Econometrica, Vol. 53, No. 1 (January, 1985)

REPEATED MORAL HAZARD[1]

BY WILLIAM P. ROGERSON

This paper considers a repeated principal agent relationship where the principal is risk neutral, the agent is risk averse, the principal can borrow or save at a fixed interest rate, and the agent discounts future consumption. It is shown that memory plays a very strong role in every Pareto-optimal contract. Sufficient conditions for Pareto-optimal contracts to exhibit rising or falling wages are identified. Finally, it is shown that the restriction of the agent's access to credit is necessary to achieve a Pareto-optimal outcome. In particular, under every Pareto-optimal contract for every outcome of every period the agent would choose to save some of his wage if he could.

1. INTRODUCTION

THIS PAPER CONSIDERS a repeated principal agent relationship with a risk neutral principal and risk averse agent. It is assumed that the principal can borrow and save at a fixed interest rate but the agent can neither borrow nor save. The agent discounts future consumption. It is shown that for any optimal contract, a simple relationship must always hold between the wages offered in any two adjacent periods. The inverse of the agent's marginal utility of income evaluated at any wage must be equal to the conditional expected value of the inverse of next period's marginal utility of income. Although this relationship does not fully characterize the optimal contract, it does imply a number of interesting properties of it.

First, it is shown that memory plays a role in every Pareto-optimal contract. Whenever an outcome affects the current wage it also affects the future periods' wages. Second, it is shown that the expected wage payment will rise (fall) over time if the inverse of the agent's marginal utility is concave (convex). For the class of HARA utility functions, expected wages are shown to increase or decrease over time depending upon whether the agent's risk tolerance increases at a rate greater than or less than 1. Expected wages are constant if the agent's utility is logarithmic. Third, it is shown that the restriction of the agent's access to credit is necessary to achieve a Pareto-optimal outcome. In particular, under every Pareto-optimal contract for every outcome of every period the agent would choose to save some of his wage if he could.

The literature on repeated moral hazard has been primarily focussed on showing that a first best solution can be achieved if the agent does not discount future income. See, for example, Radner [**6**, **7**, **8**], Rubinstein [**12**], and Rubinstein and Yaari [**13**]. However, none of these papers has developed any qualitative propositions about the structure of the contract for cases where the agent discounts future income. Related results to those of this paper have been independently

[1] This research was supported in part by Jet Propulsion Laboratories and in part by National Science Foundation Grant SES 8006654 at the Institute for Mathematical Studies in the Social Sciences, Stanford University. I would like to thank Joel Demski, Kathleen Hagerty, David Kreps, Eric Maskin, Dilip Mookherjee, Mitchel Polinsky, and the editors for helpful comments and discussions.

obtained by Lambert [3]. Using the first-order approach, he shows that the agent's compensation in one period must depend on his performance in prior periods as well as the current period. This paper does not use the first-order approach and therefore the result on memory is much more general.[2] Rogerson [9] develops some qualitative propositions but the class of contracts is limited to those with a constant wage; only the decision when to fire the agent is allowed to vary. Diamond and Mirrlees [2] derive a result on savings behavior similar to that of this paper in a model with private information. It will be shown in Section 6 that the results of this paper generalize the Diamond–Mirrlees result. Braverman and Stiglitz [1] also derive a similar savings result in a model with linear contracts that are not fully repeated; the agent works and consumes in the second period but only consumes in the first period.

Aside from the specific technical results outlined above, this analysis suggests a useful intuition for thinking about repeated incentive problems, especially when the relationship is finite or the agent discounts future income. In these cases the intuitions developed for the no-discounting infinite-horizon case are not applicable. The repetition of a moral hazard relationship creates the opportunity for intertemporal risk sharing. The optimal contract always takes advantage of this, i.e., memory plays a role in the optimal contract. However, because of the incentive problem the agent is not fully insured and the agent is left with a residual desire to intertemporally self-insure through the use of credit markets.

2. THE BASIC RESULT

For ease of exposition it will be assumed that the relationship between the principal and agent lasts two periods.[3] Each period the agent chooses an action, a, from a set A. The outcome can be one of N dollar returns to the principal, $\{x_1, \ldots, x_N\}$. Let $p_j(a)$ denote the probability of x_j occurring given that action a is taken.

The principal is risk-neutral and can borrow or save at the interest rate r. He therefore desires to maximize expected discounted income using a discount rate of $1/(1+r)$. Let α denote this discount rate. The agent's preferences are represented by the discounted sum of period-by-period utility given by

$$v(y) - c(a) \tag{2.1}$$

where y denotes the dollar value of consumption. Assume that v is twice differentiable, strictly increasing, and strictly concave. Let β denote the agent's discount rate.

A contract, w, is a set of $N + N^2$ contingent wages. Let w_i denote the wage paid in period 1 if x_i occurs and w_{ij} denote the wage paid in period 2 if x_i occurs in period 1 and x_j occurs in period 2. A contract can be thought of as a collection

[2] See Rogerson [11] for a fuller discussion of the first order approach and for conditions under which it is valid.

[3] The results for the two-period case apply to any two adjacent periods in a longer relationship. See Rogerson [10] for a proof of this.

of N branches, the jth branch being those wages which apply conditional on x_j occurring in period 1. It will sometimes be useful to view a contract as a specification of contingent utilities for the agent. Let z denote such a specification. The corresponding utility contract, z, for any wage contract, w, is given by $z_i = v(w_i)$ and $z_{ij} = v(w_{ij})$. A strategy, s, is a set of $N+1$ contingent actions. Let s_0 denote the agent's period 1 action and let s_i denote the agent's period 2 action if x_i occurs in period 1.

The basic relationship which must hold between contingent wages across periods can now be presented. Properties of the optimal contract presented in later sections are all derived from this relationship.

PROPOSITION 1:[4,5] *Suppose that w is a Pareto-optimal contract for the strategy s. Then w must satisfy*

$$\frac{1}{v'(w_j)} = \frac{\alpha}{\beta} \sum_{k=1}^{N} \frac{p_k(s_j)}{v'(w_{jk})} \tag{2.2}$$

for every $j \in \{1, \ldots, N\}$.

PROOF: For the purpose of this proof it is convenient to think of a contract as a specification of contingent utilities for the agent. Let z be any contract for which s is an optimal strategy for the agent. Now construct a new contract, z^*, as follows where y is any real number and j is any element of $\{1, \ldots, N\}$:

$$\text{(i)} \quad z_i^* = z_i \quad \text{for } i \neq j, \tag{2.3}$$

$$z_{ik}^* = z_{ik} \quad \text{for } i \neq j \text{ and } k \in \{1, \ldots, N\}, \tag{2.4}$$

and

$$\text{(ii)} \quad z_j^* = z_j - y, \tag{2.5}$$

$$z_{jk}^* = z_{jk} + y/\beta \quad \text{for } k \in \{1, \ldots, N\}. \tag{2.6}$$

Condition (i) says that utilities outside the jth branch of the contract remain unchanged. Condition (ii) says that within the jth branch, the agent's period 1 utility is reduced by y and his period 2 utility under every outcome is increased by y/β.

It is easy to see that s is still the agent's optimal strategy under the new contract z^*. Clearly s_k is still optimal for $k \neq j$ because none of the wages facing the agent have changed. The strategy s_j is still optimal because the agent receives y/β more utils under every outcome; this leaves the relative desirability of outcomes

[4] I would like to thank James Mirrlees for suggesting a more elegant and general proof of this proposition than was contained in the original version of this paper. My original proof applied to any finite action space.

[5] If the principal is risk averse with period-by-period utility function u and discount rate α, (2.2) becomes

$$\frac{u'(x_j - w_j)}{v'(w_j)} = \frac{\alpha}{\beta} \sum_{k=1}^{N} p_k(s_j) \frac{u'(x_k - w_{jk})}{v'(w_{jk})}.$$

unchanged. Finally, s_0 is still optimal because the present discounted value of the jth branch of the contract remains unchanged. If x_j occurs in period 1 the agent receives y less utils in period 1 but y/β more utils in period 2. Therefore, the relative desirability of the various branches remains unchanged and the agent chooses the same strategy.

It is also clear that the present discounted utility of z and z^* is the same to the agent. Since the agent is indifferent between all contracts constructed as above and s is an optimal strategy under each one, a cost-minimizing contract for s must minimize the principal's expected wage payment within this set. Therefore, for z to be cost-minimizing, the value $y=0$ must minimize

$$v^{-1}(z_j+y)+\alpha \sum_{k=1}^{N} p_k(s_j)v^{-1}\left(z_{jk}-\frac{y}{\beta}\right). \tag{2.7}$$

A necessary condition for this is (2.2). *Q.E.D.*

3. THE ROLE OF MEMORY

Proposition 2 shows that if an outcome plays any role in determining current wages it must necessarily also play a role in determining future wages.

PROPOSITION 2:[6] *Suppose that w is a Pareto-optimal contract. Consider $i, j \in \{1, \dots, N\}$. Suppose that $w_i \neq w_j$. Then there must exist a $k \in \{1, \dots, N\}$ such that $w_{ik} \neq w_{jk}$.*

PROOF: The proof is by contradiction. Let w be a Pareto-optimal contract for which $w_{ik}=w_{jk}$ for every $k \in \{1, \dots, N\}$. One of the Pareto-optimal strategy choices for the agent must exhibit $s_i = s_j$. (Suppose that (w, s^*) is a Pareto-optimal contract-strategy pair and $s_i^* \neq s_j^*$. It is straightforward to show that if s^* is changed so that the agent chooses s_i^* in period 2 in both the ith and jth branches that this is also a Pareto-optimal strategy for the contract w.)

Let s be a Pareto-optimal strategy for w such that $s_i = s_j$. Therefore, $p_k(s_i) = p_k(s_j)$ for every $k \in \{1, \dots, N\}$. It now follows from (2.2) that $w_i = w_j$. *Q.E.D.*

4. THE EXPECTED WAGE PAYMENT

Proposition 3 shows that when the principal and agent have the same discount rate, whether the agent's expected wage payment rises or falls over time depends on the concavity or convexity of $1/v'$.

PROPOSITION 3: *Suppose that (w, s) is a Pareto-optimal contract-strategy pair and the principal and agent have the same discount rate.*

[6] This result continues to hold if the principal is risk averse with period-by-period utility function u and discount rate α. The proof is unchanged except that the condition in footnote 5 is substituted for (2.2).

(i) *Suppose that* $1/v'$ *is convex. Then conditional on the period* 1 *outcome, the period* 1 *wage is greater than or equal to the expected period* 2 *wage. That is, for every* $j \in \{1, \ldots, N\}$,

$$w_j \geqslant \sum_{k=1}^{N} p_k(s_j) w_{jk}. \tag{4.1}$$

Consequently the unconditional expected period 1 *wage is greater than or equal to the expected period* 2 *wage*:

$$\sum_{j=1}^{N} p_j(s_0) w_j \geqslant \sum_{j=1}^{N} \sum_{k=1}^{N} p_j(s_0) p_k(s_j) w_{jk}. \tag{4.2}$$

(ii) *If* $1/v'$ *is concave the inequalities in* (4.1) *and* (4.2) *are reversed.*

(iii) *If* v *is logarithmic* ($1/v'$ *is linear*) *expressions* (4.1) *and* (4.2) *hold with equality.*

PROOF: Only (i) will be proven; (ii) is proven similarly, and (iii) follows immediately from (i) and (ii). Since (4.2) follows immediately from (4.1) it is only necessary to prove (4.1). When $\alpha = \beta$, (2.2) becomes

$$\frac{1}{v'(w_j)} = \sum_{k=1}^{N} \frac{p_k(s_j)}{v'(w_{jk})}. \tag{4.3}$$

Since $1/v'$ is convex, Jensen's inequality implies that

$$\sum_{k=1}^{N} \frac{p_k(s_j)}{v'(w_{jk})} \geqslant \frac{1}{v'\left(\sum_{k=1}^{N} p_k(s_j) w_{jk}\right)}. \tag{4.4}$$

Taken together with the fact that $1/v'$ is increasing, (4.3) and (4.4) imply (4.1).

Q.E.D.

Note that if the agent has a stronger rate of time preference than the principal (i.e., $\beta < \alpha$) the sufficient condition for wages to fall still holds but that for wages to rise does not. That is, when the agent values current income more highly there is an additional reason for wage payments to be skewed towards the beginning of the relationship. In a similar fashion, if the principal values current income more highly there is a tendency for wages to be more skewed to the end of the relationship.

The sufficient conditions in Proposition 3 will always apply if v belongs to the broad class of utility functions called HARA utility functions. A HARA utility function exhibits linear risk tolerance. That is, for some numbers δ and γ,

$$\frac{-v'(x)}{v''(x)} = \delta + \gamma x. \tag{4.5}$$

The functional forms of the HARA utility functions are

$$(4.6)\qquad v(x)=\begin{cases}-e^{-x/\delta}, & \text{for } \gamma=0,\\ \ln(x+\delta), & \text{for } \gamma=1,\\ \dfrac{1}{\gamma-1}(\delta+\gamma x)^{\gamma-1/\gamma}, & \text{otherwise.}\end{cases}$$

Straightforward calculation shows that when the agent has a HARA utility function, expected wages will be nondecreasing (nonincreasing) if $\gamma>(<)\ 1$.

5. ACCESS TO CREDIT

This section shows that the restriction of the agent's access to credit is necessary to achieve a Pareto optimum. Perhaps more surprisingly, it is shown that the agent is *always* left with a desire to save some of his wage and *never* to borrow against future wages. The agent's saving and borrowing policy can be viewed as a second decision the agent makes in addition to his effort level. This proposition is, therefore, an example of the principle that it will generally be Pareto-improving to contractually control as many of an agent's decisions as possible in order to gain more leverage on the incentive problem over effort.[7]

PROPOSITION 4: *Suppose that w is a Pareto-optimal contract under the assumption that the agent cannot borrow or save. Suppose that after the period* 1 *outcome is announced that the agent is suddenly allowed to borrow or save at the interest rate r. Then* (i) *the agent will not wish to borrow money,* (ii) *if the agent's period* 2 *wage conditional on the period* 1 *outcome can assume two values with positive probability, the agent will wish to save money.*

PROOF: The following property is used to prove the proposition. Let $(d_1,\dots,d_N)$ be N nonnegative numbers and let $(\theta_1,\dots,\theta_N)$ be N probability weights, i.e., the θ_i's are nonnegative and sum to one. Then

$$(5.1)\qquad \frac{1}{\sum_{k=1}^{N}\frac{\theta_k}{d_k}}\geqslant\sum_{k=1}^{N}\theta_k d_k$$

with the inequality being strict if there exists $i,j\in\{1,\dots,N\}$ such that $d_i\neq d_j$, $\theta_i>0$ and $\theta_j>0$. This follows immediately from Jensen's inequality and the fact that $1/x$ is a convex function of x for $x\geqslant 0$. In particular, if we choose θ_k to be $p_k(s_j)$ and d_k to be $v'(w_{jk})$ then (5.1) becomes

$$(5.2)\qquad \frac{1}{\sum_{k=1}^{N}\frac{p_k(s_j)}{v'(w_{jk})}}\geqslant\sum_{k=1}^{N}p_k(s_j)v'(w_{jk}).$$

[7] See Braverman and Stiglitz [1] for a fuller discussion of this idea.

The agent's marginal discounted expected utility with respect to saving is

$$\frac{\beta}{\alpha}\left(\sum_{k=1}^{N} p_k(s_j)v'(w_{jk})\right) - v'(w_j). \tag{5.3}$$

The goal of the proof is to show that this is nonnegative. Invert both sides of (2.2) to yield

$$v'(w_j) = \frac{\beta}{\alpha}\frac{1}{\sum_{k=1}^{N}\frac{p_k(s_j)}{v'(w_{jk})}}. \tag{5.4}$$

The result now follows from (5.4) and (5.2). *Q.E.D.*

6. PRIVATE INFORMATION

The results of this paper generalize to situations where the agent has private information on the likelihood of various outcomes occurring. To model this, the following additions and changes in notation will be made. Let y_i be a private signal that the agent receives each period. It can assume one of M values:

$$Y = \{y_1, \ldots, y_M\}. \tag{7.1}$$

Let $p_{ji}(a)$ be the probability of outcome x_j given signal y_i is observed and action a is taken. Let q_i be the probability of signal i being observed in any period.

Two cases can now be considered depending upon whether or not the agent is allowed to transmit a report on what signal he has observed to the principal. First, the case of no information transmission will be considered. In this case, the agent's strategy space for a single period can be viewed as the set for all functions from Y to A., i.e., the agent chooses a rule that specifies his action contingent on his signal. Let F denote this strategy space and let f denote an element of F. For a single period, both the principal's and agent's expected utility can be written as functions of f and w, the wage contract. Furthermore, if the agent's utility function is separable in w and a, his expected utility is also separable in w and f. Therefore, the analysis of this paper applies immediately to this case when F is viewed as the agent's action space. Diamond and Mirrlees [**2**] derive a version of Proposition 4 in a model very similar to this under the assumption that there are two possible signals and there is no moral hazard (i.e., the action space has one element).

The second case is where the agent is allowed to communicate with the principal. Each period the agent communicates a (possibly untruthful) report of what signal he has observed to the principal. In a two period world a contract would consist of $MN+(MN)^2$ numbers, where w_j^i is the wage paid in period 1 if outcome x_j occurs and the agent reports y_i and w_{jl}^{ik} is the wage paid in period 2 if outcome x_j is followed by x_l and if the signal report y_i is followed by y_k. Let s^i be the agent's action choice in period 1 if he observes y_i; let s_j^{ik} be the agent's action choice in period 2 if he observes y_i followed by y_k and if x_j occurs in period 1.

It will be assumed without loss of generality that the wage contract is always chosen so that the agent reports his observations truthfully.[8] By using the same technique as in Proposition 1 it is straightforward to show that the analogous result still holds. Namely, for any Pareto-optimal contract it must be true that

$$\frac{1}{v'(w_j^i)} = \frac{\alpha}{\beta} \sum_{k=1}^{M} \sum_{l=1}^{N} \frac{q_k p_l(s_j^{ik})}{v'(w_{jl}^{ik})} \tag{6.2}$$

for every $i \in \{1, \ldots, M\}$ and $j \in \{1, \ldots, N\}$. Propositions 2, 3, and 4 can now be derived for this more complicated environment by using (6.2) instead of (2.2).

Stanford University

Manuscript received November, 1982; revision received January, 1984.

REFERENCES

[1] BRAVERMAN, AVISHAY, AND JOSEPH STIGLITZ: "Sharecropping and the Interlinking of Agrarian Markets," *American Economic Review*, 72(1982), 596–715.
[2] DIAMOND, PETER, AND JAMES MIRRLEES: "A Model of Social Insurance with Variable Retirement," *Journal of Public Economics*, 10(1978), 295–336.
[3] LAMBERT, RICHARD: "Long Term Contracts and Moral Hazard," mimeo, 1982.
[4] MIRRLEES, J. A.: "The Theory of Moral Hazard and Unobservable Behavior," Part I, mimeo, 1975.
[5] MYERSON, ROGER: "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 47(1979), 61–74.
[6] RADNER, ROY: "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship," Discussion Paper 184, Bell Laboratories, 1981.
[7] ———: "A Repeated Principal-Agent Game With Discounting," mimeo, 1981.
[8] ———: "Optimal Equilibria in Some Repeated Games with Imperfect Monitoring," mimeo, 1981.
[9] ROGERSON, WILLIAM P.: "The Role of Reputation in a Repeated Agency Problem Involving Information Transmission," IMSSS Technical Report No. 377, 1982.
[10] ———: "The Structure of Wage Contracts in Repeated Agency Models," IMSSS Technical Report No. 388, 1982.
[11] ———: "The First-Order Approach to Principal Agent Problems," *Econometrica*, forthcoming.
[12] RUBINSTEIN, ARIEL: "An Optimal Conviction Policy for Offenses that May Have Been Committed by Accident," Research Report No. 114, Hebrew University Department of Economics, 1979.
[13] RUBINSTEIN, ARIEL, AND MANAHEM YAARI: "Repeated Insurance Contracts and Moral Hazard," Memorandum No. 17, Centre for Research in Mathematical Economics and Game Theory, 1980.

[8] See Myerson [5] for a discussion of this point which is called the revelation principle. Of course a more notationally cumbersome version of (6.2) would still hold true for contracts which did not induce truthful revelation.