is convex in v in a 'neighborhood of infinity' – although outside that neighborhood it can have an arbitrary curvature – in which case TW's result can be applied on that restricted set, or $g(\cdot)$ is concave in a neighborhood of infinity – with arbitrary curvature outside that set – and $g''(v)/g'(v)$ can be made arbitrarily small as v grows large. This is the class of problems we focus on and for which large wealth makes the principal to prefer a poorer agent without assuming that $g(\cdot)$ is globally convex. We can, however, construct examples outside this class (e.g., with $g(\cdot)$ concave but with the limit of $-g''(v)/g'(v)$ positive) where the result fails.

[11] We are grateful to a referee and the Associate Editor for their useful comments on this issue.

[12] It is important that wages in the alternative job are independent of the agent's initial wealth, for otherwise a change in θ will also affect the wages at the outside option, and we know little about how optimal compensation schemes change with θ.

In short, the specification of the agent's reservation utility is an important modeling choice (tailored to the economic application that one has in mind) that affects the principal's preferences for agents that differ in their wealth.

4. Concluding remarks

This note analyzes the principal's preferences over agents of differing wealth. We show that TW's condition for the principal to prefer a poorer agent is tight. We also prove that the principal prefers a poorer agent if the task entails a small enough disutility of effort. An implication of this result is that there is no analogous condition such that he would prefer a richer agent. Finally, we show that for an important class of problems, if agents are rich enough, the principal prefers a relatively poorer one.

We focused on the standard case with a risk neutral principal case. It would be interesting to generalize the results to the case where the principal is risk averse.[13]

Also, we analyzed the problem assuming that the agent's outside option is retirement, which is plausible if either the agent retires upon rejecting the contract, or the participation constraint binds at the alternative job where retirement is also an option. But we also showed that alternative assumptions can affect the results. This interesting issue deserves further consideration.

Appendix A

A.1. Proof of Proposition 2

(i) Suppose that $V(\cdot)$ is such that $g''(\bar{v})<0$ for some $\bar{v}$, and set $\bar{I}$ and θ so that $\bar{v}=V(\bar{I}+\theta)$. By continuity, $g''(v)<0$ for all v in some neighborhood of $\bar{v}$. We will show that there is a principal–agent problem with $V(\cdot)$ as the agent's utility function such that, for some $\widetilde{a}>0$, $\sum_{i=1}^{n}\pi_i(\widetilde{a})g(v_i)<g(\bar{v})$ (i.e., $C(\widetilde{a},\cdot)$ strictly decreases in θ).

To this end, assume a continuum of actions $A=[0,\bar{a}]$, two output levels q_1 and q_2, with $q_2>q_1$, and $\pi_2'(a)>0$, $\pi_2''(a)\leqslant 0$ for all a. For any action $a\in(0,\bar{a})$, constraint (1) is binding and (2) can be replaced by the first-order condition, i.e., $(1-\pi_2(a))v_1+\pi_2(a)v_2-\psi(a)=\bar{v}$ and $\pi_2'(a)(v_2-v_1)=\psi'(a)$. Thus, the optimal contract that implements a is given by

$$v_1=\bar{v}+\psi(a)-\pi_2(a)\frac{\psi'(a)}{\pi_2'(a)} \tag{8}$$

$$v_2=\bar{v}+\psi(a)+\big(1-\pi_2(a)\big)\frac{\psi'(a)}{\pi_2'(a)}. \tag{9}$$

To convey the main idea of the proof, we first show the result for simple disutility of effort function that does not meet all of our assumptions. We then modify the function to make it fit our assumptions and show that the proof goes with some suitable changes.

Take a small positive action a_o and let $M>0$ and

$$\psi(a)=\begin{cases}0 & \text{if } a<a_o,\\ M(a-a_o)^2 & \text{if } a\geqslant a_o.\end{cases}$$

This function is convex but it is neither strictly convex nor strictly increasing in a.[14]

Let $\widetilde{a}>a_0$ be any positive action. Set $\Delta\equiv\widetilde{a}-a_o>0$, and observe that $\psi(\widetilde{a})=M\Delta^2$ and $\psi'(\widetilde{a})=2M\Delta$. Hence, the optimal contract that implements $\widetilde{a}$ satisfies

$$v_1-\bar{v}=M\Delta(\Delta-2\gamma_1) \tag{10}$$

$$v_2-\bar{v}=M\Delta(\Delta+2\gamma_2), \tag{11}$$

where $\gamma_1\equiv\pi_2(\widetilde{a})/\pi_2'(\widetilde{a})$ and $\gamma_2\equiv(1-\pi_2(\widetilde{a}))/\pi_2'(\widetilde{a})$.

The first-order Taylor expansion of $g(\cdot)$ about $\bar{v}$ yields

$$g(v_i)-g(\bar{v})=g'(\bar{v})(v_i-\bar{v})+\frac{1}{2}g''(\xi_i)(v_i-\bar{v})^2$$

[13] Let the principal's utility function be $U(q-I)$, with $U'(\cdot)>0$ and $U''(\cdot)<0$. He prefers a poorer agent if $\sum_{i=1}^{n}\pi_i(a)U'(q_i-I_i)(\frac{V'(\bar{I}+\theta)}{V'(I_i+\theta)}-1)\geqslant 0$. Notice that TW's condition does not suffice anymore, for the marginal utility of the principal is decreasing, and it is not known in general what the correlation between q_i-I_i and I_i is at the optimum (see Grossman and Hart, 1983, Proposition 4).

[14] Also, the second derivative does not exist at a_0. This is easy to fix since everything goes through if one instead sets $(a-a_o)^k$ with $k>2$. To avoid clutter, we proceed with $k=2$ and ignore this issue.

for some ξ_i between v_i and $\bar{v}$, $i = 1, 2$. Since $\sum_{i=1}^{2} \pi_i(\widetilde{a})(v_i - \bar{v}) = \psi(\widetilde{a})$, after multiplying both sides by $\pi_i(\widetilde{a})$ and summing over i, we obtain

$$\sum_{i=1}^{2} \pi_i(\widetilde{a}) g(v_i) - g(\bar{v}) = g'(\bar{v})\psi(\widetilde{a}) + \frac{1}{2}\sum_{i=1}^{2} \pi_i(\widetilde{a}) g''(\xi_i)(v_i - \bar{v})^2. \tag{12}$$

Using the definition of $\psi(\cdot)$ and (10)–(11), the right side becomes[15]

$$M\,\Delta^2\left[g'(\bar{v}) + \frac{1}{2}M\Big(\big(1-\pi_2(\widetilde{a})\big)g''(\xi_1)(\Delta - 2\gamma_1)^2 + \pi_2(\widetilde{a})g''(\xi_2)(\Delta + 2\gamma_2)^2\Big)\right].$$

This expression is negative for suitable choices of M and Δ. To see this, notice that

$$\begin{aligned} & g'(\bar{v}) + \frac{1}{2}Mg''(\bar{v})\big(\big(1-\pi_2(\widetilde{a})\big)(\Delta - 2\gamma_1)^2 + \pi_2(\widetilde{a})(\Delta + 2\gamma_2)^2\big) \\ &\quad = g'(\bar{v}) + \frac{1}{2}Mg''(\bar{v})\left(\Delta^2 + 4\frac{(1-\pi_2(\widetilde{a}))\pi_2(\widetilde{a})}{\pi'(\widetilde{a})^2}\right) \\ &\quad < g'(\bar{v}) + Mg''(\bar{v})2\frac{(1-\pi_2(\widetilde{a}))\pi_2(\widetilde{a})}{\pi'(\widetilde{a})^2} \end{aligned}$$

can be made negative by taking M large enough (independently of Δ) because $g''(\bar{v}) < 0$. Hence, by choosing $\Delta > 0$ small enough (this depends on the value of M) so that v_1 and v_2 are close enough to $\bar{v}$, (12) will be negative and $\sum_{i=1}^{2} \pi_i(\widetilde{a}) g(v_i) - g(\bar{v}) < 0$. We have thus constructed a principal–agent problem with $C(\widetilde{a}, \cdot)$ strictly decreasing in θ.

As mentioned, $\psi(\cdot)$ is not strictly increasing or strictly convex. To fix this, consider $\varepsilon > 0$ and define the following strictly increasing modified cost function $\psi(\cdot)$:

$$\psi(a) = \begin{cases} \varepsilon a^2 & \text{if } a < a_o, \\ \varepsilon a^2 + M(a - a_o)^2 & \text{if } a \geqslant a_o. \end{cases}$$

Notice that this function $\psi(\cdot)$ is twice-continuously differentiable (except at a_0), strictly increasing, strictly convex in a, with $\psi(0) = \psi'(0) = 0$. Also,

$$\psi(\widetilde{a}) = \varepsilon\widetilde{a}^2 + M\Delta^2 \quad \text{and} \quad \psi'(\widetilde{a}) = 2\varepsilon\widetilde{a} + 2M\,\Delta.$$

Here M, $\widetilde{a}$, a_o and $\Delta = \widetilde{a} - a_o$ are as before. If we take in particular $\varepsilon = M\,\Delta^2$, then the optimal contract that implements $\widetilde{a}$ verifies

$$v_1 - \bar{v} = \psi(\widetilde{a}) - \gamma_1\psi'(\widetilde{a}) = M\Delta\beta_1 \tag{13}$$

$$v_2 - \bar{v} = \psi(\widetilde{a}) + \gamma_2\psi'(\widetilde{a}) = M\Delta\beta_2, \tag{14}$$

where $\beta_1 \equiv \Delta(\widetilde{a}^2 + 1 - 2\gamma_1\widetilde{a}) - 2\gamma_1$, and $\beta_2 \equiv \Delta(\widetilde{a}^2 + 1 + 2\gamma_2\widetilde{a}) + 2\gamma_2$. As before, from the first-order Taylor expansion of $g(\cdot)$ about $\bar{v}$ we obtain (12), and using (13)–(14), its right side becomes

$$M\Delta^2\left(g'(\bar{v})\big(\widetilde{a}^2 + 1\big) + \frac{1}{2}M\big(\pi_1(\widetilde{a})g''(\xi_1)\beta_1^2 + \pi_2(\widetilde{a})g''(\xi_2)\beta_2^2\big)\right). \tag{15}$$

Following a similar reasoning as above, we can show that (15) is negative for M large enough (independent of Δ) and Δ small enough, with Δ depending on the large value of M. Therefore, $\sum_{i=1}^{2} \pi_i(\widetilde{a}) g(v_i) - g(\bar{v}) < 0$ for these choices of M and Δ (i.e., $C(\widetilde{a}, \cdot)$ strictly decreases in θ), thereby completing the proof of part (i).

(ii) Besides the assumptions made in part (i), choose $\pi_2(\cdot)$ and A such that, along with the $\psi(\cdot)$ constructed above, the cost function is strictly convex in a (e.g., $\pi_2(\cdot)$ linear in a and $A = [0, 0.5]$). Since there are two output levels, $B(a) = q_1 + \pi_2(a)\Delta q$, where $\Delta q = q_2 - q_1 > 0$. Moreover, $\pi_2''(a) \leqslant 0$ implies that any action can be made optimal for the principal (i.e., solve $\max_{a \in A} B(a) - C(a, \theta)$) by a judicious choice of Δq. In particular, this applies to any fixed action $\widetilde{a}$. Thus, there is an open interval of wealth levels θ for which the principal prefers a richer agent.

[15] If the function g were twice differentiable (e.g. V is four times differentiable), then we could have used the second-order Taylor expansion to get (after some algebra) $\sum_{i=1}^{2} \pi_i(\widetilde{a}) g(v_i) - g(\bar{v}) = g'(\bar{v})M\Delta^2 + g''(\bar{v})B(\widetilde{a})M^2\Delta^2 + O(M^3\Delta^3)$ for $B(\widetilde{a}) = (1 - \pi_2(\widetilde{a}))\pi_2(\widetilde{a})/2(\pi_i(\widetilde{a}))^2$, which implies that the second term dominates the other two, by choosing a convenient large M and a small Δ (possibly dependent on M).

A.2. Proof of Proposition 3

(i) Since $\partial C(a,\theta,\eta)/\partial\theta|_{\eta=0}=0$, it suffices to show that $\partial^2 C(a,\theta,\eta)/\partial\theta\partial\eta|_{\eta=0}>0$, for then $\partial C(a,\theta,\eta)/\partial\theta$ would be positive for values of η in a (right) neighborhood of zero.

Differentiating (3) yields $\partial^2 C(a,\theta,\eta)/\partial\theta\partial\eta = g(\bar v)^{-1}\sum_{i=1}^n \pi_i(a) g'(v_i)(\partial v_i/\partial\eta)$. Hence,

$$\frac{\partial^2 C(a,\theta,\eta)}{\partial\theta\partial\eta}\bigg|_{\eta=0}=\frac{g'(\bar v)}{g(\bar v)}\sum_{i=1}^n \pi_i(a)\frac{\partial v_i}{\partial\eta}\bigg|_{\eta=0}=\frac{g'(\bar v)}{g(\bar v)}\left(\frac{\partial(\sum_{i=1}^n \pi_i(a)v_i)}{\partial\eta}\bigg|_{\eta=0}\right)=\frac{g'(\bar v)}{g(\bar v)}\psi(a)>0,$$

where the third equality is due to the participation constraint $\sum_{i=1}^n \pi_i(a)v_i=\bar v+\eta\psi(a)$. By continuity, there is an $\eta_a>0$ such that $C(a,\cdot,\eta)$ strictly increases in θ if $\eta\in(0,\eta_a)$.

(ii) If A is a finite set, then part (i) implies that the cost of implementing any action $a>a_1$ is strictly increasing in θ if $\eta\in(0,\eta^*)$, where $0<\eta^*=\min_{a\in A-\{a_1\}}\eta_a$ (the cost of implementing a_1 is simply $\bar I$, which is trivially increasing in θ). Thus, $\max_{a\in A} B(a)-C(a,\theta,\eta)$ is decreasing in θ for all $\eta\in(0,\eta^*)$.

Suppose $A=[0,\bar a]$. Note that the optimal action when $\eta=0$ is $\bar a$, for it solves $\max_{a\in[0,\bar a]} B(a)-\bar I$, and MLRP and $\pi_i'(a)\neq 0$ for some i for each a imply that $B(\cdot)$ is strictly increasing in a. In particular

$$B(\bar a)-C(\bar a,\theta,0)=B(\bar a)-\bar I>B(0)-\bar I=B(0)-C(0,\theta,0).$$

Since $C(a,\theta,\eta)$ is continuous and equal to $\bar I$ when $\eta=0$, there exists an action $a_\ell>0$ and $\tilde\eta>0$ such that $B(\bar a)-C(\bar a,\theta,\eta)>B(a)-C(a,\theta,\eta)$ for all $(a,\eta)\in[0,a_\ell]\times[0,\tilde\eta]$. That is, $\bar a$ yields higher expected profits for the principal than any action $a\in[0,a_\ell]$. Hence, if $\eta\in[0,\tilde\eta]$, the optimal action for the principal must be in $[a_\ell,\bar a]$. Let $\gamma=(g'(\bar v)/g(\bar v))\psi(a_\ell)>0$. For any $a\in[a_\ell,\bar a]$, we have that $\partial^2 C(a,\theta,\eta)/\partial\theta\partial\eta\,|_{\eta=0}\geqslant\gamma>0$. By the continuity of this second derivative, there is an open neighborhood W_a of a and a non-empty neighborhood $U_a=[0,\eta_a]$ of values of η such that $\partial^2 C(a',\theta,\eta)/\partial\theta\partial\eta\geqslant\gamma/2>0$, for all $a'\in W_a$ and $\eta\in U_a$. Since $\partial C(a,\theta,\eta)/\partial\theta=0$ at $\eta=0$, it follows by the Mean Value Theorem that $\partial C(a',\theta,\eta)/\partial\theta>0$ for all $a'\in W_a$ and $\eta\in U_a$. Now, the collection $\mathcal{W}=\{W_a\colon a\in[a_\ell,\bar a]\}$ is an open covering of $[a_\ell,\bar a]$; i.e., $[a_\ell,\bar a]\subseteq\bigcup_{a\in[a_\ell,\bar a]}W_a$. The compactness of $[a_\ell,\bar a]$ implies the existence of a finite subcollection $\{W_{a_1},W_{a_2},\dots,W_{a_m}\}$ such that $[a_\ell,\bar a]\subset\bigcup_{i=1}^m W_{a_i}$. Associated with it there is a finite subcollection $\{U_{a_1},U_{a_2},\dots,U_{a_m}\}$. Setting $\eta^*=\min\{\tilde\eta,\eta_{a_1},\eta_{a_2},\dots,\eta_{a_m}\}>0$ yields $\partial C(a,\theta,\eta)/\partial\theta>0$ for all $\eta\in(0,\eta^*)$ and all $a\in[a_\ell,\bar a]$, thus completing the proof.

A.3. Proof of Proposition 4

(i) Let a be any action above the lowest one, and let $v_1,v_2,\dots,v_n$ be an optimal contract that implements a, such that $|v_i-\bar v|\leqslant K_a$, with $K_a>0$ independent of $\bar v$. We must show that if θ is large enough, then $\sum_i \pi_i(a)g(v_i)>g(\bar v)$.

Consider a Taylor expansion of $g(\cdot)$ around $v=\bar v$. Then

$$g(v_i)=g(\bar v)+g'(\bar v)(v_i-\bar v)+\frac{1}{2}g''(\xi_i)(v_i-\bar v)^2,$$

for some ξ_i between v_i and $\bar v$, $i=1,2,\dots,n$. Multiply by $\pi_i(a)$ and sum over i to obtain

$$\sum_{i=1}^n \pi_i(a)g(v_i)=g(\bar v)+g'(\bar v)\psi(a)+\frac{1}{2}\sum_{i=1}^n \pi_i(a)g''(\xi_i)(v_i-\bar v)^2,$$

and thus

$$\sum_{i=1}^n \pi_i(a)g(v_i)-g(\bar v)=g'(\bar v)\left(\psi(a)+\sum_{i=1}^n \kappa_i\frac{g''(\xi_i)}{g'(\bar v)}\right), \tag{16}$$

where $\kappa_i=\frac{1}{2}\pi_i(a)(v_i-\bar v)^2$ and therefore $0\leqslant\kappa_i\leqslant K_a^2$.

Notice that $|v_i-\bar v|\leqslant K_a$ and K_a independent of $\bar v$ imply that $v_i\to\infty$ as $\bar v\to\infty$, and so does ξ_i. If $g(\cdot)$ is convex in $v>\tilde v$, then $g''(v)\geqslant 0$ for all v large enough. Hence, $g''(\xi_i)/g'(\bar v)\geqslant 0$, and the right side of (16) is positive for $\bar v$ above a threshold $\bar v_a$. Since $\bar v=V(\bar I+\theta)$ and $V(\cdot)$ is unbounded, the result follows by taking $\theta_a=h(\bar v_a)-\bar I$.

Assume instead that $g''(\cdot)$ is concave in a neighborhood of infinity and $\lim_{v\to\infty}-g''(v)/g'(v)=0$. We now show that $\lim_{\bar v\to\infty}g''(\xi_i)/g'(\bar v)=0$. Let $0<\varepsilon<(2K_a)^{-1}$. Since $\lim_{v\to\infty}g''(v)/g'(v)=0$, there exists an $M>0$ such that, if $v>M$, then

$$\left|\frac{g''(v)}{g'(v)}\right|<\varepsilon/2. \tag{17}$$

Suppose that $\bar v>M+K_a$. If $\bar v\leqslant\xi_i\leqslant v_i$, then $g'(\xi_i)/g'(\bar v)\leqslant 1$, as $g'(\cdot)$ is positive and decreasing. If $v_i\leqslant\xi_i\leqslant\bar v$, apply a linear Taylor expansion to $g'(\cdot)$ around $v=\bar v$ to obtain $g'(\bar v-K_a)=g'(\bar v)+g''(\delta)K_a$, for some δ, $\bar v-K_a<\delta<\bar v$. Using $|g''(v)/g'(v)|<\varepsilon/2$ and $0<\varepsilon<(2K_a)^{-1}$, we obtain

$$\left|g'(\bar v-K_a)-g'(\bar v)\right|=\left|g''(\delta)\right|K_a<\varepsilon g'(\delta)K_a<\frac{1}{2}g'(\delta)<\frac{1}{2}g'(\bar v-K_a),$$

which gives $|1-(g'(\bar{v})/g'(\bar{v}-K_a))|<1/2$ and thus $2/3<g'(\bar{v}-K_a)/g'(\bar{v})<2$. In particular, if $\bar{v}-K_a\leqslant v_i\leqslant \xi_i\leqslant \bar{v}$, then $0<g'(\xi_i)/g'(\bar{v})\leqslant g'(\bar{v}-K_a)/g'(\bar{v})<2$. Therefore, for $\bar{v}>M+K_a$,

$$\left|\frac{g''(\xi_i)}{g'(\bar{v})}\right|=\left|\frac{g''(\xi_i)}{g'(\xi_i)}\right|\left|\frac{g'(\xi_i)}{g'(\bar{v})}\right|\leqslant 2\left|\frac{g''(\xi_i)}{g'(\xi_i)}\right|<\varepsilon,$$

where the last inequality follows from (17). Thus, $\lim_{\bar{v}\to\infty} g''(\xi_i)/g'(\bar{v})=0$, $i=1,2,\ldots,n$, and hence the right side of (16) is positive for $\bar{v}$ above a threshold $\bar{v}_a$. Since $\bar{v}=V(\bar{I}+\theta)$ and $V(\cdot)$ is unbounded, the result follows by taking $\theta_a=h(\bar{v}_a)-\bar{I}$.

Let $A=[0,\bar{a}]$ (the A finite case is immediate), assume that (c) holds, and consider any $\tilde{a}\in(0,\bar{a}]$. Take $0<K<\infty$ such that $\sup_{a\in A}K_a\leqslant K$ and $\bar{v}^*$ large enough so that

$$\sum_{i=1}^{n}\left|\frac{g''(\xi_i)}{g'(\bar{v})}\right|<\frac{\psi(\tilde{a})}{K^2}$$

for all $\bar{v}>\bar{v}^*$ (by using K instead of K_a in the proof of $\lim_{\bar{v}\to\infty} g''(\xi_i)/g'(\bar{v})=0$, $\bar{v}^*$ can be chosen such that it only depends on $\tilde{a}$ and K). From (16) and $\psi(a)>\psi(\tilde{a})$ for all $a>\tilde{a}$, it follows that for any $a\in[\tilde{a},\bar{a}]$ and $\bar{v}>\bar{v}^*$,

$$\sum_{i=1}^{n}\pi_i(a)g(v_i)-g(\bar{v})\geqslant g'(\bar{v})\left(\psi(\tilde{a})-K^2\sum_{i=1}^{n}\left|\frac{g''(\xi_i)}{g'(\bar{v})}\right|\right)>0.$$

Finally, letting $\theta^*=h(\bar{v}^*)-\bar{I}$ completes the proof of the result.

(ii) If A is finite, then the result is clear as the cost of implementing any action is increasing in θ if $\theta\in(\theta^*,\infty)$, where $\theta^*=\max_{a\in A-\{a_1\}}\theta_a$. This implies that $\max_{a\in A}B(a)-C(a,\theta)$ is decreasing in θ for all $\theta\in(\theta^*,\infty)$.

If $A=[0,\bar{a}]$, the result follows from part (i), since $\arg\max_{a\in A}B(a)-C(a,\theta)\geqslant\tilde{a}>0$ for all θ yields $\max_{a\geqslant\tilde{a}}B(a)-C(a,\theta)$ decreasing in θ for all $\theta\in(\theta^*,\infty)$.

Appendix B. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.geb.2014.02.012.

References

Amir R., Czupryna, M., 2004. On inverse utility and third-order effects in the economics of uncertainty. CORE Working Paper.
Chiappori, P.-A., Macho, I., Rey, P., Salanié, B., 1994. Repeated moral hazard: the role of memory, commitment, and the access to credit markets. Europ. Econ. Rev. 38, 1527–1553.
Chiu, H., 2010. Non-negativity of agent's payoffs and wealth effects in the principal–agent model. University of Manchester Working Paper.
Grossman, S., Hart, O., 1983. An analysis of the principal–agent problem. Econometrica 51, 7–45.
Holmstrom, B., 1979. Moral hazard and observability. Bell J. Econ. 10, 74–91.
Holmstrom, B., Milgrom, P., 1987. Aggregation and linearity in the provision of intertemporal incentives. Econometrica 55, 303–328.
Jewitt, I., Kadan, O., Swinkels, J., 2008. Moral hazard with bounded payments. J. Econ. Theory 143, 59–82.
Kadan, O., Swinkels, J., 2013. On the moral Hazard problem without the first order approach. J. Econ. Theory 148, 2313–2343.
Milgrom, P., Segal, I., 2002. Envelope theorems for arbitrary choice sets. Econometrica 70, 583–601.
Park, I.-U., 2004. Moral hazard contracting and private credit markets. Econometrica 72, 701–746.
Rogerson, W., 1985. The first-order approach to principal–agents problems. Econometrica 53, 1357–1367.
Spear, S., Wang, C., 2005. When to fire a CEO: optimal termination in dynamic contracts. J. Econ. Theory 120, 239–256.
Thiele, H., Wambach, A., 1999. Wealth effects in the principal agent model. J. Econ. Theory 89, 247–260.