Wealth Effects and Agency Costs *

Hector Chade† Virginia N. Vera de Serio‡

First version: June 24, 2010
This version: February 11, 2014

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.

Keywords: Moral Hazard, Principal-Agent Model, Contracts, Wealth Effects.

*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.
†Arizona State University, Department of Economics, Tempe, AZ 85287-3806.
‡Universidad Nacional de Cuyo, Facultad de Ciencias Económicas, I.C.B., Mendoza, Argentina 5500.
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.

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 princi-

\(^1\)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, Kadan, and Swinkels (2008).
pal 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 affects 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, Macho, Rey, and Salanié (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 (forthcoming) 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 the Appendix, and the rest in the Online Appendix.

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 $\theta$.

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 $V : (I, \infty) \to \mathbb{R}, \ I_{\ell} \geq -\infty$, 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 $V(\bar{I} + \theta)$.

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 $\theta$. 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:

$$
C(a, \theta) = \min_{I_1, \ldots, I_n} \sum_{i=1}^{n} \pi_i(a) I_i \\
\text{s.t.} \quad \sum_{i=1}^{n} \pi_i(a) V(I_i + \theta) - \psi(a) \geq V(\bar{I} + \theta) \tag{1}
$$

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

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 $\sum_{i=1}^{n} \pi_i(a) V(I_i + \theta) - \psi(a) \geq \sum_{i=1}^{n} \pi_i(a') V(I_i + \theta) - \psi(a')$, 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)).

An equivalent formulation of the problem with the contract written in utility units is as follows: $\min_{v_1, \ldots, v_n} \sum_{i=1}^{n} \pi_i(a) h(v_i) - \theta$ subject to $\sum_{i=1}^{n} \pi_i(a) v_i - \psi(a) \geq V(\bar{I} + \theta)$ 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

---

3 All we need is the validity of the first-order approach. We use MLRP and CDFC for simplicity.
A = [0, π] is that, for each a, π′_i(a) ≠ 0 for some i, and when A is finite, that for each k, π_i(a_k) ≠ π_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, θ), 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 the Online Appendix). 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. Under MLRP and CDFC, a simple adaptation of Lemma 2 in Jewitt, Kadan, and Swinkels (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 π(·) and ψ(·) are three times continuously differentiable.

### 2.4 TW’s Result and Intuition

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

\[
\frac{\partial C(a, \theta)}{\partial \theta} = \left( \sum_{i=1}^n \pi_i(a) \frac{1}{V'(I_i + \theta)} \right) V'(\bar{I} + \theta) - 1
\]

\[
= V'(h(\bar{v})) \left( \sum_{i=1}^n \pi_i(a) \frac{1}{V'(h(v_i))} - \frac{1}{V'(h(\bar{v}))} \right)
\]

\[
= g(\bar{v})^{-1} \left( \sum_{i=1}^n \pi_i(a)g(v_i) - g(\bar{v}) \right), \tag{3}
\]

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 \( g(\bar{v}) \).

---

4One can apply that corollary after transforming the variables from \( v_i \) to \( z_i = v_i - \bar{v} \), so that \( \theta \) 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.
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 \( P(I + \theta) \leq 3R(I + \theta) \) 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 \( \theta \). As a result, the principal’s expected profit is a decreasing function of the agent’s wealth \( \theta \).

**Proof.** Simple algebra shows that \( P(I + \theta) \leq 3R(I + \theta) \) 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

\[
\sum_{i=1}^{n} \pi_i(a)g(v_i) \geq g \left( \sum_{i=1}^{n} \pi_i(a)v_i \right) = g(\bar{v} + \psi(a)) > g(\bar{v}).
\]  

(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 \( \theta \) 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 \( \sum_{i=1}^{n} \pi_i(a)g(v_i)dv \) for the principal. When \( g(\cdot) \) is convex in \( v \), then \( \sum_{i=1}^{n} \pi_i(a)g(v_i)dv > g(\bar{v})dv \). 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 \( a \).

---

5Along the same lines, in the utility formulation of the problem it is easy to see that an increase in \( \theta \) 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 \( \sum_{i=1}^{n} \pi_i(a)g(v_i) \) (the value of the Lagrange multiplier of the participation constraint). Adding both effects yields \((-1 + g(\bar{v})^{-1}(\sum_{i=1}^{n} \pi_i(a)g(v_i)))d\theta\), which is (3).
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

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

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 $\psi(a)$. The second is the gap between

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

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 $\theta$ increases, and this is a priori ambiguous. The variance of the contract, however, increases in $\theta$ (see the Online Appendix) due to the incentive constraints. This does not mean that the principal’s cost increases in $\theta$, 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 $\theta$: 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 $\varepsilon$, with mean zero and variance $\sigma_{\varepsilon}^2$. Using the standard approximation for risk aversion in the small, her compensation 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

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

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.
The Envelope Theorem argument above reveals that the first-order effect on $\sigma_v^2$ 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.\textsuperscript{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 $\theta$ 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 the Appendix, but we sketch the main idea of part (i) here.\textsuperscript{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 $(v_1, ..., v_n)$ 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

\textsuperscript{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.

\textsuperscript{8}We are grateful to an anonymous referee for providing us with this more intuitive proof.

8
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 \( \theta \).\footnote{Obviously, to make all this precise the proof in the Appendix shows that the right side is negative
when instead of an approximation we use an exact second-order Taylor expansion.} 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?

---

9Obviously, to make all this precise the proof in the Appendix 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 \(\eta \geq 0\). Also, assume that the solution of the cost minimization problem is continuously differentiable in \(\eta\), 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 the Appendix. 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 \(\eta\) approaches zero, but the second term vanishes at a faster rate since it depends on \(\eta^2\). Thus, for \(\eta\) small enough, the first term in 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 \(\eta\) 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 \(\theta\) for all \(a\) when \(\eta\) is sufficiently small, and thus \(\max_a B(a) - C(a, \theta)\) decreases in \(\theta\), 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 $\eta$. 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 $\eta$ and $a$, which allowed us to find a uniform bound
on $\eta$ and it was sufficient for the implication mentioned above. Two comments are
in order. First, instead of finding a uniform bound on $\eta$, 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 the Online Appendix). 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 $(a, \eta)$, strictly increasing
in $\eta$ 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 $\eta$ in the previous
section, now $\theta$ 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 $\bar{v}$ such that either $g(\cdot)$ is convex in $v$ when $v \in (\bar{v}, \infty)$, or
$g(\cdot)$ is concave in $v$ when $v \in (\bar{v}, \infty)$ and $\lim_{v \to \infty} -g''(v)/g'(v) = 0$. 

11
For any \( a \in A \) there is an optimal \((v_1, v_2, \ldots, v_n)\) with \(|v_i - \bar{v}| \leq K_a\) for all \( i \), where \( K_a > 0 \) is independent of \( \bar{v} \).

\( \sup_{a \in A} K_a < \infty \).

The principal’s optimal action is bounded away from the lowest action for all \( \theta \).

**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 \( a \geq \tilde{a} \) 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 the Appendix), 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 \( \theta \) (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

\(^{10}\)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.
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 the Online Appendix 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 the Online Appendix).

Consider first a scenario where at the alternative job the agent’s action is contractible
and there is a lower bound $m \geq 0$ 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 $\theta$. This result does not obtain when

\(^{11}\)We are grateful to a referee and the Associate Editor for their useful comments on this issue.
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 \( \sum_{j} \hat{\pi}_{j}(\hat{a})V(\hat{T}_{j} + \theta) - \psi(\hat{a}) > V(\bar{I} + \theta) \), where \( \hat{T}_{j} \) does not depend on \( \theta \) 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 \( a \geq \hat{a} \), 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 \( \theta \). That is, TW’s result extends but under more \textit{stringent} conditions.

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.

\(^{12}\)It is important that wages in the alternative job are independent of the agent’s initial wealth, for otherwise a change in \( \theta \) will also affect the wages at the outside option, and we know little about how optimal compensation schemes change with \( \theta \).

\(^{13}\)Let the principal’s utility function be \( U(q - I) \), with \( U'(\cdot) > 0 \) and \( U''(\cdot) < 0 \). He prefers a poorer agent if \( \sum_{i=1}^{n} \pi_{i}(a)U'(q_{i} - I_{i}) \left( \frac{V'(I_{i} + \theta)}{V'(I_{i} + \theta)} - 1 \right) \geq 0 \). 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).
References


A Appendix

A.1 Proof of Proposition 2

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

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

\[
  v_1 = v + \psi(a) - \pi_2(a) \frac{\psi'(a)}{\pi_2'(a)} \\
  v_2 = v + \psi(a) + (1 - \pi_2(a)) \frac{\psi'(a)}{\pi_2'(a)}. 
\]

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

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

This function is convex but it is neither strictly convex nor strictly increasing in \( a \).\(^{14}\)

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

\[
  v_1 - v = M \Delta (\Delta - 2\gamma_1) \\
  v_2 - v = M \Delta (\Delta + 2\gamma_2),
\]

\(^{14}\)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.
where $\gamma_1 \equiv \frac{\pi_2(\bar{a})}{\pi'_2(\bar{a})}$ and $\gamma_2 \equiv (1 - \pi_2(\bar{a}))/\pi'_2(\bar{a})$.

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

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

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

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

Using the definition of $\psi(\cdot)$ and (10)–(11), the right-side becomes $^{15}$

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

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

$$g'(\bar{v}) + \frac{1}{2}Mg''(\bar{v})\left((1 - \pi_2(\bar{a}))(\Delta - 2\gamma_1)^2 + \pi_2(\bar{a})(\Delta + 2\gamma_2)^2\right)$$

$$= g'(\bar{v}) + \frac{1}{2}Mg''(\bar{v})\left(\Delta^2 + 4\frac{(1 - \pi_2(\bar{a}))\pi_2(\bar{a})}{\pi'(\bar{a})^2}\right)$$

$$< g'(\bar{v}) + Mg''(\bar{v})\frac{(1 - \pi_2(\bar{a}))\pi_2(\bar{a})}{\pi'(\bar{a})^2},$$

can be made negative by taking $M$ large enough (independently of $\Delta$) 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 $\sum_{i=1}^{2} \pi_i(\bar{a})g(v_i) - g(\bar{v}) < 0$. We have thus constructed a principal-agent problem with $C(\bar{a}, \cdot)$ strictly decreasing in $\theta$.

As mentioned, $\psi(\cdot)$ is not strictly increasing or strictly convex. To fix this, consider

$^{15}$If the function $g(\cdot)$ 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) $\sum_{i=1}^{2} \pi_i(\bar{a})g(v_i) - g(\bar{v}) = g'(\bar{v})M\Delta^2 + g''(\bar{v})B(\bar{a})M^2\Delta^2 + O(M^3\Delta^3)$ for $B(\bar{a}) = (1 - \pi_2(\bar{a}))\pi_2(\bar{a})/2(\pi_1(\bar{a}))^2$, which implies that the second term dominates the others two, by choosing a convenient large $M$ and a small $\Delta$ (possibly dependent on $M$).
\( \varepsilon > 0 \) and define the following strictly increasing modified cost function \( \psi(\cdot) \):

\[
\psi(a) = \begin{cases} 
\varepsilon a^2 & \text{if } a < a_0, \\
\varepsilon a^2 + M(a - a_0)^2 & \text{if } a \geq a_0.
\end{cases}
\]

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,

\[
\psi(\tilde{a}) = \varepsilon \tilde{a}^2 + M \Delta^2 \quad \text{and} \quad \psi'(\tilde{a}) = 2\varepsilon \tilde{a} + 2M \Delta.
\]

Here \( M, \tilde{a}, a_0 \) and \( \Delta = \tilde{a} - a_0 \) are as before. If we take in particular \( \varepsilon = M \Delta^2 \), then the optimal contract that implements \( \tilde{a} \) verifies

\[
v_1 - \bar{v} = \psi(\tilde{a}) - \gamma_1 \psi'(\tilde{a}) = M \Delta \beta_1 \tag{13}
v_2 - \bar{v} = \psi(\tilde{a}) + \gamma_2 \psi'(\tilde{a}) = M \Delta \beta_2, \tag{14}
\]

where \( \beta_1 = \Delta (\tilde{a}^2 + 1 - 2\gamma_1 \tilde{a}) - 2\gamma_1, \) and \( \beta_2 = \Delta (\tilde{a}^2 + 1 + 2\gamma_2 \tilde{a}) + 2\gamma_2. \) 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

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

Following a similar reasoning as above, we can show that (15) is negative for \( M \) large enough (independent of \( \Delta \)) and \( \Delta \) small enough, with \( \Delta \) depending on the large value of \( M \). Therefore, \( \sum_{i=1}^2 \pi_i (\tilde{a}) g(v_i) - g(\bar{v}) < 0 \) for these choices of \( M \) and \( \Delta \) (i.e., \( C(\tilde{a}, \cdot) \) strictly decreases in \( \theta \)), 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, \( \pi''_2(a) \leq 0 \) 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 \( \Delta q. \) In particular, this applies to any fixed action \( \tilde{a}. \) Thus, there is an open interval of wealth levels \( \theta \) for which the principal prefers a richer agent. \( \square \)
A.2 Proof of Proposition 3

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

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

\[
\frac{\partial^2 C(a, \theta, \eta)}{\partial \theta \partial \eta} \rvert_{\eta=0} = \frac{g'(\overline{v})}{g(\overline{v})} \sum_{i=1}^{n} \pi_i(a) \frac{\partial v_i}{\partial \eta} \rvert_{\eta=0} = \frac{g'(\overline{v})}{g(\overline{v})} \left( \frac{\partial (\sum_{i=1}^{n} \pi_i(a)v_i)}{\partial \eta} \rvert_{\eta=0} \right) \geq \psi(a) > 0,
\]

where the third equality is due to the participation constraint \( \sum_{i=1}^{n} \pi_i(a)v_i = \overline{v} + \eta \psi(a) \).

By continuity, there is an \( \eta_a > 0 \) such that \( C(a, \cdot, \eta_a) \) strictly increases in \( \theta \) 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 \( \theta \) if \( \eta \in (0, \eta^*) \), where \( 0 < \eta^* = \min_{a \in A \setminus \{a_1\}} \eta_a \) (the cost of implementing \( a_1 \) is simply \( \bar{I} \), which is trivially increasing in \( \theta \)). Thus, \( \max_{a \in A} B(a) - C(a, \theta, \eta) \) is decreasing in \( \theta \) 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 \( \bar{\eta} > 0 \) such that \( B(\bar{a}) - C(\bar{a}, \theta, \eta) > B(a) - C(a, \theta, \eta) \) for all \( (a, \eta) \in [0, a_\ell] \times [0, \bar{\eta}] \).

That is, \( \bar{a} \) yields higher expected profits for the principal than any action \( a \in [0, a_\ell] \). Hence, if \( \eta \in [0, \bar{\eta}] \), 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 \( \partial^2 C(a, \theta, \eta)/\partial \theta \partial \eta \rvert_{\eta=0} \geq \gamma > 0 \).

By the continuity of this second derivative, there is an open neighborhood \( U_a = [0, \eta_a] \) of values of \( \eta \) such that \( \partial^2 C(a', \theta, \eta)/\partial \theta \partial \eta \geq \gamma/2 > 0 \), for all \( a' \in U_a \) 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 \( a' \in U_a \) and \( \eta \in U_a \).

Now, the collection \( W = \{W_a : a \in [a_\ell, \bar{a}]\} \) is an open covering of \( [a_\ell, \bar{a}] \); i.e., \( [a_\ell, \bar{a}] \subseteq \cup_{a \in [a_\ell, \bar{a}]} W_a \). The compactness of \([a_\ell, \bar{a}]\) implies the existence of a finite subcollection \( \{W_{a_1}, W_{a_2}, \ldots, W_{a_m}\} \) such that \([a_\ell, \bar{a}] \subseteq \cup_{i=1}^m W_{a_i} \). Associated with it there is a finite subcollection \( \{U_{a_1}, U_{a_2}, \ldots, U_{a_m}\} \). Setting \( \eta^* = \min \{\bar{\eta}, \eta_{a_1}, \eta_{a_2}, \ldots, \eta_{a_m}\} > 0 \) 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. \( \square \)
A.3 Proof of Proposition 4

(i) Let $a$ be any action above the lowest one, and let $v_1, v_2, \ldots, v_n$ be an optimal contract that implements $a$, such that $|v_i - \bar{v}| \leq K_a$, with $K_a > 0$ independent of $\bar{v}$. We must show that if $\theta$ 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 $\xi_i$ between $v_i$ and $\bar{v}$, $i = 1, 2, \ldots, n$. 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

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

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

Notice that $|v_i - \bar{v}| \leq K_a$ and $K_a$ independent of $\bar{v}$ imply that $v_i \to \infty$ as $\bar{v} \to \infty$, and so does $\xi_i$. If $g(\cdot)$ is convex in $v > \bar{v}$, then $g''(v) \geq 0$ for all $v$ large enough. Hence, $g''(\xi)/g'(\bar{v}) \geq 0$, 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 $\lim_{v \to \infty} g''(\xi) / g'(\bar{v}) = 0$. 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

$$\left|\frac{g''(v)}{g'(v)}\right| < \varepsilon/2.$$  \hspace{1cm} (17)

Suppose that $\bar{v} > M + K_a$. If $\bar{v} \leq \xi_i \leq v_i$, then $g'(\xi_i) / g'(\bar{v}) \leq 1$, as $g'(\cdot)$ is positive and decreasing. If $v_i \leq \xi_i \leq \bar{v}$, 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 $\delta$, $\bar{v} - K_a < \delta < \bar{v}$. Using $|g''(v)/g'(v)| < \varepsilon/2$ and $0 < \varepsilon < (2K_a)^{-1}$, we obtain

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

\[
\left| \frac{g''(\xi_i)}{g'(\dot{v})} \right| = \left| \frac{g''(\xi_i)}{g'(\dot{v})} \right| \frac{g'(\dot{v})}{g'(\dot{v})} \leq 2 \left| \frac{g''(\xi_i)}{g'(\dot{v})} \right| < \varepsilon,
\]

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

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

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

for all \(\dot{v} > \bar{v}^*\) (by using \(K\) instead of \(K_a\) in the proof of \(\lim_{\dot{v} \to \infty} \frac{g''(\xi_i)}{g'(\dot{v})} = 0\), \(\bar{v}^*\) 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 \(a \in [\tilde{a}, \tilde{a}]\) and \(\tilde{v} > \bar{v}^*\),

\[
\sum_{i=1}^{n} \pi_i(a) g(v_i) - g(\tilde{v}) \geq g'(\tilde{v}) \left( \psi(\tilde{a}) - K^2 \sum_{i=1}^{n} \left| \frac{g''(\xi_i)}{g'(\dot{v})} \right| \right) > 0
\]

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 \(\theta\) if \(\theta \in (\theta^*, \infty)\), where \(\theta^* = \max_{a \in A \setminus \{a_1\}} \theta_a\). This implies that \(\max_{a \in A} B(a) - C(a, \theta)\) is decreasing in \(\theta\) for all \(\theta \in (\theta^*, \infty)\).

If \(A = [0, \tilde{a}]\), the result follows from part (i), since \(\arg\max_{a \in A} B(a) - C(a, \theta) \geq \tilde{a} > 0\) for all \(\theta\) yields \(\max_{a \geq \tilde{a}} B(a) - C(a, \theta)\) decreasing in \(\theta\) for all \(\theta \in (\theta^*, \infty)\). \(\square\)
Wealth Effects and Agency Costs: Supplementary Material
(Not Intended for Publication)

Hector Chade∗    Virginia N. Vera de Serio†

January 24, 2014

∗Arizona State University, Department of Economics, Tempe, AZ 85287-3806.
†Universidad Nacional de Cuyo, Facultad de Ciencias Económicas, I.C.B., Mendoza, Argentina 5500.
1 Introduction

In these notes we provide detailed proofs of results omitted from the text. We start with a comment on the constraint qualification in the cost minimization problem, and then move on to a long section on differentiability of the cost function and the optimal contract. Then we construct an \( n \)-outcome example of a principal-agent problem that illustrates the tightness of TW’s condition. Next we justify an assertion about the variance of the contract made in the text, as well as one about the small disutility of effort case. Finally, we prove two claims about alternative assumptions on the agent’s outside option, and provide proofs of a couple of assertions made about large wealth and a continuum of actions. The Appendix contains a proof of the continuity properties of the cost function and the optimal contract—which requires only first-order continuous differentiability of the primitives—that is used to apply an envelope theorem, and that can be of independent interest.

2 A Remark on the Constraint Qualification

Notice that the cost minimization problem amounts to minimizing a strictly convex function subject to linear constraints. Thus, the ‘refined Slater condition’ (see Boyd and Vandenbergue (2006) pp. 226-7) reduces to feasibility plus open domain of the objective function in this case (which holds in this problem). Therefore, when the feasible set is nonempty, then the unique optimum is characterized by the Kuhn-Tucker conditions. In short, either the constraint set is empty and there is nothing to solve (the action is not implementable or, equivalently, its cost is infinite), or it is nonempty and no additional regularity condition is needed.

In the continuum of actions case there is a simple sufficient condition to ensure that the feasible set is nonempty: for each \( a \) there is an outcome \( q_i \) such that \( \pi'_i(a) \neq 0 \) (this is weaker than assuming the strict MLRP). Similarly, in the finite case with the MLRP and CDFC conditions — in which case only the local downward incentive constraint binds (see Grossman and Hart (1983) p. 34) — it suffices to assume that for each \( a_k \), there is an outcome \( q_i \) such that \( \pi_i(a_k) \neq \pi_i(a_{k-1}) \). To avoid dealing with the uninteresting case in which the constraint set is empty, we make these mild assumptions in these cases.
3 Differentiability Properties of $C$ and $v_i$

3.1 Continuous Differentiability of $C$ with respect to $\theta$

Both in Thiele and Wambach (1999) and in our paper, the derivative of the cost function with respect to $\theta$ plays a fundamental role. We now justify it from primitives requiring only first-order continuous differentiability of $V(\cdot)$, $\pi(\cdot)$, and $\psi(\cdot)$.

Consider the following transformation of variables: $z_i = v_i - V(I + \theta)$, $i = 1, ..., n$. The problem then becomes:

$$C(a, \theta) = \min_{z_1, ..., z_n} \sum_i \pi_i(a) h(z_i + V(I + \theta)) - \theta$$

subject to $\sum_i \pi_i(a) z_i - \psi(a) \geq 0$, and $a \in \arg\max_{a'} \sum_i \pi_i(a') z_i - \psi(a')$ (this is a finite number of inequalities when $A$ is finite, or a first order condition when $A = [0, \bar{a}]$). Notice that the parameter $\theta$ appears now only in the objective function.

Grossman and Hart (1983) plus strict convexity of $h$ imply that, for each $\theta$, there is a unique solution, $Z^*(\theta) = \{z^*(\theta)\}$, to the above problem. We show in the Appendix below that $z^*(\cdot)$ is continuous in $\theta$.

To see that the cost function is continuously differentiable in $\theta$, restrict the domain of $\theta$ to some compact set $[0, \bar{\theta}]$, $\bar{\theta} < \infty$, and consider the set $Z^*([0, \bar{\theta}])$.

Clearly, both the value function and solution to the cost minimization problem are the same if we replace the feasible set by $Z^*([0, \bar{\theta}])$, that is, if we solve

$$C(a, \theta) = \min_{(z_1, ..., z_n) \in Z^*([0, \bar{\theta}])} \sum_i \pi_i(a) h(z_i + V(I + \theta)) - \theta.$$ 

Moreover, since $z^*(\cdot)$ is continuous in $\theta$, it follows that $Z^*([0, \bar{\theta}]) \subset \mathbb{R}^n$ is compact.

Therefore, it is now easy to verify that all the assumptions of Corollary 4 part $(iii)$ in Milgrom and Segal (2002) are satisfied. Hence, $C(a, \cdot)$ is continuously differentiable in
\( \theta \), and for any interior \( \theta \) the derivative is given by

\[
\frac{\partial C(a, \theta)}{\partial \theta} = \sum_i \pi_i(a) h'(z^*(a, \theta) + V(I + \theta))V'(I + \theta) - 1 \\
= \sum_i \pi_i(a) \frac{1}{V'(I_i + \theta)} V'(I + \theta) - 1.
\]

**3.2 Continuous Differentiability of \( C \) and \( v_i \) in \((a, \eta, \theta)\)**

In the proof of Proposition 3 we assume that \( v_i \) is continuously differentiable in \( \eta \). We now show, using a straightforward adaptation of Lemma 2 of Jewitt, Kadan, and Swinkels (2008), that if \( V(\cdot), \pi(\cdot) \) and \( \psi(\cdot) \) are \( C^2 \), then both \( C \) and \( v_i \) are continuously differentiable in \((a, \theta, \eta)\) under MLRP and CDFC. These two properties ensure that, when \( A = [0, \bar{a}] \), the first-order approach is valid, and when \( A \) is finite, only the local downward incentive constraint is binding when implementing any given action (Grossman and Hart (1983) p. 34). In both cases the constraint set can, without loss of generality, be reduced to two equality constraints.

We will just prove the result for \( A = [0, \bar{a}] \), as the finite case is analogous. Consider the cost minimization problem

\[
(P) \quad C(a, \theta, \eta) = \min \sum_i \pi_i(a) h(v_i) - \theta \\
\text{s.t.} \quad \sum_i \pi_i(a) v_i = v + \eta \psi(a), \\
\sum_i \pi'_i(a) v_i = \eta \psi'(a).
\]

As mentioned, problem \((P)\) has a unique solution \( v_i = v_i(a, \theta, \eta) \), which is characterized by the following system of equations:

\[
-\pi_i(a) h'(v_i) + \pi_i(a) \lambda(a, \theta, \eta) + \pi'_i(a) \mu(a, \theta, \eta) = 0, \quad i = 1, \ldots, n, \\
\sum_i \pi_i(a) v_i = v + \eta \psi(a), \\
\sum_i \pi'_i(a) v_i = \eta \psi'(a), \quad (1)
\]
where \( \lambda(a, \theta, \eta) \) and \( \mu(a, \theta, \eta) \) are Lagrange multipliers. Therefore

\[
v_i = V \left( (V')^{-1} \left( \frac{1}{\lambda(a, \theta, \eta) + \pi'(a) / \pi(a) \mu(a, \theta, \eta)} \right) \right),
\]

and \( \lambda(a, \theta, \eta) \) and \( \mu(a, \theta, \eta) \) are implicitly defined by the last two equations in the system (1). As in Lemma 2 in Jewitt, Kadan, and Swinkels (2008), we can apply the Implicit Function Theorem to conclude that \( \lambda(a, \theta, \eta) \) and \( \mu(a, \theta, \eta) \) are continuously differentiable for any \( a \in A, \theta > 0 \) and \( \eta \geq 0 \). Therefore the optimal contract \( v_i = v_i(a, \theta, \eta) \) is \( C^1 \); obviously, this implies that the same property holds for \( C(a, \theta, \eta) \).

3.3 Twice Continuous Differentiability of \( C \) and \( v_i \) in \( (a, \eta, \theta) \)

In Section 4 of this appendix, we assume the existence of the second derivative of the optimal contract with respect to \( a \). To justify this property, we provide here a simple set of conditions that will make the contract twice continuously differentiable. This simple argument we use might be of some independent interest.

We claim that MLRP, CDFC, plus \( V(\cdot), \pi(\cdot), \psi(\cdot) \) three times continuously differentiable imply that the optimal contract is twice continuously differentiable. To justify this assertion, notice that following the same steps as in the previous section, we can apply the Implicit Function Theorem for \( C^k \) functions with \( k = 2 \) (see Fiacco (1983), Theorem 2.4.1, or Dontchev and Rockafellar (2009), Proposition 1B.5) given that the primitives are three times continuously differentiable. Then the corresponding Lagrange multipliers and the optimal contract \( v_i = v_i(a, \theta, \eta) \) are twice continuously differentiable. Moreover, \( C(a, \theta, \eta) \) is twice continuously differentiability as well.

4 Tightness of TW’s Condition: \( n \)–Outcome Case

In the paper, we provide an intuitive proof of Proposition 2 (based on a suggestion by one of the referees), which shows that TW’s condition is tight. That proof exploits in a crucial way the assumption of two outcomes, since this allows for a closed form solution

\footnote{Moreover, notice that the cross partial \( \partial C / \partial \theta \partial \eta \) exists in this case, which we use in the proof of Proposition 3. To see this, apply Corollary 4 in Milgrom and Segal (2002) as above to obtain \( \partial C / \partial \theta \), and then differentiate with respect to \( \eta \); only \( \partial v_i / \partial \eta \) appears in the resulting expression.}
of the optimal contract that implements any action.

In this section we assume \( n \) outcomes, and prove that if the principal’s cost of implementing an action higher than the lowest one is increasing in the agent’s wealth \( \theta \) for all choices of \((\psi(\cdot), \pi(\cdot), A, Q, \hat{I}, \theta)\), then \( V(\cdot) \) satisfies TW’s condition. We do so by generalizing the argument that we used in our original proof of the Proposition 2 in the paper (which was done for two possible outcomes only).

Suppose that \( V(\cdot) \) is such that \( P(\hat{I} + \hat{\theta}) > 3R(\hat{I} + \hat{\theta}) \) for some \( \hat{I} \) and \( \hat{\theta} \). We will show that there exists a principal-agent problem with \( V(\cdot) \) as the agent’s utility function such that, for some action \( a \), \( \partial C(a, \theta)/\partial \theta < 0 \) in an open neighborhood of \( \hat{\theta} \).

To this end, assume \( A = [0, \bar{a}]; q_1, \ldots, q_n \), with \( q_i < q_{i+1} \); \( \pi(\cdot) \) three times continuously differentiable with \( \pi_i(a) > 0 \) and \( \pi_i'(a) \) not all null for all \( a \); \( \psi(\cdot) \) three times continuously differentiable, with \( \psi''(0) > 0 \); \( \bar{I} = \hat{I} \) and \( \theta = \hat{\theta} \).

The optimal contract that implements \( a \) is characterized by the following system:

\[
\begin{align*}
-\pi_i(a)h'(v_i) + \lambda(a)\pi_i(a) + \mu(a)\pi_i'(a) &= 0, \quad i = 1, \ldots, n, \\
\sum_{i=1}^{n} \pi_i(a)v_i &= \mathcal{V} + \psi(a), \\
\sum_{i=1}^{n} \pi_i'(a)v_i &= \psi'(a), \quad (2)
\end{align*}
\]

where \( \lambda(a) \) and \( \mu(a) \) are Lagrange multipliers. Now, reasoning as before in Subsections 3.2 and 3.3 we may apply the Implicit Function Theorem for \( C^k \) functions with \( k = 2 \), to get that \( v_i(a) \) is twice continuously differentiable in \( a = 0 \). Lemma 1 below then shows that there is an \( \tilde{a} > 0 \) (which depends on \( \theta \)) such that \( \partial C(a, \theta)/\partial \theta < 0 \) for all \( \theta \) when \( a \in (0, \tilde{a}) \) if

\[
P(\bar{I} + \theta) < 3R(\bar{I} + \theta) + \frac{\psi''(0)V'(\bar{I} + \theta)}{\sum_{i=1}^{n} \pi_i(0) (v_i'(0))^2}, \quad (3)
\]

and, conversely, if there is such an action \( \tilde{a} \), then (3) holds with less than or equal to.

Now, the derivatives \( v_i'(0) \) can be calculated by differentiating (2) and solving the corresponding linear system to obtain \( v_i'(0) = \Gamma_i\psi''(0) \), where

\[
\Gamma_i = \frac{\sum_{i \neq j} \left( \pi_i \pi_j' - \pi_i' \pi_j \right) \left. \pi_k \right|_{a=0}}{\pi_i \sum_{j < k} \left( \pi_k \pi_j' - \pi_k' \pi_j \right) \pi_j \pi_k} = \frac{-\pi_i'}{\pi_i \sum_{j < k} \left( \pi_k \pi_j' - \pi_k' \pi_j \right) \pi_j \pi_k} \left. \pi_k \right|_{a=0}.
\]
By our assumptions on $\pi'$, it follows that at least one $\Gamma_i$ is not null.

To complete the proof, recall that $P(\hat{I} + \hat{\theta}) > 3R(\hat{I} + \hat{\theta})$ for $\hat{I}$ and $\hat{\theta}$, and we have set $\bar{I} = \hat{I}$ and $\bar{\theta} = \hat{\theta}$. Then there exists a threshold $\bar{k} > 0$ such that if $\psi''(0) > \bar{k}$,

$$P(\hat{I} + \hat{\theta}) > 3R(\hat{I} + \hat{\theta}) + \frac{\psi''(0)V'(\hat{I} + \hat{\theta})}{\sum_{i=1}^{n} \pi_i(0)(v_i'(0))^2} = 3R(\hat{I} + \hat{\theta}) + \frac{V'(\hat{I} + \hat{\theta})}{\psi''(0)\sum_{i=1}^{n} \pi_i(0)\Gamma_i^2},$$

and therefore $\partial C(\hat{a}, \hat{\theta})/\partial \theta < 0$ for some action $\hat{a} \in (0, \bar{a})$. But then $\partial C(\hat{a}, \hat{\theta})/\partial \theta < 0$ for all levels of wealth $\theta$ in an open neighborhood of $\hat{\theta}$.

To finish the construction of a principal-agent problem where the principal prefers a richer agent, notice that MLRP, CDFC, and the assumption that for each $a$, $\pi_i'(a) \neq 0$ for some $q_i$, imply that $B_a(a) > 0$, and it is nonincreasing in $a$. Assume that $\pi(\cdot)$ is such that the cost function is convex in $a$ (e.g., it follows from Jewitt, Kadan, and Swinkels (2008), adapted to a finite number of outcomes, that $C$ is convex in $a$ if $\pi(\cdot)$ is such that $L(a) = \min_{t} \sum_{i=1}^{n} \pi_i'(a) - t \sum_{i=1}^{n} \pi_i(a)$ is convex in $a$ for all $t$). Then 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 $(q_1, ..., q_n)$. In particular, this applies to any action in $(0, \bar{a})$. Thus, there is an open interval of wealth levels $\theta$ for which the principal prefers a richer agent.

\begin{proof}

Set $m(a) = \sum_{i=1}^{n} (\pi_i(a)g(v_i(a)))$, and note that $m(0) = g(\bar{v})$. Thus, $\partial C(a, \theta)/\partial \theta$ is positive if and only if $m(a) > m(0)$.

Differentiating $m(a)$, we obtain $m'(a) = \sum_{i=1}^{n} \pi_i'(a)g(v_i(a)) + \sum_{i=1}^{n} \pi_i(a)g'(v_i(a))v_i'(a)$. Notice that $m'(0) = 0$. For $m'(0) = g(\bar{v})\sum_{i=1}^{n} \pi_i'(0) + g'(\bar{v})\sum_{i=1}^{n} \pi_i(0)v_i'(0)$, and $\sum_{i=1}^{n} \pi_i'(0) = 0$ while $\sum_{i=1}^{n} \pi_i(0)v_i'(0) = (\sum_{i=1}^{n} \pi_i(a)v_i(a))'|_{a=0} = (\bar{v} + \psi(a))'|_{a=0} = 0$, where the second equality follows from the participation constraint.

Thus, to assess the behavior of $m(a)$ near $a = 0$, we look at $m''(0)$. If it is positive at $a = 0$, then so it is (by continuity) in a right neighborhood of zero, i.e., for $a$ small.

\end{proof}
Differentiating $m'(a)$, we obtain after some algebra:

$$m''(0) = g'(\bar{v}) \left( 2 \sum_{i=1}^{n} \pi'_i(0)v'_i(0) + \sum_{i=1}^{n} \pi_i(0)v''_i(0) \right) + g''(\bar{v}) \sum_{i=1}^{n} \pi_i(0)(v'_i(0))^2.$$

Differentiating $\sum_{i=1}^{n} \pi_i(a)v_i(a) = \bar{v} + \psi(a)$ yields $\sum_{i=1}^{n} \pi'_i(a)v'_i(a) + \sum_{i=1}^{n} \pi_i(a)v'_i(a) = \psi'(a)$. Since $\sum_{i=1}^{n} \pi'_i(a)v_i(a) = \psi'(a)$, it follows that $\sum_{i=1}^{n} \pi_i(a)v'_i(a) = 0$, and its derivative yields $\sum_{i=1}^{n} \pi_i(a)v''_i(a) = -\sum_{i=1}^{n} \pi'_i(a)v'_i(a)$. Also, the derivative of the incentive constraint is $\sum_{i=1}^{n} \pi''_i(a)v_i(a) + \sum_{i=1}^{n} \pi'_i(a)v'_i(a) = \psi''(a)$, which converges to $\sum_{i=1}^{n} \pi'_i(0)v'_i(0) = \psi'(0)$ as $a$ goes to zero since $v_i(a)$ converges to $\bar{v}$. Hence,

$$m''(0) = g'(\bar{v}) \left( \psi''(0) + \frac{g''(\bar{v})}{g'(\bar{v})} \sum_{i=1}^{n} \pi_i(0)(v'_i(0))^2 \right). \quad (5)$$

Since $\bar{v} = V(\bar{I} + \theta)$, (5) and the definition of $g(\cdot)$ imply that $m''(0) > 0$ if and only if (4) holds. Also, $m''(0) > 0$ implies that $m'(a) > m'(0) = 0$ for $a$ near zero. Thus, there is an $\tilde{a} > 0$ such that $\partial C(a, \theta)/\partial \theta > 0$ for all $a \in (0, \tilde{a})$. To prove the converse, if (4) did not hold with $\leq$, then $m''(0) < 0$, and thus $\partial C(a, \theta)/\partial \theta < 0$ for $a$ near zero. \hfill \Box

## 5 Agent’s Wealth and the Variance of the Contract

In Section 2.4 we asserted that an increase in $\theta$ increases the variance of the contract that implements any given action $a$ above the lowest one. We now prove this assertion.

Let $I = (I_1, ..., I_n)$ be the contract that minimizes the cost of implementing action $a$. Since $E[I|a, \theta] = C(a, \theta)$, it follows that

$$\frac{\partial E[I|a, \theta]}{\partial \theta} = \sum_{i=1}^{n} \pi_i(a) \left( \frac{V'(\bar{I} + \theta)}{V'(I_i + \theta)} - 1 \right),$$

which we know can be positive or negative.
In turn, the variance of $I$ is
\[
\text{var}[I|a, \theta] = E[I^2|a, \theta] - (E[I|a, \theta])^2,
\]
and thus
\[
\frac{\partial \text{var}[I|a, \theta]}{\partial \theta} = 2 \sum_{i=1}^{n} \pi_i(a) I_i \frac{dI_i}{d\theta} - 2(E[I|a, \theta]) \sum_{i=1}^{n} \pi_i(a) \frac{dI_i}{d\theta},
\]
where the last line follows by rearranging terms and using the expression above for the derivative of the mean of the contract.

To sign (6), rearrange the $I_i$’s in increasing order and reinterpret the sum in (6) as already being reordered in this way. Then this is of the form $\sum_i \pi_i(a) f_i m_i$, with $f_i = I_i - E[I|a, \theta]$ and $m_i = (V'(\bar{I} + \theta)/V'(I_i + \theta)) - 1$. Now, $f_i$ is increasing in $i$ and crosses zero from negative to positive; moreover, $\sum_i \pi_i(a) f_i = 0$. Also, $m_i$ is increasing in $i$. Hence, it follows from Lemma 1 in Persico (2000) that $\sum_i \pi_i(a) f_i m_i \geq 0$, thereby proving that the variance of the contract increases in $\theta$.

### 6 Small Disutility of Effort

At the end of Section 3.2 we asserted that in the continuum of actions case, under some differentiability assumptions the principal prefers a poorer agent if the equilibrium disutility of effort is small. We now prove this assertion.

Let $a(\theta, \eta)$ be the optimal action implemented by the principal when the agent’s wealth is $\theta$ and the disutility of effort parameter is $\eta$. Assume that $a(\theta, \cdot)$ is continuously differentiable in $\eta$ for each $\theta$, which holds if the contract that solves the cost minimization problem for each action is twice continuously differentiable (see Section 3.3 of this appendix for a rigorous justification). Let $\pi(\theta, \eta) \equiv B(a(\theta, \eta)) - C(a(\theta, \eta), \theta, \eta)$ the be principal’s expected profit at the optimal contract. By the Envelope Theorem, $\frac{\partial \pi}{\partial \theta} = -\partial C(a(\theta, \eta), \theta, \eta)/\partial \theta$, and it easily follows that $\frac{\partial \pi}{\partial \theta}|_{\eta=0} = 0$. Thus, to prove the assertion it suffices to show that $\frac{\partial^2 \pi}{\partial \theta \partial \eta}|_{\eta=0} < 0$. Now
\[
\frac{\partial^2 \pi}{\partial \theta \partial \eta}|_{\eta=0} = -\left(\frac{\partial^2 C(a(\theta, \eta), \theta, \eta) \partial a(\theta, \eta)}{\partial \theta \partial \eta}\right)|_{\eta=0} = \frac{\partial^2 C(a(\theta, \eta), \theta, \eta)}{\partial \theta \partial \eta}|_{\eta=0}.
\]

We know from the proof of Proposition 3 that the second term on the right side is
negative. Regarding the first term, the derivative of the action converges to \( \frac{\partial a(\theta, 0)}{\partial \eta} \), which is finite by continuous differentiability. Also,

\[
\frac{\partial^2 C(a(\theta, \eta), \theta, \eta)}{\partial \theta \partial a} \bigg|_{\eta=0} = \frac{1}{g(\bar{v})} \left( \sum_i \pi'_i(a)g(v_i) \bigg|_{\eta=0} + \sum_i \pi_i(a)g'(v_i) \frac{\partial v_i}{\partial a} \bigg|_{\eta=0} \right)
\]

\[
= \frac{1}{g(\bar{v})} \left( g(\bar{v}) \sum_i \pi'_i(a) + g'(\bar{v}) \sum_i \pi_i(a) v_i \bigg|_{\eta=0} \right)
\]

\[
= \frac{g'(\bar{v})}{g(\bar{v})} \eta \psi'(a) \bigg|_{\eta=0} = 0,
\]

where the last equality follows from \( \sum_i \pi'_i(a) = 0 \) and \( \sum_i \pi_i(a)v_i = \bar{v} + \eta \psi(a) \). Thus, \( \partial^2 \pi/\partial \theta \partial \eta \bigg|_{\eta=0} < 0 \), and this completes the proof of the assertion.

### 7 The Agent’s Outside Option

In Section 3.4 we made two claims on the agent’s outside option. We now prove them.

Consider first the case where the agent’s outside option is \( V(m + \theta) - \psi(\hat{a}) > V(I + \theta) \), where \( \hat{a} \) is the action implemented at the alternative job, and assume that the action the principal wants to implement is \( a \).

**Claim 1** If \( a < \hat{a} \) and \( g(\cdot) \) is concave in \( v \), then the principal prefers a richer agent.

**Proof.** Let \( \bar{v} \equiv V(m + \theta) \) and consider any action \( a < \hat{a} \). Proceeding as usual, the condition for \( C(a, \cdot) \) to be decreasing in \( \theta \) is \( \sum_i \pi_i(a)g(v_i) \leq g(\bar{v}) \). Now

\[
\sum_i \pi_i(a)g(v_i) \leq g \left( \sum_i \pi_i(a)v_i \right)
\]

\[
= g \left( \bar{v} + \psi(a) - \psi(\hat{a}) \right)
\]

\[
\leq g(\bar{v}) ,
\]

where the first inequality follows from \( g(\cdot) \) concave, the equality from the binding participation constraint, and the second inequality from the premise that \( \hat{a} > a \). This shows that \( C(a, \theta) \) is decreasing in \( \theta \) for any \( a < \hat{a} \). Hence, if the principal optimally implements an action lower than \( \hat{a} \), then he prefers a richer agent. \( \square \)
Consider now the case where the agent’s outside option is $\sum_j \hat{\pi}_j(\hat{a}) V(T_j + \theta) - \psi(\hat{a}) > V(\bar{T} + \theta)$. Assume that the action the principal wants to implement is $a$.

**Claim 2** If $a \geq \hat{a}$, $V(\cdot)$ satisfies DARA, and $g(\cdot)$ is convex in $v$, then the principal prefers a poorer agent.

**Proof.** Let $\bar{\pi}_j \equiv V(\bar{T}_j + \theta)$ for all $j$, and consider any action $a \geq \hat{a}$. Proceeding as usual, the condition for $C(a, \cdot)$ to be increasing in $\theta$ is $\sum_i \pi_i(a) g(v_i) \geq \frac{1}{\sum_j \hat{\pi}_j(\hat{a}) \frac{1}{g(\bar{\pi}_j)}}$.

Notice that DARA is equivalent to $1/g(\cdot) = V'(h(\cdot))$ convex in $v$, for

$$\frac{d^2V'(h(v))}{dv^2} = \frac{R}{V'(P - R)} \geq 0 \Leftrightarrow P \geq R,$$

which proves the assertion.

The convexity of $1/g$ (DARA) implies that

$$g \left( \sum_j \hat{\pi}_j(\hat{a}) \bar{v}_j \right) \geq \frac{1}{\sum_j \hat{\pi}_j(\hat{a}) \frac{1}{g(\bar{\pi}_j)}}. \quad (7)$$

In turn, $g$ convex implies

$$\sum_i \pi_i(a) g(v_i) \geq g \left( \sum_i \pi_i(a) v_i \right) \quad (8)$$

Finally, the binding participation constraint and $a \geq \hat{a}$ imply

$$g \left( \sum_i \pi_i v_i \right) = g \left( \sum_j \hat{\pi}_j \bar{v}_j + \psi(a) - \psi(\hat{a}) \right) \geq g \left( \sum_j \hat{\pi}_j \bar{v}_j \right). \quad (9)$$

From (7)–(9) we obtain

$$\sum_i \pi_i(a) g(v_i) \geq \frac{1}{\sum_j \hat{\pi}_j(\hat{a}) \frac{1}{g(\bar{\pi}_j)}}$$

and thus $C(a, \cdot)$ increases in $\theta$ for any $a \geq \hat{a}$. Hence, if the principal optimally implements an action bigger than $\hat{a}$, then he prefers a poorer agent. \hfill $\square$
8 Large Wealth with \( n = 2 \)

Recall the conditions we imposed in Section 3.4 of the paper:

(a) There is a threshold \( \bar{v} \) such that either \( g(\cdot) \) is convex in \( v \) when \( v \in (\bar{v}, \infty) \), or \( g(\cdot) \) is concave in \( v \) when \( v \in (\bar{v}, \infty) \) 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 \(|v_i - \bar{v}| \leq K_a\) 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 \( \theta \).

Assume that condition (a) holds. We will show that if \( n = 2 \) (so there are two possible outcomes), and \( \pi(\cdot) \) satisfy the strict MLRP and CDFC, then conditions (b)–(d) hold when the action set is finite or is an interval.

If \( A = \{a_1, a_2, \ldots, a_m\} \), and the principal wants to implement \( a_k > a_1 \), then, under the general assumptions made in Section 2 plus MLRP and CDFC, only the incentive constraint corresponding to \( a_{k-1} \) binds (Grossman and Hart (1983) p. 34). Hence,

\[
\begin{align*}
v_1 &= \bar{v} + \psi(a_k) - \pi_2(a_k) \frac{\psi(a_k) - \psi(a_{k-1})}{\pi_2(a_k) - \pi_2(a_{k-1})}, \\
v_2 &= \bar{v} + \psi(a_k) + (1 - \pi_2(a_k)) \frac{\psi(a_k) - \psi(a_{k-1})}{\pi_2(a_k) - \pi_2(a_{k-1})}.
\end{align*}
\]

Since \( \pi_2(\cdot) \) is strictly increasing, it follows that \( \pi_2(a_k) - \pi_2(a_{k-1}) > 0 \). Thus, we can set

\[
K_{a_k} = \psi(a_k) + \max \left\{ \pi_2(a_k) \frac{\psi(a_k) - \psi(a_{k-1})}{\pi_2(a_k) - \pi_2(a_{k-1})} , (1 - \pi_2(a_k)) \frac{\psi(a_k) - \psi(a_{k-1})}{\pi_2(a_k) - \pi_2(a_{k-1})} \right\} > 0.
\]

Clearly, \( \sup_{a_k \in A} K_{a_k} < \infty \), showing that (b)–(c) hold ((d) is not needed here).

Let now \( A = [0, \bar{a}] \). The optimal contract that implements an action \( a > 0 \) is given

\[
\begin{align*}
v_1 &= \bar{v} + \psi(a) - \pi_2(a) \frac{\psi(a)}{\pi_2'(a)}, \\
v_2 &= \bar{v} + \psi(a) + (1 - \pi_2(a)) \frac{\psi(a)}{\pi_2'(a)}.
\end{align*}
\]
Therefore, we can set the value of $K_a$ as

$$K_a = \psi(a) + \max \{ \pi_2(a)(\psi'(a)/\pi_2'(a)), (1 - \pi_2(a))(\psi'(a)/\pi_2'(a)) \} > 0.$$ 

The continuity of the functions involved in the definition of $K_a$ and the fact that $\pi_2'(a) > 0$, yield $\sup_{a \in A} K_a < \infty$. Hence, conditions (b)–(c) hold in this case as well.

Regarding condition (d), it holds if MLRP is strict, the vector of outcomes is large enough (which is a plausible assumption in, e.g., the CEO application), and there is an arbitrarily large upper bound on wealth. Alternatively, one can show that if $V(\cdot)$ is such that $g''(\cdot)$ is increasing in $v$, then condition (d) is not needed in this case ((a)–(c) suffice). We prove both assertions now:

**CONDITION (d) FROM PRIMITIVES.** Assume two outcomes, $\pi_2'(a) > 0$ for all $a$, and $\theta \in [0, \bar{\theta}]$ for $\bar{\theta} > 0$ arbitrarily large. We will show that if $\Delta q = q_2 - q_1$ is sufficiently large, then condition (d) holds (i.e., there is an action $\tilde{a}$ such that the principal’s optimal action is greater than or equal to $\tilde{a}$ for all $\theta \in [0, \bar{\theta}]$).

Recall that the optimal contract that implements action $a$ in this case is

$$v_1 = \overline{v} + \psi(a) - \pi_2(a) \frac{\psi'(a)}{\pi_2'(a)},$$

$$v_2 = \overline{v} + \psi(a) + (1 - \pi_2(a)) \frac{\psi'(a)}{\pi_2'(a)},$$

and thus the cost function is $C(a, \theta) = \pi_2(a)h(v_2) + (1 - \pi_2(a))h(v_1) - \theta$, while the expected revenue is $B(a) = q_1 + \pi_2(a)\Delta q$. It is straightforward to show that $C_a(0, \theta) = 0$ and thus $B_a(0) - C_a(0, \theta) = \pi'_2(0)\Delta q > 0$. In general

$$B_a(a) - C_a(a, \theta) = \pi'_2(a)\Delta q - \pi'_2(a) (h(v_2) - h(v_1)) - \frac{\partial}{\partial a} \left( \frac{\psi'(a)}{\pi_2'(a)} \right) (1 - \pi_2(a))\pi_2(a) (h'(v_2) - h'(v_1)).$$

Notice that $C_a(a, \theta)$ is continuous on $[0, \overline{a}] \times [0, \bar{\theta}]$. Let $\rho \in (0, \infty)$ be

$$\rho = \max_{[0, \overline{a}] \times [0, \bar{\theta}]} \pi'_2(a) (h(v_2) - h(v_1)) + \frac{\partial}{\partial a} \left( \frac{\psi'(a)}{\pi_2'(a)} \right) \pi_1(a)\pi_2(a) (h'(v_2) - h'(v_1)).$$
Take \( \tilde{a} \in (0, \bar{a}] \), and consider \( \Delta q = q_2 - q_1 > 0 \) large enough so that

\[
\pi_2'(\tilde{a}) (q_2 - q_1) > \rho.
\]

Then \( B_a(a) - C_a(a, \theta) > 0 \) for all \( a \in (0, \tilde{a}] \) and \( \theta \in [0, \bar{\theta}] \), and hence \( \arg \max_{a \in [0, \tilde{a}]} B(a) - C(a, \theta) \geq \tilde{a} > 0 \) for all \( \theta \in [0, \bar{\theta}] \), thereby proving that condition (d) holds.

**Increasing \( g''(\cdot) \).** When \( g''(\cdot) \) is increasing in \( v \), the principal prefers a poorer agent in the two-outcome case without imposing condition (d). We show this by proving that \( C(a, \cdot) \) is increasing in \( \theta \) for all \( a \in [0, \tilde{a}] \) when \( \theta \) is large enough. As a result, \( \max_{a \in [0, \tilde{a}]} B(a) - C(a, \theta) \) is decreasing in \( \theta \) when \( \theta \) is sufficiently large.

Assume \( n = 2 \), and recall that Section 3.3 assumes that \( g(\cdot) \) is concave in \( v \) and \( g''(v)/g'(v) \) converges to zero as \( v \) goes to infinity. We also assume here that \( g''(\cdot) \) is increasing in \( v \). The derivative of the cost function with respect to \( \theta \) satisfies the following inequality:

\[
C_\theta(a, \theta) = \frac{1}{g(\bar{v})} (\pi_2(a)g(v_2) + (1 - \pi_2(a))g(v_1) - g(\bar{v}))
\]

\[
= \frac{1}{g(\bar{v})} \left( (\pi_2(a)g(v_2) + (1 - \pi_2(a))g(v_1) - g(\bar{v} + \psi(a))) + (g(\bar{v} + \psi(a)) - g(\bar{v})) \right)
\]

\[
\geq \frac{1}{g(\bar{v})} \left( (\pi_2(a)g(v_2) + (1 - \pi_2(a))g(v_1) - g(\bar{v} + \psi(a))) + g'(\bar{v} + \psi(a))\psi(a) \right),
\]

where the inequality uses the concavity of \( g \).

The assumptions on \( g(\cdot) \) imply that \( g''(v_1) \leq g''(v) \leq 0 \), for all \( v \geq v_1 \).

Using a second order Taylor expansion around \( \bar{v} + \psi(a) \), we obtain

\[
\pi_2(a)g(v_2) + (1 - \pi_2(a))g(v_1) - g(\bar{v} + \psi(a)) \geq 0.5g''(v_1)\pi_2(a)(1 - \pi_2(a)) \frac{\psi'(a)^2}{\pi_2^2(a)}.
\]

Thus,

\[
C_\theta(a, \theta) \geq \frac{\psi(a)}{g(\bar{v})} \left( 0.5g''(v_1)\pi_2(a)(1 - \pi_2(a)) \frac{\psi'(a)^2}{\pi_2^2(a)} + g'(\bar{v} + \psi(a)) \right)
\]

\[
\geq M \frac{g'(\bar{v} + \psi(a))}{g(\bar{v})} \psi(a) \left( g'' \left( \bar{v} + \psi(a) - \pi_2(a) \frac{\psi'(a)}{\pi_2(a)} \right) \frac{\pi_2(a)^2 \psi'(a) \psi(a)}{g(\bar{v} + \psi(a))} + M^{-1} \right),
\]

13
where $M = \max_{a \in [0, \bar{a}]} 0.5(\pi_2(a)(1 - \pi_2(a))/\pi'_2(a)^2)(\psi'(a)^2/\psi(a)) \in (0, \infty)$. (Notice that since $\psi(0) = \psi'(0) = 0$, L'Hopital's Rule yields $\lim_{a \to 0} \psi'(a)^2/\psi(a) = 2\psi''(0) > 0$.)

Since $\lim_{v \to \infty} g''(v)/g'(v) = 0$, the first term in parenthesis goes to zero as $\bar{v}$ goes to infinity. Since $V(\cdot)$ is unbounded, it follows that there exists a threshold $\theta^* < \infty$ such that, if $\theta \geq \theta^*$, then $C_0(a, \theta) \geq 0$ for all $a$ and hence the principal prefers a poorer agent.

9 Appendix: Continuity of the Optimal Contract

In this appendix we prove that the optimal contract is continuous in agent’s wealth. The proof only requires that $V(\cdot)$, $\pi(\cdot)$, and $\psi(\cdot)$ be $C^2$. We proceed in terms of the transformed variables $z_i = v_i - V(I, \theta), i = 1, ..., n$ defined above in Section 3.1. For definiteness, we will focus on the cost minimization problem for $A = [0, a], a > 0, \bar{a}$ fixed. But it will be clear below that the same results hold in the finite case where the first-order condition of the agent’s problem is replaced by a finite number of inequalities (see the parenthetical remark in the proof of Claim 2).

Consider

$$C(a, \theta) = \min \sum_i \pi_i(a) h \left(z_i + V(I + \theta)\right) - \theta$$

s.t.

$$\sum_i \pi_i(a) z_i = \psi(a),$$

$$\sum_i \pi'_i(a) z_i = \psi'(a).$$

Set $f(z; \theta) = f(z_1, ..., z_n; \theta) := \sum_i \pi_i(a) h \left(z_i + V(I + \theta)\right) - \theta$. If $z^*(\theta)$ denotes the unique solution of this problem, then $C(a, \theta) = f(z^*(\theta); \theta)$.

Claim 3 The cost function $C(a, \theta)$ is upper semi-continuous on $\theta$.

Proof: Fix any $\theta^*$ and take the unique solution $z^*(\theta^*)$. Then, for any $\varepsilon > 0$, consider $\delta > 0$ such that

$$|f(z; \theta) - f(z^*(\theta^*); \theta^*)| < \varepsilon \quad \text{if} \quad \|z, \theta\) - (z^*(\theta^*), \theta^*)\| < \delta.$$ 

Observe that for all $\theta \in (\theta^* - \delta, \theta^* + \delta)$, we have

$$\|z^*(\theta^*), \theta\) - (z^*(\theta^*), \theta^*)\| = |\theta - \theta^*| < \delta,$$
so
\[ |f(z^*(\theta^*); \theta) - f(z^*(\theta^*); \theta^*)| < \varepsilon, \]
consequently
\[ f(z^*(\theta); \theta) \leq f(z^*(\theta^*); \theta) < f(z^*(\theta^*); \theta^*) + \varepsilon. \]
Hence
\[ C(a, \theta) < C(a, \theta^*) + \varepsilon, \]
which gives
\[ \limsup_{\theta \to \theta^*} C(a, \theta) \leq C(a, \theta^*) + \varepsilon \]
for all \( \varepsilon > 0 \). Therefore
\[ \limsup_{\theta \to \theta^*} C(a, \theta) \leq C(a, \theta^*). \]
\[ \square \]

**Claim 4** The optimal contract \( z^*(\theta) \) is continuous.\(^2\)

**Proof.** Take any \( \theta^* \) and any sequence \( \{\theta^k\} \) converging to \( \theta^* \). Write \( z^* = z^*(\theta^*) \) and \( z^k = z^*(\theta^k), \ k = 1, 2, \ldots \). We need to show that \( z^k \to z^* \).

Observe that if \( z^k \to z \), then the continuity of \( f \) and Claim 1 give
\[ f(z, \theta^*) = \lim_{k \to \infty} f(z^k, \theta^k) = \lim_{k \to \infty} C(a; \theta^k) \leq C(a; \theta^*) = f(z^*; \theta^*), \]
which implies that \( z = z^* = z^*(\theta^*) \) because of the uniqueness of the solution of the cost minimization problem. Hence it is enough to show that the sequence \( \{z^k\} \) is bounded, because then any of its convergent subsequence has to converge to \( z^* \).

Towards a contradiction, suppose that \( \{z^k\} \) is unbounded, and assume (by passing to a subsequence if necessary) that \( z^k \neq z^* \) and
\[ \frac{z^k - z^*}{\|z^k - z^*\|} \to d, \]
for some vector \( d \in \mathbb{R}^n \) with \( \|d\| = 1 \). We will show that \( z^* + d \in Z^* := \{z^*(\theta^*)\} = \{z^*\} \), thereby reaching an obvious contradiction.

\(^2\)This proof of this claim builds on (Gaya, Lopez, and Vera de Serio 2003) Lemma 3.6, Proposition 4.4, and Theorem 4.8; and (Goberna and Lopez 1998), Theorems 10.3 (ii)-(iii), 10.4 (i)-(ii).
Since \( z^* \) and all \( z^k \) satisfy the constraint restrictions, we have

\[
0 = \frac{1}{\|z^k - z^*\|} \sum_i \pi_i(a) \left( z_i^k - z_i^* \right) = \frac{\pi(a) \cdot \frac{z^k - z^*}{\|z^k - z^*\|}}{\pi' \cdot \frac{z^k - z^*}{\|z^k - z^*\|}} \rightarrow \pi(a) \cdot \frac{z^*}{d}, \tag{10}
\]

\[
0 = \frac{1}{\|z^k - z^*\|} \sum_i \pi_i'(a) \left( z_i^k - z_i^* \right) = \frac{\pi'(a) \cdot \frac{z^k - z^*}{\|z^k - z^*\|}}{\pi' \cdot \frac{z^k - z^*}{\|z^k - z^*\|}} \rightarrow \pi'(a) \cdot \frac{z^*}{d}, \tag{11}
\]

which implies that \( z^* + d \) satisfies the participation constraint and the incentive constraint, i.e. it is an element of the feasible set of the cost minimization problem.

(In the finite case we have to consider the finite set of inequalities given by the incentive constraints. For any \( a' \in A \),

\[
\sum_i \left( \pi_i(a) - \pi_i(a') \right) \left( \frac{z_i^k - z_i^*}{\|z^k - z^*\|} \right) = \frac{1}{\|z^k - z^*\|} \sum_i \left( \pi_i(a) - \pi_i(a') \right) z_i^k \tag{12}
\]

\[
= \frac{\pi(a) \cdot \frac{z^k - z^*}{\|z^k - z^*\|}}{\pi' \cdot \frac{z^k - z^*}{\|z^k - z^*\|}} \rightarrow \pi(a) \cdot \frac{z^*}{d} \]

\[
\leq \frac{1}{\|z^k - z^*\|} (\psi(a) - \psi(a'))
\]

\[
\geq \frac{\pi(a) \cdot \frac{z^k - z^*}{\|z^k - z^*\|}}{\pi' \cdot \frac{z^k - z^*}{\|z^k - z^*\|}}.
\]

Thus, taking limit as \( k \rightarrow \infty \), we obtain \( \pi(a) \cdot d \geq \pi(a') \cdot d \). Hence \( \pi(a) \cdot (z^* + d) - \psi(a) \geq \pi(a') \cdot (z^* + d) - \psi(a') \), and again \( z^* + d \) is an element of the feasible set. This is the only significant modification needed in the finite case. The rest holds with minor changes.)

Now, let \( c^* = C(a; \theta^* \right) \) and observe that

\[
Z^* = \left\{ z : \sum_i \pi_i(a) \ z_i = \psi(a), \sum_i \pi_i'(a) \ z_i = \psi'(a), \ f(z; \theta^*) \leq c^* \right\}
\]

\[
= \left\{ z : \sum_i \pi_i(a) \ z_i = \psi(a), \sum_i \pi_i'(a) \ z_i = \psi'(a); \ u \cdot z \leq F(w; \theta^*) + c^* \right\}, \tag{13}
\]

16
where $F(u; \theta^*)$ is the Fenchel conjugate of $f(z; \theta^*)$, defined for any $u \in \mathbb{R}^n$ by

$$F(u; \theta^*) = \sup_{z \in \mathbb{R}^n} \{u \cdot z - f(z; \theta^*)\},$$

which in particular satisfies that $F(u; \theta^*) + f(z; \theta^*) = u \cdot z$ whenever $u = \nabla_z f(z; \theta^*)$.\(^3\)

By inspection of (10), (11), and (13), it is apparent that $z^* + d$ would be a solution of the cost minimization problem, i.e. $z^* + d \in Z^*$, if $d \cdot \nabla_z f(z; \theta^*) \leq 0$ for all $z \in \mathbb{R}^n$.

Observe that the sequence $\{f(\cdot; \theta^k)\}$ converges uniformly to $f(\cdot; \theta^*)$ on compact sets.\(^4\) Hence, we can apply Theorem 4.2 in (Attouch and Beer 1993). For any $z \in \mathbb{R}^n$ and $\pi = \nabla_z f(z; \theta^*)$, this Theorem gives the existence of sequences $\{u^k\}$ and $\{y^k\}$ in $\mathbb{R}^n$ such that $u^k = \nabla_z f(y^k; \theta^k)$, $u^k \rightarrow \pi$, and $y^k \rightarrow z$. It follows that

$$u^k \cdot \frac{z^k - z^*}{\|z^k - z^*\|} + u^k \cdot \frac{z^* - y^k}{\|z^k - z^*\|} = \frac{1}{\|z^k - z^*\|} u^k \cdot (z^k - y^k) \leq \frac{1}{\|z^k - z^*\|} \left( f(z^k; \theta^k) - f(y^k; \theta^k) \right) = \frac{1}{\|z^k - z^*\|} \left( C(a, \theta^k) - f(y^k; \theta^k) \right),$$

where the $\leq$ sign in the third line is justified by the convexity of the function $f(\cdot; \theta^k)$. Letting $k \rightarrow \infty$, the first and last lines yield

$$\pi \cdot d \leq 0,$$

which holds because $u^k \cdot \frac{z^k - z^*}{\|z^k - z^*\|} \rightarrow \bar{u} \cdot d$, $\frac{1}{\|z^k - z^*\|} \rightarrow 0$, and all the other expressions are bounded since the sequence $(u^k, y^k, \theta^k)$ is convergent and $\lim \sup_{k \rightarrow \infty} C(a, \theta^k) \leq C(a, \theta^*) < \infty$. This completes the proof that $z^* + d \in Z^* = \{z^*(\theta^*)\} = \{z^*\}$. Since $\|d\| = 1$, we have reached the sought after contradiction.

---

\(^3\)In general, $u \in \partial f(z; \theta^*)$, if $f$ is not differentiable. See (Rockafellar 1970) Section 12 for further properties of the Fenchel conjugate of a convex function.

\(^4\)We can apply Arzelà–Ascoli theorem to the convergent sequence $\{f(\cdot; \theta^k)\}$, which has uniformly bounded derivatives, besides being uniformly bounded itself on compact sets. To see this it is enough to consider the restriction of the continuously differentiable function $f$ to compact sets of the form $B_1 \times B_2$, where $B_1$ is a compact box in $\mathbb{R}^n$ and $B_2$ is the compact set formed by $\theta^*$ and all $\theta^k$, $k = 1, 2, \ldots$. Then, take bounds for $f$ and $\|\nabla_z f\|$ on $B_1 \times B_2$. 

---
Therefore \( \{z^k\} \) is bounded and thus \( z^* (\theta^k) = z^k \rightarrow z^* = z^* (\theta^*) \).

**Corollary:** The set \( Z^* ([0, \bar{\theta}]) \) is compact for any fixed \( \bar{\theta} > 0 \).

**Proof:** Since \( Z^* ([0, \bar{\theta}]) \) is the image of a compact set through the continuous function \( z^* (\theta) \), it follows that it is compact.

**Corollary:** The cost function \( C (a, \theta) \) is continuous on \( \theta \).

**Proof:** \( C (a, \theta) = f (z^* (\theta); \theta) \) is a composition of continuous functions.

**Remark:** We have kept the action \( a \) fixed to avoid complicating the notation. Notice, however, that we can repeat all the steps above considering, instead of \( \{ \theta^k \} \), a sequence \( \{(a^k, \theta^k)\} \) converging to \((a^*, \theta^*)\). This way we would obtain the continuity of the optimal contract \( z^* (a, \theta) \) and of the cost \( C (a, \theta) \) as functions of the couple \((a, \theta)\). The only significant change needed is in obtaining (10) and (11) (and the incentive constraints in the finite case), e.g.

\[
\begin{align*}
\frac{\psi (a^k) - \psi (a^*)}{\|z^k - z^*\|} &= \frac{1}{\|z^k - z^*\|} \left( \sum_i \pi_i (a^k) \ z_i^k - \sum_i \pi_i (a^*) \ z_i^* \right) \\
&= \frac{1}{\|z^k - z^*\|} \left( \pi (a^k) \cdot (z^k - z^*) + \left( \pi (a^k) - \pi (a^*) \right) \cdot z^* \right) \\
&= \pi (a^k) \cdot \frac{z^k - z^*}{\|z^k - z^*\|} + \frac{\left( \pi (a^k) - \pi (a^*) \right) \cdot z^*}{\|z^k - z^*\|};
\end{align*}
\]

letting \( k \rightarrow \infty \) we get \( 0 = \pi (a^*) \cdot d \), by the \( C^1 \) property of the functions \( \psi \) and \( \pi \) (in the finite case, consider \( \sum_i (\pi_i (a^k) - \pi_i (a')) [z_i^k - z_i^*] /\|z^k - z^*\| \) in (12)).

**References**


