Forthcoming in *Econometrica*

IMPLEMENTING THE FIRST BEST IN AN AGENCY RELATIONSHIP WITH RENEGOTIATION: A CORRIGENDUM[1]

BY AARON S. EDLIN AND BENJAMIN E. HERMALIN

Abstract:

The proof of Proposition 4 in Hermalin and Katz (1991) is incorrect, because it fails to check post-renegotiation utilities against the incentive compatibility constraints. This note states and proves a comparable proposition with a slightly stronger assumption regarding the monotonicity of bargaining. This result vindicates the central intuition of Hermalin and Katz about the potential *in*significance of the observable, but unverifiable distinction in contracting.

Keywords: Contract renegotiation, bargaining.

JEL: C78, D23, K12, L23.

1. INTRODUCTION

Hermalin and Katz (1991) investigate the outcome of agency games in which the principal observes information about the agent's action that she is *un*able to verify. Even when critical variables are not verifiable, parties can often present some relevant evidence, and Hermalin and Katz show that such noisy but informative signals can allow parties to write contracts *as if* the "critical" variables were verifiable.[2] The key is that the parties can renegotiate their contract after the principal observes the observable, but unverifiable information (e.g., the agent's action), but before the realization of the state of the world (e.g., output) upon which the contract is contingent. Indeed, when the agent's action itself is observable, but unverifiable, then the parties can use the initial

[1]This paper draws from Section 5 of a 1996 working paper, "Contract Renegotiation in Agency Problems." The authors thank Yeon-Koo Che, Preston McAfee, Paul Milgrom, Andrew Postlewaite, Stefan Reichelstein, Bill Rogerson, Alan Schwartz, Ilya Segal, Chris Shannon, Lars Stole, an anonymous referee, and participants at conferences or seminars at the University of Texas, the 1997 ALEA meetings, the 1997 NBER Summer Institute, the University of British Columbia, Tel Aviv University, Cornell University, the University of Maryland, and the Olin School for their comments and suggestions. They also thank Eric Emch for research assistance. Edlin enjoyed financial support from a Sloan Faculty Fellowship and from an Olin Fellowship at the Law & Economics Program at Georgetown Law Center. Hermalin enjoyed financial support from the NSF under grant SBR-9616675;JGSM, Cornell University; COR & the Willis H. Booth Chair in Banking & Finance, University of California, Berkeley.

[2]Contrast this to the view in much of the contract-theory literature (e.g., Grossman and Hart, 1986 and its successors) in which no verifiable signals exist, and a variety of inefficiencies result.

contract, written contingent on the verifiable outcome, to supply the agent with incentives and, then, renegotiate to eliminate the inefficient risk sharing. In this way, the first best can be achieved.

Hermalin and Katz establish their results assuming take-it-or-leave-it bargaining in renegotiation. However, because bargaining power is more typically shared, they recognize that their results would have more economic significance if they were extended beyond take-it-or-leave-it bargaining. They attempt such an extension in their Proposition 4. Unfortunately, their proof is wrong because it does not properly fold the renegotiation outcomes back into the incentive-compatibility constraints.[3] Here, we provide a correct proof of essentially the same proposition. Our proposition imposes only a slightly stronger monotonicity assumption; requiring that, as the agent's expected wage under the initial contract increases, his renegotiated wage increases at a rate bounded from zero.

2. Model

A principal employs an agent to take an action a chosen from $\mathcal{A}$, a finite subset of $\mathbb{R}$. The principal observes this action, but it is not verifiable. After the principal observes the agent's action, the parties can renegotiate the contract. Ultimately, a verifiable signal $x \in \mathcal{X}$ is realized. Let $G(\cdot|a)$ be a probability measure conditional on a that maps some σ-algebra of $\mathcal{X}$ into $\mathbb{R}$.

The agent's utility is $u(y) - a$, where y is what he's paid and $u(\cdot)$ is strictly increasing and strictly concave.[4] Assume that the domain of $u^{-1}(\cdot)$ is $\mathbb{R}$. The principal's utility is $b(a) - y$, where $b(\cdot)$ is the benefit, possibly an expected value, enjoyed by the principal.

A contract is a mapping from the signal to wages: $w : \mathcal{X} \to \mathbb{R}$. The initial contract remains in force unless renegotiated.

The first-best action maximizes the principal's expected utility subject to the constraint that the agent receive at least his reservation utility, which we normalize to 0. This constraint must bind; hence, this program reduces to

$$\max_{a \in \mathcal{A}} b(a) - u^{-1}(a).$$

[3] Since we first drafted this paper, Ishiguro (1998) has independently found that the proof of Hermalin and Katz's Proposition 4 is wrong. He does not, however, comment on whether the Proposition itself could be correct. For a specific example, he shows that the first best can be achieved for fixed-shares bargaining.

[4] Hermalin and Katz assume the utility function is $u(y) - c(a)$; but, there is no loss of generality in letting $c(a) \equiv a$.

We assume that this program has a unique solution, a^*. So that the agency problem is meaningful, assume this action is *not* a least-cost action for the agent; that is, $a^* > \min \mathcal{A}$.

3. Analysis

The solution to the agency problem *absent* renegotiation typically involves compensation contingent on the signal, x. This will fail to achieve the first best, however, because the risk-averse agent bears risk in equilibrium. Renegotiation can improve the allocation of risk by shifting it from the agent to the risk-neutral principal; that is, by renegotiating a contingent-wage contract into a flat-wage contract after the agent has acted. Moreover, since that flat wage depends on what *would have happened* under the initial contingent contract given the agent's action, the agent will, therefore, still have incentives to work hard. In fact, as Hermalin and Katz (1991) showed, if we assume extreme bargaining power in the renegotiation game, then the first best is attainable.

Here we extend the result to "intermediate" bargaining. Following Hermalin and Katz, we assume that the bargaining outcome is unique and entirely determined by the agent's expected utility $\tilde{u}$ exclusive of action costs and the principal's expected wage payment $\tilde{w}$ should the *initial* contract remain in force. In particular, we assume that bargaining results in the original contingent contract being replaced with a nonrandom payment by the principal to the agent in the amount of $h(\tilde{u}, \tilde{w})$, thereby yielding the agent a certain utility of $u[h(\tilde{u}, \tilde{w})]$. We assume, moreover, that h satisfies the following properties:

Individual rationality: $u[h(\tilde{u}, \tilde{w})] \geq \tilde{u}$ and $h(\tilde{u}, \tilde{w}) \leq \tilde{w}$ for all $\tilde{u}$, $\tilde{w}$.

Uniform monotonicity: $h_1(\tilde{u}, \tilde{w}) \geq 0$ and $h_2(\tilde{u}, \tilde{w}) > \eta > 0$, for some $\eta > 0$ for all $\tilde{u}, \tilde{w}$.

Observe that individually rational bargaining satisfies $h\left[\tilde{u}, u^{-1}(\tilde{u})\right] = u^{-1}(\tilde{u})$ for all $\tilde{u}$. Observe also that both properties are satisfied by any constant-shares bargaining game over the surplus, $\tilde{w} - u^{-1}(\tilde{u})$; that is, by $h(\tilde{u}, \tilde{w}) = \bar{\sigma} \times \left[\tilde{w} - u^{-1}(\tilde{u})\right] + u^{-1}[\tilde{u}]$, where $\bar{\sigma} \in (0, 1)$ is a constant.

For individually-rational uniformly-monotonic bargaining games, we can establish sufficient conditions for the first best to be attainable.[5]

Proposition 1: *Suppose that:*

[5] Relatedly, Ishiguro (1998) shows that if $x = a + \nu$, where ν is a mean-zero normal random variable, and utility is negative exponential, then the first best can be achieved precisely with a continuum of actions.

(i) the principal's payment from the renegotiation bargaining game is uniquely given by $h(u, w)$, where u and w are the agent's expected utility (exclusive of action costs) and the principal's expected wage payment, respectively, should the initial contract remain in force, and h satisfies both individual rationality and uniform monotonicity; and

(ii) there exists a subset $\mathcal{X}^*$ of $\mathcal{X}$ such that $G(\mathcal{X}^*|a^*) > G(\mathcal{X}^*|a)$ for all $a \in \mathcal{A}\backslash\{a^*\}$;

then action a^* is implementable at first-best cost with renegotiation.

PROOF: Define $\bar{\pi}(a) \equiv G(\mathcal{X}^*|a)$ and consider the contract

$$w(x) = \begin{cases} w_1, \text{ if } & x \in \mathcal{X}^* \\ w_2, \text{ if } & x \notin \mathcal{X}^* \end{cases}.$$

Define $\hat{h}(w_1, w_2, \pi) \equiv h(\pi u(w_1) + [1-\pi]u(w_2), \pi w_1 + [1-\pi]w_2)$.

After the agent chooses a, the parties will renegotiate. By assumption, the agent's resulting utility is $u\left[\hat{h}(w_1, w_2, \bar{\pi}(a))\right]$. A contract implements a^* at first-best cost with renegotiation if and only if

$$\begin{aligned} u\left[\hat{h}\left(w_1, w_2, \bar{\pi}(a^*)\right)\right] - a^* &= 0; \text{ and} \\ u\left[\hat{h}\left(w_1, w_2, \bar{\pi}(a)\right)\right] - a &\leq u\left[\hat{h}\left(w_1, w_2, \bar{\pi}(a^*)\right)\right] - a^* \quad \text{for all} \quad a \neq a^*. \end{aligned} \tag{1}$$

Observe that, using (1), these expressions can be rewritten as

$$\hat{h}\left(w_1, w_2, \bar{\pi}(a^*)\right) = u^{-1}(a^*); \text{ and} \tag{2}$$

$$\hat{h}\left(w_1, w_2, \bar{\pi}(a^*)\right) - \hat{h}\left(w_1, w_2, \bar{\pi}(a)\right) \geq u^{-1}(a^*) - u^{-1}(a) \quad \text{for all} \quad a \neq a^*. \tag{3}$$

Define $w_2^*(\cdot)$ so that $\langle w_1, w_2^*(w_1)\rangle$ satisfies (2) for all w_1. To see that $w_2^*(\cdot)$ is well-defined, observe, first, that $\hat{h}$ has positive first derivatives with respect to both w_1 and w_2. Moreover, the individual rationality of bargaining implies that

$$w_1 = w_2 = u^{-1}(a^*)$$

satisfies (2), because $h(a^*, u^{-1}(a^*)) = u^{-1}(a^*)$. These observations together imply that $w_2^*(\cdot)$ indeed

exists and, moreover, is decreasing and continuous. Note that for $w_1 > u^{-1}(a^*)$, $w_2^*(w_1) < w_1$. Henceforth, we will consider only $w_1 > u^{-1}(a^*)$.

Observe that

$$\frac{\partial \hat{h}\left(w_1, w_2^*(w_1), \bar{\pi}\right)}{\partial \bar{\pi}} = h_1 \times [u(w_1) - u(w_2^*[w_1])] + h_2 \times (w_1 - w_2^*[w_1]).$$

Invoking uniform monotonicity, we have

$$\frac{\partial \hat{h}}{d\bar{\pi}} \geq \eta\left(w_1 - w_2^*(w_1)\right) \tag{4}$$

for all $\bar{\pi}$ and $w_1 > w_2^*(w_1)$. Combining the mean value theorem with inequality (4), yields

$$\hat{h}\left(w_1, w_2^*(w_1), \bar{\pi}(a^*)\right) - \hat{h}\left(w_1, w_2^*(w_1), \bar{\pi}(a)\right) \geq \left[\bar{\pi}(a^*) - \bar{\pi}(a)\right] \eta\left(w_1 - w_2^*(w_1)\right)$$

for all a. Since $\mathcal{A}$ is finite, we can define $\hat{a} = \arg\min_{\mathcal{A}\setminus\{a^*\}} \bar{\pi}(a^*) - \bar{\pi}(a)$. Hence,

$$\hat{h}\left(w_1, w_2^*(w_1), \bar{\pi}(a^*)\right) - \hat{h}\left(w_1, w_2^*(w_1), \bar{\pi}(a)\right) \geq \left[\bar{\pi}(a^*) - \bar{\pi}(\hat{a})\right] \eta\left(w_1 - w_2^*(w_1)\right) \tag{5}$$

for all $a \neq a^*$. Because $w_2^*(\cdot)$ is a continuous and decreasing function and $\bar{\pi}(a^*) > \bar{\pi}(\hat{a})$, we can find a w_1^* sufficiently large that the right-hand side of (5) exceeds $u^{-1}(a^*) - u^{-1}(\underline{a})$, where $\underline{a} \equiv \min \mathcal{A}$. Transitivity then yields

$$\hat{h}\left(w_1^*, w_2^*(w_1^*), \bar{\pi}(a^*)\right) - \hat{h}\left(w_1^*, w_2^*(w_1^*), \bar{\pi}(a)\right) \geq u^{-1}(a^*) - u^{-1}(a)$$

for all $a \neq a^*$. Hence, (3) holds. By construction, (1) holds. Therefore, the contract paying w_1^* if $x \in \mathcal{X}^*$ and $w_2^*(w_1^*)$ if $x \notin \mathcal{X}^*$ is efficient.

■

There are three main differences between this proposition and the one Hermalin and Katz state. First, we introduce individual rationality, an innocuous restriction that might be viewed as implicit in Hermalin and Katz's statement. Second, our proof requires the stronger monotonicity assumption that $h_2 > \eta > 0$, whereas Hermalin and Katz assume that $h_2 > 0$ and that h goes to infinity as expected wages approach infinity. We do not know if the proposition holds without

uniform monotonicity. Third, we state the proposition in terms of the principal's post-renegotiation payment instead of the agent's post-renegotiation utility; this formulation allows us to impose uniform monotonicity in a way that is satisfied by constant-shares bargaining.

Our proof is intuitively understood by considering an initial contract that pays the agent w_1 if $x \in \mathcal{X}^*$ and w_2 if $x \notin \mathcal{X}^*$. As the agent switches from a to a^* he raises the probability that $x \in \mathcal{X}^*$. This increases the amount the principal is willing to pay the agent to buy out the contract at a rate of $w_1 - w_2$. The agent's bargaining fall-back (i.e., his certainty equivalent) also improves, but—even ignoring this—the agent receives at least a share η of the principal's extra willingness to pay under uniformly monotonic sharing. Hence, by driving the wedge $w_1 - w_2$ sufficiently large (while maintaining the agent's participation constraint), we can induce an arbitrarily large post-renegotiation pay difference between actions a^* and a.

Finally, we should point out two potential limitations and one extension of this analysis. First, there is an implicit assumption in the above bargaining game that the outcome will be independent of bygone actions. To illustrate why this matters and why efficiency is not obtainable for all bargaining games, assume $u(y) = \ln(y)$, $\mathcal{A} = \{0, 1\}$, and two possible signals, x_1 and x_2. Let the probability of x_1 given a be $\frac{a+2}{4}$. Suppose that $a^* = 1$ is the first-best action. It is implementable *without* renegotiation by Hermalin and Katz's Proposition 2. Suppose that the harder the agent works, the less energy he has for bargaining, so that the principal has all the bargaining power when $a = 1$, but no bargaining power when $a = 0$. Let $U_a(w_1, w_2)$ be the *equilibrium* utility of the agent conditional on a and the contract $w_i = w(x_i)$. Then, $U_1(w_1, w_2) = \frac{3}{4}\ln(w_1) + \frac{1}{4}\ln(w_2) - 1$ and $U_0(w_1, w_2) = \ln\left(\frac{1}{2}w_1 + \frac{1}{2}w_2\right)$. The first-best action $a = 1$ is not implementable with this bargaining game, because there exist no w_1 and w_2 such that $U_1(w_1, w_2) \geq U_0(w_1, w_2)$.[6] To be sure, we are *not* suggesting this bargaining game is more reasonable than the one considered above; its purpose is simply to demonstrate that *some* constraints on the bargaining game are necessary for efficiency. A second potential limitation is the model's realism when the signals (i.e., the x 's) are relatively uninformative: The resulting large spread between the wages w_1 & w_2 increases the amount at stake in renegotiation, which could raise doubts about our assumption of efficient bargaining; and it could result in $w_2 \ll 0$, which courts might interpret as an invalid penalty. Lastly, we note that an earlier version of this paper (Edlin and Hermalin, 1997) extended the analysis to

[6] Maximizing $U_1(w_1, w_2) - U_0(w_1, w_2)$ with respect to w_1 and w_2 reveals that any w_1 and w_2 pair on the line $w_1 = 3w_2$ is optimal. When $w_1 = 3w_2$, $U_1(w_1, w_2) - U_0(w_1, w_2)$ reduces to $\ln\left(3^{3/4}/2\right) - 1$, which is negative.

an interval of actions, $[\underline{a}, \infty)$, and showed that it was possible to induce an action arbitrarily close to the first-best action at arbitrarily close to first-best cost.

Department of Economics and School of Law, University of California, 549 Evans Hall #3880, Berkeley, CA 94720-3880, U.S.A.; http://emlab.berkeley.edu/users/edlin/index.html.

and

Walter A. Haas School of Business, University of California, 545 Student Services Bldg. #1900, Berkeley, CA 94720-1900, U.S.A.; hermalin@haas.berkeley.edu, http://haas.berkeley.edu/~hermalin.

4. References

Edlin, Aaron S. and Benjamin E. Hermalin, "Contract Renegotiation in Agency Problems," National Bureau of Economic Research Working Paper, 1997, W6086.

Grossman, Sanford and Oliver D. Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 1986, *94*, 691–719.

Hermalin, Benjamin E. and Michael L. Katz, "Moral Hazard and Verifiability: The Effects of Renegotiation in Agency," *Econometrica*, 1991, *59*, 1735–1753.

Ishiguro, Shingo, "A Note on the Effects of Renegotiation and Verifiability in Agency," 1998. Working paper, School of Business Administration, Nanzan University.