On the Performance of Linear Contracts

by

Arup Bose*, Debashis Pal**, and David E. M. Sappington***

Abstract

Linear contracts are common in practice, even though they are seldom optimal. This observation suggests that linear contracts may closely approximate the performance of fully optimal contracts in many settings of economic importance. Our investigation of the canonical moral hazard setting provides support for this possibility. In the broad class of environments we consider, the optimal linear contract always secures for the principal at least 90% of the expected profit she secures with a fully optimal contract, as long as the productivity of the agent's effort is not too small.

July 2007

* Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India.

** Department of Economics, University of Cincinnati, Cincinnati, Ohio 45221, USA.

*** Department of Economics, University of Florida, Gainesville, Florida 32611, USA.

We thank Rick Harbaugh for very helpful comments and observations. Arup Bose and Debashis Pal thank Taft Research Center for its generous support.

1 Introduction.

Linear contracts are common in practice. From sharecropping contracts to sales commissions to profit-sharing arrangements in regulated industries, linear contracts are ubiquitous.[1] Linear contracts can be fully optimal contracts under special circumstances, as, for example, in a pure moral hazard setting where a risk neutral agent is not wealth constrained (e.g., Shavell, 1979) or where a risk-averse agent with a negative exponential utility function controls the drift of a Brownian motion process (e.g., Holmstrom and Milgrom, 1987; Sung, 1995).[2] However, despite their widespread use, linear contracts typically are not optimal more generally. Why, then, are linear contracts so common in practice?

The widespread use of linear contracts likely is explained in part by their simple structure.[3] However, contracting parties would be paying a large premium for "simplicity" if linear contracts were employed in settings where nonlinear contracts would secure substantially more surplus. The purpose of this research is to determine whether the use of linear contracts typically entails a substantial loss of surplus in the canonical moral hazard setting. We compare the principal's expected profit under the optimal linear contract (π^L) with her expected profit under a fully optimal contract (π) in a broad class of environments. We find that π^L is more than 90% of π as long as the productivity of the agent's effort is not too small.[4] Thus, in many settings of practical economic importance, linear contracts enable the principal to secure a large fraction of the expected profit that she secures with a fully optimal contract.

[1] Milgrom and Roberts (1992, p. 216) note that "Linear compensation formulas are commonly observed in the form of commissions paid to sales agents, contingency fees paid to attorneys, piece rates paid to tree planters or knitters, crop shares paid to sharecropping farmers, and so on." Schmalensee (1989, p. 418) observes that "most incentive schemes observed in practice are linear." Bhattacharyya and Lafontaine (1995, pp. 763-4) observe that "Linear pricing rules have been found in a number of diverse areas such as, but not limited to, sales force compensation, sharecropping, leasing arrangements, author's fees, legal fees, licensing agreements, commercial real estate rental fees, and franchising."

[2] Linear contracts also can be an optimal means to resolve double moral hazard problems (e.g., Bhattacharyya and Lafontaine, 1995; Kim and Wang, 1998; Corbett et al., 2005) or to motivate efficient investment (e.g., Pfeiffer and Velthius, 2005). Menus of linear contracts can constitute optimal contracts in adverse selection settings (e.g., Laffont and Tirole, 1986; Reichelstein, 1992) and in settings with both adverse selection and moral hazard (e.g., Guesnerie et al., 1989), as long as agents are risk neutral.

[3] As Schmalensee (1989, p. 418) notes, "Linear schemes ... seem to be more readily understood and administered in practice than nonlinear regimes." In their analysis of regulatory policies, Gasmi et al. (1999, p. 91) note the "simplicity and practicality of ... price-cap regulation with [linear] taxation of earnings."

[4] As explained in section 2, we consider settings in which the agent's stochastic performance reflects a gamma density with mean pa, where a is the agent's effort. We find that π^L is more than 90% of π whenever $p \geq 1$ and the agent's preferences satisfy standard sufficient conditions for the validity of the first-order approach to solving the moral hazard problem.

We employ specific representations of preferences and the stochastic relationship between the agent's unobserved effort and his observed performance in order to characterize completely both fully optimal contracts and optimal linear contracts.[5] While not entirely general, the functional forms we employ are highly flexible and admit a great variety of preferences and stochastic production functions, as we explain in more detail in section 2.

Section 2 describes the key elements of our model. Sections 3 and 4 characterize the fully optimal contract and the optimal linear contract, respectively. Section 5 compares the performance of the two types of contracts and identifies conditions under which linear contracts approximate the performance of fully optimal contracts particularly closely. Section 6 summarizes our key findings, discusses additional links with the literature, provides further evidence regarding the robustness of our findings, and suggests directions for future research. The proofs of all formal conclusions are presented in the Appendix.

2 Elements of the Model.

We analyze the canonical moral hazard model (e.g., Holmstrom, 1979) in which a risk neutral principal contracts with a single risk averse agent to produce output, x. Increased effort (a) by the agent stochastically increases the output he produces. The impact of the agent's effort on output is captured by the two-parameter gamma density:

$$f(x|a) = \frac{x^{p-1}\, e^{-x/a}}{a^p\, \Gamma(p)} \quad \text{for} \quad x \in [0, \infty), \quad \text{where} \quad \Gamma(p) = \int_0^\infty e^{-u}\, u^{p-1}\, du. \tag{1}$$

The mean of this density is pa, so p can be viewed as a parameter that reflects the productivity of the agent's effort.

The use of the gamma density to capture the stochastic relationship between the agent's unobserved effort and his observed performance offers three important advantages. First and foremost, the density is highly flexible, and so permits an analysis of a great variety of relationships between effort and output. Suitable values of p and a admit close approximations to many unimodal densities.[6] Figure 1 illustrates some of the many shapes that the gamma

[5] As Guo and Yang (2006, p. 150) note, "Using general utility functions as well as general output processes, it is usually not possible to characterize explicitly optimal contracts and other properties that can be tested empirically."

[6] The exponential and the chi-square densities are special cases of the gamma density. The normal distribution can be approximated by an appropriately centered gamma density with a large value for the shape

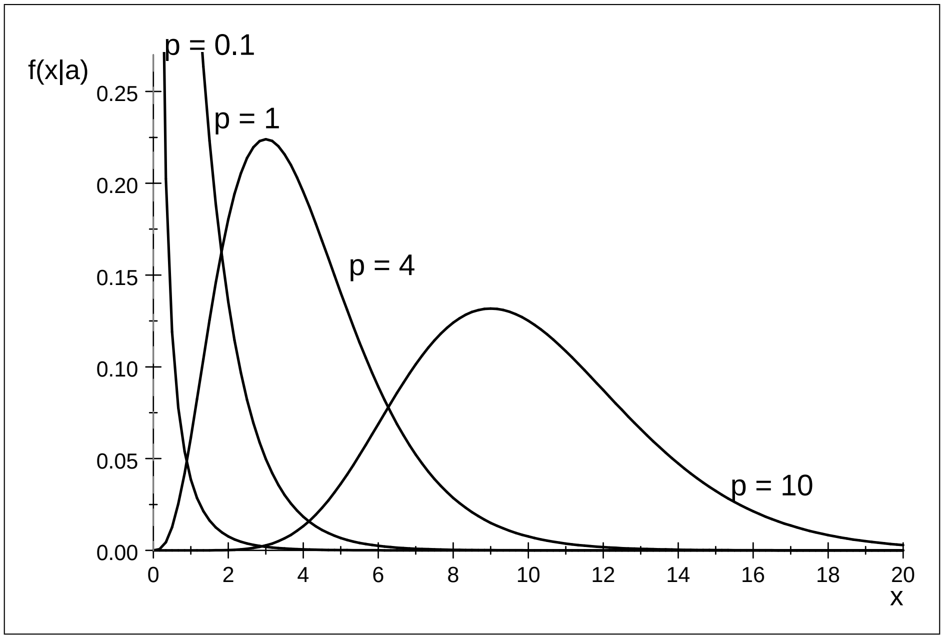

Figure 1: The Gamma density for $p = 0.1, 1.0, 4.0$ and 10.0 ($a = 1$).

density can assume as p increases from 0.1 to 1 to 4 to 10, holding a constant at 1. Second, the gamma density admits closed-form analytic solutions to the principal's problem in cases of interest. Third, use of the gamma density facilitates the identification of conditions under which the tractable "first-order approach" can be employed to solve the principal's problem (e.g., Rogerson, 1985; Jewitt, 1988; Mirrlees, 1999).

The agent's utility when he is paid (wage) w and he delivers effort a is $U^A(w, a) = 2w^\theta - a^\delta$, where $\theta \in (0, 1/2]$ and $\delta \geq 1$.[7] The agent has no wealth initially. Because θ is less than 1, the agent is averse to risk. His degree of relative risk aversion $(1 - \theta)$ declines as θ increases.[8] The presumed separability of $U^A(\cdot)$ in w and a is standard in the literature. By varying θ and δ, we are able to consider wide variations in the agent's degree of relative risk aversion and his aversion to effort.

The risk neutral principal designs a payment schedule, $w(x)$, that specifies the payment

parameter, p. The gamma distribution is used extensively in applied work in many disciplines, including meteorology, ecology, and economics (Johnson et al., 1994).

[7] As explained further below, θ is assumed to be no greater than $\frac{1}{2}$ in our primary analysis to ensure that the first-order approach can be employed to solve for the fully optimal contract. Section 6 considers the setting where $\theta \in (\frac{1}{2}, 1)$.

[8] The agent's degree of relative risk aversion is $-w[\frac{\partial^2}{\partial w^2} U^A(w, a)]/[\frac{\partial}{\partial w} U^A(w, a)] = -w[2\theta(\theta - 1)w^{\theta-2}]/[2\theta w^{\theta-1}] = 1 - \theta$.

the agent will receive for each level of output x he produces. The principal designs this schedule to maximize the difference between the agent's expected output and the expected payment from the principal to the agent.[9] The agent will only work for the principal if he anticipates expected utility in excess of his reservation utility $\overline{U} \geq 0$.

The timing in the model is as follows. First, the principal specifies the payment schedule, or contract, $w(x)$. Second, the agent decides whether he will accept the specified contract and work for the principal. If the agent declines to do so, his interaction with the principal is forever terminated. If the agent accepts the contract, he next chooses his effort supply. Finally, realized output is observed and the principal delivers the corresponding promised payment to the agent. This interaction is not repeated.

This timing, the stochastic relationship between x and a, the agent's utility function, and his reservation utility are all common knowledge. The output produced by the agent is observed publicly, but the agent's effort supply is not. The principal's task is to motivate the agent to deliver effort without imposing undue risk on the agent. Formally, the principal's problem, labeled [P], is the following:

$$\underset{w(x)\geq 0,\ a}{\text{Maximize}} \quad \int_x [x - w(x)]\, f(x|a)\ dx \tag{2}$$

$$\text{subject to:} \quad \int_x U^A(w(x), a)\, f(x|a)\ dx \ \geq \ \overline{U}\ , \ \text{ and} \tag{3}$$

$$a \in \arg\max_{\tilde{a}} \ \{ \int_x U^A(w(x), \tilde{a})\, f(x|\tilde{a})\ dx \}. \tag{4}$$

Expression (2) reflects the principal's desire to maximize the difference between expected output and payments to the agent. Expression (3) is the participation constraint, which requires the principal to design $w(x)$ to ensure the agent at least $\overline{U}$ in expected utility.[10] Expression (4), the effort selection constraint, indicates that the agent will choose his effort to maximize his expected utility, given the prevailing contract, $w(x)$. It is readily verified

[9] Thus, for simplicity, we normalize the unit value of output to 1, so x denotes both the amount of output produced and the value of that output to the principal.

[10] We assume that $\overline{U}$ is sufficiently small relative to the total expected surplus from the agency relationship that the principal always finds it optimal to induce the agent to contract with the principal and to deliver a strictly positive level of effort.

that when $\theta \leq \frac{1}{2}$, $\delta \geq 1$, and $f(\cdot)$ is as specified in equation (1), the first-order approach to solving $[P]$ is valid (Jewitt, 1988, Theorem 1). Under these conditions (which are maintained throughout the ensuing analysis, unless otherwise noted), the effort selection constraint (4) can be replaced by the following first-order condition for the agent's self-interested choice of a in deriving the solution to $[P]$:

$$\int_x \{U^A(w(x),a)\, f_a(x|a) \;+\; \partial U^A(.)/\partial a\; f(x|a)\}\; dx \;=\; 0. \tag{5}$$

3 The Fully Optimal Contract.

Using standard techniques, it is readily shown that the solution to [P] is characterized by the following four equations:

$$w(x) = \begin{cases} 0 & \text{if} \quad x < \widehat{x} \\ \left[2\theta\left(\lambda + \mu\left[\frac{f_a(x|a)}{f(x|a)}\right]\right)\right]^{\frac{1}{1-\theta}} & \text{if} \quad x \geq \widehat{x}\ ; \end{cases} \tag{6}$$

$$\int_{\widehat{x}}^{\infty} 2[w\,(x)]^{\theta}\; f(x|a)dx - a^{\delta} = \overline{U}\ ; \tag{7}$$

$$\int_{\widehat{x}}^{\infty} 2[w\,(x)]^{\theta}\, f_a(x|a)dx - \delta a^{\delta-1} = 0\ ; \quad \text{and} \tag{8}$$

$$\int_0^{\widehat{x}} x\; f_a(x|a)dx \;+\; \int_{\widehat{x}}^{\infty} [x - w(x)]\;\; f_a(x|a)dx$$
$$+\,\mu\;\left[\int_{\widehat{x}}^{\infty} 2[w\,(x)]^{\theta}\, f_{aa}(x|a)dx - \delta\,(\delta-1)\,a^{\delta-2}\right] = 0\ , \tag{9}$$

$$\text{where} \quad \widehat{x} = \min\left\{x \geq 0 \mid \lambda + \mu\left[\frac{f_a(x|a)}{f(x|a)}\right] \geq 0\right\}. \tag{10}$$

For given values of p, θ, δ and $\overline{U}$, equations (6) - (9) can be solved numerically to characterize the fully optimal contract.[11]

[11] An analytic upper bound for the principal's expected payoff under the fully optimal contract can be identified when $\theta = \frac{1}{2}$. The upper bound in this case is $\frac{1}{2\delta}\left[a^* p[2\delta - 1] \;-\; \delta\lambda^*\overline{U}\right]$, where $\lambda^* = [(a^*)^{\delta} + \overline{U}]/2$ and a^* is the solution of $\delta^3\,(a^*)^{2\delta-1} + p[(a^*)^{\delta} + \overline{U}]\delta\,(a^*)^{\delta-1} - 2p^2 = 0$.

4 The Optimal Linear Contract.

The optimal linear contract is the solution to problem $[P-L]$, which is problem $[P]$ with the additional constraint that payments to the agent must be a piecewise linear function of output, i.e.,

$$w(x) = \begin{cases} 0 & \text{if} \quad x < x_0 \\ \beta_0 + [x - x_0]\beta_1 & \text{if} \quad x \geq x_0 \ , \end{cases} \tag{11}$$

where $x_0 \geq 0$, and where β_0 and β_1 are non-negative constants.[12] Thus, the agent receives the minimum possible payment (zero) if realized output falls below a specified threshold, x_0. The agent's payment for output in excess of x_0 increases linearly with output at the rate β_1. If β_0 is strictly positive, the payment to the agent increases discontinuously at x_0.[13]

Lemma 1. *Suppose* $(\widetilde{x}_0, \widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{a})$ *is a solution to* $[P-L]$*. Then* $\widetilde{x}_0 = 0$ *and/or* $\widetilde{\beta}_0 = 0$*. Furthermore,* $\widetilde{x}_0$ *and* $\widetilde{\beta}_0$ *can both be zero for at most a single value of* $\overline{U}$ *when the participation constraint (3) binds.*

Lemma 1 indicates that the payment to the agent will not increase discontinuously at any $x_0 > 0$. A zero payment for output below some level, $x_0 > 0$, coupled with strictly positive payment for all higher output realizations is inefficient. A discontinuous increase in payment of this nature imposes excessive risk on the agent relative to any increased incentive for expanded effort supply that this payment structure might provide. The principal can create commensurate incentives for expanded effort supply while imposing less risk on the agent by altering the range of the smallest outputs for which the agent receives no payment, and then providing payments that do not increase discontinuously at the upper bound of this modified range of outputs.

Lemma 1 also implies that the optimal linear contract will assume one of the three forms depicted in Figure 2. The contract shown in Figure 2a delivers the smallest possible payment ($w(\cdot) = 0$) to the agent over a nontrivial range of the smallest output realizations ($x \in [0, x_0)$). Payment then increases linearly with x for output in excess of x_0. In Figure 2b, $x_0 = 0$, and so the agent is paid 0 if and only if he produces no output. In Figure 2c,

[12] The total payment from the principal to the agent cannot be negative because the agent has no wealth. Therefore, to ensure $w(x) \geq 0$ at $x = x_0$, it must be the case that $\beta_0 \geq 0$. β_0 and/or β_1 must be strictly positive to induce the agent to deliver strictly positive effort.

[13] For expositional ease, we refer to contracts of the form specified in equation (11) as "linear contracts" even though the term "piecewise linear" might be more accurate. As we explain below, the potential discontinuity at x_0 is an important feature of the contracts that we analyze.

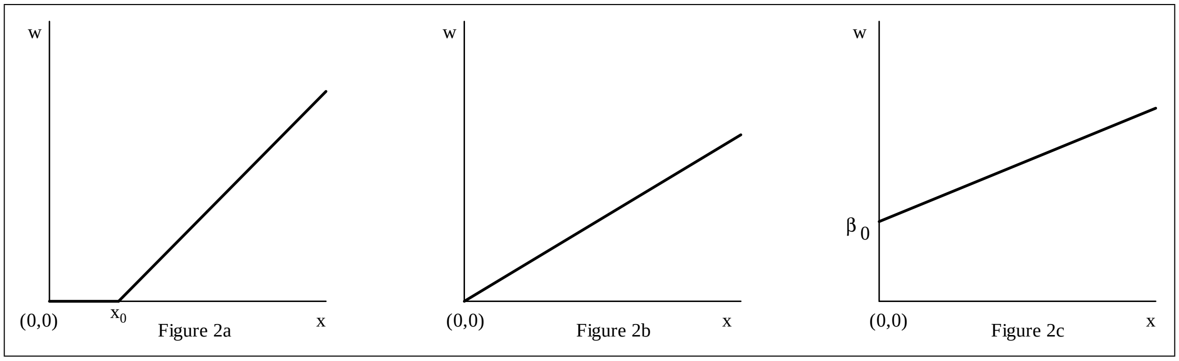

Figure 2: Possible Solutions to [P - L].

the agent's compensation is always strictly positive.

The linear contract illustrated in Figure 2a employs both a "stick" and a "carrot" to motivate the agent to work diligently. The stick takes the form of no payment for the smallest outputs ($x \in [0, x_0)$). The carrot is the fraction (β_1) of the output above x_0 that is awarded to the agent. Increased effort reduces the likelihood of incurring the stick and increases the likelihood of receiving the carrot.

Of course, extensive use of the stick under a contract like the one in Figure 2a may fail to provide the compensation required to ensure the agent's participation. Therefore, as the agent's reservation utility level ($\overline{U}$) increases, the principal may be compelled to limit her use of the stick by reducing the range of output realizations for which the agent receives no payment. For some intermediate reservation utility level, the principal will optimally deliver no payment to the agent only when he produces no output, as in Figure 2b. For larger values of $\overline{U}$, the principal will be forced to forego use of the stick altogether, and deliver strictly positive payment to the agent for all output realizations, as in Figure 2c.

Theorem 1 employs Lemma 1 to characterize optimal linear contracts by providing explicit solutions to $[P - L]$. The explicit solutions are challenging to derive because, in contrast to the standard approach used to solve $[P]$, pointwise optimization cannot be employed to solve $[P - L]$. The solutions presented in Theorem 1 are useful in part because they permit a complete characterization of optimal linear contracts through the use numerical solutions rather than programming-intensive simulations. Theorem 1 employs the following definitions:

$$f\left(\frac{x}{a}\right) = a\, f\,(x|a)\,. \qquad \alpha(c) = \int_c^\infty [y-c]^\theta\, f\,(y)\; dy. \qquad g(c) = \left[\frac{2\theta}{\delta}\right] \int_c^\infty [y-c]^{\theta-1} y\, f\,(y)\; dy.$$

$$g_2(c) = \int_c^{\infty} [y{-}c]\, f(y)\ dy. \qquad f_1(c_3) = \frac{2\alpha(c_3) - g(c_3)}{g(c_3)}. \qquad f_2(c_3) = \left[\frac{\delta g_2(c_3)}{p\theta [g(c_3)]^{\frac{1}{\theta}}}\right]^{\frac{\delta\theta}{\delta-\theta}}.$$

$\widehat{c}_3$ is the unique solution of $2\alpha(c) = g(c)$.

c_3^* is the value of c_3 at which $\sup_{c_3 \in [0,\ \widehat{c}_3]} \left[\frac{1}{f_2(c_3)}\right]^{\frac{1}{\delta}}$ is attained.

$$J(c_3) = \frac{f_2(c_3)}{f_1(c_3)}. \qquad \overline{U}_3 = \sup_{c_3\ <\ \widehat{c}_3} \frac{1}{J(c_3)}. \qquad S(x, c_3) = p\left[1 - \frac{\theta x^{\frac{\delta-\theta}{\delta\theta}}}{\delta}\right]\left[\frac{x}{f_2(c_3)}\right]^{\frac{1}{\delta}}.$$

$$A = \{c_3 \mid \overline{U} J(c_3) > 1\} \cap [0,\ \widehat{c}_3]. \qquad A^c = \{c_3 \mid \overline{U} J(c_3) \leq 1\} \cap [0,\ \widehat{c}_3].$$

$$\pi^{LA} = \max_{c_3 \in A} \{S(\overline{U} J(c_3),\ c_3)\}. \qquad \pi^{LA^c} = \max_{c_3 \in A^c} \{S(1,\ c_3)\}. \qquad \pi^{LM} = \max\ \{\pi^{LA},\ \pi^{LA^c}\}.$$

c_3^{**} is the value of c_3 at which the maximum value of π^{LM} is attained.

$$c_1(\alpha) = \int_0^{\infty} [\alpha{+}y]^{\theta} f(y)\ dy. \qquad c_2(\alpha) = \int_0^{\infty} y[\alpha{+}y]^{\theta-1} f(y)\ dy.$$

$$s(\alpha_1) = \frac{\theta c_2(\alpha_1)}{\delta c_1(\alpha_1) - \theta c_2(\alpha_1)}. \qquad \psi(\alpha_1) = \left[\frac{p + \alpha_1}{p}\right]\left[\frac{\delta}{2\theta c_2(\alpha_1)}\right]^{\frac{1}{\theta}}.$$

$$L(\overline{U}, \alpha_1) = p\left\{\left[\overline{U} s(\alpha_1)\right]^{\frac{1}{\delta}} - \psi(\alpha_1)\left[\overline{U} s(\alpha_1)\right]^{\frac{1}{\theta}}\right\}. \qquad \overline{U}_2 = \inf\ \{\overline{U} \mid \sup_{\alpha_1} L(\overline{U},\ \alpha_1) > L(\overline{U},\ 0)\}.$$

Theorem 1. *The solution to [P − L] has the following properties:*

(i) If $\overline{U} = 0$*, then* $\beta_0 = 0$*,* $x_0 = ac_3^*$*,* $\beta_1 = \left[\frac{a^{\delta-\theta}}{g(c_3^*)}\right]^{\frac{1}{\theta}}$*,* $a = \left[\frac{1}{f_2(c_3^*)}\right]^{\frac{1}{\delta}}$*, and* $\pi^L = ap\left[\frac{\delta-\theta}{\delta}\right]$*. The participation constraint binds in this case if and only if* $c_3^* = \widehat{c}_3$*.*

(ii) If $\overline{U} < \overline{U}_2$*, then* $\beta_0 = 0$*,* $x_0 = ac_3^{**}$*, and* $\beta_1 = \left[\frac{a^{\delta-\theta}}{g(c_3^{**})}\right]^{\frac{1}{\theta}}$*, where: (a)* $a = \left[\frac{\overline{U}}{f_1(c_3^{**})}\right]^{\frac{1}{\delta}}$*,* $\pi^L = \pi^{LA}$*, and the participation constraint binds if* $\pi^{LA} \geq \pi^{LA^c}$*; and (b)* $a = \left[\frac{1}{f_2(c_3^{**})}\right]^{\frac{1}{\delta}}$*,* $\pi^L = \pi^{LA^c}$*, and the participation constraint does not bind if* $\pi^{LA^c} > \pi^{LA}$*.*

(iii) If $\overline{U} \geq \overline{U}_2$*, then: (a)* β_0*,* x_0*,* β_1*,* a*, and* π^L *are as specified in (ii.a) above if* $\pi^{LM} \geq \widehat{\pi}^L$*; and (b)* $\beta_0 = \alpha\alpha_1$*,* $x_0 = 0$*,* $\beta_1 = \frac{\alpha}{a}$*,* $a = \widehat{a}$*,* $\pi^L = \widehat{\pi}^L$*, and the participation constraint binds if* $\widehat{\pi}^L > \pi^{LM}$*, where* $\widehat{a} = \left[\overline{U} s(\alpha_1)\right]^{\frac{1}{\delta}}$*,* $\widehat{\pi}^L = [p(1-\beta_1)]\,\widehat{a}$*,* $\alpha_1 = \arg\max_{\widetilde{\alpha}_1} L\left(\overline{U}, \widetilde{\alpha}_1\right)$*, and* $\alpha = \left[\frac{\delta \widehat{a}^{\delta}}{2\theta c_2(\alpha_1)}\right]^{\frac{1}{\theta}}$*.*

(iv) If $\overline{U} > \max\ \{\overline{U}_2,\ \overline{U}_3\}$*, then* β_0*,* x_0*,* β_1*,* a*, and* π^L *are as specified in (iii.b) above if*

$\widehat{\pi}^L \geq \pi^{LA}$, and as specified in (ii.a) above if $\pi^{LA} > \widehat{\pi}^L$. The participation constraint always binds in this case.

5 The Relative Performance of Linear Contracts.

Having identified some key features of both the fully optimal contract and the optimal linear contract, it remains to compare the performances of the two types of contracts. We implement this comparison by examining $R \equiv \frac{100[\pi - \pi^L]}{\pi}$, which is the percentage reduction in the principal's expected payoff from employing the optimal linear contract rather than the fully optimal contract.[14] This percentage measure of the loss from employing a linear contract has the advantage of being robust to units of measurement. The percentage measure can be misleading, though, when the absolute level of profit under the fully optimal contract approaches zero. In this case, the percentage loss from employing a linear contract can be quite large even though the absolute loss is very small.[15] To avoid such issues, we focus the ensuing analysis on settings in which the principal's expected profit under the fully optimal contract (π) is at least one-tenth of the agent's reservation utility level.[16] In addition to ruling out uninteresting reasons for a large percentage loss from employing a linear contract, this focus seems consistent with the structure of agency relationships. In a setting where one party – the principal – designs the contract and has the power to make a take-it-or-leave-it offer to the other party – the agent – it seems natural to presume that the principal typically would secure a substantial portion of the expected surplus from the agency relationship.[17]

We employ as a benchmark for the ensuing discussion the setting in which $\theta = 0.5$, $\delta = 2$, $\overline{U} = 0$, and $p = 4$. In this setting, the utility the agent derives from wealth w is $2\sqrt{w}$, the agent's (quadratic) cost of supplying effort a is a^2, and the agent's reservation utility level is 0. In addition, the mean of the output distribution is $4a$ when the agent supplies effort a.

Figure 3 illustrates both the fully optimal contract and the optimal linear contract in this benchmark setting. Under the fully optimal contract, the payment to the agent is an

[14] Recall that π denotes the principal's expected payoff at the solution to $[P]$ and π^L denotes the principal's expected payoff at the solution to $[P - L]$.

[15] Notice, for example, that R will be 100 when π^L is zero while π is positive but arbitrarily close to zero.

[16] It can be shown that $\pi > 0$ when $\overline{U} = 0$ for all parameter values under consideration. Therefore, π always exceds one-tenth of $\overline{U}$ when $\overline{U}$ is zero.

[17] If attention were restricted to settings where the principal secures a larger (and arguably more realistic minimum) share of the expected surplus, the maximum percentage losses from employing a linear contract would be smaller than those reported below. In contrast, in settings where $\pi/\overline{U} < 0.1$ so that the agency relationship admits little more than $\overline{U}$ in surplus, the percentage losses from employing linear contracts can exceed the losses reported below.

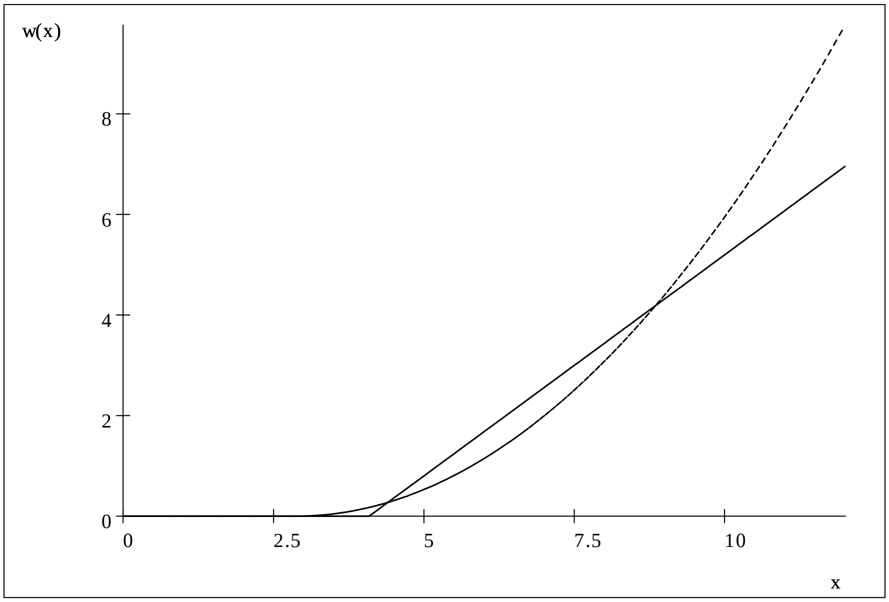

Figure 3: The Fully Optimal and Optimal Linear Contracts (when $p = 4$, $\theta = 0.5$, $\delta = 2$, and $\overline{U} = 0$.)

increasing, convex function of his performance. The payment to the agent is strictly positive for all $x \geq 2.875$. The agent delivers effort 1.26 under this contract, and the principal's expected profit is 3.77. Under the optimal linear contract, the agent receives no payment if $x < 4.09$, but receives 87.9% of the output above 4.09. The agent delivers effort 1.24 under this linear contract, and the principal's expected profit is 3.73. Thus, the optimal contract secures for the principal almost 99% of the expected profit she achieves under the fully optimal contract in this benchmark setting.

As is apparent from Figure 3, payments under the optimal linear contract in this benchmark setting diverge substantially from the corresponding payments under the fully optimal contract over broad ranges of output – particularly the highest output levels. Nevertheless, the principal's loss from employing a linear contract is small in this setting. This is the case because the outputs for which the payments diverge substantially under the two contracts are unlikely to occur in equilibrium. To illustrate, the equilibrium probability of an output in excess of 12.0 is less than 0.015 under both contracts. The strong performance of the optimal linear contract in this benchmark setting (and more generally) stems from its ability to approximate the fully optimal contract over the range of outputs that are most likely to occur in equilibrium, rather than over the entire range of outputs. By tailoring the threshold performance level (x_0) and the agent's share of output (β_1) for performance above the threshold to the prevailing environment, the principal often is able to replicate closely the

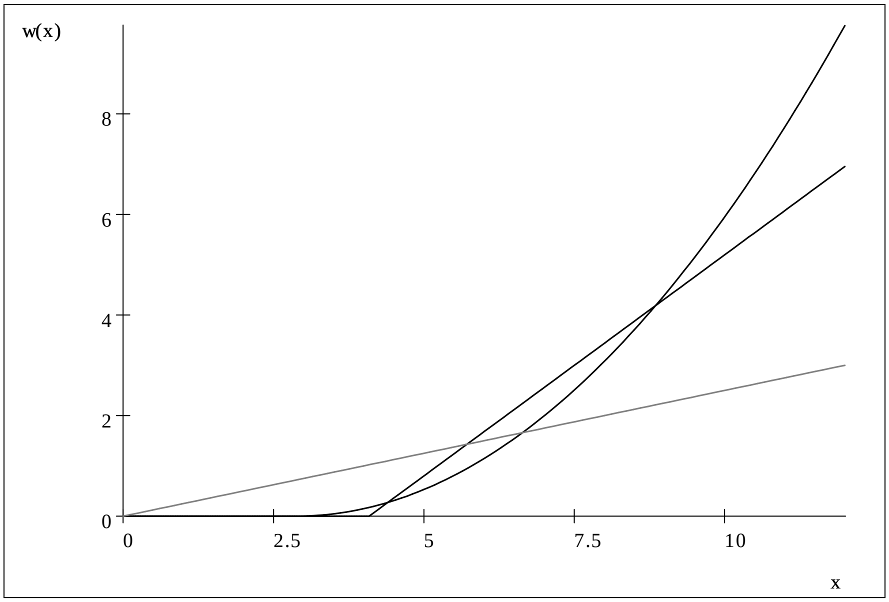

Figure 4: The Fully Optimal, Optimal (Pure) Linear, and Optimal (Piecewise) Linear Contracts (when $p = 4$, $\theta = 0.5$, $\delta = 2$, and $\overline{U} = 0$).

incentives that arise under the fully optimal contract.

The principal's ability to implement both a carrot and an effective stick is of central importance in the benchmark setting and more generally. Absent this ability, linear contracts often are unable to approximate closely the incentives created by fully optimal contracts. To illustrate this more general conclusion, suppose x_0 is constrained to be 0 in the benchmark setting where $\overline{U} = 0$, $\delta = 2$, $\theta = 0.5$, and $p = 4$. It can be shown that the principal optimally awards the agent 25% of the output he produces in this setting, as illustrated in Figure 4. The agent supplies 0.617 units of effort under this reward structure, which generates an expected profit of 1.851 for the principal. This expected profit is less than one half ($1.851/3.766 = .492$) of the principal's expected profit under the fully optimal contract in this benchmark setting.

Figure 4 helps to understand the relatively poor performance of a "pure" linear contract (i.e., a linear contract in which x_0 is constrained to be 0). When x_0 is constrained to be 0, the principal can only select the constant share of output that she delivers to the agent for all output realizations. A single straight line often cannot closely approximate a strictly convex curve like the one depicted in Figure 4 over an extended region. When x_0 is not constrained to be 0, the principal is able to select both the range of outputs over which the stick will be applied ($[0, x_0)$) and the size of the carrot that is delivered selectively, when output is sufficiently large. The ability to select both the length of a lower segment and the slope of an upper segment of a piecewise linear reward structure often increases substantially the principal's ability to approximate an optimal convex reward structure, as it does in the benchmark setting.[18]

Tables 1 through 4 provide additional information about how the principal tailors the carrot and the stick in the optimal linear contract to the prevailing economic environment. Table 1 illustrates how the optimal linear contract and its performance change as the agent's reservation utility ($\overline{U}$) increases. As $\overline{U}$ increases, the principal's expected profit declines under both the fully optimal contract (π) and the optimal linear contract $\left(\pi^L\right)$. As $\overline{U}$ (and thus the agent's equilibrium expected utility) increases, the range over which the agent receives strictly positive payment ((x_0, ∞)) increases.

Initially, as $\overline{U}$ increases from 0 to 1.0, the slope of the increasing portion of the linear contract (β_1) declines as the range of strictly positive payments ((x_0, ∞)) increases. As $\overline{U}$ increases between 1.0 and 2.5, this slope and the range of positive payments both increase. As $\overline{U}$ increases above 2.5, the payment the agent receives for zero output (β_0) increases above zero.[19] As the principal delivers more generous compensation to the agent in response to his increasing reservation utility, the agent's marginal utility of income declines. Consequently, it becomes more costly for the principal to induce expanded effort from the agent, and so the agent's equilibrium effort under the optimal linear contract (a^L, as reported in the sixth column of Table 1) declines as $\overline{U}$ increases, just as the agent's effort under the fully optimal contract (a^*, as reported in the fifth column of Table 1) declines.

The last three columns in Table 1 reveal that the identified changes in the optimal linear contract allow the principal to approximate the incentives of the fully optimal contract closely and consistently as $\overline{U}$ increases. As $\overline{U}$ increases from 0 to 2.7 (the value of $\overline{U}$ at which $\pi/\overline{U}$ declines to 0.1), the principal's loss from employing a linear contract is always less than 2% of her expected profit under a fully optimal contract. This loss increases from 1.1% to 1.3% as $\overline{U}$ increases from 0 to 1.2, and then declines from 1.3% to 0.9% as $\overline{U}$ increases from

[18] As Table 1 suggests, the principal typically is compelled to deliver greater compensation to the agent as $\overline{U}$ increases, even for the smaller output realizations. Consequently, the principal's gain from being able to employ both a stick and a carrot can be limited (or even zero) when the agent's reservation utility is large.

[19] As $\overline{U}$ increases above 2.7, the principal is compelled to deliver such generous compensation to the agent that the principal's expected profit even under the fully optimal contract (π) is less than one-tenth of $\overline{U}$. The principal's expected profit under both the fully optimal contract and the optimal linear contract declines rapidly toward zero as $\overline{U}$ increases above 2.7. For example, when $\overline{U} = 2.895$, π is .00448 and π^L is .00121.

1.2 to 2.5. The principal's loss from employing a linear contract increases above 0.9% as $\overline{U}$ increases above 2.5, as the rising $\overline{U}$ causes both π and π^L to decline toward zero.

Table 2 reveals how the optimal linear contract and its performance change as the agent's aversion to effort (δ) changes (holding $\overline{U}$ at 0, p at 4, and θ at 0.5). For $\delta \in [1.0, 2.0]$, the principal induces the agent to deliver more than one unit of effort (i.e., $a^* > 1$ and $a^L > 1$). Consequently, effort becomes more costly for the agent to deliver as δ increases in this range. The principal optimally induces less effort from the agent as effort becomes more onerous for him to supply. The principal induces less effort under the optimal linear contract by employing the stick less extensively, i.e., by reducing the range of output realizations for which the agent receives no payment.

As δ increases above 2.5, the principal induces the agent to supply less than one unit of effort under the optimal linear contract. When a^L is less than 1, the disutility the agent incurs from delivering effort a^L declines as δ increases. Consequently, for δ sufficiently large, the principal optimally induces more effort from the agent as δ increases.

The variations in x_0 and β_1 that the principal implements as δ changes enable the principal to closely approximate the incentives of the fully optimal contract in the settings analyzed in Table 2. As the last three columns in the table reveal, the principal's expected profit under both the optimal linear contract and the fully optimal contract initially declines and subsequently increases as δ increases. The profit variation is always similar under the two contracts, though, rendering the loss from a linear contract systematically below 2% for all values of δ. The loss declines below 1% as δ increases above 2.

Table 3 reveals how the optimal linear contract and its performance change as the agent's relative risk aversion changes (holding $\overline{U}$ at 0, p at 4, and δ at 2). As θ declines, the agent's degree of relative risk aversion increases. The principal responds to this increased risk aversion by implementing a payment structure that is less risky in the sense that the maximum penalty ($w(\cdot) = 0$) is imposed over a smaller range (i.e., x_0 declines) and payments increase less rapidly with performance (i.e., β_1 declines) in the range where payments are strictly positive. The agent delivers less effort under this less risky payment structure (i.e., a^L declines). The agent's effort (a^*) also declines under the fully optimal contract as θ declines, though, and the principal's profit under the optimal linear contract is always at least 98% of the principal's profit under the fully optimal contract as θ varies between 0.10 and 0.95.

Table 4 reveals how the optimal linear contract and its performance change as the productivity of the agent's effort (p) changes (holding $\overline{U}$ at 0, θ at 0.5, and δ at 2). As p increases, the productivity of the agent's effort increases, and so the principal induces greater effort from the agent under both the fully optimal contract and the optimal linear contract. The increased effort is secured under the optimal linear contract in large part via expanded use of the stick, i.e., by expanding the range of outputs ($[0, x_0)$) in which the agent receives no payment. When the productivity of the agent's effort is limited ($p < 1$), the smallest output realizations are relatively likely, even when the agent works diligently. (As Figure 1 suggests, $\lim_{x \to 0} f(x|a) = \infty$ and the density of the output realizations is everywhere declining when $p < 1$.) Consequently, the stick is not a particularly effective instrument for the principal when p is small, and so she optimally increases her use of the (costly) carrot (increasing β_1) to induce greater effort from the agent as p increases. In contrast, when the productivity of the agent's effort is more pronounced ($p > 1$), the lowest output realizations become fairly unlikely when the agent works diligently. Consequently, as p increases, the principal motivates the agent to work diligently by increasing her use of the stick (i.e., increasing x_0) substantially and reducing her use of the costly carrot (i.e., reducing β_1).

This increased use of the stick and reduced use of the carrot when p is large enables the principal to closely approximate the incentives that arise under the fully optimal contract. When p is at least 4, the principal is able to secure with a linear contract more than 99% of the profit she secures under the fully optimal contract in the benchmark setting. The optimal linear contract performs less well as the productivity of the agent's effort declines below 1. However, even when $p = 0.10$, the principal can achieve with a linear contract more than 97% of the profit she enjoys under the fully optimal contract in this setting.

The data presented in Tables 1 – 4 illustrate how the optimal linear contract changes as key elements of the agency relationship change. The data also provide examples of the losses that can arise from employing linear contracts in the benchmark setting. Conclusions 1 and 2 proceed beyond these examples to examine the principal's losses from employing a linear contract more generally. The conclusions provide upper bounds on these losses in the setting of primary interest where the productivity of the agent's effort is sufficiently pronounced (i.e., where $p \geq 1$) that the lowest output realizations are not particularly likely in equilibrium. The losses were calculated by allowing p to vary between 1 and 100, δ to vary between 1 and 20, θ to vary between 0.1 and 0.5, and $\overline{U}$ to vary between 0.0 and $\widetilde{U}$, where $\widetilde{U}$ is such that $\frac{\pi}{\widetilde{U}} = 0.1$. Initially, p was varied in increments of 0.1 for $p \in [1, 2]$, and then, p was taken to be 4, 10, 20, 40 and 100. Initially, δ was varied in increments of 0.1 for $\delta \in [1, 2.5]$ and then, δ was taken to be 5, 10 and 20. θ was varied in increments of 0.1 for $\theta \in [0.1, 0.5]$ and $\overline{U}$ was varied in increments of 0.1 for $\overline{U} \in \left[0.0, \widetilde{U}\right]$. Smaller increments were employed in all instances where the patterns produced by the larger increments were not entirely systematic.[20]

Conclusion 1. $R \leq 9.98\%$ *when* $p \geq 1$.

Conclusion 1 reflects the numerical results derived from the formulae presented in sections 3 and 4. The conclusion reports that when the productivity of the agent's effort is not too meager (so $p \geq 1$), the optimal linear contract always secures for the principal more than 90% of the expected profit that she secures under a fully optimal contract. When $p \geq 1$, the lowest output realizations are not particularly likely when the agent works diligently. Consequently, the principal can deter shirking by threatening to deliver no payment to the agent when the smallest output realizations are observed. Thus, the combination of an effective stick and a carrot enables the principal to approximate fairly closely the incentives created by the fully optimal contract in this setting.

The numerical results that underlie Conclusion 1 reveal that the percentage loss from employing a linear contract achieves its upper bound ($R = 9.98\%$) when p is 1, $\overline{U}$ is 0.7241, δ is 1, and θ is 0.5. $\overline{U}$ is large relative to the total potential surplus from the agency relationship at these parameter values. Consequently, the principal is compelled to cede most of the available surplus to the agent, even under the fully optimal contract. In particular, $\pi/\overline{U} = 0.1$, so the principal's expected profit under the fully optimal contract is the smallest permissible fraction of the agent's reservation utility. In this case, small absolute reductions in the principal's expected profit under the optimal linear contract translate into relatively large percentage reductions.

The finding that the principal's loss from employing a linear contract is most pronounced when the productivity of the agent's effort is small ($p = 1$) is not surprising. As explained above, the principal's ability to employ the stick to motivate the agent declines as p becomes small, and the principal suffers when she is forced to rely heavily on a single instrument (the costly carrot, β_1) to motivate the agent.

Notice that the maximum percentage loss identified in Conclusion 1 occurs when the agent's aversion to effort is least pronounced ($\delta = 1$). In this case, the total potential surplus from the agency relationship is most pronounced, and the inability to fine-tune the reward structure to induce the desired effort from the agent is relatively constraining.

The finding that the maximum loss from employing a linear contract arises for an inter-

[20] Numerical solutions indicate that the principal's losses from employing a linear contract decline systematically as p increases, *ceteris paribus*. We conjecture that this conclusion is true in general.

mediate level of relative risk aversion ($\theta = 0.5$) is consistent with the results presented in Table 3. When the agent's relative risk aversion is pronounced (so θ is close to 0), the agent is induced to deliver little effort even under the fully optimal contract. Therefore, limited ability to fine-tune the reward structure is not particularly constraining for the principal. Also, when the agent is not very averse to risk (so θ is close to 1), the fully optimal contract imposes substantial risk on the agent. The optimal risk structure can be approximated relatively closely with a (piecewise) linear contract that couples a fairly pronounced stick (a relatively large x_0) with a sizable carrot (a relatively high β_1).

Conclusion 2 provides an indication of the extent to which the relative performance of the optimal linear contract improves as the productivity of the agent's effort increases.

Conclusion 2. $R \leq 5.8\%$ *when* $p \geq 4$. $R \leq 2.98\%$ *when* $p \geq 10$.

As p increases above 1, the equilibrium density of outputs becomes less highly concentrated near zero. Consequently, the principal enjoys greater flexibility to employ both a stick (zero payment for $x \in [0, x_0)$) and a carrot to induce the agent to work diligently. This flexibility increases as p increases. As Conclusion 2 reports, when p is at least 4, the principal can always secure more than 94% of the expected profit that she can secure under a fully optimal contract. The corresponding percentage exceeds 97% when p is at least 10.

Linear contracts perform less well when the productivity of the agent's effort is smaller ($p < 1$), and so the equilibrium density of output is highly concentrated near zero. To illustrate, when p is 0.1, the principal can only be certain to secure with an optimal linear contract at least 82% of the expected profit that she secures with a fully optimal contract in the settings under consideration. When the productivity of the agent's effort is this limited, the pronounced likelihood of the smallest output realizations limits the principal's ability to employ both a stick and a carrot to motivate the agent to work diligently. Nevertheless, even under these arguably stringent conditions, the principal always secures with a linear contract a sizable fraction of the expected profit that she secures with a fully optimal contract.

6 Conclusions.

In an attempt to help understand why linear contracts are prevalent in practice, we have examined the ability of linear contracts to replicate the performance of fully optimal contracts in a broad class of economic environments. We have identified plausible conditions under which the principal can always secure with a linear contract more than 90% of the expected profit that she can secure with a fully optimal contract. Linear contracts perform less well when the agent's effort is relatively unproductive (i.e., when the agent's expected output given effort a is pa, where $p < 1$). Even in this case, though, the principal is often able to secure with a linear contract a sizable fraction (e.g., more than 82% when $p = 0.1$) of the expected profit that she secures with a fully optimal contract.

The linear contracts that we analyzed achieved this strong performance in part because they included a performance threshold (x_0) below which the agent received no payment. The combination of no payment for the smallest output realizations and a constant share of the realized output above a threshold admitted a piecewise linear reward structure that often approximated the optimal reward structure quite closely for the key output realizations. The approximation can be much less accurate under pure linear contracts, which deliver a constant share of output to the agent for all output realizations. Thus, a "stick" (in the form of no payment for the smallest outputs) often serves as a highly productive complement to the carrot (i.e., the agent's share of output for the larger output realizations) under linear reward structures. This observation suggests the practical importance of being able to dismiss an agent without pay in settings where simple (piecewise linear) contracts are employed.

Although our analysis considered a broad class of economic environments, the class was nevertheless structured. We have considered alternative environments in order to assess the robustness of our findings. For example, we have calculated the losses that arise from the use of linear contracts in settings where the agent's degree of relative risk aversion is less pronounced (so $\theta > \frac{1}{2}$). These losses increase slightly as θ increases above $\frac{1}{2}$ in the benchmark setting (where $p = 4$, $\delta = 2$, and $\overline{U} = 0$) and then decline as θ increases further toward 1. In all cases that we considered, the losses from employing a linear contract were below the bound reported in Conclusion 1. The same is true in the settings that we examined with alternative conditional densities for x. In the case of a triangular density, no losses arise from the use of linear contracts.

Several avenues for research remain to be explored. Alternative functional forms merit consideration, for example, as do settings with multiple agents, a risk averse principal, and double moral hazard concerns. Future research also might examine the performance of other types of "simple" contracts.[21] The simple moral hazard setting considered here also might be extended to admit adverse selection concerns. For example, the agent might acquire superior information about the stochastic production technology before contracting with

[21] See Gjesdal (1988), Reichelstein (1992), and Zhou and Swan (2003), for example.

the principal.[22] The adverse selection literature has identified circumstances under which a small number of contracts and/or simple contracts can perform nearly as well as the complete menu of comparatively complex contracts that are fully optimal.[23] It would be useful to determine whether similar conclusions arise in the presence of adverse selection and moral hazard concerns with risk averse agents.[24]

[22] Zabojnik (1996) examines a moral hazard setting in which $x = ba + \varepsilon$, where the agent learns the value of b after signing a contract but before choosing his effort, a. Zabojnik finds that the slope of the optimal linear contract declines as the variance of the normally distributed random variable ε increases, but may increase as the variance of the random variable b increases.

[23] See, for example, Bower (1993), McAfee (2002), Rogerson (2003), and Chu and Sappington (2007).

[24] Guesnerie et al. (1989), for example, demonstrate that optimal contracts may consist of menus of linear contracts in settings with adverse selection, moral hazard, and risk neutral agents.

$\overline{U}$	β_0	x_0	β_1	a^*	a^L	π	π^L	$R \equiv 100\left[\frac{\pi-\pi^L}{\pi}\right]$
0.0	0.0	4.09	0.879	1.255	1.241	3.766	3.725	1.097%
0.2	0.0	3.61	0.795	1.231	1.217	3.609	3.569	1.115%
0.4	0.0	3.18	0.744	1.206	1.192	3.434	3.394	1.156%
0.6	0.0	2.79	0.715	1.180	1.167	3.243	3.204	1.201%
0.8	0.0	2.44	0.700	1.154	1.141	3.037	2.999	1.246%
1.0	0.0	2.12	0.696	1.128	1.114	2.816	2.780	1.283%
1.2	0.0	1.82	0.701	1.102	1.088	2.582	2.548	1.301%
1.4	0.0	1.54	0.712	1.076	1.062	2.333	2.303	1.290%
2.0	0.0	0.76	0.773	1.000	0.986	1.500	1.485	1.033%
2.5	0.0	0.1	0.835	0.939	0.932	0.703	0.697	0.870%
2.6	0.044	0.0	0.845	0.927	0.921	0.532	0.526	0.999%
2.7	0.179	0.0	0.854	0.915	0.910	0.357	0.352	1.236%

Table 1. The Effects of Varying $\overline{U}$ (when $\delta = 2$, $\theta = 0.5$, and $p = 4$).

δ	β_0	x_0	β_1	a^*	a^L	π	π^L	$R \equiv 100\left[\frac{\pi-\pi^L}{\pi}\right]$
1.0	0.0	10.57	0.850	6.400	6.280	12.800	12.560	1.872%
1.2	0.0	6.33	0.827	3.113	3.061	7.263	7.142	1.656%
1.4	0.0	5.01	0.821	2.110	2.079	5.425	5.345	1.477%
1.6	0.0	4.45	0.829	1.659	1.636	4.561	4.500	1.335%
1.8	0.0	4.19	0.849	1.410	1.393	4.074	4.025	1.196%
2.0	0.0	4.09	0.879	1.255	1.241	3.766	3.727	1.097%
2.5	0.0	4.10	0.997	1.046	1.036	3.347	3.317	0.892%
3.0	0.0	4.27	1.181	0.942	0.935	3.139	3.115	0.757%
3.5	0.0	4.22	1.171	0.882	0.876	3.022	3.003	0.644%
4.0	0.0	4.06	1.024	0.848	0.843	2.967	2.950	0.552%
4.5	0.0	3.97	0.911	0.828	0.824	2.944	2.930	0.483%
5.0	0.0	3.91	0.819	0.816	0.813	2.939	2.926	0.430%
10.0	0.0	3.91	0.410	0.814	0.813	3.094	3.088	0.204%
20.0	0.0	4.12	0.205	0.858	0.857	3.345	3.342	0.099%

Table 2. The Effects of Varying δ (when $\overline{U} = 0$, $\theta = 0.5$, and $p = 4$).

θ	β_0	x_0	β_1	a^*	a^L	π	π^L	$R \equiv 100\left[\frac{\pi-\pi^L}{\pi}\right]$
0.1	0.0	3.44	0.254	0.861	0.860	3.2701	3.2697	0.011%
0.2	0.0	3.49	0.476	0.903	0.903	3.252	3.250	0.057%
0.3	0.0	3.61	0.661	0.981	0.982	3.334	3.325	0.271%
0.4	0.0	3.80	0.777	1.095	1.088	3.503	3.481	0.616%
0.5	0.0	4.09	0.879	1.255	1.242	3.766	3.725	1.097%
0.6	0.0	4.51	0.956	1.479	1.456	4.142	4.076	1.583%
0.7	0.0	5.11	1.012	1.794	1.760	4.664	4.574	1.918%
0.8	0.0	5.95	1.051	2.245	2.205	5.388	5.290	1.804%
0.9	0.0	7.21	1.076	2.921	2.887	6.425	6.351	1.161%
0.95	0.0	8.07	1.084	3.393	3.371	7.125	7.080	0.632%

Table 3. The Effects of Varying θ (when $\overline{U} = 0$, $\delta = 2$, and $p = 4$).

p	β_0	x_0	β_1	a^*	a^L	π	π^L	$R \equiv 100\left[\frac{\pi-\pi^L}{\pi}\right]$
0.10	0.0	0.052	0.587	0.132	0.129	0.0099	0.0097	2.298%
0.25	0.0	0.140	0.709	0.238	0.233	0.0446	0.0437	1.935%
0.50	0.0	0.330	0.869	0.369	0.362	0.138	0.136	1.792%
0.75	0.0	0.565	1.003	0.475	0.467	0.267	0.263	1.667%
1.0	0.0	0.839	1.120	0.569	0.560	0.427	0.420	1.600%
2.0	0.0	1.77	0.981	0.863	0.851	1.294	1.276	1.370%
4.0	0.0	4.09	0.879	1.255	1.241	3.766	3.727	1.097%
10.0	0.0	13.74	0.820	1.925	1.912	14.440	14.341	0.688%
20.0	0.0	35.79	0.827	2.554	2.544	38.315	38.156	0.416%
50.0	0.0	128.56	0.882	3.590	3.585	134.652	134.454	0.147%

Table 4. The Effects of Varying p (when $\overline{U} = 0$, $\delta = 2$, and $\theta = 0.5$).

7 References.

Bhattacharyya, Sugato and Francine Lafontaine, "Double-Sided Moral Hazard and the Nature of Share Contracts," *Rand Journal of Economics,* 26(4), Winter 1995, 761-781.

Bower, Anthony, "Procurement Policy and Contracting Efficiency," *International Economic Review,* 34(4), November 1993, 873-901.

Chu, Leon Yang and David Sappington, "Simple Cost-Sharing Contracts," *American Economic Review*, 97(1), March 2007, 419-428.

Corbett, Charles, Gregory DeCroix, and Albert Ha, "Optimal Shared-Savings Contracts in Supply Chains: Linear Contracts and Double Moral Hazard," *European Journal of Operational Research*, 163(3), June 2005, 653-667.

Gasmi, Farid, Jean-Jacques Laffont, and William Sharkey, "Empirical Evaluation of Regulatory Regimes in Local Telecommunications Markets," *Journal of Economics & Management Strategy*, 8(1), Spring 1999, 61-93.

Gjesdal, Froystein, "Piecewise Linear Incentives Schemes," *Scandinavian Journal of Economics*, 90(3), September 1988, 305-328.

Guesnerie, Roger, Pierre Picard, and Patrick Rey, "Adverse Selection and Moral Hazard with Risk Neutral Agents", *European Economic Review*, 33(4), April 1989, 807-823.

Guo, Ming and Hui-Ou Yang, "Incentives and Performance in the Presence of Wealth Effects and Endogenous Risk," *Journal of Economic Theory*, 129(1), July 2006, 150-191.

Holmstrom, Bengt, "Moral Hazard and Observability," *Bell Journal of Economics*, 10(1), Spring 1979, 74-91.

Holmstrom, Bengt and Paul Milgrom, "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55(2), March 1987, 303-328.

Jewitt, Ian, "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56(5), September 1988, 1177-1190.

Johnson, Norman, Samuel Kotz and N. Balakrishnan. *Continuous Univariate Distributions, Volume 1.* New York: Wiley Interscience, New York, 1994.

Kim, Son Ku and Susheng Wang, "Linear Contracts and the Double Moral-Hazard," *Journal of Economic Theory*, 82(2), October 1998, 342-378.

Laffont, Jean-Jacques and Jean Tirole, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94(3), June 1986, 614-641.

McAfee, R. Preston, "Coarse Matching," *Econometrica*, 70(5), September 2002, 2025-2034.

Milgrom, Paul and John Roberts, *Economics, Organization & Management*. Englewood Cliffs, NJ: Prentice Hall, 1992.

Mirrlees, James, "The Theory of Moral Hazard and Unobservable Behavior: Part I," *Review of Economic Studies*, 66(1), January 1999, 3-22.

Pfeiffer, Thomas and Louis Velthius, "On the Optimality of Linear Contracts to Induce Goal-Congruent Investment Behaviour," *Applied Economics Letters*, 12(4), March 2005, 207-211.

Reichelstein, Stefan, "Constructing Incentive Schemes for Government Contracts: An Application of Agency Theory," *Accounting Review*, 67(4), October 1992, 712-731.

Rogerson, William, "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53(6), November 1985, 1357-1368.

Rogerson, William, "Simple Menus of Contracts in Cost-Based Procurement and Regulation," *American Economic Review*, 93(3), June 2003, 919-926.

Schmalensee, Richard, "Good Regulatory Regimes," *Rand Journal of Economics*, 20(3), Autumn 1989, 417-436.

Shavell, Steven, "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, 10(1), Spring 1979, 55-73.

Sung, Jaeyoung, "Linearity with Project Selection and Controllable Diffusion Rate in Continuous Time Principal-Agent Problems," *Rand Journal of Economics*, 26(4), Winter 1995, 720-743.

Zabojnik, Jan, "Pay-Performance Sensitivity and Production Uncertainty," *Economics Letters*, 53(3), December 1996, 291-296.

Zhou, Xianming and Peter Swan, "Performance Thresholds in Managerial Incentive Contracts," *Journal of Business*, 76(4), 2003, 665-696.

8 Appendix

Proof of Lemma 1.

The Lagrangian function corresponding to $[P-L]$ is:

$$\begin{aligned}
\pounds = & \int_0^{x_0} xf\left(x|a\right)dx + \int_{x_0}^{\infty}\left[x - w(x)\right]f\left(x|a\right)dx \\
& + \lambda\left[\int_{x_0}^{\infty} 2\left[w(x)\right]^{\theta} f\left(x|a\right)dx - a^{\delta} - \overline{U}\right] + \mu\left[\int_{x_0}^{\infty} 2\left[w(x)\right]^{\theta} f_a\left(x|a\right)dx - \delta a^{\delta-1}\right].
\end{aligned} \tag{A.1}$$

Define:

$$B = \int_{x_0}^{\infty} f\left(x|a\right)dx, \quad B_1 = \int_{x_0}^{\infty} 2\theta\left[w(x)\right]^{\theta-1} f\left(x|a\right)dx, \quad B_2 = \int_{x_0}^{\infty} 2\theta\left[w(x)\right]^{\theta-1} f_a\left(x|a\right)dx,$$

$$G = \beta_0 f\left(x_0|a\right) - 2\lambda[\beta_0]^{\theta} f\left(x_0|a\right) - 2\mu[\beta_0]^{\theta} f_a\left(x_0|a\right). \tag{A.2}$$

From (A.1) and (A.2):

$$\begin{aligned}
\frac{\partial \pounds}{\partial x_0} & = x_0 f\left(x_0|a\right) + \beta_1 \int_{x_0}^{\infty} f\left(x|a\right)dx - [x_0 - \beta_0]f\left(x_0|a\right) \\
& + \lambda\left[\int_{x_0}^{\infty} 2\theta\left[w(x)\right]^{\theta-1}[-\beta_1]f\left(x|a\right)dx - 2[\beta_0]^{\theta} f\left(x_0|a\right)\right] \\
& + \mu\left[\int_{x_0}^{\infty} 2\theta\left[w(x)\right]^{\theta-1}[-\beta_1]f_a\left(x|a\right)dx - 2]\beta_0]^{\theta} f_a\left(x_0|a\right)\right] \\
& = -\beta_1[-B + \lambda B_1 + \mu B_2] + G.
\end{aligned} \tag{A.3}$$

$$\begin{aligned}
\frac{\partial \pounds}{\partial \beta_0} & = -\int_{x_0}^{\infty} f\left(x|a\right)dx + \lambda\left[\int_{x_0}^{\infty} 2\theta\left[w(x)\right]^{\theta-1} f\left(x|a\right)dx\right] + \mu\left[\int_{x_0}^{\infty} 2\theta\left[w(x)\right]^{\theta-1} f_a\left(x|a\right)dx\right] \\
& = -B + \lambda B_1 + \mu B_2.
\end{aligned} \tag{A.4}$$

(A.3) and (A.4) imply:
$$\beta_1 \frac{\partial \pounds}{\partial \beta_0} = -\frac{\partial \pounds}{\partial x_0} + G. \tag{A.5}$$

Suppose $x_0 > 0$ and $\beta_0 > 0$ is a local maximum. Then $\frac{\partial \pounds}{\partial \beta_0} = \frac{\partial \pounds}{\partial x_0} = 0$ at this point, and so $G = 0$. Because $\widetilde{\beta}_0 > 0$, by assumption:

$$f\left(\widetilde{x}_0|\widetilde{a}\right) - 2\lambda\left[\widetilde{\beta}_0^{\theta-1}\right] f\left(\widetilde{x}_0|\widetilde{a}\right) - 2\mu\left[\widetilde{\beta}_0\right]^{\theta-1} f_a\left(\widetilde{x}_0|\widetilde{a}\right) = 0. \tag{A.6}$$

Hence:
$$\lambda + \mu\left[\frac{f_a\left(\widetilde{x}_0|a\right)}{f\left(\widetilde{x}_0|a\right)}\right] > 0 \text{ and } \widetilde{\beta}_0 = \left[2\left(\lambda + \mu\frac{f_a\left(\widetilde{x}_0|\widetilde{a}\right)}{f\left(\widetilde{x}_0|\widetilde{a}\right)}\right)\right]^{1-\theta}. \tag{A.7}$$

$(A.2)$ and $(A.6)$ imply that $\beta_0 = 0$ and $\beta_0 = \widetilde{\beta}_0 > 0$ are two solutions to $\frac{\partial \pounds}{\partial \beta_0} = 0$. It is readily verified that the values of λ and μ are unique at the optimum, so there are exactly two possible values of β_0.

For $\widetilde{\beta}_0$ to be a global maximum, we must have $\frac{\partial \pounds}{\partial \beta_0} > 0$ for all $\beta_0 \in (0, \widetilde{\beta}_0)$. Fix any such β_0, and let $x_0 = \widetilde{x}_0$. Then we know that $G \neq 0$. From (A.5), $\widetilde{\beta}_1 \neq 0$ and

$$\begin{aligned} \frac{\partial \pounds}{\partial \beta_0} &= \frac{\beta_0}{\widetilde{\beta}_1} \left[f\left(\widetilde{x}_0|\widetilde{a}\right) - 2\lambda \left(\beta_0\right)^{\theta-1} f\left(\widetilde{x}_0|\widetilde{a}\right) - 2\mu \left(\beta_0\right)^{\theta-1} f_a\left(\widetilde{x}_0|\widetilde{a}\right) \right] \\ &= \frac{\beta_0}{\widetilde{\beta}_1} f\left(\widetilde{x}_0|\widetilde{a}\right) \left[1 - 2\left(\beta_0\right)^{\theta-1} \left(\lambda + \mu \frac{f_a\left(\widetilde{x}_0|\widetilde{a}\right)}{f\left(\widetilde{x}_0|\widetilde{a}\right)} \right) \right]. \end{aligned} \tag{A.8}$$

$f\left(\widetilde{x}_0|\widetilde{a}\right) > 0$ because $\widetilde{x}_0 > 0$. Also, $(A.7)$ implies $\lambda + \mu \frac{f_a(\widetilde{x}_0|a)}{f(\widetilde{x}_0|a)} > 0$. Therefore, when β_0 is close to 0, $(A.8)$ implies $\frac{\partial \pounds}{\partial \beta_0} < 0$, which is a contradiction. Hence $\widetilde{x}_0 = 0$ and/or $\widetilde{\beta}_0 = 0$ at the solution to $[P - L]$.

If $\widetilde{x}_0 = 0$ and $\widetilde{\beta}_0 = 0$, then $\frac{\partial \pounds}{\partial x_0} \leq 0$ at $x_0 = 0$ and $\frac{\partial \pounds}{\partial \beta_0} \leq 0$ at $\beta_0 = 0$. If $\frac{\partial \pounds}{\partial x_0} < 0$ at $x_0 = 0$, then $(A.5)$ implies that $\frac{\partial \pounds}{\partial \beta_0} > 0$ at $\beta_0 = 0$, which is a contradiction. Similarly, if $\frac{\partial \pounds}{\partial \beta_0} < 0$ at $\beta_0 = 0$, then $(A.5)$ implies that $\frac{\partial \pounds}{\partial x_0} > 0$ at $x_0 = 0$, which is a contradiction. Therefore, it must be the case that $\frac{\partial \pounds}{\partial x_0} = 0$ at $x_0 = 0$, and $\frac{\partial \pounds}{\partial \beta_0} = 0$ at $\beta_0 = 0$. Consequently:

$$\begin{aligned} -\int_0^\infty f\left(x|\widetilde{a}\right) dx + \lambda \int_0^\infty 2\theta \left[x\widetilde{\beta}_1\right]^{\theta-1} f\left(x|\widetilde{a}\right) dx + \mu \left[\int_0^\infty 2\theta \left[x\widetilde{\beta}_1\right]^{\theta-1} f_a\left(x|\widetilde{a}\right) dx\right] &= 0; \text{ and} \\ -\int_0^\infty x f\left(x|\widetilde{a}\right) dx + \lambda \int_0^\infty 2\theta \left[x\widetilde{\beta}_1\right]^{\theta-1} x f\left(x|\widetilde{a}\right) dx + \mu \left[\int_0^\infty 2\theta \left[x\widetilde{\beta}_1\right]^{\theta-1} x f_a\left(x|\widetilde{a}\right) dx\right] &= 0. \end{aligned} \tag{A.9}$$

Further, $\frac{\partial \pounds}{\partial \mu} = 0$ yields:

$$2\int_0^\infty \left[x\beta_1\right]^\theta f_a\left(x|\widetilde{a}\right) dx = \delta a^{\delta-1}. \tag{A.10}$$

If the participation constraint is satisfied with equality, then:

$$\int_0^\infty 2\left[x\widetilde{\beta}_1\right]^\theta f\left(x|\widetilde{a}\right) dx - \widetilde{a}^\delta = \overline{U}. \tag{A.11}$$

It is readily verified that (A.9) - (A.11) can be solved uniquely for β_1, a, λ and μ, in terms of $\overline{U}$. (For example, β_1 can be eliminated from (A.10) and (A.11) to yield $a^\delta = \left[\frac{\theta}{\delta-\theta}\right]\overline{U}$.) It is also readily verified that the equation $\frac{\partial \pounds}{\partial a} = 0$ is satisfied by at most one value of $\overline{U}$.

The Optimal Linear Contract when x_0 is Exogenously Fixed at Zero

Define: $\quad c_1(\alpha) \equiv \frac{1}{\Gamma(p)} \int_0^\infty [\alpha{+}y]^\theta \, e^{-y} y^{p-1} \, dy, \quad$ and $\quad c_2(\alpha) \equiv \frac{1}{\Gamma(p)} \int_0^\infty y[\alpha{+}y]^{\theta-1} \, e^{-y} y^{p-1} \, dy.$

It is readily verified that:

$$c_1(0) = c_2(0) = \frac{\Gamma(p+\theta)}{\Gamma(p)} \quad \text{and} \quad c_1'(\alpha_1) > 0, \;\; c_2'(\alpha_1) < 0 \;\; \text{for all } \alpha_1 \geq 0. \tag{A.12}$$

The following re-parameterizations are helpful in the ensuing analysis:

$$\alpha = a\beta_1, \quad \alpha_1 = \frac{\beta_0}{a\beta_1}, \quad \text{so} \;\; \beta_0 = \alpha\alpha_1, \quad \text{for} \;\; \beta_1 > 0, \; a > 0, \; \alpha > 0. \tag{A.13}$$

Using (A.13) and the change of variable $\frac{x}{a} = y$, the incentive compatibility constraint can be expressed as:

$$-\left[\frac{2\alpha^\theta}{a}\right] \int_0^\infty [\alpha_1 + y]^\theta \; S'(y) \; dy = \delta a^{\delta-1} \quad \text{where} \;\; S(y) = \frac{e^{-y} y^p}{\Gamma(p)}. \tag{A.14}$$

Integration by parts provides:

$$\int_0^\infty [\alpha_1 + y]^\theta \; S'(y) \; dy \;\; = \;\; -\int_0^\infty \theta[\alpha_1 + y]^{\theta-1} \; S(y) \; dy. \tag{A.15}$$

Substituting (A.15) into (A.14), the incentive compatibility constraint reduces to:

$$2\theta\alpha^\theta c_2(\alpha_1) = \delta a^\delta \quad \Rightarrow \quad \alpha = \left[\frac{\delta a^\delta}{2\theta c_2(\alpha_1)}\right]^{\frac{1}{\theta}}. \tag{A.16}$$

Similarly, the participation constraint can be expressed as:

$$2\alpha^\theta c_1(\alpha_1) \geq a^\delta + \overline{U}. \tag{A.17}$$

Substituting α from (A.16) into (A.17) provides:

$$2\left[\frac{\delta a^\delta}{2\theta c_2(\alpha_1)}\right] c_1(\alpha_1) \geq a^\delta + \overline{U} \quad \text{or} \quad \delta a^\delta c_1(\alpha_1) \geq \theta c_2(\alpha_1) \left[a^\delta + \overline{U}\right]. \tag{A.18}$$

Using (A.13), the principal's payoff function L can be written as:

$$L^P = \int_0^\infty [x - \beta_0 - \beta_1 x] \, f(x|a) dx \;\; = \;\; ap - p\alpha - \alpha_1\alpha. \tag{A.19}$$

Substituting α from $(A.16)$ into (A.19) and simplifying provides:

$$L^P = p\left[a - \psi(\alpha_1) a^{\frac{\delta}{\theta}}\right] \quad \text{where} \quad \psi(\alpha_1) = \left[\frac{p+\alpha_1}{p}\right]\left[\frac{\delta}{2\theta c_2(\alpha_1)}\right]^{\frac{1}{\theta}}. \tag{A.20}$$

Therefore, $[P-L]$ can be written as:

$$\underset{\alpha_1,\ a}{\text{Maximize}}\ \ p\left[a - \psi(\alpha_1) a^{\frac{\delta}{\theta}}\right] \quad \text{subject to} \quad \delta a^\delta c_1(\alpha_1) \geq \theta c_2(\alpha_1)\left[a^\delta + \overline{U}\right]. \tag{A.21}$$

It is readily verified that for every fixed α_1, L^P is concave in a, and its unconstrained maximum with respect to a is obtained by solving $\frac{\partial L}{\partial a} = 0$. This gives:

$$1 - \psi(\alpha_1)\frac{\delta}{\theta} a^{\frac{\delta}{\theta}-1} = 0 \quad \Rightarrow \quad a = \left[\frac{\theta}{\delta \psi(\alpha_1)}\right]^{\frac{\theta}{\delta-\theta}} \equiv h(\alpha_1). \tag{A.22}$$

For fixed α_1, L^P is increasing for all $a \leq h(\alpha_1)$ and decreasing otherwise. Also, for any fixed a, L^P is a decreasing function of α_1.

Define $\quad \overline{U}_1 = \left[\frac{\delta-\theta}{\theta}\right]\left[\frac{1}{2c_2(0)}\left(\frac{\delta}{\theta}\right)^{1+\theta}\right]^{\frac{\delta}{\theta-\delta}}.$

Proposition 1. *If $x_0 = 0$ and $\overline{U} \in \left[0,\ \overline{U}_1\right)$, the agent's expected utility at the solution to $[P-L]$ is $\overline{U}_1$, and so the participation constraint does not bind. Furthermore:*

$$\beta_0 = 0, \quad \beta_1 = \frac{\theta}{\delta}, \quad a = \left[\frac{\theta \overline{U}_1}{\delta-\theta}\right]^{\frac{1}{\delta}}, \ \textit{and}\ \ \pi^L = p[1-\beta_1]\left[\frac{\theta \overline{U}_1}{\delta-\theta}\right]^{\frac{1}{\delta}} = p\left[1-\frac{\theta}{\delta}\right]\left[\frac{\theta \overline{U}_1}{\delta-\theta}\right]^{\frac{1}{\delta}}.$$

Proof. Ignoring the participation constraint (A.18), L^P is maximized at $\hat{\alpha}_1 = 0$ for any fixed a. In this case, the global optimum with respect to a occurs at $\hat{a} = h(0) = \left[\left(\frac{\theta}{\delta-\theta}\right)\overline{U}_1\right]^{\frac{1}{\delta}}$.

It remains to check that these values satisfy the participation constraint. Since $c_1(0) = c_2(0)$, it is readily verified that the participation constraint can be written as:

$$[\hat{a}]^\delta \geq \left[\frac{\theta}{\delta-\theta}\right]\overline{U}. \tag{A.23}$$

But (A.23) is equivalent to $\overline{U} \leq \overline{U}_1$. The Proposition then follows by solving for the relevant values of β_0, β_1, and a. □

Now, suppose $\overline{U} > \overline{U}_1$. For convenience, write L^P in (A.20) as $L(a^\delta, \alpha_1)$ so that:

$$L(a^\delta, \alpha_1) = p\left[a - \psi(\alpha_1)\, a^{\frac{\delta}{\theta}}\right].$$

Consider any pair $\left(a^\delta, \alpha_1\right)$ that satisfies the constraint in (A.21). This implies:

$$a^\delta \geq \frac{\overline{U}\theta c_2(\alpha_1)}{\delta c_1(\alpha_1) - \theta c_2(\alpha_1)} \equiv \overline{U}\, s(\alpha_1).$$

First suppose $\alpha_1 = 0$. In this case, $a^\delta \geq \overline{U}s(0) \geq \overline{U}_1 s(0) = [h(0)]^\delta$. However, $L(a^\delta, 0)$ is a decreasing function of a in this range. Thus, for $\alpha_1 = 0$, it is sufficient to consider $L(\overline{U}s(0), 0)$. On the other hand, for fixed a, L is a decreasing function of α_1. Further, $s(\alpha_1)$ is a decreasing function of α_1. Hence, to maximize L, it is sufficient to consider only those α_1 that satisfy either $a^\delta = \overline{U}s(\alpha_1)$ or $\alpha_1 = 0$. Together, these two observations imply that it is sufficient to consider $L(\overline{U}s(\alpha_1), \alpha_1)$.

Define: $\quad D(\overline{U}, \alpha_1) = L(\overline{U}s(\alpha_1), \alpha_1) - L(\overline{U}s(0), 0).$

Lemma 2. *(i) Suppose $D(\overline{U}, \alpha_1) < 0$. Then $D(\overline{U}_A, \alpha_1) < 0$ for $\overline{U}_A < \overline{U}$.*

(ii) For $\overline{U}$ sufficiently large, there exists an $\alpha_1 > 0$ such that $D(\overline{U}, \alpha_1) > 0$.

Proof. First note that

$$\begin{aligned} p^{-1}D(\overline{U}) &= (\overline{U})^{\frac{1}{\delta}}\left[(s(\alpha_1)^{\frac{1}{\delta}} - (s(0))^{\frac{1}{\delta}}\right] - (\overline{U})^{\frac{1}{\theta}}\left[\psi(\alpha_1)(s(\alpha_1))^{\frac{1}{\theta}} - \psi(0)(s(0))^{\frac{1}{\theta}}\right] < 0 \\ &\Leftrightarrow (\overline{U})^{\frac{1}{\delta}-\frac{1}{\theta}}\left[(s(\alpha_1))^{\frac{1}{\delta}} - (s(0))^{\frac{1}{\delta}}\right] < \left[\psi(\alpha_1)(s(\alpha_1))^{\frac{1}{\theta}} - \psi(0)(s(0))^{\frac{1}{\theta}}\right]. \qquad \text{(A.24)} \end{aligned}$$

Now note that $s(\alpha_1)$ is decreasing in α_1 and $\delta > \theta$. Therefore, the expression to the left of the inequality in (A.24) is increasing in $\overline{U}$, which proves part (i) of the lemma.

To prove part (ii), first note that for any fixed $\alpha_1 > 0$, the expression to the left of the inequality in (A.24) tends to zero as $\overline{U} \to \infty$. It is also readily verified that there exists $\alpha_1 > 0$ such that the expression to the right of the inequality in (A.24) is negative. Part (ii) of the lemma follows from these observations. □

Next, define $\overline{U}_2 = \inf\{\overline{U} \mid \sup_{\alpha_1} D(\overline{U}, \alpha_1) > 0\}$, and note that $\overline{U}_2 \geq \overline{U}_1$. The following proposition characterizes the solution to $[P-L]$ when $\overline{U} \geq \overline{U}_1$. The proposition follows directly from Lemma 2.

Proposition 2. *Suppose $x_0 = 0$. Then the participation constraint binds at the solution to $[P-L]$ for all $\overline{U}_1 \leq \overline{U}$. In addition:*
(i) For all $\overline{U}_1 \leq \overline{U} \leq \overline{U}_2$, $L(s(\alpha_1), \alpha_1)$ is maximized at $\alpha_1 = 0$. Furthermore, at the solution to $[P-L]$:

$$\beta_0 = 0, \quad a = \left[\frac{\theta \overline{U}}{\delta - \theta}\right]^{\frac{1}{\delta}}, \quad \beta_1 = \left[\frac{1}{\sigma c_2(0)}\left(\frac{\delta}{\theta}\right)\right]^{\frac{1}{\theta}} \left[\frac{\theta \overline{U}}{\delta - \theta}\right]^{\frac{\delta-\theta}{\theta\delta}} = \frac{\theta}{\delta}\left[\frac{\overline{U}}{\overline{U}_1}\right]^{\frac{\delta-\theta}{\theta\delta}}, \text{ and}$$

$$\pi^L = p\left[1 - \beta_1\right]\left[\frac{\theta \overline{U}}{\delta - \theta}\right]^{\frac{1}{\delta}} = p\left[1 - \frac{\theta}{\delta}\left(\frac{\overline{U}}{\overline{U}_1}\right)^{\frac{\delta-\theta}{\theta\delta}}\right]\left[\frac{\theta \overline{U}}{\delta - \theta}\right]^{\frac{1}{\delta}}.$$

Further, π^L is a decreasing function of $\overline{U}$ for all $\overline{U} \geq \overline{U}_1$.
(ii) For $\overline{U} > \overline{U}_2$, there exist $\widetilde{\alpha}_1 > 0$, $\widetilde{\alpha} > 0$, and $\widetilde{a} > 0$, that solve $[P-L]$ and the corresponding optimal outcomes are:

$$\beta_0 = \widetilde{\alpha}\widetilde{\alpha}_1, \quad \beta_1 = \frac{\widetilde{\alpha}}{a}, \quad a = \left[\frac{\overline{U}\theta c_2(\widetilde{\alpha}_1)}{\delta c_1(\widetilde{\alpha}_1) - \theta c_2(\widetilde{\alpha}_1)}\right]^{\frac{1}{\delta}}, \text{ and } \pi^L = p\left[1 - \left(\frac{p + \widetilde{\alpha}_1}{p}\right)\beta_1\right] a,$$

$$\text{where } \widetilde{\alpha}_1 = \arg\max_{\alpha_1} L\left(\overline{U}s(\alpha_1), \alpha_1\right), \quad \text{and } \widetilde{\alpha} = \left[\frac{\delta a^\delta}{\sigma\theta c_2(\widetilde{\alpha}_1)}\right]^{\frac{1}{\theta}}.$$

The Optimal Linear Contract when β_0 is Exogenously Fixed at Zero

When β_0 is constrained to be 0, $[P-L]$ can be written as:

$$\underset{x_0,\, a,\, \beta_1}{\text{Maximize}} \quad \widetilde{L} = \left[\int_0^{x_0} x f(x|a)\,dx + \int_{x_0}^{\infty} [x - (x - x_0)\beta_1] f(x|a)\,dx\right] \tag{A.25}$$

$$\text{subject to} \quad \int_{x_0}^{\infty} 2[w(x)]^\theta f(x|a)\,dx - a^\delta \geq \overline{U}, \quad \text{and} \quad \int_{x_0}^{\infty} 2[w(x)]^\theta f_a(x|a)\,dx - \delta a^{\delta-1} = 0. \tag{A.26}$$

Define:

$$\alpha(c) \equiv \int_c^{\infty} [y-c]^\theta f(y)\,dy, \quad g(c) \equiv \frac{2\theta}{\delta}\int_c^{\infty} [y-c]^{\theta-1} y f(y)\,dy, \quad g_2(c) \equiv \int_c^{\infty} [y-c] f(y)\,dy.$$

Using the change of variable, $y = \frac{x}{a}$ and letting $c_3 = \frac{x_0}{a}$, the participation constraint can be written as:

$$2a^\theta \beta_1^\theta \int_{c_3}^{\infty} [y - c_3]^\theta f(y)\,dy \geq a^\delta + \overline{U} \;\Leftrightarrow\; 2a^\theta \beta_1^\theta \alpha(c_3) \geq a^\delta + \overline{U} \;\Leftrightarrow\; [\beta_1]^\theta \geq \frac{a^\delta + \overline{U}}{2a^\theta \alpha(c_3)}. \tag{A.27}$$

Note that:

$$\frac{\partial f(x|a)}{\partial a} = f_a\left(x|a\right) = -\frac{1}{a^2} f\left(\frac{x}{a}\right) - \frac{x}{a^3} f'\left(\frac{x}{a}\right). \tag{A.28}$$

Substituting from $(A.28)$ into $(A.26)$, the incentive compatibility constraint can be written as:

$$\int_{x_0}^{\infty} 2[(x-x_0)\beta_1]^{\theta}\left[-\frac{1}{a^2} f\left(\frac{x}{a}\right) - \left[\frac{x}{a^3}\right] f'\left(\frac{x}{a}\right)\right] dx = \delta a^{\delta-1}. \tag{A.29}$$

Substituting $y = \frac{x}{a}$ into $(A.29)$ provides:

$$\begin{aligned} \delta a^{\delta-1} &= \int_{\frac{x_0}{a}}^{\infty} 2[(ay-x_0)\beta_1]^{\theta}\left[-\frac{1}{a^2} f\left(y\right) - \left[\frac{ay}{a^3}\right] f'\left(y\right)\right][a]dy \\ \Rightarrow\ \beta_1^{\theta} &\int_{c_3}^{\infty} [y-c_3]^{\theta}[-1]\left[f\left(y\right) + y f'\left(y\right)\right] dy = \frac{\delta a^{\delta-\theta}}{2}. \end{aligned} \tag{A.30}$$

Integrating by parts and using the fact that $f\left(y\right)$ decays exponentially, $(A.30)$ implies:

$$\beta_1^{\theta}\theta \int_{c_3}^{\infty} [y-c_3]^{\theta-1} y f\left(y\right) dy = \frac{\delta a^{\delta-\theta}}{2} \quad \Rightarrow \quad \beta_1^{\theta} g\left(c_3\right) = a^{\delta-\theta}. \tag{A.31}$$

Using the change of variable $y = \frac{x}{a}$, $(A.31)$ and (A.25) imply:

$$\widetilde{L} = ap - \int_{c_3}^{\infty} [ay-x_0]\beta_1 f\left(y\right) dy = a[p - \beta_1 g_2\left(c_3\right)] = a\left[p - \left(\frac{g_2(c_3)}{g(c_3)^{\frac{1}{\theta}}}\right) a^{\frac{\delta-\theta}{\theta}}\right]. \tag{A.32}$$

$(A.27)$, $(A.31)$, and $(A.32)$ imply that $[P-L]$ can be written as:

$$\underset{a,\, \beta_1,\, c_3}{\text{Maximize}}\ a\left[p - \left(\frac{g_2(c_3)}{g(c_3)^{\frac{1}{\theta}}}\right) a^{\frac{\delta-\theta}{\theta}}\right] \quad \text{subject to} \quad \frac{a^{\delta}}{g(c_3)} \geq \frac{a^{\delta}+\overline{U}}{2\alpha\left(c_3\right)} \quad \text{and} \quad [\beta_1]^{\theta} g\left(c_3\right) = a^{\delta-\theta}. \tag{A.33}$$

To identify relevant properties of $\alpha(\cdot)$, $g(\cdot)$, and $g_2(\cdot)$, note that because $\Gamma(p) = [p-1]\Gamma(p-1)$:

$$f_p'(y) = -f_p(y) + f_{p-1}(y). \tag{A.34}$$

Since $\alpha(\cdot)$ depends on p, we can write:

$$\alpha(c) = \alpha_p(c) = \int_{c}^{\infty} [y-c]^{\theta} f_p(y) dy = \int_{0}^{\infty} t^{\theta} f_p(t+c) dt. \tag{A.35}$$

Taking the derivative inside the integral sign (which is valid when $p > 1-\theta$) provides:

$$\alpha_p'(c) = \int_{c}^{\infty} t^{\theta}[-f_p(c+t) + f_{p-1}(c+t)]dt = \alpha_{p-1}(c) - \alpha_p(c). \tag{A.36}$$

Now, let:

$$g_p(c) = \left[\frac{\delta}{2\theta}\right] g(c) = \int_c^\infty [y-c]^{\theta-1} y f_p(y)dy = \int_0^\infty t^{\theta-1}[c+t] f_p(t+c)dt.$$

Again, taking the derivative inside the integral and using equation $(A.34)$ provides:

$$\begin{aligned} g_p'(c) &= \int_0^\infty t^{\theta-1}[(c+t)f_p'(t+c) + f_p(c+t)]dt \\ &= \int_0^\infty t^{\theta-1}[(c+t)(f_{p-1}(c+t) - f_p(t+c)) + f_p(c+t)]dt \\ &= \int_c^\infty (y-c)^{\theta-1}[y f_{p-1}(y) - y f_p(y) + f_p(y)]dy \\ &= g_{p-1}(c) - g_p(c) + \int_c^\infty [y-c]^{\theta-1} f_p(y)]dy = \left[\frac{p}{p-1}\right] g_{p-1}(c) - g_p(c). \end{aligned} \tag{A.37}$$

Also, observe that:

$$g_p\,(c) = \int_0^\infty t^{\theta-1}[c+t] f_p\,(c+t)\,dt = \int_0^\infty t^\theta f_p(c+t)dt \;+\; c\int_0^\infty t^{\theta-1} f_p\,(c+t)\,dt. \tag{A.38}$$

$(A.35)$ and (A.38) imply:

$$g_p(c) = \alpha_p(c) + \left[\frac{c}{p-1}\right] g_{p-1}(c). \tag{A.39}$$

Lemma 3. *(i) $\alpha\,(c)$ and $g_2\,(c)$ are decreasing functions of c.*

(ii) $2\alpha(0) > g(0)$. Further, $\frac{\alpha_p(c)}{g_p(c)}$ is a decreasing function of c with a maximum of 1 at $c = 0$.

Proof. Property (i) follows from the definitions of $\alpha\,(c)$ and $g_2\,(c)$, since $\delta > 1$ and $\theta \leq \frac{1}{2}$.

It is apparent that $\alpha_p(0) = g_p(0)$. We will now demonstrate that:

$$g_{p-1}^2(c) \leq g_p(c) g_{p-2}(c) \left[\frac{p-1}{p-2}\right]. \tag{A.40}$$

To show that the inequality in (A.40) holds, let Y be a gamma random variable with parameter $(p-1)$ and let $I(z)$ be the indicator function that takes on the value 1 if z is true and the value 0 otherwise. Then

$$g_{p-1}(c) = E[(Y-c)^{\theta-1} Y \; I(Y > c)] = E(X),$$

where

$$X = [Y - c]^{\theta-1} Y I(Y > c) = Y^{(\theta-1)/2}\ I(Y > c) \times [Y - c]^{(\theta-1)/2}\ I(Y > c) = X_1 X_2.$$

By the Cauchy - Schwartz inequality:

$$E(X) = E(X_1 X_2) < [E(X_1^2)]^{1/2}[E(X_2^2)]^{1/2}. \tag{A.41}$$

But

$$E(X_1^2) = \int_c^\infty [y - c]^{\theta-1} y^2 f_{p-1}(y) dy = \int_c^\infty [y - c]^{\theta-1} y f_p(y)[p - 1] dy = g_{p-1}(c)[p - 1], \text{ and} \tag{A.42}$$

$$E(X_2^2) = \int_c^\infty [y - c]^{\theta-1} f_{p-1}(y) dy = \int_c^\infty [y - c]^{\theta-1} y \left[\frac{f_{p-2}(y)}{p - 2}\right] dy = \frac{g_{p-2}(c)}{p - 2}. \tag{A.43}$$

Inequality (A.40) follows from equations $(A.41)$, $(A.42)$ and $(A.43)$.

To complete the proof of the Lemma, it suffices to show that $D_p(c) = \alpha_p'(c) g_p(c) - \alpha_p(c) g_p'(c) < 0$. Using equations $(A.36)$, (A.37) and $(A.39)$:

$$\begin{aligned}
D_p(c) &= g_p(c)\left[\alpha_{p-1}(c) - \alpha_p(c)\right] - \alpha_p(c)\left[\left(\frac{p}{p-1}\right) g_{p-1}(c) - g_p(c)\right] \\
&= g_p(c)\left[\frac{g_{p-1}(c) - c g_{p-2}(c)}{p-2}\right] - \left[\frac{p}{p-1}\right] g_{p-1}(c)\left[\frac{g_p(c) - c g_{p-1}(c)}{p-1}\right] \\
&= -\frac{g_p(c) g_{p-1}(c)}{p-1} - \frac{c g_p(c) g_{p-2}(c)}{p-2} + \frac{c p g_{p-1}^2(c)}{[p-1]^2} \\
&\le g_p(c)\left[-\frac{g_{p-1}(c)}{p-1} - \frac{c g_{p-2}(c)}{p-2} + c p \frac{g_{p-2}(c)}{(p-1)(p-2)}\right] \\
&= g_p(c)\left[-\frac{g_{p-1}(c)}{p-1} + c \frac{g_{p-2}(c)}{(p-2)(p-1)}\right] = g_p(c)\left[-\frac{\alpha_{p-1}(c)}{(p-1)}\right] < 0.
\end{aligned}$$

□

Let $\widehat{c}_3$ be the unique solution of $2\alpha(c_3) = g(c_3)$. When $\overline{U} = 0$, (A.33) implies:

$$\frac{a^\delta}{2a^\theta \alpha(c_3)} \le \frac{a^{\delta-\theta}}{g(c_3)} \quad \Rightarrow \quad 2\alpha(c_3) \ge g(c_3) \quad \Rightarrow \quad c_3 \le \widehat{c}_3. \tag{A.44}$$

Hence the principal's problem can be written as:

$$\underset{a,\ c_3 \le \widehat{c}_3}{\text{Maximize}}\ \widehat{L} = a\left[p - \frac{g_2(c_3)}{[g(c_3)]^{\frac{1}{\theta}}} a^{\frac{\delta-\theta}{\theta}}\right].$$

Define $f_2(c_3) = \left[\frac{\delta g_2(c_3)}{p\theta[g(c_3)]^{\frac{1}{\theta}}}\right]^{\frac{\delta\theta}{\delta-\theta}}$. Because $\widehat{L}$ is concave in a, the unconstrained optimum for a

occurs where $\frac{\partial L}{\partial a} = 0$, which yields:

$$\widehat{a} = \left[\frac{p\theta}{\delta}\frac{[g(c_3)]^{\frac{1}{\theta}}}{g_2(c_3)}\right]^{\frac{\theta}{\delta-\theta}} = \left[\frac{1}{f_2(c_3)}\right]^{\frac{1}{\delta}}. \tag{A.45}$$

Thus, the solution is given by:

$$a_0 = \left[\frac{p\theta}{\delta}\frac{[g(c_3^*)]^{\frac{1}{\theta}}}{g_2(c_3^*)}\right]^{\frac{\theta}{\delta-\theta}} \quad \text{and} \quad \widehat{L}_{opt} = a_0\left[p - \frac{p\theta}{\delta}\right] = a_0 p\left[\frac{\delta-\theta}{\delta}\right], \tag{A.46}$$

where c_3^* is the point at which $f_2(c_3)$ achieves its minimum value in the range $[0, \hat{c}_3]$.

Now consider $\overline{U} > 0$. Note that $2\alpha(c_3) \leq g(c_3)$ for all $c_3 \geq \widehat{c}_3$. Hence, when $\overline{U} > 0$, $(A.33)$ implies:

$$\frac{a^{\delta-\theta}}{g(c_3)} \geq \frac{a^{\delta}+\overline{U}}{2a^{\theta}\alpha(c_3)} \Rightarrow c_3 < \widehat{c}_3 \quad \text{and} \quad a^{\delta} \geq \frac{\overline{U}g(c_3)}{2\alpha(c_3) - g(c_3)} \equiv \frac{\overline{U}}{f_1(c_3)}. \tag{A.47}$$

Hence, the principal's problem can be stated as:

$$\underset{a,\ c_3 < \ \widehat{c}_3}{\text{Maximize}}\ \ \widehat{L} = a\left[p - \frac{g_2(c_3)}{[g(c_3)]^{\frac{1}{\theta}}}a^{\frac{\delta-\theta}{\theta}}\right] \quad \text{subject to} \quad a^{\delta} \geq \frac{\overline{U}}{f_1(c_3)}.$$

Recall that for any fixed c_3, $\widehat{L}$ is increasing in a for $a^{\delta}f_2(c_3) \leq 1$ and is decreasing in a if $a^{\delta}f_2(c_3) \geq 1$. For $c_3 \in [0, \widehat{c}_3]$, define the following two sets:

$$A = A(\overline{U}) = \{c_3 \,|\, f_1(c_3) < \overline{U}f_2(c_3)\} \cap [0,\ \widehat{c}_3], \quad A^c = A^c(\overline{U}) = \{c_3 \,|\, f_1(c_3) \geq \overline{U}f_2(c_3)\} \cap [0,\ \widehat{c}_3].$$

By the definition of $\hat{c}_3$, $f_1(\hat{c}_3) = 0$. Furthermore, $f_2(c_3)$ is always positive. Therefore, $f_1(\hat{c}_3) - \overline{U}f_2(\widehat{c}_3) < 0$ and $\widehat{c}_3 \in A$. Then, by the the continuity of $f_1(c_3) - \overline{U}f_2(c_3)$, there exists an interval $[c_3^*, \hat{c}_3]$ such that $[c_3^*, \hat{c}_3] \subseteq A$.

First suppose $c_3 \in A$ is fixed. Then the optimal a and corresponding value of $\widehat{L}$ are given by:

$$a_{01}^{\delta}f_1(c_3) = \overline{U} \quad \text{and} \quad L(\overline{U}, c_3) = p\left[1 - \frac{\theta}{\delta}\left(\frac{\overline{U}f_2(c_3)}{f_1(c_3)}\right)^{\frac{\delta-\theta}{\delta\theta}}\right]\left[\frac{\overline{U}}{f_1(c_3)}\right]^{\frac{1}{\delta}}. \tag{A.48}$$

Let $L_1(\overline{U}) = \sup_{c_3 \in A}\{L(\overline{U}, c_3)\}$, and suppose $c_3 \in A^c$ is fixed. Then the optimal a and corresponding value of $\widehat{L}$ are given by:

$$a_{02}^{\delta} f_2(c_3) = 1 \text{ and } L(\overline{U}, c_3) = a_0 p - a_0 \frac{p\theta}{\delta} = p\left[1 - \frac{\theta}{\delta}\right]\left[\frac{1}{f_2(c_3)}\right]^{\frac{1}{\delta}}. \tag{A.49}$$

Let $L_2(\overline{U}) = \sup_{c_3 \in A^c} \{L(\overline{U}, c_3)\}$. If $\overline{U}_B < \overline{U}_A$, then $A(\overline{U}_B) \subset A(\overline{U}_A)$ and hence $L_2(\overline{U}_A) \leq L_2(\overline{U}_B)$.

Finally, define $\overline{U}_3 = \sup_{c_3 < \widehat{c}_3} \frac{f_1(c_3)}{f_2(c_3)}$, and $\overline{U}_4 = \overline{U}_3 \left[\frac{\delta}{\theta}\right]^{\frac{\delta\theta}{\delta-\theta}}$, and observe that $\overline{U}_4 > \overline{U}_3$. Furthermore, for all $\overline{U} > \overline{U}_3$, $\overline{U} f_2(c_3) > f_1(c_3)$ for all $c_3 \leq \widehat{c}_3$, and so A^c is empty. In addition, the optimum over A^c is always positive. This discussion provides the following proposition.

Proposition 3. *Let $w(x) = 0$ if $x < x_0$ and $w(x) = [x - x_0]\beta_1$ if $x \geq x_0$. Then:*
(i) The value of c_3 that maximizes the principal's objective subject to the prevailing constraints is the value of c_3 at which $\max\{L(\overline{U}, c_3), c_3 \in A; \ L(\overline{U}, c_3), c_3 \in A^c\}$ is maximized. The corresponding optimal value of a is the solution to $a^{\delta} f_1(c_3) = \overline{U}$ if $\max\{L(\overline{U}, c_3), c_3 \in A\} \geq \max\{L(\overline{U}, c_3), c_3 \in A^c\}$. The optimal value of a is the solution to $a^{\delta} f_2(c_3) = 1$ if $\max\{L(\overline{U}, c_3), c_3 \in A^c\} > \max\{L(\overline{U}, c_3), c_3 \in A\}$. The optimal value of a is positive if $\overline{U} \leq \overline{U}_4$.
(ii) If $\overline{U} > \overline{U}_3$, the optimal c_3 is the value of c_3 at which $L(\overline{U}, c_3)$ achieves its maximum on $[0, \ \widehat{c}_3]$. The optimal a is the solution to $a^{\delta} f_1(c_3) = \overline{U}$. If $\overline{U} > \overline{U}_4$, then the principal's payoff, $L_1(\overline{U})$, is negative.
(iii) If $\overline{U} = 0$, then the solution to $[P - L]$ is given by:

$$a_0 = \left[\frac{p\theta}{\delta} \frac{[g(c_3^*)]^{\frac{1}{\theta}}}{g_2(c_3^*)}\right]^{\frac{\theta}{\delta-\theta}} \quad \text{and} \quad \pi^L = a_0\left[p - \frac{p\theta}{\delta}\right] = a_0 p\left[1 - \frac{\theta}{\delta}\right], \tag{A.50}$$

where c_3^* is the point at which $f_2(c_3)$ attains its minimum value in the range $[0, \ \widehat{c}_3]$.

The proof of Theorem 1 now follows from Propositions 1, 2 and 3.