Working Paper Series

THE FEDERAL RESERVE BANK OF RICHMOND

RICHMOND ■ BALTIMORE ■ CHARLOTTE

Moral Hazard and Persistence[*]

Hugo Hopenhayn
UCLA

Arantxa Jarque
FRB– Richmond
and U. Carlos III de Madrid

Abstract

We study a multiperiod principal-agent problem with moral hazard in which effort is persistent: the agent is required to exert effort only in the initial period of the contract, and this effort determines the conditional distribution of output in the following periods. We provide a characterization of the optimal dynamic compensation scheme. As in a static moral hazard problem, consumption —regardless of time period— is ranked according to likelihood ratios of output histories. As in most dynamic models with asymmetric information, the inverse of the marginal utility of consumption satisfies the martingale property derived in Rogerson (1985). Under the assumption of i.i.d. output we show that (i) incentives are concentrated in the later periods of the contract, implying an increase of the variance of compensation over time; (ii) the cost of implementing high effort decreases when there is an increase in either the duration or the intensity of persistence (i.e., how long and how strongly effort affects the distribution of output, respectively); and (iii) under infinite duration the cost gets arbitrarily close to that of the first best.

Journal of Economic Literature Classification Numbers: D80, D82.
Key Words: mechanism design; moral hazard; persistence

1 Introduction

There is a large literature on dynamic contracts that analyzes problems of repeated moral hazard. In the canonical model, a risk neutral principal and a risk averse agent commit to a long term contract in order to solve an incentive problem: each period, the unobservable effort of the agent determines the probability distribution over the observable contemporaneous output. Current effort choices affect *only* current output, i.e., effort does not have persistent effects in time. The solution to this problem specifies the contingent consumption transfers that 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.[1] 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, and still allows us to learn about the implications of persistence. The key simplification is that the agent takes only one action, at the beginning of the contract, with persistent effects. This model can be understood as a complement to the recent literature on repeated moral hazard with persistence, since it isolates a subset of the effects of persistence and studies the properties of optimal consumption paths.[2]

The model is as follows. The contract lasts for an exogenously specified number of periods. At the beginning of the relationship, the principal offers a contract to the agent, specifying consumption in each period contingent on a publicly observable history of output realizations. If the agent accepts, they both commit to the contract. The distribution over the possible output histories is determined by the agent's choice of effort in the first period, which can take two values: low or high. Every period, the agent consumes according to the contingent scheme specified in the contract, but he does not exert any further effort. The agent has time separable, strictly concave utility with discounting. The principal is risk neutral. For simplicity we assume the principal and the agent have the same discount factor, and the agent is not allowed to save. The problem faced by the principal is to design a contract that implements high effort at the lowest expected discounted cost.

This simple dynamic problem with persistence captures essential features of many important long term relationships. One example is investment in human capital. A private firm may offer wage profiles that encourage firm–specific human capital investment, or the government may want to design the tax system in order to provide incentives for high human capital accumulation.[3] Miller (1999) originally used a variation of the model presented here to analyze a two period problem of a car insurance contract in which agents can affect their probability of being in an accident by exerting effort when learning how to drive. In another example, Jarque (2007) shows that repricing CEO stock options may be optimal when the actions of CEOs affect the output of their firms for a

---

[1]Examples of these applications include problems of incomplete insurance due to asymmetric information (see, for example, Atkeson (1991) and Phelan, 1995), CEO's optimal compensation (Wang, 1997), optimal design of loans for entrepreneurs (Albuquerque and Hopenhayn, 2004), or the study of optimal unemployment insurance programs (Shavell and Weiss (1979) and Hopenhayn and Nicolini, 1997).

[2]See the next section for the related literature.

[3]See Grochulski and Piskorski (2006) for a recent contribution to the "new public finance" literature that explicitly models schooling effort as an unobservable investment in human capital at the beginning of life, affecting future productivity of the agents.

number of periods. Her paper uses our model to describe a benchmark for the optimal compensation scheme.

It stems from our analysis that, in spite of its dynamic structure, our moral hazard problem with persistence formally reduces to a static moral hazard case. In the optimal compensation scheme, 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 (see Grossman and Hart, 1983), compensation will be monotone in the past realizations of output if and only if the likelihood ratios satisfy some appropriately modified version of the Monotone Likelihood Ratio Property. Our characterization of the optimal contract has implications for the dynamics of consumption. The inverse of the marginal utility of consumption satisfies the martingale property derived in Rogerson (1985). This implies that, as in most dynamic problems with asymmetric information, including standard repeated moral hazard models, the agent would like to save if he were allowed to do so, and the evolution of his expected consumption through time depends on the concavity or convexity of the inverse of his marginal utility of consumption.

When realizations are i.i.d. over time, our model provides some stark predictions. The contract takes a simple form: the current consumption of the agent depends only on consumption in the previous period of the contract, the number of periods he has been in the contract already (his tenure), and the current output realization. Longer histories contain more information, so the dispersion of likelihood ratios and the variance of compensation increases over time.

We define two measures of effort persistence and we perform comparative static exercises. The first measure is the duration of persistence. It is defined as the number of consecutive periods in which effort affects the distribution of output; in any period after that, output contains no information about effort. Increasing the duration of persistence decreases the cost of implementing high effort. Using the closed form solution for the case in which the utility of the agent is given by the square root of consumption, we show that an increase in duration not only decreases the average variance of the per–period compensation, bringing the cost down, but in particular it decreases the need to spread consumption in earlier periods. For any utility function that allows for unlimited punishment, we show that for a contract that lasts for an infinite number of periods (with infinite duration of persistence) the cost of implementing high effort is arbitrarily close to that of the First Best. 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, in the spirit of Mirrlees (1974).

The second measure with which we perform comparative statics is the intensity of persistence. We modify the i.id. framework and allow the distribution over output to be a weighted sum of a probability determined by effort and an exogenous one. For decreasing sequences of weights, the effect of the initial action depreciates over time. Intensity ranks sequences of weights according to vector dominance. We show that lower intensity of persistence implies a higher cost of the contract. When the agent has square root utility, we can show that for lower intensity average variance of compensation is higher, although the allocation of that increase in variability is not necessarily concentrated in the initial periods of the contract, as it is with duration.

3

1.1 Related Literature on Moral Hazard and Persistence

There are a few papers that tackle the problem of moral hazard and persistence in the context of a repeated action model. In a repeated action model, persistence changes the problem in two dimensions: it introduces a richer information structure and it complicates the incentive problem. Information is richer because the principal observes more than one signal containing information about the same past action. The incentive problem worsens because "joint" deviations of effort may be profitable in the presence of persistence: when the effort of the agent today affects the conditional distribution of output tomorrow, the agent can substitute effort across periods (for example, accumulating it today in order to work less tomorrow). The changes along these two dimensions complicate both the characterization and the numerical computation of the optimal contract. The existing literature in repeated moral hazard with persistence includes some partial characterizations. These results are derived under different proposals of assumptions aimed, mainly, at simplifying the joint deviations problem. What follows is a short summary of that literature that tries to highlight the way in which our model, with its different set of assumptions, complements the existing results.

Fernandes and Phelan (2000) provided the first recursive treatment of agency problems with effort persistence. In their paper, the current effort of the agent affects output in the current period and in the following one. Their setup is characterized by three parameters: the number of periods the effect of effort lasts, the number of possible effort levels, and the number of possible outcome realizations. All three parameters are set to two and this makes their formulation and their computational approach feasible. The optimal contract is found by checking, one by one, all possible joint deviations of effort. The curse of dimensionality applies whenever any of the three parameters is increased. Moreover, no results are given in their paper on the properties of the optimal contract.

Mukoyama and Sahin (2005) show in a two period contract that, when the principal wants to implement high effort every period and persistence is high, it may be optimal for the principal to perfectly insure the agent in the first period. By restricting the number of possible efforts and the length of the contract, they manage to find conditions on the conditional probabilities of output such that only a limited number of joint deviations are relevant. Under those sets of parameters, they prove the optimality of perfect insurance in the first period. In a related model, also with two possible levels of effort but for a $T-$period model, Kwon (2006) assumes the probability distribution over output is concave in the sum of past effort. This provides an equivalent characterization to that in Mukoyama and Sahin (2005): the optimal contract exhibits perfect insurance until period $T-1$, and a contingent increase in consumption in the last period of the contract. He identifies this increase in consumption with a promotion and tests the model using wage and promotions data from health insurance claim processors in a large U.S. insurance company.

Jarque (2005) assumes a continuum of effort choices, and imposes two simplifying assumptions that allow for a complete characterization of the optimal contract. First, utility is linear in effort. Second, the distribution of output depends on the sum of all past discounted effort. This simplifies the problem of joint deviations by making the marginal disutility of effort independent of the actual level of effort chosen each period, and its marginal benefit a function of a summary of all

past histories of effort. Under these assumptions, the solution to the optimal contract can be found using an auxiliary problem without persistence. This implies that consumption in the problem with persistence exhibits the same properties as consumption in a repeated moral hazard problem without persistence.

The assumptions in these four papers simplify the joint deviations problem, but at the same time they impose restrictions on the information structure: they limit the amount and the structure of information about current effort that is embedded in future realizations of output. The model we propose in this paper complements the existing literature by exploring the implications of the richest possible information structure under persistence. We maximize the information about past effort contained in each output realization, since the conditional distribution of all future output is determined by the initial effort. However, we completely eliminate the possibility of joint deviations, and its interaction with the information structure, since the agent only exerts effort once.

The paper is organized as follows. The model is presented in the next section. A characterization of the optimal contract is given for the general model in section 2. Results and numerical examples for the i.i.d. case are discussed in section 4. In section 5 we present the comparative statics on the duration of persistence, and section 6 includes the asymptotic result. In section 7 we present the case of decreasing persistence. Section 8 concludes.

2   The Model

The relationship between the principal and the agent lasts for T periods, where T is finite.[4] The principal is risk neutral, and the agent has strictly concave utility of consumption $u(c)$. There is the same finite set of possible outcomes each period, $Y_t = \{y_i\}_{i=1}^n$, with $y_i < y_{i+1}$ for all $i = 1, \ldots, n$. Let Y^t denote the set of histories of outcome realizations up to time t, with typical element $y^t = \{y_1, y_2, ..., y_t\}$. This history of outcomes is assumed to be common knowledge. The agent's effort can take two possible values, $e \in \{e_L, e_H\}$.[5] A contract prescribes an effort to the agent at time 1, as well as a transfer c_t from the principal to the agent for every period of the contract, contingent on the history of outcomes up to t: $c_t : Y^t \rightarrow \mathbb{R}_+$, for $t = 1, 2, ..., T$.[6] Each period, the probability of a given history of outcomes is conditional on the effort level chosen at the beginning of the first period: $\Pr(y^t | e)$. With this specification, we allow the distribution of the period outcome to change over time, including the possibility that realizations are not independent across periods (i.e., persistent output). We assume $\Pr(y^t | e)$ strictly positive for all possible histories and for both levels of effort, and that there exists at least one t such that $\Pr(y^t | e_H) \neq \Pr(y^t | e_L)$.

---

[4]The solution to the problem presented here is not well defined when $T = \infty$. The case of infinite T is dealt with in the last section of the paper, where an asymptotic approximation result is presented.

[5]As it becomes clear in the core of the paper, the results presented here generalize to the case of multiple effort levels as do the results in 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 for our characterization of the optimal contract to be complete.

[6]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.

Both the agent and the principal discount cost and utility at the same rate β. The agent cannot privately save. Commitment to the contract is assumed on both parts.

As in most principal–agent models, the objective of the principal is to choose the level of effort and the contingent transfers that maximize her expected profit, i.e., the difference between the expected stream of output and the contingent transfers to the agent. In the context of a static moral hazard problem, Grossman and Hart (1983) showed in their seminal paper that this problem can be solved in two steps. The same procedure applies in our dynamic setting: first, for any possible effort level, choose the sequence of contingent transfers that implements that level of effort in the cheapest way. The cost of implementing effort e in a T period contract is just the expected discounted stream of consumption to be provided to the agent:

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

Second, choose among the possible efforts the one that gives the biggest difference between expected output and cost of implementation. Note that, as it is the case in static models, implementing the lowest possible effort is trivial: it entails providing the agent with a constant wage each period such that he gets as much utility from being in the contract as he could get working elsewhere. Since the interesting problem is the one of implementing e_H, we assume throughout the paper that parameters are such that in the second step the principal always finds it profitable to implement e_H. We focus on the problem of minimizing the cost of implementing high effort and, to simplify notation, we drop the dependence of total cost on the effort level: $K\left(T\right) = K\left(T,e_H\right)$. We also assume unlimited resources on the part of the principal, so we do not need to carry his balances throughout the contract. A contract is then simply stated as a sequence of contingent consumptions, $\left\{c\left(y^t\right)\right\}_{t=1}^{T}$.

The Participation Constraint (PC) states that the expected utility that the agent gets from a given contract, contingent on his choice of effort, should be at least equal to the agent's outside utility, $\underline{U}$:

$$\underline{U} \leq \sum_{t=1}^{T}\sum_{y^t}\beta^{t-1}u\left(c\left(y^t\right)\right)\Pr\left(y^t|e_H\right) - e_H, \tag{PC}$$

where e denotes both the choice of effort and the disutility implied by it. As a benchmark, we consider the case of effort being observable. The optimal contract in this case (sometimes referred to as the First Best) is the solution to the following cost minimization problem:

$$\min_{\left\{c(y^t)\right\}_{t=1}^{T}} \sum_{t=1}^{T}\sum_{y^t}\beta^{t-1}\left\{c\left(y^t\right)\right\}\Pr\left(y^t|e_H\right)$$
$$s.t. \ \ \text{PC}$$

It is easy to show that the First Best calls for perfect insurance of the agent: when effort is observable, a constant wage minimizes the cost of delivering the outside utility level. The constant wage c^* in the First Best satisfies:

$$\underline{U} + e_H = \frac{1 - \beta^T}{1 - \beta} u\left(c^*\right).$$

Later in the paper we use the cost of the first best scheme, $K^*\left(T\right) \equiv \frac{1-\beta^T}{1-\beta} c^*$, as a benchmark for evaluation of the severity of the incentive problem when effort is not observable.

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

$$\sum_{t=1}^{T} \sum_{y^t} \beta^{t-1} u\left(c\left(y^t\right)\right) \Pr\left(y^t | e_H\right) - e_H$$

$$\geq \sum_{t=1}^{T} \sum_{y^t} \beta^{t-1} u\left(c\left(y^t\right)\right) \Pr\left(y^t | e_L\right) - e_L. \tag{IC}$$

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 high effort than under low effort. Formally, the optimal contract (often referred to as the Second Best) is the solution to the following cost minimization problem:

$$\min_{\{c(y^t)\}_{t=1}^T} \sum_{t=1}^{T} \sum_{y^t} \beta^{t-1} \left\{c\left(y^t\right)\right\} \Pr\left(y^t | e_H\right) \tag{CM}$$

$$s.t. \ \ PC \text{ and } IC$$

3 Characterization of the Optimal Contract for a General Process for Output

The optimal contract can be characterized by looking at the first order conditions of the cost minimization problem in (CM). As in the static moral hazard case, an important term in these first order conditions is the Likelihood Ratio. The Likelihood Ratio of a history y^t, denoted as $LR\left(y^t\right)$, can be defined as the ratio of the probability of observing y^t under a deviation, to the probability under the recommended level of effort:

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

Proposition 1 *The optimal sequence $\{c_\tau\left(y^\tau\right)\}_{\tau=1}^T$ of contingent consumption in the Second Best contract is ranked according to the likelihood ratios of the histories of output realizations, i.e., for any two histories y^t and $\widetilde{y}^{t'}$ of (possibly) different lengths t and 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 are binding. From the FOCs,

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

where $\lambda > 0$ and $\mu > 0$ are the multipliers associated with the PC and the IC respectively. Since $u'\left(\cdot\right)$ is decreasing, the result follows from the above set of equations. $\blacksquare$

We now argue that this characterization implies that, in spite of its dynamic structure, this problem can be reduced to a standard static moral hazard case. This is true because the agent chooses effort only once, at the beginning of the relationship. Incentives are smoothed over time, but they are evaluated only once by the agent, at the moment of choosing his action. This means that the principal is indifferent between minimizing the total cost of the contract or minimizing its average discounted per period cost. The "averaged" problem looks as follows:

$$\min_{\{c(y^t)\}_{t=1}^T} \frac{1-\beta}{1-\beta^T} \sum_{t=1}^T \sum_{y^t} \beta^{t-1} \left\{c\left(y^t\right)\right\} \Pr\left(y^t|e_H\right)$$

$$s.t. \quad \frac{1-\beta}{1-\beta^T}\underline{U} \leq \frac{1-\beta}{1-\beta^T}\left\{\sum_{t=1}^T \sum_{y^t} \beta^{t-1} \left\{u\left(c\left(y^t\right)\right)\right\} \Pr\left(y^t|e_H\right)\right\} - \frac{1-\beta}{1-\beta^T}e_H$$

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

$$\geq \quad \frac{1-\beta}{1-\beta^T}\left\{\sum_{t=1}^T \sum_{y^t} \beta^{t-1} \left\{u\left(c\left(y^t\right)\right)\right\} \Pr\left(y^t|e_L\right)\right\} - \frac{1-\beta}{1-\beta^T}e_L$$

The one to one mapping between this averaged alternative specification of the dynamic problem and a static cost minimization problem is as follows. In the averaged formulation, the original probability of each history y^t appears adjusted by the corresponding discount factor, β^{t-1}, and divided by the averaging term $\frac{1-\beta^T}{1-\beta}$. We can rename a history y^t of arbitrary length as $h_i \in H_T$, $i = 1, \ldots 2^T$, where $H_T \equiv \cup_{t=1}^T Y^t$ is the set of all possible histories in a T–period problem. History y^t corresponding to h_i happens with "normalized" probability

$$P\left(h_i\right) \equiv \frac{1-\beta}{1-\beta^T}\beta^{t-1}\Pr\left(y^t|e_H\right), \tag{2}$$

or

$$\widehat{P}\left(h_i\right) \equiv \frac{1-\beta}{1-\beta^T}\beta^{t-1}\Pr\left(y^t|e_L\right). \tag{3}$$

These normalized probabilities add up to one. Thus, we may think of the set H_T as the set of possible signals in a static problem. Notice that the utility levels $\underline{U}$, e_L and e_H in the PC and the IC are normalized as well to their per period value in the averaged problem, and that the constraints are equivalent to the dynamic ones.

With the averaged formulation of our problem in mind, we can now discuss the intuition for the characterization result. The reason for the similarity with a static moral hazard characterization becomes clear. The information structure for a standard static 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. Consider now the dynamic problem. The states in the dynamic case are all histories in H_T. Each $h_i \in H_T$ happens with probability $P(h_i)$ under e_H, and with probability $\widehat{P}(h_i)$ under e_L. The expected discounted utility of any contingent consumption plan reduces to a convex combination of the utilities in each of these states, with these adjusted weights. Hence, in the dynamic problem 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 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, i.e., to those that have a high likelihood ratio.

From the simple characterization of the solution we can derive a number of properties of the optimal contract. The first set refers to the relationship between output and compensation. The second refers to intertemporal properties of compensation.

Output and compensation

Consumption in the optimal contract depends on the whole history of output realizations, and it is, typically, a non linear function of total output. We now go on to identify necessary and sufficient conditions for some form of monotonicity of consumption in output. Our results parallel standard results in the static moral hazard literature.[7] Throughout the paper, we use $LR\left(y_t|y^{t-1}\right)$ to denote the likelihood ratio of an individual realization of output at time t, conditional on history y^{t-1}:

$$LR\left(y_t|y^{t-1}\right) = \frac{\Pr\left(y_t|e_L, y^{t-1}\right)}{\Pr\left(y_t|e_H, y^{t-1}\right)}.$$

Definition 2 *The **Monotone Likelihood Ratio Property** holds **at the individual output level** in period t if, for any y^{t-1}, the individual output likelihood ratio at t, $LR\left(y_t|y^{t-1}\right)$, decreases monotonically with y_t.*

Corollary 3 *For any given finite history y^{t-1}, $c\left(y^{t-1}, y_i\right) > c\left(y^{t-1}, y_j\right)$, for all $i > j$, if and only if the MLRP holds at the individual output level in period t.*

In our dynamic setup, the MLRP for individual outcomes is satisfied by a large set of processes for output. A natural example is the case of independently and identically distributed (i.i.d.) output, which we study in depth in the next section. On the other hand, we can find examples that violate MLRP at the individual output level, as when the probability of outcomes depends on

---

[7]See Grosman and Hart (83) for a complete treatment of a general case with multiple effort levels, and Holmstrom (79) and Mirrlees (76) for the case of continuous effort under the assumption of the validity of the First Order Approach.

an exogenous parameter over which both the principal and the agent have some prior distribution, which they update according to Bayes' rule.[8]

We can think of a stronger restriction, which applies to a smaller set of stochastic processes and assures a stronger form of monotonicity, based on ranking histories according to vector dominance. A history $y^t = (y_1, y_2, \ldots, y_t)$ vector-dominates a history $\widetilde{y}^t = (\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_t)$ if $y_\tau \geq \widetilde{y}_\tau$ for all $\tau = 1, \ldots, t$, with strict inequality for at least one τ.

Definition 4 *The **Monotone Likelihood Ratio Property** holds **at the history level** in period t if, for any two histories y^t, and $\widetilde{y}^t$ such that y^t vector-dominates $\widetilde{y}^t$, the likelihood ratio of y^t is smaller than that of $\widetilde{y}^t$, i.e. $LR\left(y^t\right) < LR\left(\widetilde{y}^t\right)$.*

Corollary 5 *For any two histories y^t and $\widetilde{y}^t$ such that y^t vector-dominates $\widetilde{y}^t$, the optimal contract implies $c\left(y^t\right) > c\left(\widetilde{y}^t\right)$ if and only if the MLRP holds at the history level in period t.*

Note that since the likelihood ratio of a history is the product of the likelihood ratios of the individual outputs in the history, if MLRP holds at the individual outcome level for all periods up to t, it also holds at the history level in period t. The MLRP at the history level rules out stochastic processes in which the effect of effort on later periods depends on the history of outputs in very different magnitudes for each of the two levels of effort, i.e., the informativeness of individual outcomes about effort depends strongly on the past history of realizations, yet without violating the MLRP at the individual output level. It is also possible to find stochastic processes that violate both forms of monotonicity. An example such a process is one in which low effort implements always the same probability over a good outcome, while under high effort the probability is higher when last period's output is low but it is lower for one period after a high output is observed.

In some settings it may be natural to compare histories based on their cumulative output. We can construct a weaker monotonicity condition for histories based on this ranking:

Definition 6 *The **Monotone Likelihood Ratio Property** holds **at the cumulative output level** in period t if, for any two histories y^t, and $\widetilde{y}^t$ such that $\sum_{\tau=1}^{t} y_\tau > \sum_{\tau=1}^{t} \widetilde{y}_\tau$, the likelihood ratio of y^t is smaller than that of $\widetilde{y}^t$, i.e. $LR\left(y^t\right) < LR\left(\widetilde{y}^t\right)$.*

Corollary 7 *For any two histories y^t and $\widetilde{y}^t$ such that y^t has higher cumulative output than $\widetilde{y}^t$, the optimal contract implies $c\left(y^t\right) > c\left(\widetilde{y}^t\right)$ if and only if the MLRP holds at the cumulative output level in period t.*

Monotonicity on cumulative output is implied by monotonicity at the individual output level, but is weaker than monotonicity at the history level. Although it is not an easy condition to satisfy in general, it holds in the case of i.i.d. output that we study in the next section.

Intertemporal properties of optimal compensation

In spite of the parallel that we established with the solution to a static moral hazard problem, our problem is one of dynamic provision of incentives. Some of the features that we observe in

---

[8]See Miller (1999) and Jarque (2007) for details of this example.

dynamic asymmetric information models are present also in our problem with persistence. The FOCs of the problem described in equation 1 can be combined to get the condition on the inverse of the marginal utility derived by Rogerson (1985) in the context of a two period repeated moral hazard problem:

$$\frac{1}{u'\left(c\left(y^t\right)\right)} = \sum_{y_{t+1}} \frac{1}{u'\left(c\left(y^t, y_{t+1}\right)\right)} \Pr\left(y_{t+1}|e_H, y^t\right). \tag{4}$$

This equality shows that our problem with persistence exhibits the standard dynamic trade–off between incentives and consumption smoothing. As in Rogerson (1985), this property implies that 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 8 *In the Second Best contract, the expected consumption of the agent 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 (1985), (4) holds. Take the case of $\frac{1}{u'(c)}$ being concave. By Jensen's inequality,

$$
\begin{aligned}
\frac{1}{u'\left(c\left(y^t\right)\right)} &> \frac{1}{u'\left(E\left[c\left(y^t, y_{t+1}\right)\right]\right)} \\
u'\left(E\left[c\left(y^t, y_{t+1}\right)\right]\right) &> u'\left(c\left(y^t\right)\right).
\end{aligned}
$$

Since utility is concave,

$$c\left(y^t\right) < E\left[c\left(y^t, y_{t+1}\right)\right].$$

A similar argument applies for the other two cases. ∎

3.1 Application: Sorting Types

With a simple relabeling of terms, our model applies to adverse selection problems. In these situations, there is no unobservable effort to be exerted at the beginning of the contract. Instead, there is asymmetry of information about the productivity of the agent, i.e. about the probability distribution over output that he induces by working at the firm. The agent may be of high productivity, θ_H, in which case the probability of a given outcome is determined by the conditional distribution $\Pr\left(y_t|\theta_H\right)$, or he may be of low productivity, θ_L, in which case output follows $\Pr\left(y_t|\theta_H\right)$. Assume that an agent with high productivity has an outside utility (an opportunity cost of working for the principal) of $\widetilde{\underline{U}}_H$. The low productivity worker, instead, has an outside utility of $\widetilde{\underline{U}}_L < \widetilde{\underline{U}}_H$.

In order for the high ability workers to accept the contract, the following participation constraint must hold:

$$\widetilde{\underline{U}}_H \leq \sum_{t=1}^{T} \sum_{y^t} \beta^{t-1} u\left(c\left(y^t\right)\right) \Pr\left(y^t|\theta_H\right).$$

Relabeling $\underline{U} + e_H = \widetilde{\underline{U}}_H$, this equation is equivalent to our original PC. If the contract offered by the principal is to be accepted only by high productivity workers, the following sorting constraint must hold:

$$\widetilde{\underline{U}}_L \geq \sum_{t=1}^{T} \sum_{y^t} \beta^{t-1} u\left(c\left(y^t\right)\right) \Pr\left(y^t | \theta_L\right).$$

Letting $\underline{U} + e_L = \widetilde{\underline{U}}_L$, we can rewrite the sorting constraint as

$$\underline{U} \geq \sum_{t=1}^{T} \sum_{y^t} \beta^{t-1} u\left(c\left(y^t\right)\right) \Pr\left(y^t | \theta_L\right) - e_L,$$

or, substituting $\underline{U}$ from the PC,

$$\sum_{t=1}^{T} \sum_{y^t} \beta^{t-1} u\left(c\left(y^t\right)\right) \Pr\left(y^t | \theta_H\right) - e_H \geq \sum_{t=1}^{T} \sum_{y^t} \beta^{t-1} u\left(c\left(y^t\right)\right) \Pr\left(y^t | \theta_L\right) - e_L.$$

This last equation is equivalent to our original IC, which is reinterpreted here as a sorting constraint: the difference in expected utilities under the two possible processes for output should be equal to the difference in outside utilities. The optimal contract is signed in equilibrium only by high productivity agents.

This extends the scope of our analysis, for example, to the design of wage profiles when firms face potential workers who have private information about their own abilities.

4 Outcomes Independently and Identically Distributed

In this section, we study a particular specification of the probability distribution of outcomes: an i.i.d. process. We have

$$\Pr\left(y_t | e, y^{t-1}\right) = \Pr\left(y_t | e\right) \quad \forall y_t, \quad t = 1, ..., T.$$

This assumption puts additional structure on the probability distribution of histories, and allows for the optimal contract to be further characterized.

For the rest of the paper, we analyze the two outcomes case, $Y_t = \{y_L, y_H\}$. To simplify notation, let

$$\begin{aligned}
\Pr\left(y_H | e_H\right) &= \pi \\
\Pr\left(y_H | e_L\right) &= \widehat{\pi},
\end{aligned}$$

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

Note that

$$LR(y_H) = \frac{\widehat{\pi}}{\pi} < 1 \text{ and } LR(y_L) = \frac{1 - \widehat{\pi}}{1 - \pi} > 1.$$

For any history y^t, the length and the fraction of high realizations in of the history are a sufficient statistic for the history's probability. Denote the number of high realizations contained in a given history as $x\left(y^t\right)$. The likelihood ratio of the history can be written as

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

and the first order conditions of problem 1 for a history of length t read:

$$\left(c\left(y^{t-1}, y_H\right)\right) \quad : \quad \frac{1}{u'\left(c\left(y^{t-1}, y_H\right)\right)} = \lambda + \mu\left[1 - LR\left(y^{t-1}\right)\frac{\widehat{\pi}}{\pi}\right]$$

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

We can easily see that in the two outcome setup the MLRP holds trivially at the individual outcome level. Hence, we can simplify the general results on monotonicity of consumption given in the previous section:

Corollary 9 *Assume output can only take two values, $\{y_L, y_H\}$, and it is i.i.d. over time. Given any history y^t of any finite length t, $c\left(y^t, y_H\right) > c\left(y^t, y_L\right)$. In other words, the consumption of the agent increases when a new high realization is observed. Moreover, for any two histories of the same length y^t and $\widetilde{y}^t$, $c\left(y^t\right) \geq c\left(\widetilde{y}^t\right)$ if and only if $x\left(y^t\right) \geq x\left(\widetilde{y}^t\right)$, regardless of the sequence in which the realizations occurred in each of the histories.*

The second part of the corollary can be rephrased as perfect substitutability of output realizations across time. This implies that the tuple $\left\{t, x\left(y^t\right)\right\}$ contains all the information about history y^t that is used in the optimal contract. 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 consumption. Simply put, the consumption received by the agent in the previous period, together with his tenure in the contract are sufficient to determine his current consumption.

Under the i.i.d. assumption, we can study the evolution of the values of the likelihood ratios at each period. We can translate this into formal properties of the support of the contract, that is, the evolution of the possible values of consumption through time. Denote $\overline{y}^t$ the history at t with all high outcomes, that is, the history of length t that satisfies $x\left(\overline{y}^t\right) = t$. Similarly, $\underline{y}^t$ denotes the history with all low outcomes, with $x\left(\underline{y}^t\right) = 0$. The following simple proposition says that the support of consumption values within a certain period increases with t.

Proposition 10 *As t increases, $c\left(\overline{y}^t\right)$ increases and $c\left(\underline{y}^t\right)$ decreases. Hence, as t increases, $d_t = c\left(\overline{y}^t\right) - c\left(\underline{y}^t\right)$ increases.*

Proof. In the i.i.d. case, for a length t, the lowest likelihood ratio is that of the history with t high realizations of output:

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

13

For the highest, instead, it is

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

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

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

The result follows from the First Order Conditions. ∎

We can as well characterize the distribution of the likelihood ratios over time and use its moments to study stochastic properties of consumption over time in the optimal contract. To each history y^t corresponds a likelihood ratio $LR\left(y^t\right)$. The probability of observing that particular likelihood ratio is the probability of history y^t. In equilibrium, under high effort choice,

$$\Pr\left(LR\left(y^t\right)|e_H\right) = \Pr\left(y^t|e_H\right) = \left(\begin{array}{c} x\left(y^t\right) \\ t \end{array}\right) \pi^{x(y^t)} (1-\pi)^{1-x(y^t)},$$

where $\left(\begin{array}{c} x \\ t \end{array}\right)$ denotes the standard combinatorial function. We can calculate the expectation and the variance of the likelihood ratios at each t. The expectation in equilibrium is constant over time, and equal to one:

$$E\left[LR\left(y^t\right)|e_H\right] = \sum_{y^t} \Pr\left(y^t|e_H\right) \frac{\Pr\left(y^t|e_L\right)}{\Pr\left(y^t|e_H\right)} = \sum_{y^t} \Pr\left(y^t|e_L\right) = 1 \quad \forall t.$$

The variance of the likelihood ratios at time t can be written in terms of the expectation of the likelihood ratio off the equilibrium path, $E\left[LR\left(y^t\right)|e_L\right]$. Let $\widehat{E}$ be the expectation of the likelihood ratio under low effort at $t=1$:

$$\widehat{E} \equiv E\left[LR\left(y^1\right)|e_L\right] = \widehat{\pi}\frac{\widehat{\pi}}{\pi} + (1-\widehat{\pi})\frac{(1-\widehat{\pi})}{(1-\pi)}.$$

Any values of π and $\widehat{\pi}$ that satisfy our initial assumption of $\pi > \widehat{\pi}$ imply $\widehat{E} > 1$. It is easy to check that

$$E\left[LR\left(y^t\right)|e_L\right] = \widehat{E}^t.$$

After some algebra, we get the following expression for the variance of the likelihood ratios under high effort, which we denote by v_t :

$$v_t \equiv Var\left(LR\left(y^t\right)|e_H\right) = \widehat{E}^t - 1.$$

Lemma 11 *The variance of the likelihood ratios increases with t. The one–period increase in the variance also increases with t.*

14

Proof. Since $\widehat{E} > 1$, the first part follows immediately from the expression for v_t derived above. For any two periods t and $t+1$, for $t = 1, \ldots, T-1$, the one–period increase in the variance equals

$$
\begin{aligned}
v_t - v_{t-1} &= \widehat{E}^t - 1 - \left(\widehat{E}^{t-1} - 1 \right) \\
&= \left(\widehat{E} - 1 \right) \widehat{E}^{t-1},
\end{aligned}
$$

which increases with t. ∎

As we have already mentioned for the case of more general processes for output, the inverse of the marginal utility of consumption satisfies the martingale property. From the same first order conditions of the problem (Eq.1), in the i.i.d. case, we can characterize the evolution of the variance of the inverse of the marginal utility of consumption. Although this is not, in general, a direct measure of the evolution of consumption, it is closely related to it, and provides us with a measure of risk smoothing across periods in the optimal contract.

Proposition 12 *The variance of $\frac{1}{u'(c(y^t))}$ increases with t. The one–period increase in the variance also increases with t.*

Proof. From the first order conditions in Eq. 1 we have that

$$
E \left(\frac{1}{u' \left(c \left(y^t \right) \right)} | e_H \right) = \lambda.
$$

and

$$
Var \left(\frac{1}{u' \left(c \left(y^t \right) \right)} | e_H \right) = \mu^2 v_t.
$$

The result in the proposition follows from the previous lemma. ∎

We postpone the discussion of this proposition to state two corollaries that rely on the same intuition. The first one applies to the special case in which the agent has logarithmic utility.

Corollary 13 *When the agent has $u(c) = \ln(c)$, the variance of $c(y^t)$ increases with t. The one–period increase in this variance also increases with t.*

Proof. It follows from the above proposition and $u'(c) = \frac{1}{c}$. ∎

There is a second functional form for utility that provides an intuitive measure for the optimal way of smoothing incentives over time: $u(c) = 2\sqrt{c}$, which corresponds to a CRRA utility function with coefficient of risk aversion equal to $\frac{1}{2}$. For this specification, the inverse of the marginal utility of a given level of consumption is proportional to the level of utility implied by the consumption.

Corollary 14 *When the agent's utility is given by $u(c) = 2\sqrt{c}$, the variance of $u(c(y^t))$ increases with t. The one–period increase in this variance also increases with t.*

Proof. Since $\frac{1}{u'(c)} = \sqrt{c}$, we have that

$$
Var \left[u \left(c \left(y^t \right) \right) | e_H \right] = Var \left[2\sqrt{c \left(y^t \right)} | e_H \right] = 4 Var \left[\frac{1}{u'(c)} | e_H \right].
$$

15

By Prop.12, we know that $Var\left[\frac{1}{u'(c)}|e_H\right]$ increases with t. The second part of the corollary follows in the same way. ∎

As we discussed after stating Prop.1, the contract optimally places incentives in periods when information is more precise. In the i.i.d. case, these correspond to the latter periods of the contract. The higher precision of the information corresponds to a higher variance of the likelihood ratios. This, in turn, implies higher variation in consumption in late periods. In other words, the optimal contract provides higher insurance and less incentives in the initial periods, and decreases insurance in favor of incentives later on, when punishments can be placed more efficiently in light of the richer information.

1.a) $u(c) = \ln(c)$

$c^* = 9.9742$

$\frac{K(4)}{K_*} = 1.32$

$\lambda = 13.14$

$\mu = 18.31$

	$E[c_t]$	$\frac{E[c_t]}{c^*}$	σ_t^2	$\frac{\sigma_t}{E[c_t]}$	d_t	$\frac{d_t}{c_t}$
$t=1$	13.14	1.32	15.97	0.30	8.72	0.8
$t=2$	13.14	1.32	32.70	0.44	15.78	2.1
$t=3$	13.14	1.32	50.23	0.54	21.91	5.3
$t=4$	13.14	1.32	68.60	0.63	27.63	132.1

1.b) $u(c) = 2\sqrt{c}$

$c^* = 1.3225$

$\frac{K(4)}{K_*} = 1.14$

$\lambda = 1.15$

$\mu = 1.23$

	$E[c_t]$	$\frac{E[c_t]}{c^*}$	σ_t^2	$\frac{\sigma_t}{E[c_t]}$	d_t	$\frac{d_t}{c_t}$
$t=1$	1.40	1.05	0.47	0.49	1.49	1.6
$t=2$	1.47	1.11	0.86	0.63	2.77	4.7
$t=3$	1.55	1.17	1.17	0.70	3.78	12.9
$t=4$	1.63	1.24	1.41	0.73	4.50	58.1

Table 1. Numerical examples

In Table 1 we report values for two examples, with the two functional forms of our corollaries, as an illustration of the results.[9] Both contracts last four periods and share all the parameters. In the first column of each matrix we report expected consumption per period, and in the second the normalized value we obtain when dividing it by the First Best period consumption. As predicted by Prop.8, expected consumption is constant across periods for the logarithmic utility, while it is increasing for the square root case (which has the inverse of the marginal utility concave in consumption.) In the third column we report the variance of consumption, and in the fourth the standard deviation of consumption divided by the expected consumption in the period so that numbers are comparable across periods and utility specifications. As predicted in Corollary 13 and paralleling the prediction about variance of utility in Corollary 14, both measures increase with t. For the logarithmic case, for example, in the first period of the contract the scaled standard deviation of consumption is 0.3. At the fourth period of the contract, the value increases to 0.63.

---

[9]Parameters of the example: $T = 4$, $\beta = .95$, $\underline{U} = 7.42$, $e_H = 1.11$, $e_L = 0$, $\pi = .3, \hat{\pi} = .2$.

We observe the same trend for the square root utility. The last two columns report the difference between the highest and the lowest consumption levels in a given period, d_t, and the value of this difference relative to the lowest consumption:

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

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. Consistently with the result in Prop. 10, d_t increases with t. For the logarithmic case, in the first period the scaled difference of consumption is less than one, and it increases to more than 132 by the fourth period. We observe the same trend for the square root utility.

Due to the range of potential curvatures of the utility function, the increase over time in the variance of the likelihood ratios does not always translate directly in a proportional change in the variance of the optimal consumption, so a general result for the variance of consumption cannot be stated. However, in all our numerical examples the general intuition about concentrating incentives in the later periods appears robust to different specifications of utility.[10]

5   Changes in the Duration of Persistence

In this section, we consider T–period contracts in which the effect of effort on the probability of observing high output dies out completely before the end of the contract. We introduce the following terminology:

Definition 15 *An outcome realization* y_t *is **informative** whenever*

$$\Pr_t(y_t|e_H) \neq \Pr_t(y_t|e_L).$$

For informative outcomes, we maintain the i.i.d. assumption. To model changes in the duration of persistence, we consider stochastic processes that are i.i.d. up to period $\tau \geq 1$, and for any $t > \tau$ they satisfy:

$$\Pr_t(y_t|e_H) = \Pr_t(y_t|e_L) = \overline{\pi} \ \ \forall y_t \in Y_t,$$

i.e., outcomes after period τ are not informative. When the effect of effort dies out the probability of the individual period realizations is the same, $\overline{\pi}$, independently of whether the agent chose high or low effort at the beginning of the contract. We refer to τ as the *duration* of persistence.

We use the FOCs of the problem to derive the form of the optimal contract when $\tau < T$. Letting $I_t = I(y_t = y_H)$ be an indicator function that takes value 1 when $y_t = y_H$, and zero otherwise, we can write the following probabilities for histories corresponding to $t > \tau$,

$$\Pr(y^t|e_H) = \prod_{j=1}^{\tau} [\pi I_t + (1 - \pi)(1 - I_t)] \prod_{j=\tau+1}^{t} [\overline{\pi} I_t + (1 - \overline{\pi})(1 - I_t)],$$

---

[10] We computed examples for CRRA utility with different degrees of risk aversion.

while

$$\Pr\left(y^t|e_L\right) = \prod_{j=1}^{\tau} \left[\widehat{\pi}I_t + (1-\widehat{\pi})(1-I_t)\right] \prod_{j=\tau+1}^{t} \left[\overline{\pi}I_t + (1-\overline{\pi})(1-I_t)\right].$$

From these history probabilities we construct the corresponding likelihood ratios, and we easily see that they remain constant and equal to $LR\left(y^\tau\right)$ for any uninformative history following y^τ :

$$LR\left(y^t\right) \equiv \begin{cases} \frac{\Pr\left(y^t|e_L\right)}{\Pr\left(y^t|e_H\right)} & \text{for } t \leq \tau \\\\ \frac{\Pr\left(y^\tau|e_L\right)}{\Pr\left(y^\tau|e_H\right)} & \text{for } t > \tau. \end{cases}$$

In other words, the likelihood ratio of individual realizations of output takes a constant value of one after period τ:

$$LR\left(y_t\right) \equiv 1 \quad \text{for } t > \tau.$$

Since, by the FOCs, the optimal contract ranks consumption according to likelihood ratios, consumption is constant from τ until T. With these observations we are now ready to state the main result of this section.

Since output is assumed to be i.i.d. up to τ, contracts with higher duration have a richer information structure. This allows us to show, in the next proposition, that a longer duration of persistence allows the implementation of high effort at a lower cost.

Proposition 16 *The cost of a contract strictly decreases if the duration of persistence, τ, increases.*

Proof. Denote by $C_1 = \left\{c_1\left(y^t\right)\right\}_{t=1}^{T}$ the optimal contract corresponding to a persistence of duration τ_1. Consider a change in duration from τ_1 to τ_2, where $\tau_2 > \tau_1$. Denote the corresponding new optimal contract as $C_2 = \left\{c_2\left(y^t\right)\right\}_{t=1}^{T}$. First, note that C_1 is feasible and incentive compatible under τ_2: both the PC and the IC of the problem under τ_2 are satisfied by the C_1 contract. However, C_1 does not satisfy the first order conditions of C_2 for any strictly positive value of λ and μ : at any t such that $\tau_1 < t \leq \tau_2$ the FOC corresponding to τ_2 implies a different consumption following y_L than following y_H, for any y^{t-1}, since $LR\left(y^{t-1}, y_L\right) \neq LR\left(y^{t-1}, y_H\right)$ for all y^{t-1}. Contract C_1, however, implies a constant consumption for those histories, since outcomes in that period range are not informative and hence $LR\left(y^{t-1}, y_L\right) = LR\left(y^{t-1}, y_H\right)$ for all y^{t-1}. Hence, although C_1 is feasible and incentive compatible under τ_2, it is not the solution to the cost minimization problem under τ_2: this establishes that the total cost of C_2 is strictly smaller than that of C_1. ∎

The intuition for this result hinges on the better quality of the signal structure of the problem when the duration of persistence is longer. As already established by Holmström (1979), any informative signal is valuable. Intuitively, when information quality increases incentives can be given more efficiently, lowering the cost of the contract.

If the utility of the agent is $u\left(c\right) = 2\sqrt{c}$ we can characterize analytically the solution for the optimal contract. When doing comparative statics with respect to τ, we can show both the effect of duration on cost and the implied changes in the variance of compensation of each individual period. When duration increases, less variation is needed in the early periods, given that, with

higher duration, more informative realizations are available in late periods, when punishments are exercised with lower probability on the equilibrium path.

Let $Var_\tau \left[u \left(c \left(y^t \right) \right) | e_H \right]$ denote the variance of utility in period t when the contract is of duration τ.

Proposition 17 *If the agent's utility is given by $u = 2\sqrt{c}$, an increase in the duration of the contract from τ_1 to $\tau_2 > \tau_1$ implies a lower cost of the contract, lower average variance of utility, and lower variance of utility in any period $t \leq \tau_1$, i.e.,*

$$Var_2 \left[u \left(c \left(y^t \right) \right) | e_H \right] < Var_1 \left[u \left(c \left(y^t \right) \right) | e_H \right] \quad \forall t \leq \tau_1 \Leftrightarrow \tau_2 > \tau_1.$$

Proof. Given a set of parameters e, $\underline{U}$, π, $\widehat{\pi}$, τ and T, the explicit solutions for the multipliers, using the FOCs and the constraints of the problem, are:

$$\lambda = \frac{(\underline{U} + e_H)}{2} \frac{1 - \beta}{1 - \beta^T}$$

and

$$\mu = \frac{e_H/2}{\bar{v}},$$

where $\bar{v}$ denotes the average across periods of the variances of the likelihood ratios:

$$\bar{v} \equiv \frac{\sum_{t=1}^{T} \beta^{t-1} v_t}{\frac{1 - \beta^T}{1 - \beta}}.$$

The solution for consumption at each t is:

$$c \left(y^t \right) = \left[\lambda + \mu \left(1 - \frac{\Pr \left(y^t | e_L \right)}{\Pr \left(y^t | e_H \right)} \right) \right]^2.$$

With this we can write expected consumption at each t as:

$$E \left[c_t | e_H \right] = \lambda^2 + \mu^2 v_t.$$

The average per period cost of the contract is then easily written as:

$$
\begin{aligned}
k \left(T \right) &= \frac{1 - \beta^T}{1 - \beta} \sum_{t=1}^{T} \beta^{t-1} E \left[c_t | e_H \right] \\
&= \lambda^2 + \mu^2 \bar{v} \\
&= \frac{1}{4} \left(\left[(\underline{U} + e) \frac{1 - \beta^T}{1 - \beta} \right]^2 + \frac{e^2}{\bar{v}} \right).
\end{aligned}
\tag{5}
$$

Let subscript i denote variables corresponding to a contract of duration τ_i. We have

$$v_{ti} = \begin{cases} \widehat{E}^t - 1 & \text{for } t \leq \tau_i \\ \widehat{E}^{\tau_i} - 1 & \text{for } t > \tau_i. \end{cases}$$

19

Hence,

$$\bar{v}_i \equiv \frac{\sum_{t=1}^{\tau_i} \beta^{t-1} \left(\widehat{E^t} - 1 \right)}{\frac{1-\beta^T}{1-\beta}} + \beta^{\tau_i - 1} \frac{1 - \beta^T}{1 - \beta} \left(\widehat{E}^{\tau_i} - 1 \right).$$

Since $\tau_1 < \tau_2$, it follows that $\bar{v}_1 < \bar{v}_2$. This implies $\mu_1 > \mu_2$ and, by Eq. (5), $k_2(T) < k_1(T)$. This confirms the general result of the previous proposition. For the second part of the proposition, we can express variance of utility as a function of $\bar{v}$ using the solution for μ :

$$Var_i \left[u \left(c \left(y^t \right) \right) | e_H \right] = e_H^2 \frac{v_{ti}}{\bar{v}_i^2}.$$

For every $t \leq \tau_1$ we have $v_{t1} = v_{t2}$; since $\bar{v}_1 < \bar{v}_2$, this makes the variance of utility lower under duration τ_2 for those periods. ∎

The average variance of the likelihood ratios is a measure of informativeness of the stochastic process, or information structure, that the principal is facing. With this particular curvature of the utility function, for a given value of μ, consumption in a given period is a convex function of the likelihood ratios. Moreover, an increase in the average variance of the likelihood ratios decreases the value of μ, and leaves unchanged λ. Hence, an increase in the variance of the likelihood ratios translates into cost savings.

$\tau = 1$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t = 1$	2.73	1.30	19.10
$t = 2$	2.73	1.30	19.10
$t = 3$	2.73	1.30	19.10
$t = 4$	2.73	1.30	19.10

$\lambda = 27.27 \quad \mu = 25.56$

$\tau = 2$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t = 1$	1.61	0.60	2.14
$t = 2$	1.61	0.85	14.46
$t = 3$	1.61	0.85	14.46
$t = 4$	1.61	0.85	14.46

$\lambda = 16.05 \quad \mu = 43.9$

$\tau = 3$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t = 1$	1.40	0.40	1.18
$t = 2$	1.40	0.57	3.60
$t = 3$	1.40	0.71	22.73
$t = 4$	1.40	0.71	22.73

$\lambda = 13.94 \quad \mu = 18.31$

$\tau = 4$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{\underline{c}_t}$
$t = 1$	1.32	0.30	0.83
$t = 2$	1.32	0.44	2.10
$t = 3$	1.32	0.54	5.33
$t = 4$	1.32	0.63	132.1

$\lambda = 13.14 \quad \mu = 18.31$

Table 2. Effect of changes in τ on variability measures: numerical example with logarithmic utility

For logarithmic utility, we do not have closed form solutions that allow us to determine the effect of varying τ on the variance of compensation in the optimal contract. Table 2 presents a numerical example that illustrates this effect. Each matrix corresponds to a different duration τ,

going from $\tau = 1$ to $\tau = T = 4$. We set $\overline{\pi} = 0.5$ and all other parameters are kept equal to the values in the previous examples. In the first column of each matrix, we report the expected consumption of a period proportional to that of the First Best; this per–period cost decreases with duration, as predicted by Prop. 16. The two last columns present the two standardized measures of variability of consumption presented in the previous section. Looking at the same column for each of the matrices, we can observe the effect of increasing τ in the variability measures corresponding to a given period t. For period one, for example, the normalized standard deviation of consumption, $\frac{\sigma_1}{E[c_1]}$, falls from 1.3 to 0.3 when comparing a $\tau = 1$ contract with a $\tau = 4$ contract, confirming the pattern proved for the square root specification in Prop.17. The same pattern applies for $\frac{d_1}{\underline{c}_1}$: it falls from 19.1 to 0.83.

Under each matrix in Table 2 we report the value of the multipliers for the corresponding τ. Both multipliers decrease when τ increases. The sharper decrease corresponds to the multiplier of the IC, μ. Although we only have a formal proof for $u(c) = \sqrt{c}$, all of our numerical examples with CRRA utility, for different degrees of risk aversion, show the same negative relation between μ and τ. This decrease in μ means that, as the duration of persistence increases, the IC is easier to satisfy. The availability of better quality information is materialized in more extreme values of the likelihood ratios. Rearranging the FOCs of the Second Best we have that for any two histories y^t and $\widetilde{y^t}$ of any length,

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

The patterns for the variability of compensation we just described can be understood in terms of a decrease in μ. For the logarithmic utility, in particular, this means that the difference in consumption is proportional to the difference in the likelihood ratios; for the square root, it is the difference in utility levels. The factor of proportionality between differences in likelihood ratios and differences in compensation 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.

6   Asymptotic Optimal Contract

Assume, as in Section 3, that output is distributed i.i.d. and $\tau = T$. If the principal and the agent can commit to an infinite contractual relationship ($T = \infty$) and utility is unbounded below (as in the logarithmic case, for example) the cost of the contract under moral hazard can get arbitrarily close to that of the First Best, i.e., under observable effort.

The First Best contract implies

$$c^* = u^{-1}\left(\left(\underline{U} + e_H\right)\left(1 - \beta\right)\right)$$

and the First Best cost is

$$K^*\left(\infty\right) = \frac{1}{1 - \beta}c^*.$$

The Second Best is not well defined for an infinite number of periods. In this section, we present an alternative feasible and incentive compatible contract, which we call the "one–step" contract. This contract is not necessarily optimal, but it is a useful benchmark to study because we can get an upper bound on its cost. This bound is, in turn, an upper bound on the cost of the Second Best contract. In the next proposition we show that the upper bound on the cost of the "one–step" contract can get arbitrarily close to the cost of the First Best when contracts last an infinite number of periods.

A "one–step" contract is a tuple $(c_0, \underline{c}, L)$ of two possible consumption levels c_0 and $\underline{c}$ plus a threshold L for the Likelihood Ratio. The contract is defined in the following way:

$$c(y^t) = \begin{cases} c_0 & if \ \ LR\left(y^t\right) < L \\ \underline{c} & if \ \ LR\left(y^t\right) \geq L \end{cases}.$$

Proposition 18 *Assume output is distributed i.i.d. and the agent has a utility function that satisfies $\lim_{c \to 0} u(c) = -\infty$. For any $\beta \in (0,1]$ and any $\varepsilon > 0$, there exists a one–step contract $(c_0, \underline{c}, L)$ such that the principal can implement high effort at a cost $K(\infty) < K^*(\infty) + \varepsilon$, where $K^*(\infty)$ is the cost when effort is observable.*

Proof. Let δ and P satisfy the following two equations:

$$u(c_0) = u_0 = u(c^*) + \delta,$$

where c^* is the level of consumption provided in the First Best, and

$$u(\underline{c}) = u_0 - P.$$

For a given L and for each possible date t, denote by $A_t(L)$ the set including all histories of length t such that their likelihood ratio is lower than the threshold L, so they are assigned a consumption equal to c_0. Denote by $A_t^c(L)$ the complement of that set; that is:

$$\begin{aligned} A_t(L) &= \left\{y^t \mid LR\left(y^t\right) \leq L\right\} \ \ \text{and} \\ A_t^c(L) &= \left\{y^t \mid LR\left(y^t\right) > L\right\} \ \ \forall t. \end{aligned}$$

Define $F_t(L)$ and $\widehat{F}_t(L)$ as the total probability of observing a history in $A_t(L)$ for high and low effort, correspondingly:

$$\begin{aligned} F_t(L) &= \sum_{y^t \in A_t(L)} \Pr\left(y^t | e_H\right) \\ \widehat{F}_t(L) &= \sum_{y^t \in A_t(L)} \Pr\left(y^t | e_L\right). \end{aligned}$$

Given this one–step contract, the expected utility of the agent from choosing high effort is

$$\frac{u_0}{1-\beta} - P \sum_t \beta^{t-1} (1 - F_t(L)) - \frac{e_H}{1-\beta}.$$

We can find the maximum $\underline{c}$ –or, equivalently, the minimum punishment P – that satisfies the IC:

$$-P \sum_t \beta^{t-1} \left(1 - F_t(L)\right) - \frac{e_H}{1-\beta} = -P \sum_t \beta^{t-1} \left(1 - \widehat{F}_t(L)\right) - \frac{e_L}{1-\beta}$$

so we can write

$$P(L) = \frac{e_H - e_L}{(1-\beta) \sum_t \beta^{t-1} \left(F_t(L) - \widehat{F}_t(L)\right)}.$$

Now we can write the PC substituting $P(L)$, which pins down u_0 :

$$\underline{U} + e_H = \frac{u_0}{1-\beta} - (e_H - e_L) \frac{\sum_t \beta^{t-1} \left(1 - F_t(L)\right)}{(1-\beta) \sum_t \beta^{t-1} \left(F_t(L) - \widehat{F}_t(L)\right)}$$

Since $u(c^*) = (\underline{U} + e_H)(1-\beta)$ and $u_0 = u(c^*) + \delta$,

$$\delta(L) = (e_H - e_L) \frac{\sum_t \beta^{t-1} \left(1 - F_t(L)\right)}{\sum_t \beta^{t-1} \left(F_t(L) - \widehat{F}_t(L)\right)}. \tag{6}$$

Consider the following upper bound for the cost of the two-step contract:

$$K(\infty) < \frac{c_0}{1-\beta} = \frac{u^{-1}\left(u(c^*) + \delta(L)\right)}{1-\beta}.$$

The actual cost will be strictly lower than $\frac{c_0}{1-\beta}$ since, with probability $(1 - F_t(L)) > 0$ the agent receives $\underline{c}$. The final step of the proof is to show that by increasing L we can decrease the cost of the contract, since $\delta(L)$ is decreasing in L. When L increases, $\sum_t \beta^{t-1}(1 - F_t(L))$ decreases. Both $F_t(L)$ and $\widehat{F}_t(L)$ increase, but we have:

$$\begin{aligned}
\frac{1 - \widehat{F}_t(L)}{1 - F_t(L)} &\geq L \\
1 - \widehat{F}_t(L) &\geq L(1 - F_t(L)) \\
1 - \widehat{F}_t(L) - (1 - F_t(L)) &\geq L(1 - F_t(L)) - (1 - F_t(L)) \\
F_t(L) - \widehat{F}_t(L) &\geq (1 - F_t(L))(L - 1).
\end{aligned}$$

This implies

$$\sum_t \beta^{t-1} \left(F_t(L) - \widehat{F}_t(L)\right) \geq (L - 1) \sum_t \beta^{t-1} \left(1 - F_t(L)\right).$$

Substituting this inequality in expression (6),

$$\delta(L) < \frac{1}{L-1}(e_H - e_L).$$

We have that $\delta(L)$ is decreasing in L as long as $\sum_t \beta^{t-1}\left(F_t(L) - \widehat{F}_t(L)\right) > 0$ for L. From the above inequalities, this will hold whenever $1 - \widehat{F}_t(L) > 0$ for some t. For the discrete case, this

23

holds if there exists a path y^t such that $L\left(y^t\right) > L$, which is guaranteed in the i.i.d. case. [11] Hence, for any $\varepsilon > 0$ we can find an L low enough so that $K\left(\infty\right) < K^*\left(\infty\right) + \varepsilon$. $\blacksquare$

When we increase L, the $P\left(L\right)$ (i.e., the decrease in utility) that needs to be imposed so the contract is incentive compatible increases. However, increasing L also shrinks the sets $A_t^c\left(L\right)$ of histories that have the punishment attached: the probability of those histories in equilibrium, $\sum_t \beta^t \left(1 - F_t\left(L\right)\right)$, decreases with an increase in L. In the last step of the proof we show that, when we increase L, the decrease in $\sum_t \beta^t \left(1 - F_t\left(L\right)\right)$ is bigger than the corresponding increase needed in P, so the expected punishment decreases, allowing us to decrease δ. This is true because the term

$$\frac{\sum_t \beta^{t-1} \left(1 - F_t\left(L\right)\right)}{\sum_t \beta^{t-1} \left(F_t\left(L\right) - \widehat{F}_t\left(L\right)\right)}$$

is decreasing in L. This term is the inverse of the increase in the proportional probability of receiving a punishment if the agent were to change from high to low effort, which increases with L under our assumptions about the stochastic process for output. The intuition for this result parallels that of Mirrlees (74); in his paper, the richness of information is due to having infinitely many agents, while here we have infinitely many periods.

It should be noticed that for this result to hold the principal must have unlimited punishment power; i.e., the utility of the agent can be made as low as needed. Also, it is derived under the assumption of extreme persistence of effort, since output was assumed to be i.i.d. and the duration $\tau = T = \infty$. This is in fact what allows the quality of the information to keep growing and reach levels that permit tailoring punishments so that they are almost surely not exercised in equilibrium.

7   Changes in the Intensity of Persistence

The i.i.d. assumption allows us to characterize the optimal contract in interesting ways, 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 effort on the probability distribution of output may now decrease as time passes. We make the probability of observing y_H 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 $\overline{\pi}$:

$$\begin{aligned}
p_t\left(y_H|e_H\right) &= \alpha_t \pi + \left(1 - \alpha_t\right)\overline{\pi} \\
p_t\left(y_H|e_L\right) &= \alpha_t \widehat{\pi} + \left(1 - \alpha_t\right)\overline{\pi}.
\end{aligned}$$

The sequence of weights, $\{\alpha_t\}_{t=1}^T$ with $0 \le \alpha_t \le 1$ for every t, with $\alpha_t > 0$ for at least one t, represents the intensity of the persistence of effort at t: $\alpha_t = 1$ for all t corresponds to the

---

[11]This proof would go through for a more general assumption about the output stochastic process, as long as this condition is met. If the condition were not met, the maximum value of the Likelihood Ratio over all possible histories would determine the lower bound on ε.

i.i.d. case of perfect persistence, while $\alpha_t = 0$ for all t would imply that effort does not affect the distribution of output. We refer to $\{\alpha_t\}$ as a *persistence sequence*. Whenever the persistence sequence is decreasing, the effect of effort decreases over time. The implications of the duration of persistence described in section 5 still hold as long as $\alpha_t > 0$ for all $t \leq \tau$, i.e., as long as there is *some* information contained in realizations, the principal is better off when duration is longer.

To each persistence sequence $\{\alpha_t\}$ correspond two sequences of probabilities over outcomes $\{p_t (y_t|e_H)\}$ and $\{p_t (y_t|e_L)\}$. Using these probabilities, we can construct the corresponding probabilities over histories in the usual way. It is convenient for the analysis that follows to normalize probabilities over histories as in Eq. 2. This way, our problem can be interpreted as a static one, and each history of any possible length can be treated just as one of the possible signals h_i, with $i = 1, \ldots, 2^T$, in our (static) normalized problem. The normalized probability of signal $h_i = y^t$ is:

$$P(h_i) = \frac{1-\beta}{1-\beta^T}\beta^{t-1}\text{Pr}_t\left(y^t|e_H\right) = \frac{1-\beta}{1-\beta^T}\beta^{t-1}\prod_{j=1}^{t} p_t\left(y_t|e_H\right),$$

and $\widehat{P}(h_i)$ correspondingly for low effort.

Lowering persistence worsens the quality of information available. In the results that follow, we show that a decrease in the intensity of persistence increases the cost of implementing high effort. The main argument behind the results is similar to that in Prop. 13 in Grossman and Hart (83). We follow them in defining *information systems*.

Definition 19 *The information system defined by $\left(\pi, \widehat{\pi}, \overline{\pi}, \{\alpha_t\}_{t=1}^{T}\right)$ is described as the pair of vectors P (for high effort) and $\widehat{P}$ (for low effort) containing the normalized probabilities of all possible histories under the corresponding effort choice.*

To prove our main proposition, we show that any information system corresponding to higher intensity of persistence is sufficient, in the sense of Blackwell, for another system corresponding to lower intensity of persistence; i.e., we can find a stochastic matrix R (i.e., a matrix with all entries between zero and one, and with each of its columns summing up to one) such that any vector in the lower persistence system can be written as the corresponding vector in the higher persistence system times the matrix R.[12] In doing so, it is useful to first establish the following lemma:

Lemma 20 *Consider any two sequences of individual outcome probabilities $\{p_t (y_t)\}$, $\{p_t^0 (y_t)\}$ (all strictly positive,) where*

$$p_t^0 (y_t) = \gamma_t p_t (y_t) + (1 - \gamma_t) q_t (y_t), \quad for\ 0 \leq \gamma_t \leq 1$$

and $\{q_t (y_t)\}$ is some strictly positive probability sequence. Let $P_0 \left(y^t\right)$ and $P \left(y^t\right)$ be the corresponding probability distributions over histories for these two processes, respectively. We can find a stochastic matrix R such that $P^0 = RP$.

---

[12] See Blackwell and Girshick (54).

Proof. The probability of a history y^t corresponding to individual outcome probabilities $\{p_t^0(y_t)\}$ is:

$$P^0\left(y^t\right) = \Pi_j\left[\gamma_j p_j\left(y_j\right) + (1 - \gamma_j)q_j(y_j)\right].$$

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

$$\gamma_\theta \Pi_{j \in \theta} p_j\left(y_j\right) \tag{7}$$

where γ_θ is some coefficient that varies with the subset θ of terms considered, and $\theta \in \left(\bigcup_{j=1}^t 2^j\right) \cup \varnothing$ (i.e., all possible combinations of individual outcome probabilities in groups of size 1 to t). The constant term corresponds to $\theta = \varnothing$. Note that the individual probability terms that appear in Eq. (7), $p_j\left(y_j\right)$, can be expressed as the sum of the probability of all length t histories that coincide on this subset of realizations $\{y_j | j \in \theta\}$. Hence, each term multiplying γ_θ can be expressed as a linear combination of probabilities $P\left(y^t\right)$, as can the constant term as well. It follows that the vector P^0 containing all history probabilities $P^0\left(y^t\right)$ is also a linear combination of the vector P of all history probabilities defined by $P\left(y^t\right)$. Together, the coefficients $\{\gamma_\theta\}$ define a matrix R. Denote the entries of this matrix as r_{ij}. Since all histories have positive probability in every period, the sum of the elements of both P and P^0 is equal to T, and thus $\sum_{i=1}^{2^T} r_{ij} = 1$. Hence, R is a stochastic matrix such that $P_0 = RP$. ∎

In this context, the problem for the principal is completely described by the tuple of outside utility, effort disutility and primitives of the information system: $\left(\underline{U}, e, \pi, \widehat{\pi}, \overline{\pi}, \{\alpha_t\}_{t=1}^T\right)$. As in Grossman and Hart (83), sufficiency of one system for another in the sense of Blackwell implies a particular ranking of costs, as stated in the following proposition.[13]

Proposition 21 *Consider two problems* $\left(\underline{U}, e, \pi, \widehat{\pi}, \overline{\pi}, \{\alpha_t\}_{t=1}^T\right)$ *and* $\left(\underline{U}, e, \pi, \widehat{\pi}, \overline{\pi}, \{\alpha'_t\}_{t=1}^T\right)$, *where* $\alpha_t \geq \alpha'_t$ *for all t, with at least one strict inequality. The cost of the contract is strictly lower for the problem with the higher persistence sequence,* $\{\alpha_t\}_{t=1}^T$.

Proof. Consider the two information systems corresponding to the two normalized problems. In the previous lemma, let $p_t\left(y_t\right)$ denote the probabilities defined by the $\{\alpha_t\}$ sequence:

$$p_t\left(y_t\right) = \alpha_t \pi + (1 - \alpha_t)\overline{\pi},$$

and $p_t^0\left(y_t\right)$ denote the probabilities defined by $\{\alpha'_t\}_{t=1}^T$:

$$p_t^0\left(y_t\right) = \alpha'_t \pi + \left(1 - \alpha'_t\right)\overline{\pi}.$$

Let $\{q_t\left(y_t\right)\} = \overline{\pi}$ for all t in both cases. Let P and P^0 be the vectors containing the normalized probabilities (under high effort) of all possible histories h_i, with $i = 1, \ldots, 2^T$. The typical element in these vectors, for $h_i = y^t$, is

$$P(h_i) = \frac{1 - \beta}{1 - \beta^T}\beta^{t-1} P\left(y^t | e_H\right),$$

[13]Kim (95) provides a sufficient condition to rank incentive problems which is weaker than Blackwell sufficiency. He looks at the distribution function of the likelihood ratios of different problems, and shows that if one distribution is a Mean Preserving Spread of another, then in the first case effort is less costly to implement.

and

$$P^0\left(h_i\right) = \frac{1-\beta}{1-\beta^T}\beta^{t-1}P^0\left(y^t|e_H\right),$$

correspondingly. With $\gamma_t = \alpha'_t/\alpha_t$, the previous lemma applies: $\exists R$ such that $P^0 = RP$. The lemma can be applied also for the probabilities that are constructed for each α sequence under the assumption of low effort:

$$\widehat{p}_t\left(y_t\right) = \alpha_t\widehat{\pi} + \left(1-\alpha_t\right)\overline{\pi},$$

$$\widehat{p}_t^0\left(y_t\right) = \alpha'_t\widehat{\pi} + \left(1-\alpha'_t\right)\overline{\pi},$$

with vectors of history probabilities $\widehat{P}$ and $\widehat{P}^0$. Moreover, the matrix R that satisfies $\widehat{P}^0 = R\widehat{P}$ is the same as the one in $P^0 = RP$. We conclude that the first information structure is sufficient for the second. Denote by $C = \left\{c_1\left(y^t\right)\right\}_{t=1}^T$ the optimal contract corresponding to persistence sequence $\{\alpha_t\}$, and C' the contract corresponding to $\{\alpha'_t\}$. Following the proof of Grossman and Hart (83), Prop. 13, we first show that C' can be replicated under the $\{\alpha_t\}$ information system. After observing realization $y^t = h_i$, the principal performs a randomization across all 2^T possible histories with the probabilities determined by the ith column of matrix R. It follows that a payment $c'\left(h_i\right)$ is provided to the agent with probability

$$\sum_{j=1}^{2^T} r_{ji}p_i = P^0\left(h_i\right).$$

By construction, C' satisfies the PC and the IC of the agent. This establishes that the cost of the optimal contract under the $\{\alpha_t\}$ system is never greater than that under the $\{\alpha'_t\}$ system.

Since there exists at least one t such that $\alpha'_t < \alpha_t$, the R matrix is not equal to the unit diagonal matrix (i.e., some randomizing is needed to replicate C' in the above proposed scheme.) Since the agent has strictly concave utility, this implies that the principal can implement high effort at a lower cost by offering a payment at each y^t that provides the agent with the same expected utility as the randomization in R, without uncertainty. Hence, the cost of the optimal contract under the $\{\alpha_t\}$ information system must be strictly lower. ∎

As a way of illustration, we can prove the above result for the case of $u\left(c\right) = 2\sqrt{c}$ using the closed form solution. In this case we can also derive implications for variability of compensation: we can determine how each individual v_t depends on the difference $\left(\alpha_t - \alpha'_t\right)$, and we know that the variance of the inverse of the marginal utility of consumption at time t is inversely proportional to the variance of the likelihood ratios, v_t. We can establish the following result:

Proposition 22 *Assume the agent's utility is given by $u\left(c\right) = 2\sqrt{c}$. Consider two possible persistence sequences $\left(\alpha_1, ..., \alpha_T\right)$ and $\left(\alpha'_1, ..., \alpha'_T\right)$ where $\alpha_t \geq \alpha'_t$ for all t, with strict inequality for at least one t. The average variance of utility and the cost of the contract are strictly lower for the problem with higher persistence, $\left(\alpha_1, ..., \alpha_T\right)$.*

Proof. For a given persistence sequence, we have

$$E\left[LR\left(y^t\right)|e_H\right] = 1 \quad \forall t$$

and

$$E\left[LR\left(y^t\right)|e_L\right] \equiv \prod_{\tau=1}^{t} \widehat{E}_\tau,$$

where

$$\widehat{E}_t \equiv E\left[LR\left(y_t\right)|e_L\right] = \frac{\Pr_t\left(y_H|e_L\right)^2 + \Pr_t\left(y_H|e_H\right) - 2\Pr_t\left(y_H|e_L\right)\Pr_t\left(y_H|e_H\right)}{\Pr_t\left(y_H|e_H\right)\left(1 - \Pr_t\left(y_H|e_H\right)\right)}$$

is the expectation of the likelihood ratio of an individual output in period t, when effort is low. Note that

$$\widehat{E}_t - 1 = \frac{\left[\Pr_t\left(y_H|e_H\right) - \Pr_t\left(y_H|e_L\right)\right]^2}{\Pr_t\left(y_H|e_H\right)\left(1 - \Pr_t\left(y_H|e_H\right)\right)},$$

so it is easy to see that $\widehat{E}_t$ is increasing in $\Pr_t\left(y_H|e_H\right) - \Pr_t\left(y_H|e_L\right)$ for all t. The variance of the likelihood ratios at any t is

$$\begin{aligned} v_t &= E\left[LR\left(y^t\right)|e_L\right] - 1 \\ &= \prod_{\tau=1}^{t} \widehat{E}_\tau - 1, \end{aligned}$$

which is also increasing in any of the $\Pr_t\left(y_H|e_H\right) - \Pr_t\left(y_H|e_L\right)$. When comparing the two persistence sequences, we can write

$$\Pr_t\left(y_H|e_H\right) - \Pr_t\left(y_H|e_L\right) = \alpha_t\left(\pi - \widehat{\pi}\right)$$

and, for the second sequence,

$$\Pr'_t\left(y_H|e_H\right) - \Pr'_t\left(y_H|e_L\right) = \alpha'_t\left(\pi - \widehat{\pi}\right)$$

Since $\alpha_t \geq \alpha'_t$, it follows that $v_t \geq v'_t$ for all t, with strict inequality for at least one t, so the average variance of the Likelihood Ratios corresponding to the first sequence is strictly higher:

$$\bar{v} > \bar{v}'.$$

As shown in the proof of Prop. 17, Eq. (5), a higher average variance of the likelihood ratios results in lower cost. The expression for average variance of utility is

$$\frac{1-\beta}{1-\beta^T}Var_i\left[u\left(c\left(y^t\right)\right)|e_H\right] = \frac{1-\beta}{1-\beta^T}e_H^2\frac{v_{ti}}{\bar{v}^2} = \frac{e_H^2}{\bar{v}},$$

which is clearly decreasing in $\bar{v}_i$. ∎

A smaller α_t represents a more severe incentive problem. The distributions of output under high and low effort are more difficult to discriminate statistically under the α' sequence, even as we get to later periods of the contract. There is less benefit in waiting to provide incentives, so the spread of consumption may be more even across periods. The higher average variance of utility is due to the higher value of μ in the contract with less intensity of persistence, which dominates over the relatively lower variance of the likelihood in this contract. The effect on individual period variance of compensation is not determined, since each v_t varies with its corresponding α_t.

We now present some numerical examples for logarithmic utility that allow us to discuss more explicitly how decreasing the intensity of persistence influences the optimal contract. Similar results to those of the square root specification hold in our examples. For these exercises, 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 two different levels of persistence. Table 3 contains two matrices: the first one reproduces the results in Table 1 for the logarithmic utility, corresponding to the case of $\alpha = 1$. The second one presents the results for lower intensity of persistence, with $\alpha = 0.9$ and $\overline{\pi} = 0.5$ (all other parameters are kept equal.)

$\alpha = 1$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{c_t}$		$\alpha = .9$	$\frac{E[c_t]}{c^*}$	$\frac{\sigma_t}{E[c_t]}$	$\frac{d_t}{c_t}$
$t = 1$	1.32	0.30	0.83		$t = 1$	1.48	0.34	1
$t = 2$	1.32	0.44	2.10		$t = 2$	1.48	0.46	2.5
$t = 3$	1.32	0.54	5.33		$t = 3$	1.48	0.54	6.70
$t = 4$	1.32	0.63	132.1		$t = 4$	1.48	0.60	4097

$$\lambda = 13.14 \quad \mu = 18.31 \qquad\qquad \lambda = 14.75 \quad \mu = 26.26$$

Table 3. Changes in persistence of effort: effect on variability of
consumption

When α is lower, the effect on the cost of the contract parallels that of a decrease in τ. The expected consumption increases when α decreases (it goes from the original 1.3 of the First Best cost when $\alpha = 1$ to 1.48 of the First Best when $\alpha = .9$), reflecting the increase in the risk premium due to the higher average variability. The effect on the variability of consumption in each period, as mentioned above, depends on the combination of two factors: the increase in μ (it goes from 18.31 to 26.26) and the change in the variance of the likelihood ratios at every period. Looking at the scaled standard deviation, we can see that for periods one to three the increase in the multiplier dominates and $\frac{\sigma_t}{E[c_t]}$ increases or stays the same. We can see, however, that $\frac{\sigma_4}{E[c_4]}$ is lower for $\alpha = .9$, implying a significant drop in the variance of the likelihood ratios in period four. This is consistent with the faster decrease in informativeness of the fourth period when α is lower. The value of $\frac{d_t}{c_t}$ increases in every period, including the last, since this measure does not take into account changes in the distribution over consumption values.

8   Conclusions

We study a simple representation of a moral hazard problem with persistence in which only one effort is taken by the agent at the beginning of the contract. This effort determines the probability distribution of outcomes in all the periods to come. In principle, the implications of our model apply to a large class of environments. For example: the design of compensation in firms where

an initial investment in human capital is needed, or when sorting of high skilled workers is to be done at the time of hiring; the design of a tax scheme or an unemployment program that would provide incentives for acquiring skills early in the lifetime of agents; or the design of optimal compensation for CEOs and hiring committees of sports clubs, editorial and record companies. The optimal contract derived in this paper suggests that, whenever commitment to long term contracts is available, the efficient provision of incentives calls for an increase in the variability of consumption over time. Moreover, it suggests that the stronger the importance for production of the unobserved effort (or the unobserved skills or unobserved investment in human capital), the bigger the efficiency gains from postponing incentives, and the higher the level of insurance provided to the agent in early periods (or the lower the variance of compensation within cohorts of agents).

Our model is a partial approximation to the problem of compensation design in those complex environments. In its simplicity, it abstracts from many important elements that may change the form of the optimal contract. In particular, in most of the examples the agents may be able, or required, to exert further unobservable efforts during the whole relationship with their employers — efforts that may or may not be persistent. Combining a repeated effort incentive problem with the persistence framework presented here is a natural next step towards understanding the importance of persistence in many relevant contracting environments.

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] Blackwell, D., and M. A. Girshick. *Theory of Games and Statistical Decisions*. New York. John Wiley and Sons, Inc., 1954.

[4] Fernandes, A. and C. Phelan. "A Recursive Formulation for Repeated Agency with History Dependence," *Journal of Economic Theory,* **91** (2000): 223-247.

[5] Grochulski, B. and T. Piskorski. "Risky Human Capital and Deferred Capital Income Taxation." Mimeo (2006)

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

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

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

[9] Jarque, A. "Repeated Moral Hazard with effort Persistence". Mimeo (2005)

[10] Jarque, A. "Optimal Stock Option Repricing: Incentives and Learning". Mimeo (2007)

[11] Kim, S. K. "Efficiency of an Information System in an agency Model". Econometrica, vol **63**(1), pages 89–102 (1995)

[12] Kwon, I. "Incentives, Wages, and Promotions: Theory and Evidence". *Rand Journal of Economics*, **37** (1), 100-120 (2006)

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

[14] 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)

[15] Mukoyama, T. and A. Sahin, "Repeated Moral Hazard with Persistence," *Economic Theory,* vol. **25**(4), pages 831-854, 06 (2005)

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

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

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

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