Moral Hazard and Persistence

Hugo Hopenhayn	Arantxa Jarque
Department of Economics	Department of Economics
UCLA	U. of Alicante

PRELIMINARY AND INCOMPLETE

Abstract

We study a multiperiod principal-agent problem with moral hazard in which the agent is required to exert effort only in the initial period of the contract. The effort choice of the agent in this first period determines the conditional distribution of output in the following periods. The paper characterizes the optimal compensation scheme. We find that the results for the static moral hazard problem extend to this setting: consumption at each point in time is ranked according to the likelihood ratio of the corresponding history. As the length of the contract increases, the cost of implementing effort decreases, and consumption on the equilibrium path becomes less volatile. If the contract lasts for an infinite number of periods, assuming the effect of effort does not depreciate with time, the cost of the principal gets arbitrarily close to that of the first best.

1 Introduction

There is by now a fairly large literature on dynamic contracts. Most of this literature relies on a model of repeated moral hazard: each period, an agent takes an unobservable action that only affects the contemporaneous outcome, which is observed by the principal. The problem is to bring the agent to exert a certain level of effort, every period, at a minimum cost. There is a wide array of applications of these models in macroeconomics, industrial organization, or public finance. The lack of persistence of effort is an important limitation in some of these applications, such as in the literature on incomplete insurance due to asymmetric information (see, for example, [2] and [10]), CEO's optimal compensation ([13]), optimal design of loans for entrepreneurs ([1]), or the study of optimal unemployment insurance programs ([12],[7]). There is a reason for this gap in the literature: it is considered a very difficult problem. This paper studies a special problem of moral hazard with persistence that turns out to have an elementary solution. The key simplification is that the agent takes only one action, at time zero, with persistent effects.

The model is as follows. At time zero, a principal offers a contract to an agent, specifying consumption contingent on a publicly observable history of states. The conditional distribution for these states is affected by the agent's choice of effort at time zero, which can take two values: low and high. The agent has time separable, strictly concave utility with discounting. The principal is risk neutral. For simplicity we assume principal and agent have the same discount factor. The problem is to implement high effort at the lowest expected discounted cost.

In spite of its dynamic structure, this problem reduces to a standard static moral hazard case. The intuition for this result is the following. The information structure for the standard moral hazard case is given by a set of states and probability distributions over these states, conditional on the actions. The agent maximizes expected utility, which is a convex combination of the utility associated to each state with the corresponding probabilities. Let the state in the dynamic case be the set of all histories (all possible nodes in the tree) with probabilities that are adjusted by the corresponding discount factors and normalized to add up to one. The expected discounted utility of any contingent consumption reduces to a convex combination of the utilities in each of these states, with these adjusted weights.

The optimal compensation scheme is derived as in the static moral hazard problem: all histories -regardless of time period- are ordered by likelihood ratios, and the assigned consumption is a monotone function of this ratio. As in the static case, compensation will be monotone in the past realizations of output only if likelihood ratios are so, i.e. if the monotone likelihood ratio property holds for all histories.

The model has some simple predictions. Longer histories contain obviously more information, so the dispersion of likelihood ratios and compensation increases over time. Extending the number of periods reduces the cost of implementing high effort and reduces the variance of compensation in earlier periods. When realizations are iid over time, the wage given to the agent in the previous period of the contract, his tenure and current output are sufficient to determine his current compensation. Moreover, in this iid framework the cost of the contract approaches the first best as the number of time periods goes to infinity. This result is explained by the fact that the variance of likelihood ratios goes to infinity with time, so asymptotically, deviations can be statistically discriminated at no cost. The paper considers the effect of reducing persistence by examining a specification where the effect of the initial action depreciates over time. As persistence decreases, the variability of compensation increases and so does the cost of implementation.

The paper is organized as follows. A characterization of the optimal contract is given for the general model in the next section. Results and numerical examples are discussed in sections 3 and 4, for the iid case and the model with decreasing persistence respectively. In section 5 the case of infinite contracts is discussed.

2 The Model

The relationship between the principal and the agent lasts for T periods, where T is possibly infinite. There is the same finite set of possible outcomes each period, $\{y_i\}_{i=1}^{n}$. Let Y^t denote the set of histories of outcome realizations up to time t, with typical element $y^t = \{y_{0,}y_1, ..., y_t\}$. This history of outcomes is assumed to be common knowledge. The agent can exert two possible levels of effort, $\{e_L, e_H\}$, with $e_L < e_H$.[1] A contract prescribes an effort to the agent at time 0, as well as a transfer c_t from the principal to the agent, contingent on the history of outcomes up to the present time: $c_t : Y^t \rightarrow \mathbb{R}_+$,for $t = 1, 2, ..., T$.[2] Each period, the probability of a given history of outcomes is conditional on the effort level chosen at time zero: $\Pr(y^t)$ if the effort is e_H, and $\widehat{\Pr}(y^t)$ if the effort is e_L. Both the agent and the principal discount cost and utility at the same rate β. Commitment to

[1]As it becomes clear in the core of the paper, the results presented here generalize to the case of multiple effort levels as much as those of a static moral hazard problem. That is, it may be that some of the levels are not implementable, and for a continuum of efforts we would need to rely on the validity of the first order approach to find the optimal contract.

[2]Even though unlimited punishments are needed for the asymptotic results of the paper, the restriction on consumption is without loss of generality; we only need utility to be unbounded below.

the contract is assumed on both parts.

Assuming that parameters are such that it is profitable for the principal to implement the high level of effort, the cost of the contract is simply the expected discounted stream of consumption to be provided to the agent:

$$K = \sum_{t=0}^{\infty} \sum_{y^t} \beta^t \left\{ c\left(y^t\right) \right\} \Pr\left(y^t\right).$$

The expected utility that the agent gets from a given contract (an effort recommendation, e_H and contingent consumtion $\{c(y^t)\}_{t=0}^{\infty}$) that implements the high level of effort, equals his expected utility of consumption, minus the disutility of effort.

$$U\left(e; \left\{c\left(y^t\right)\right\}_{t=0}^{\infty}\right) = \sum_{t=0}^{\infty} \sum_{y^t} \beta^t u\left(c\left(y^t\right)\right) \Pr\left(y^t\right) - e_H.$$

Letting $\underline{U}$ denote the initial outside utility, the Participation Constraint (PC) reads:

$$\underline{U} \leq \sum_{t=0}^{\infty} \sum_{y^t} \beta^t u\left(c\left(y^t\right)\right) \Pr\left(y^t\right) - e_H, \tag{PC}$$

Given the moral hazard problem due to the unobservability of the effort taken by the agent, the standard Incentive Compatibility (IC) condition further constrains the choice of the contract:

$$\begin{aligned} & \sum_{t=0}^{\infty} \sum_{y^t} \beta^t u\left(c\left(y^t\right)\right) \Pr\left(y^t\right) - e_H \\ \geq & \sum_{t=0}^{\infty} \sum_{y^t} \beta^t u\left(c\left(y^t\right)\right) \widehat{\Pr}\left(y^t\right) - e_L. \end{aligned} \tag{IC}$$

In words, the expected utility of the agent when choosing the high level of effort should be at least as high as the one from choosing the low effort. In order to satisfy this constraint, the difference in costs of effort should be compensated by assigning higher consumption to histories that are more likely under the required action than under the deviation. Formally, the optimal contract is the solution to the following cost minimization problem:

$$\begin{aligned} & \min_{\{c(y^t)\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \sum_{y^t} \beta^t \left\{ c\left(y^t\right) \right\} \Pr\left(y^t\right) \\ & s.t. \;\; \text{PC and IC} \end{aligned}$$

The analysis of the infinite period contract is postponed until Section 5. The finite contract can be characterized by looking at the first order conditions of this problem. As in the static moral hazard case, the Likelihood Ratio of a history $LR(y^t)$ can be defined as the ratio of the probability of observing that history under a deviation, to the probability under the recommended level of effort:

$$LR\left(y^t\right) \equiv \frac{\widehat{\Pr}\left(y^t\right)}{\Pr\left(y^t\right)}.$$

Proposition 1 *The optimal sequence* $\{c_\tau\left(y^\tau\right)\}_{\tau=0}^{T}$ *of contingent consumption in the Second Best contract is ranked according to the likelihood ratios of the histories of profit realizations up to time* T*:*

$$c\left(y^t\right) > c\left(\widetilde{y}^t\right) \Leftrightarrow LR\left(y^t\right) < LR\left(\widetilde{y}^t\right)$$

Proof. Since utility is separable in consumption and effort, both the PC and the IC will be binding. From the FOC's,

$$\left(c\left(y^t\right)\right) : \frac{1}{u'\left(c\left(y^t\right)\right)} = \lambda + \mu\left[1 - \frac{\widehat{\Pr}\left(y^t\right)}{\Pr\left(y^t\right)}\right] \qquad \forall y^t, \tag{1}$$

where λ and μ are the multipliers associated with the PC and the IC respectively. Since $u'(\cdot)$ is decreasing, the result follows from the above set of equations. ■

As in the static problem, the contract tries to balance insurance and incentives. To achieve this optimally, punishments (lower consumption levels) are assigned to histories of outcomes that are more likely under a deviation than under the recommended effort.

The FOC's of the problem described in equation 1 can be combined to get the following expression:

$$\frac{1}{u'\left(c\left(y^t\right)\right)} = \sum_{y_t} \frac{1}{u'\left(c\left(y^t, y_{t+1}\right)\right)} \Pr\left(y^t, y_{t+1}\right). \tag{2}$$

As in Rogerson , this property implies that whenever the inverse of the marginal utility of consumption is convex, the agent, if allowed, would like to save part of his wage every period in order to smooth his consumption over time. Other properties discussed by Rogerson are also true in this setup, as indicated in the next proposition

Proposition 2 *In the Second Best contract, the expected wage decreases with time whenever* $\frac{1}{u'(\cdot)}$ *is convex, increases if it is concave, and is constant whenever utility is logarithmic. (Rogerson)*

Proof. As in the dynamic moral hazard problem studied by Rogerson, (2) holds. Take the case of $\frac{1}{u'(\cdot)}$ being convex. By Jensen's inequality,

$$u'\left(c\left(y^t\right)\right) < \sum_{y_t} u'\left(c\left(y^t, y_{t+1}\right)\right) \Pr\left(y^t, y_{t+1}\right).$$

Since utility is concave, this implies

$$c\left(y^t\right) > \sum_{y_t} c\left(y^t, y_{t+1}\right) \Pr\left(y^t, y_{t+1}\right).$$

A similar argument applies for the other two cases. ■

3 Outcomes Independently and Identically Distributed

In this section, we study a particular specification of the probability distribution of outcomes: an iid process. This assumption puts additional structure on the probability distribution of histories, and allows for the optimal contract to be further characterized.

Suppose there are two possible outcomes, $\{y_L, y_H\}$. If effort is high, let

$$\begin{aligned} \Pr\left(y_H\right) &= \pi \\ \Pr\left(y_L\right) &= 1-\pi \end{aligned}$$

If the low effort is chosen instead,

$$\begin{aligned} \widehat{\Pr}\left(y_H\right) &= \widehat{\pi} \\ \widehat{\Pr}\left(y_L\right) &= 1-\widehat{\pi}, \end{aligned}$$

with $\pi > \widehat{\pi}$.

In this specific setup, the optimal contract can be further characterized. First, we establish monotonicity of wages.

Proposition 3 *Given any history y^t of any finite length t, $c(y^t, y_H) > c(y^t, y_L)$. In other words, the wage of the agent increases when a new high realization is observed.*

Proof. The likelihood ratio that determines the ordering of consumptions in the optimal contract is also the product of the likelihood ratios of the individual one–period output realizations. Since

$$LR(y_H) = \frac{\widehat{\pi}}{\pi} < 1$$

and

$$LR(y_H) = \frac{1-\widehat{\pi}}{1-\pi} > 1$$

for a given history y^t, a high realization at time $t+1$ makes the likelihood ratio smaller than a low realization, implying higher consumption:

$$LR\left(y^t, y_H\right) = LR\left(y^t\right)\frac{\widehat{\pi}}{\pi} < LR\left(y^t\right)\frac{1-\widehat{\pi}}{1-\pi} = LR\left(y^t, y_L\right) \qquad \forall y^t.$$

■

Moreover, in this simple setup, the fraction of high realizations of the output in a history is a sufficient statistic for the history's probability. For any history $y^t = (y_1, y_2, \ldots, y_t)$, denote the number of high realizations contained in the history as $x(y^t)$.

Proposition 4 *For any two histories of the same length y^t and $\widetilde{y}^t$, $c(y^t) \geq c(\widetilde{y}^t)$ if and only if $x(y^t) \geq x(\widetilde{y}^t)$, regardless of the sequence in which the realizations occurred in each of the histories.*

Proof. Under the iid assumption, the probability of a given history is just the product of the probability of each of the realizations that conform it. In the iid case, the likelihood ratio of a history can be written as

$$LR\left(y^t\right) = \left(\frac{\widehat{\pi}}{\pi}\right)^{x\left(y^t\right)} \left(\frac{1-\widehat{\pi}}{1-\pi}\right)^{t-x\left(y^t\right)}.$$

Since $\frac{\widehat{\pi}}{\pi} < 1$ and $\frac{1-\widehat{\pi}}{1-\pi} > 1$, $x(y^t) \geq x(\widetilde{y}^t)$ implies $LR(y^t) < LR(\widetilde{y}^t)$. The result follows from the ordering of consumption imposed by the first order conditions. ■

In other words, there is perfect substitutability of output realizations across time. This implies, in fact, that we do not need to carry all the information contained in the history of realizations. Faced with a current output realization y_t following a given history y^{t-1}, we only need to know $x\left(y^{t-1}\right)$ to determine current compensation. Simply put, the wage received by the agent in the previous period, together with his tenure in the contract are sufficient to determine his current wage.

Under the iid assumption, time gives us some information about the ranking of the likelihood ratios of histories of different length. We can translate this temporal structure into two formal properties of the contract. First, the range of variation in wages within a certain period is larger in the latter periods of the contract. Denote $\overline{y}^t$ the history at t with all high outcomes, that is, $x\left(\overline{y}^t\right) = t$. Similarly, $\underline{y}^t$ denotes the one with all low outcomes, with $x\left(\underline{y}^t\right) = 0$:

$$\begin{aligned}\overline{y}^t &= \left(y_H, y_H, \ldots, y_H\right) \\ \underline{y}^t &= \left(y_L, y_L, \ldots, y_L\right),\end{aligned}$$

both of length t.

Proposition 5 *As t increases, $c\left(\overline{y}^t\right)$ increases and $c\left(\underline{y}^t\right)$ decreases.*

Proof. In the IID case, for a length t, the lowest likelihood ratio will be that of the history with t high realizations of output:

$$LR\left(\overline{y}^t\right) = \frac{\widehat{\pi}^t}{\pi^t}.$$

For the highest, instead, will be

$$LR\left(\underline{y}_t\right) = \frac{\left(1-\widehat{\pi}\right)^t}{\left(1-\pi\right)^t}.$$

Given that $\frac{\widehat{\pi}}{\pi} < 1$ and $\frac{(1-\widehat{\pi})}{(1-\pi)} > 1$,

$$\begin{aligned}LR\left(\overline{y}^t\right) &> LR\left(\overline{y}^{t+1}\right), \\ LR\left(\underline{y}_t\right) &< LR\left(\underline{y}_{t+1}\right).\end{aligned}$$

The result follows from the ranking established by the first order conditions. ■

Corollary 6 *As t increases, $d_t = c\left(\overline{y}^t\right) - c\left(\underline{y}^t\right)$ increases.*

The difference between the highest and the lowest wage within a period, denoted here by d_t, increases with time; this follows simply from Prop. 5. For the case of logarithmic utility, which we use in the numerical examples reported later in the paper, consumption changes linearly with likelihood ratios:

$$\begin{aligned}
\overline{c}_{t+1} - \overline{c}_t &= LR\left(\overline{y}^{t+1}\right) - LR\left(\overline{y}^t\right) \\
\underline{c}_t - \underline{c}_{t-1} &= LR\left(\underline{y}_t\right) - LR\left(\underline{y}_{t+1}\right).
\end{aligned}$$

3.1 Changes in the Length of the Contract

In the iid context analyzed here, longer contracts have a better information structure that allows the implementation of high effort at a lower cost per period. We make explicit this point in the next proposition. Below, we present numerical results for an example that illustrates the effect of contract length on the variation in the optimal consumption.

To isolate the effect of better information, we conduct comparative statics on the number of periods in the contract. We wish to abstract from the fact that it is always cheaper to provide a given level of utility in several periods than in one period, simply because the agent's utility function is concave. We therefore change the specification of the model, so that the participation constraint does not become looser when increasing the contract length T. As T grows, we renormalize the agent's outside utility, $\underline{U}(T)$, so that the per period utility $\underline{U}$ remains constant

$$\underline{U}(T) = \frac{1-\beta^{T-1}}{1-\beta}\underline{U}.$$

The same normalization is used for the disutility of effort, which now depends on T:

$$e_H(T) = \frac{1-\beta^{T-1}}{1-\beta}e_H.$$

The per period cost of the contract is defined in the obvious way:

$$K(T) = \frac{1-\beta}{1-\beta^{T-1}}\sum_{t=1}^{T}\beta^t c\left(y^t\right)\Pr\left(y^t\right).$$

Proposition 7 *Longer contracts have a lower per period cost.*

Proof. Let $n(t)$ denote the number of time-t histories and $y_{jt},\ j = 1, ..., n(t)$ a typical t-period history. Let p_{jt} and $\tilde{p}_{jt}$ denote the probabilities under high and low efforts for these histories. We will now define a static moral hazard problem that is identical to this dynamic one. Let the state space $S = \cup_t \left\{y_{1t}, ..., y_{n(t)t}\right\}$. Define probabilities $q_{jt}^T = \beta^t p_{jt}/\Delta(T)$, where

$$\Delta(T) = \frac{1-\beta^T}{1-\beta}.$$

Similarly define probabilities $\tilde{q}_{jt}^T$ with distribution $\tilde{p}_{jt}$. The optimal contract can be rewritten as:

$$K(T) = \frac{1}{\Delta(T)} \min \sum_{j,t} q_{jt}^T w_{jt}$$

subject to:

$$\begin{aligned} \sum_{j,t} u(w_{jt}) q_{jt}^T - e_H(T)/\Delta(T) &\geq \underline{U}(T)/\Delta(T) \\ \sum_{j,t} u(w_{jt}) \left(q_{jt}^T - \tilde{q}_{jt}^T\right) &\geq \left(e_H(T) - e_L(T)\right)/\Delta(T). \end{aligned}$$

Now suppose we decrease the number of time periods to $T-1$. Define a static problem by using the same state space as above, but zero probability to those states that correspond to T–period histories and probabilities q_{jt}^{T-1} and $\tilde{q}_{jt}^{T-1}$ to all other histories as done above, but using $\Delta(T-1)$ instead of $\Delta(T)$. The corresponding $T-1$ period moral hazard problem is equivalent to the above, replacing $\Delta(T-1)$ for $\Delta(T)$. (It follows from our definitions that $e(T)\Delta(T) = e(T-1)\Delta(T-1)$ and the same holds for $\underline{U}(T)$.) It is easy to show that the information structure $\left(q_{jt}^T, \tilde{q}_{jt}^T\right)$ is sufficient for $\left(q_{jt}^{T-1}, \tilde{q}_{jt}^{T-1}\right)$ (in the sense of Blackwell), so by Grossman and Hart [3] it follows that:

$$\Delta(T) K(T) \leq \Delta(T-1) K(T-1),$$

which completes the proof. ■

The intuition for this result hinges on the better quality of the signal structure of the problem when the contract is longer. As already established by Holmstrom [5], any informative signal is valuable. Under the assumption of an iid process determined by the initial effort, a longer contract translates into a greater number of informative signals. When more periods are available, information quality increases and incentives can be given more efficiently: the wage scheme calls for less variation in the early periods, since punishments in later periods are exercised with lower probability on the equilibrium path.

T = 1	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t=1$	2.7	1.30	19.1
$t=2$		-	-
$t=3$		-	-
$t=4$		-	-

T = 2	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t=1$	1.7	0.66	2.5
$t=2$		0.94	32.7
$t=3$		-	-
$t=4$		-	-

T = 3	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t=1$	1.4	0.43	1.3
$t=2$		0.61	4.6
$t=3$		0.75	57.9
$t=4$		-	-

T = 4	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t=1$	1.3	0.30	0.8
$t=2$		0.44	2.1
$t=3$		0.54	5.3
$t=4$		0.63	132.1

Table 1. Changes in length of contract: effect on variability of consumption.

Table 1 presents an example[3] of the effect of increasing T on expected wage and on two different measures of variability of consumption. The solution for consumption under observable effort can be used as a benchmark to evaluate the changes in cost, as the contract length is modified. When effort is observable, the optimal contract (sometimes referred to as the First Best) can achieve perfect insurance, which minimizes the cost of delivering the outside utility level. The constant wage c^* in the First Best satisfies:

$$\widetilde{\underline{U}} + e_H = u\left(c^*\right).$$

Expected consumption represents as well the per period cost of the contract. Since we use logarithmic utility in the example, as established in Prop. 2, $E\left[c_t\right]$ is constant for a given T.

[3] Parameters of the example: $T=4$, $\beta=.95$, $\underline{U}=1$, $e_H=.7$, $e_L=.4$, $\pi=.3, \widehat{\pi}=.2$. Utility is logarithmic.

The table contains four matrices, each corresponding to a different length T, going from one to four. All other parameters are kept equal. Expected consumption is reported in the first column of each matrix. For comparison purposes, we divide $E[c]$ by consumption in the First Best contract. The expected consumption decreases from 2.7 times the constant consumption in the First Best, when $T = 1$, to 1.3 times when $T = 4$.

The second and the third column present the variability measures. Again, to keep the values comparable, we divide the standard deviation by expected consumption in the contract, $\frac{\sigma_t}{E[c]}$. The third column reports the difference between the highest and the lowest consumption levels in a given period, d_t, relative to the lowest consumption:

$$\frac{d_t}{\underline{c}_t} = \frac{\overline{c}_t - \underline{c}_t}{\underline{c}_t}.$$

This normalization captures the fact that, since utility is concave, the same d_t represents more variation in utility when the levels of consumption are lower. Both σ and d vary across periods within the same matrix, and across matrices. The pattern is the same for both: when T increases, they decrease (σ_1 goes from 1.30 in the $T = 1$ contract to 0.3 in the $T = 4$ one; for the same observations, d_1 changes from 19.1 to 0.8). Looking at any given matrix instead, for example the one corresponding to $T = 4$, we see that both measures of variability increase with t. In the first period of the contract, the scaled standard deviation of consumption is 0.3, and the scaled difference of consumption is 0.8. At the fourth period of the contract, the values increase to 0.63 and 132.1 correspondingly.

Both the change across different lengths and within contracts respond to the better quality of information. Variation in consumption is necessary for implementing high effort. It is efficient to concentrate most of this variation in longer histories for which relative likelihoods are more informative. In particular, the bigger change in the scaled d_t compared to that of the scaled σ_t is explained by the fact that d_t measures maximum variation within a period, and not expected variation. For long histories, the probability of the history associated with $\min(c_t)$ becomes much smaller under the equilibrium effort than under the deviation. Punishments are exercised less often in equilibrium, lowering the risk premium and thus reducing the cost of the contract.

In Fig.1 the effect of time on the multiplier of the IC, μ, is plotted for the same example as analyzed in Table 1. The numerical examples all obtain a

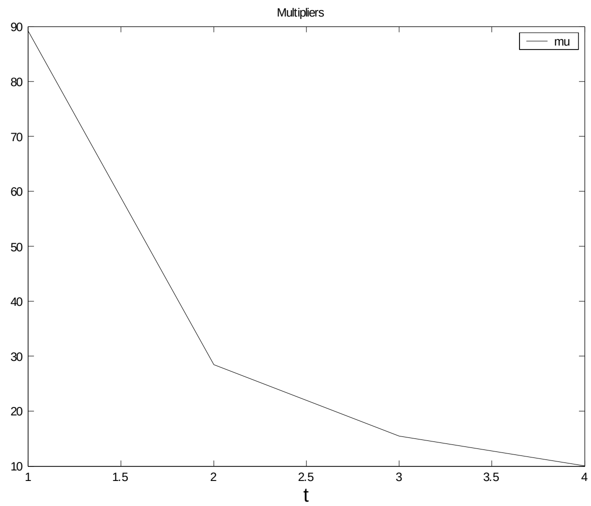

Figure 1: Effect of time on the multiplier of the IC.

decrease in μ with time. Although we do not have a formal proof for this effect, the discussion above supports it intuitively. The IC is easier to satisfy as the length of the contract increases. The availability of better quality information is materialized in more extreme values of the likelihood ratios. In fact, the patterns for the variability of consumption we just described can be understood in terms of decrease of μ. Just recall the FOC's of the Second Best. Now, rearranging them, we have that for any two histories y^t and $\widetilde{y^t}$ of any length,

$$\frac{1}{u'\left(c\left(y^t\right)\right)} - \frac{1}{u'\left(c\left(\widetilde{y^t}\right)\right)} = \mu\left(\frac{\widehat{\Pr}\left(\widetilde{y^t}\right)}{\Pr\left(\widetilde{y^t}\right)} - \frac{\widehat{\Pr}\left(y^t\right)}{\Pr\left(y^t\right)}\right).$$

For the logarithmic utility used in the example, this means that the difference in consumption is proportional to the difference in the likelihood ratios. The factor of proportionality is μ. A lower multiplier for longer contracts delivers the general decrease in variability, since the sensitivity of compensation to the likelihood ratios is smaller. However, likelihood ratios take on more extreme values for long histories. For the most favorable history with T successes,

$$LR\left(y^T\right) = \left(\frac{\widehat{\pi}}{\pi}\right)^T,$$

and since $\frac{\widehat{\pi}}{\pi} < 1$,this likelihood ratio tends to zero. For the one with zero high realizations,

$$LR\left(y^T\right) = \left(\frac{1-\widehat{\pi}}{1-\pi}\right)^T,$$

and since $\frac{1-\widehat{\pi}}{1-\pi} > 1$, the likelihood ratio tends to infinity. This differentiated effect in the likelihood ratios implies that variability is concentrated in latter periods, and the lowest consumption is associated with very unlikely histories.

4 Decrease in persistence of effort

The iid assumption allows for a more complete characterization of the contract, but it implies a very strong concept of effort persistence. In this section, we propose a modified stochastic structure that still preserves the tractability of the solution, but relaxes the assumption of "perfect" persistence.

The effect of the action may now decrease as time passes. We model this by making the probability of the high level of output a convex combination of the effort-determined probability, π or $\widehat{\pi}$, and an exogenously determined probability (i.e., independent of the agent's effort choice), denoted by $\overline{\pi}$:

$$\Pr_t\left(y_H\right) = \left(\alpha_t \pi + \left(1-\alpha_t\right)\overline{\pi}\right).$$

The sequence of weights, $\{\alpha_t\}_{t=1}^T$ with $0 \leq \alpha_t \leq 1$ for every t, represents the rate at which the effect of effort diminishes. This gives a measure of persistence of effort: $\alpha_t = 1$ for all t corresponds to the iid case of perfect persistence, while $\alpha_t = 0$ for all t implies that effort does not affect the distribution of output. We consider a sequence for α_t such that $0 < \alpha_t < 1$ for every t, and decreases over time. The effects of time on cost described in the previous section still hold as long as $\alpha_t > 0$, i.e., as long as there is some information contained in new realizations, the principal is better off when contracts are longer. Also, the changes in the properties of the contract with T are as described in the previous section.

Lowering persistence worsens the quality of information available. The effect on cost parallels that of shortening the contract, as established in the following proposition:

Proposition 8 *Consider two possible persistence sequences* $(\alpha_1, ..., \alpha_T)$ *and* $(\alpha'_1, ..., \alpha'_T)$ *where* $\alpha_t \geq \alpha'_t$ *for all* t*. The cost of the contract is lower for the problem with higher persistence,* $(\alpha_1, ..., \alpha_T)$*.*

Proof. Consider any two sequences of probabilities $\{P_t(y_t)\}$, $\{P_t^0(y_t)\}$ where $p_t^0(y_t) = \gamma_t p_t(y_t) + (1-\gamma_t)\pi_t(y_t)$, for $0 \leq \gamma_t \leq 1$. Let $P_0(y_1,...,y_t)$ and $P(y_1,...,y_t)$ be the corresponding probability distributions over histories for these processes with probabilities $\{p_t^0\}$ and $\{p_t\}$, respectively. It follows that:

$$P_0(y_1,...,y_t) = \Pi_j \left[\gamma_j p_j(y_j) + (1-\gamma_j)\pi_j(y_j)\right].$$

A typical element in the expansion of this product has the form:

$$\gamma_\theta \Pi_{j\in\Theta\subset\{1,..,t\}} p_j(y_j)$$

where γ_θ is some coefficient that varies with the subset θ of terms considered (the constant term corresponds to $\theta = \varnothing$.) This term can be obtained integrating out all histories that coincide on this subset of realizations $\{y_j | j \in \Theta\}$ and can thus be expressed as a linear combination of probabilities $P(y_1,...,y_t)$ for all these realizations. It follows that $P_0(y_1,...,y_t)$ is also a linear combination of probabilities defined by P and thus the information system defined by P is sufficient for P_0 and the corresponding implementation cost to the principal is lower. Now consider two information structures with same baseline probabilities but different weights $(\alpha_1,...,\alpha_T), (\alpha_1',...,\alpha_T')$ where $\alpha_t \geq \alpha_t'$ for all t. Letting $p_t(y_t)$ denote the probability defined by the first information structure and $\gamma_t = \alpha_t'/\alpha_t$ the previous result can be applied showing that the the first information structure is sufficient for the second and the corresponding cost of implementation lower. ■

For the numerical examples presented, we choose to have α_t decrease exponentially:

$$\alpha_t = \alpha^t \ \forall t.$$

Fixing $T = 4$, we describe changes in the optimal contracts for the example of Table 1 under different levels of persistence. Table 2 contains four matrices, with α ranging from 1 to 0.48; all other parameters are kept equal.

$\boldsymbol{\alpha = 1.00}$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t=1$	1.3	0.30	0.8
$t=2$		0.44	2.1
$t=3$		0.54	5.3
$t=4$		0.63	132.1

$\boldsymbol{\alpha = 0.83}$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t=1$	1.5	0.42	1.3
$t=2$		0.55	3.3
$t=3$		0.63	8.9
$t=4$		0.68	415.5

$\boldsymbol{\alpha = 0.65}$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t=1$	1.7	0.58	2.0
$t=2$		0.69	5.7
$t=3$		0.74	16.5
$t=4$		0.76	953.2

$\boldsymbol{\alpha = 0.48}$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t=1$	2.0	0.78	3.5
$t=2$		0.87	11.1
$t=3$		0.89	36.5
$t=4$		0.89	754.2

Table 2. Changes in persistence of effort: effect on variability of consumption.

When persistence is lower, the effect on the variability of consumption parallels that of a decrease in T. When α is lower, the quality of information decreases as it does when the length is shorter. Both the scaled standard deviation and the difference between highest and lowest consumption increase significantly when persistence is lower. The expected consumption increases when α decreases (it goes from the original 1.3 of the First Best cost when $\alpha = 1$ to being twice of the First Best when α=.48), reflecting the increase in the risk premium due to the higher variability.

5 Asymptotic Optimal Contract

If the principal and the agent could commit to an infinite contractual relationship, assuming that output is distributed iid, and utility is unbounded below (as in the logarithmic case) the cost of the contract under moral hazard could get arbitrarily close to that of the First Best, i.e., under observable effort.

Consider a “one-step" contract, a tuple $(c^*, \underline{c}, L)$ of two possible consumption levels c^* and $\underline{c}$ plus a threshold L for the Likelihood Ratio. The contract is defined in the following way:

$$c(y^t) = \begin{cases} c^* & if \;\; LR(y^t) < L \\ \underline{c} & if \;\; LR(y^t) \geq L \end{cases} .$$

Proposition 9 *For any $\beta \in (0,1]$ and any $\varepsilon > 0$, there exists a one-step contract $(c^*, \underline{c}, L)$ such that the principal can implement high effort at a cost $C < C^o + \varepsilon$, where C^o is the cost when effort is observable.*

Proof. Let

$$\begin{aligned} F_t(L) &= P\left\{ y^t | L\left(y^t\right) \equiv \frac{\tilde{P}(y^t)}{P(y^t)} \leq L \right\} \\ \tilde{F}_t(L) &= \tilde{P}\left\{ y^t | L\left(y^t\right) \equiv \frac{\tilde{P}(y^t)}{P(y^t)} \leq L \right\} \end{aligned}$$

It is easy to verify by definition that $\left(1 - \tilde{F}_t(L)\right) \geq L\left(1 - F_t(L)\right)$. This implies that

$$\begin{aligned} F_t(L) - \tilde{F}_t(L) &= \left(1 - \tilde{F}_t(L)\right) - \left(1 - F_t(L)\right) \\ &\geq (L-1)\left(1 - F_t(L)\right) \end{aligned}$$

and also

$$F_t(L) - \tilde{F}_t(L) \geq \frac{\left(1 - \tilde{F}_t(L)\right)}{L}.$$

Let $\delta > 0$. For any L, define P so that[4]:

$$e_H - e_L = P \sum_{t=0}^{\infty} \beta^t \left(F_t(L) - \widetilde{F}_t(L) \right) \tag{3}$$

From the above inequalities, this will hold whenever $1 - \widetilde{F}_t(L) > 0$ for some t. For the discrete case, this occurs if there exists a path y^t such that $L\left(y^t\right) > L$, which is guaranteed in the iid case. Let c_t be the consumption plan where $u(c_t(y^t)) = u_0 + \delta$ whenever $L\left(y^t\right) \leq L$ and $u\left(c_t\left(y^t\right)\right) = u_0 + \delta - P$ otherwise.

[4] This definition requires that $\sum \beta^t F_t(L) > \sum \beta^t \widetilde{F}_t(L)$.

It is easy to verify that by the above definition of P, this plan is incentive compatible. Moreover, the utility of the agent is equal to:

$$U = \frac{u_0 + \delta}{1 - \beta} - P \sum_{t=0}^{\infty} \beta^t \left(1 - F_t(L)\right) \tag{4}$$

Rewriting the IC constraint and using the above inequality:

$$\begin{aligned} e_H - e_L &= P \sum_{t=0}^{\infty} \beta^t \left(F_t(L) - \widetilde{F}_t(L) \right) \\ &= P \sum_{t=0}^{\infty} \beta^t \left(1 - \widetilde{F}_t(L) - (1 - F_t(L)) \right) \\ &\geq P \sum_{t=0}^{\infty} \beta^t \left(1 - F_t(L)\right)(L - 1) \end{aligned}$$

Combining this inequality with equation (4) it follows that:

$$U \geq \frac{u_0 + \delta}{1 - \beta} - \frac{e_H - e_L}{L - 1}.$$

The cost of this plan $C(\delta, P) \leq \frac{c^{-1}(u_0+\delta)}{(1-\beta)}$. For any δ, take sufficiently large L so that $U > u_0/(1-\beta)$. This gives a plan that is incentive compatible and satisfies the participation constraint. Taking the limit as δ goes to zero, the cost of the plan converges to the cost of the first best plan. ∎

This result requires the assumption that the principal has unlimited punishment power, that is, that the utility of the agent can be made as low as wanted. Also, it is derived under the assumption of extreme persistence of effort (i.e., $\alpha = 1$). This is in fact what allows the quality of the information to keep on growing and reach levels that permit to tailor punishments so that they are almost surely not exercised in equilibrium.

6 Conclusions

We study a simple representation of persistence in which only one effort is taken by the agent at the begining of the contract. This effort determines the probability distribution of outcomes in all the periods to come. In spite of its dynamic structure, the contract has a very simple solution. Time is just an extra element in the signal structure, and the problem can be understood as a static moral hazard: the optimal compensation scheme determines that all histories -regardless of time period- are ordered by likelihood ratios, and the assigned consumption is a monotone function of this ratio. As in the static case, compensation will be monotone in the past realizations of output only if likelihood ratios are so, i.e. if the monotone likelihood ratio property holds for all histories.

Relying on comparisons of the quality, or informativeness, of probability distributions, several conclusions are reached: longer contracts have a lower cost of implementing high effort, and a lower variance of compensation in earlier periods. When realizations are iid over time, the cost of the contract approaches the first best as the number of time periods goes to infinity. As persistence decreases, the variability of compensation increases and so does the cost of implementation.

References

[1] Albuquerque, R. and H. Hopenhayn, 2004. "Optimal Lending Contracts and Firm Dynamics," *Review of Economic Studies*, vol. 71(2), pages 285-315.

[2] Atkeson, A. "International Lending with Moral Hazard and Risk of Repudiation", *Econometrica*, **59** (1991), 1069-1089.

[3] Grossman, Sanford and Oliver D. Hart. "An Analysis of the Principal–Agent Problem." *Econometrica* 51, Issue 1 (Jan.,1983), 7-46.

[4] Holmström, B. and P. Milgrom. "Aggregation and Linearity in the Provision of Intertemporal Incentives", *Econometrica*, 55, 303-328. (1987)

[5] Bengt Holmstrom. "Moral Hazard and Observability," *Bell Journal of Economics*, Vol. 10 (1) pp. 74-91. (1979)

[6] Holmström, B. "Managerial Incentive Problems: a Dynamic Perspective." *Review of Economic Studies*, 66, 169-182. (1999)

[7] Hopenhayn, H. and J.P. Nicolini: "Optimal Unemployment Insurance", *Journal of Political Economy*, **105** (1997), 412-438.

[8] Miller, Nolan."Moral Hazard with Persistence and Learning", Manuscript. (1999)

[9] Mirrlees, James. "Notes on Welfare Economics, Information and Uncertainty", in M. Balch, D. McFadden, and S. Wu (Eds.), *Essays In Economic Behavior under Uncertainty*, pgs.. 243-258. (1974)

[10] Phelan, C., Repeated Moral Hazard and One–Sided Commitment. *J. Econ. Theory* **66** (1995), 468-506.

[11] Rogerson, William P. "Repeated Moral Hazard", *Econometrica*, Vol. 53, No. 1. (1985), pp. 69-76.

[12] Shavell, S. and L. Weiss: "The Optimal Payment of Unemployment Insurance Benefits over Time", *Journal of Political Economy*, **87** (1979), 1347-1362.

[13] Wang, C. "Incentives, CEO compensation, and Shareholder Wealth in a Dynamic Agency Model," *Journal of Economic Theory*, 76, 72-105 (1997)

7 Algorithm for Numerical Examples

The program computes the optimal wage contract that the principal should set in order to implement the high level of effort at the minimum cost. The set of parameters consists of the physical environment and the preferences. The first can be described by the length of the contract (T), the probability structure (π and $\widehat{\pi}$, as well as $\overline{\pi}$) and the persistence sequence $(\{\alpha^t\}_{t=1}^T$, where α takes values between 0 and 1). The outside utility of the agent $(\underline{U})$, the discount factor (β), the disutility of both levels of effort (e_L and e_H), together with the choice of the logarithmic utility function, conform the choice for the preferences of the agent and the principal.

The code starts by computing the probabilities of all the possible histories given the length T. Then for each $t < T$ it uses the foc's of the minimization problem and the two binding constraints to find the solution for the optimal contract of length t. A solution exists for the contract for any given set of parameters, since the equality constraints are linear in the level of utility provided after each history and the objective function is strictly convex in it.

Given a guess for the multipliers (λ_0, μ_0), using the foc's a corresponding guess for all contingent consumptions can be calculated from the probabilities, using the likelihood ratios:

$$c_0\left(y^t\right) = \lambda_0 + \mu_0 \left(1 - \frac{\widehat{\Pr}\left(y^t\right)}{\Pr\left(y^t\right)}\right) \quad \forall y^t.$$

This set of implied consumptions can in turn be plugged into the PC and the IC. Since (λ_0, μ_0) do not typically coincide with the real solution for the multipliers, the two constraints are not met for the guess. The first step of the computational strategy is to construct a grid for possible values of μ. For each of this values, the value of λ that makes the PC closest to hold with equality is calculated using the bisection method; denote this temporary solution as $\lambda_{PC}(\mu_i)$, where μ_i is an arbitrary value on the grid. As an illustration, the PC for a one period contract is:

$$0 = \underline{U} + e_H - \pi \ln\left(\lambda + \mu\left(1 - \frac{\widehat{\pi}}{\pi}\right)\right) - (1-\pi)\ln\left(\lambda + \mu\left(1 - \frac{1-\widehat{\pi}}{1-\pi}\right)\right).$$

Holding μ constant,

$$\frac{\partial PC}{\partial \lambda_{PC}} = -\frac{\pi}{\lambda + \mu\left(1 - \frac{\widehat{\pi}}{\pi}\right)} - \frac{(1-\pi)}{\lambda + \mu\left(1 - \frac{1-\widehat{\pi}}{1-\pi}\right)} < 0,$$

showing that the error in the PC, denoted Δ_{PC}, changes monotonically with λ and thus the bisection method is appropriate. For an arbitrary μ_i,

$$\begin{aligned} \Delta_{PC} &= \underline{U} + e_H - \pi \ln\left(\lambda_{PC}(\mu_i) + \mu_i\left(1 - \frac{\widehat{\pi}}{\pi}\right)\right) \\ &\quad - (1-\pi)\ln\left(\lambda_{PC}(\mu_i) + \mu_i\left(1 - \frac{1-\widehat{\pi}}{1-\pi}\right)\right) \end{aligned}$$

There is a nonnegativity constraint on consumption, so the initial lower bound for the bisection, $\underline{\lambda}(\mu_i)$, is chosen the smallest possible; that is, to make the lowest consumption an ε bigger than zero. The history that receives the lowest consumption is that with the highest value for the likelihood ratio (in the case of an iid process and one period duration, it corresponds to $c(y_L)$):

$$\varepsilon = \underline{\lambda}(\mu_i) + \mu_i\left(1 - \max_{y^t}\left\{\frac{\widehat{\Pr}\left(y^t\right)}{\Pr\left(y^t\right)}\right\}\right).$$

The error Δ_{PC} is positive for such low value of λ; the upper bound of the bisection, $\overline{\lambda}(\mu_i)$,is chosen arbitrarily big in order to be sure that Δ_{PC} is negative and the bisection works properly. Starting with a $\overline{\lambda}(\mu_i)$ very far from the final solution does not seem to increase the computation time considerably.

The same is done for the IC, obtaining a (typically different) value $\lambda_{IC}(\mu_i)$. The program then proceeds to select the μ_i that minimizes the difference $[\lambda_{PC}(\mu_i) - \lambda_{IC}(\mu_i)]$, since in the solution they ought to be equal. The grid is refined around the selected μ_i and the loop is repeated until the two constraints are satisfied with equality up to numerical precision. By the implicit function theorem,

$$\frac{\partial \lambda_{PC}}{\partial \mu} = -\frac{\frac{\partial PC}{\partial \mu}}{\frac{\partial PC}{\partial \lambda_{PC}}} = \frac{\mu \frac{(\pi-\widehat{\pi})^2}{\pi(1-\pi)}}{\lambda + \mu \frac{(1-2\pi)(\pi-\widehat{\pi})}{\pi(1-\pi)}} > 0,$$

since it was already established that $\frac{\partial PC}{\partial \lambda_{PC}} < 0$ and

$$\frac{\partial PC}{\partial \mu} = -\frac{(\pi - \widehat{\pi})}{\lambda + \mu\left(1 - \frac{\widehat{\pi}}{\pi}\right)} + \frac{(\pi - \widehat{\pi})}{\lambda + \mu\left(1 - \frac{1-\widehat{\pi}}{1-\pi}\right)} > 0.$$

In the same way,

$$\frac{\partial IC}{\partial \lambda_{IC}} = (\pi - \widehat{\pi})\left[\frac{1}{\lambda + \mu\left(1 - \frac{\widehat{\pi}}{\pi}\right)} - \frac{1}{\lambda + \mu\left(1 - \frac{1-\widehat{\pi}}{1-\pi}\right)}\right] < 0$$

assures the monotonicity in the bisection search for $\lambda_{IC}(\mu_i)$. Together with the fact that

$$\frac{\partial IC}{\partial \mu} = (\pi - \widehat{\pi})^2\left[\frac{1}{\pi\left(\lambda + \mu\left(1 - \frac{\widehat{\pi}}{\pi}\right)\right)} + \frac{1}{(1-\pi)\left(\lambda + \mu\left(1 - \frac{1-\widehat{\pi}}{1-\pi}\right)\right)}\right] > 0,$$

it implies that

$$\frac{\partial \lambda_{IC}}{\partial \mu} = -\frac{\frac{\partial IC}{\partial \mu}}{\frac{\partial IC}{\partial \lambda_{IC}}} = \frac{\lambda}{\mu} > 0.$$

This monotonicity and that established by $\frac{\partial \lambda_{PC}}{\partial \mu} > 0$, ensure that for a big enough starting grid for μ, the procedure chooses the right pair of $\lambda's$ that minimize $[\lambda_{PC}(\mu_i) - \lambda_{IC}(\mu_i)]$, regardless of how fine the initial grid is around the solution point.

The graphics coming out of the simulations suggest that this monotonic patterns are also true for longer contracts.

The set of consumptions calculated with the two resulting multipliers is the optimal contract. The cost of such contract is simply the sum of the expected consumption in every period. Measures of dispersion such as standard deviation or difference between the highest and the lowest consumption within a period are calculated in the obvious way.