



UNIVERSIDAD CARLOS III DE MADRID

working
papers

Working Paper 09-27
Economic Series (17)
March 2009

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

Unobservable Persistent Productivity and Long Term Contracts^{*}

Hugo Hopenhayn
UCLA

Arantxa Jarque
Universidad Carlos III

Abstract

We study the problem of a firm that faces asymmetric information about the productivity of its potential workers. In our framework, a worker's productivity is either assigned by nature at birth, or determined by an unobservable initial action of the worker that has persistent effects over time. We provide a characterization of the optimal dynamic compensation scheme that attracts only high productivity workers: consumption –regardless of time period– is ranked according to likelihood ratios of output histories, and the inverse of the marginal utility of consumption satisfies the martingale property derived in Rogerson (1985). However, in the case of i.i.d. output and square root utility we show that, contrary to the features of the optimal contract for a repeated moral hazard problem, the level and the variance of consumption are negatively correlated, due to the influence of early luck into future compensation. Moreover, in this example long-term inequality is lower under persistent private information.

Journal of Economic Literature Classification Numbers: D80, D82.

Key Words: mechanism design; moral hazard; persistence

* An earlier version of this paper circulated under the title "Moral Hazard and Persistence." We would like to thank two anonymous referees and the editor for their valuable comments. We would also like to thank Árpád Ábrahám, Hector Chade, Lola Collado, Huberto Ennis, Borys Grochulski, Juan Carlos Hatchondo, Leonardo Martínez, Ned Prescott, Michael Raith and seminar audiences at Universidad Carlos III, Universidad de Alicante, the 2006 Wegmans Conference in Rochester, the Richmond Fed, the 2006 Summer Meetings of the Econometric Society in Minnesota, the 2006 Meetings of the SED in Vancouver, and the Ente Einaudi. All remaining errors are ours. Jarque gratefully acknowledges financial support from the Research Projects financed by the Ministry of Education, Culture and Sports of Spain, SEJ2004 - 08011ECON and SEJ2007-62656. Correspondence: ajarque@econ.uc3m.es.

1 Introduction

We study the problem of a firm that faces asymmetric information about the productivity of its potential workers. The output of the firm depends stochastically but positively on the productivity of its employed worker. We assume that after signing the contract all work effort is observable, so there is no further asymmetry of information. In our framework, there are two alternative ways of interpreting the origin of the difference in productivity across workers. In one alternative, productivity is assigned randomly by nature at birth, and it affects not only the distribution of output in the firm but also the worker's outside value. In a second alternative, productivity is determined by an unobservable action of the worker taken prior to the relationship with the firm, which has persistent effects in time. This action makes the worker more productive in the specific job that he performs in the firm, but does not affect his outside value. We show that both formulations fit into our model. We study how the persistence of productivity can be exploited in long term contracts to sort workers.¹ We characterize the contract that attracts only high productivity workers at the minimum cost, and compare its properties to optimal contracts in the presence of non-persistent asymmetric information.

The model is as follows. The contract lasts for an exogenously specified number of periods. At the beginning of the relationship, the firm (principal) offers a contract to the worker (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 stochastic and depends positively (first-order stochastic dominance) on the agent's productivity. Every period, the agent consumes according to the contingent scheme specified in the contract, but he does not exert any further unobservable 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 is signed in equilibrium only by high productivity types – implements high effort – at the lowest expected discounted cost.

Our simple model allows us to impose very few restrictions on the stochastic process for output. This generality is useful when studying optimal compensation contracts in professions where output informativeness about productivity varies strongly over time. Our dynamic model with persistence captures essential features of many important long term relationships in which productivity is unobservable, like the design of optimal compensation for CEOs, or for committees in charge of hiring in sports clubs, editorial and record companies, or the design of a tax system that provides incentives for human capital accumulation.² Miller (1999) originally used a variation of the model

¹See Lazear (2000) for a discussion of contingent contracts as a sorting device in a static model with risk neutral agents. See Autor (2001) for evidence of screening efforts by firms, and a model of sorting that takes competition into account. His model, however, focuses on the role played by temporary help supply firms, and hence does not study the evolution of wages on the job. In a framework with heterogeneous firms and workers, Li and Sue (2000) study the role of early matching in sorting, as observed in college admissions, summer internships in law firms, or matching of medical interns and residents with hospitals.

²See 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.

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.

It stems from our analysis that, in spite of its dynamic structure, our moral hazard problem with persistence can be formally studied as a static moral hazard situation (Holmström (1979), Grossman and Hart, 1983.) 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. Because the agent consumes every period, our model can be understood as a modification of Holmström and Milgrom (1987). Hence, we confirm their conjecture that consumption in the optimal contract is not, in general, a linear function of output, and that the principal can do better than in the repeated effort case, because he faces less incentive constraints.

Our characterization of the optimal contract has also important implications for the dynamics of consumption. In spite of the incentive problem being inherently static, the inverse of the marginal utility of consumption satisfies the martingale property derived in Rogerson (1985) (see also Golosov, Kocherlakota and Tsyvinski, 2003). This implies that, as in most dynamic problems with asymmetric information, including the standard repeated moral hazard model, 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 and output takes only two values, our model provides some stark predictions. The contract takes a simple form: the current consumption of the agent depends only on the number of periods he has been in the contract (his tenure) and the number of high output realizations observed to the date. Longer histories contain more information, so the dispersion of likelihood ratios and the variance of compensation increases over time. We provide a closed form solution for the optimal contract assuming that the utility of the agent is given by the square root of consumption. We use this specification as our leading example, and complement it with numerical solutions to illustrate the main general properties of the contract.

The key difference between our model and the standard incentive problems studied in most of the literature is the persistent nature of the asymmetric information. To understand the implications of this persistence, we use our leading example of square root utility and the i.i.d. framework to compare the features of the optimal contract in our framework with those of the optimal contract in a related repeated moral hazard model.³ In the standard no-persistence setting, the productivity of a worker in a given period is determined by his hidden work effort in that particular period only. Long term contracts are used to smooth incentives over time. Our analysis shows that the long-term features of the optimal contracts with and without persistence are very different, and hence persistence cannot be safely ignored in the study of optimal compensation contracts. In our example, we find that persistence implies lower levels of long-term inequality within cohorts of workers. Moreover, when productivity is persistent, the uncertainty faced by a worker in a

³There is a large literature in labor economics developing models that include realistic features such as worker mobility, competition among firms for talented workers, promotions, task assignments, and learning on the job. This literature typically contrasts its results against data (see, for example, Gibbons and Waldman (1999a, 1999b, 2006), or Baker et al. (1994a, 1994b)). Our focus here is not to produce testable implications. Instead, we provide a simple model that allows us to understand the role of persistence of asymmetric information. Hence, we choose to compare the implications of our model with those of other theoretical models, and not with the data.

given period of his career depends strongly on the previous work history. A good stream of output realizations in early periods (early good luck) translates into low conditional variance of utility, while early bad luck translates into higher conditional variance of utility. This implies a negative correlation of the level and the variance of consumption over time. This contrasts again with the predictions of the repeated effort model, in which the uncertainty over future utility streams faced by the worker is independent of past history, implying a positive correlation of the level and the variance of consumption over time.

Under the i.i.d. assumption for output, we show that for a contract that lasts for an infinite number of periods the cost of implementing high effort is arbitrarily close to that of the contract with perfect information. 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).

To complement our analysis of the i.i.d. framework, we introduce the following variation: after an exogenous number of periods of i.i.d. realizations, the effect of effort completely dies out. We define the number of consecutive periods in which effort affects the distribution of output as the “duration” of persistence. We show that increasing the duration of persistence decreases the cost of implementing high effort. In our leading example, 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.

The paper is organized as follows. The model is presented in the next section. A characterization of the optimal contract is given for a general stochastic process in section 3. Results and numerical examples for the i.i.d. case are discussed in section 4, with a subsection devoted to the case of infinite contracts. In section 5 we analyze the implications of changes in the duration of persistence. Section 6 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 and strictly increasing utility of consumption $u(c)$. There is the same finite set of possible outcomes each period, $Y = \{y_i\}_{i=1}^n$, with $y_i < y_{i+1}$ for all $i = 1, \dots, n$. Let Y^t denote the set of histories of outcome realizations up to time t , with typical element $y^t = \{y_1, y_2, \dots, y_t\}$. This history of outcomes is assumed to be common knowledge. We model the productivity of the agent as the probability distribution over output that he induces by working at the firm. Productivity is determined by the agent’s unobservable effort in the first period of the relationship.⁵ This effort can take two possible values, $e \in \{e_L, e_H\}$.⁶ A contract

⁴A solution to the cost minimization problem presented later in this section does not exist when $T = \infty$. The case of infinite T is discussed later in the paper, where an asymptotic approximation result is presented.

⁵Section 2.1 presents the alternative interpretation in which productivity is randomly assigned by nature, as opposed to being endogenously determined by effort. We show that the two specifications are equivalent. Hence, all results in the paper hold in both frameworks.

⁶As it becomes clear in the core of the paper, the results presented here generalize to the case of multiple effort levels, just as the results in a static moral hazard problem. With finite effort levels, it may be that some of the

prescribes an effort to the agent at time 1, as well as a transfer c from the principal to the agent for every period of the contract, contingent on the history of outcomes up to t : $c : Y^t \rightarrow [c_{\min}, c_{\max}]$, for $t = 1, 2, \dots, T$. Denote the probability of a given history of outcomes, conditional on choosing effort level e in the first period, as $\Pr(y^t|e)$. 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 and one y^t such that $\Pr(y^t|e_H) \neq \Pr(y^t|e_L)$. 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.⁷

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:

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

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 consumption 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: $K(T) = K(T, e_H)$. 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, $\{c(y^t)\}_{t=1}^T$.

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} :

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

levels are not implementable. For a continuum of efforts, our characterization may fail if the problem of the agent corresponding to the incentive constraint is not strictly concave. Sufficient conditions are as in the first order approach (see Rogerson (1985b), Jewitt, 1988).

⁷See Fudenberg, Homstrom and Milgrom (1990), Example 1, for a clear explanation of the value of commitment in frameworks with persistence.

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:

$$K^*(T) = \min_{\{c(y^t)\}_{t=1}^T} K(T)$$

s.t. PC

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

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

Later in the paper we use the cost of the first best scheme, $K^*(T) \equiv \frac{1 - \beta^T}{1 - \beta} c^*$, 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:

$$\begin{aligned} & \sum_{t=1}^T \sum_{y^t} \beta^{t-1} u(c(y^t)) \Pr(y^t | e_H) - e_H \\ & \geq \sum_{t=1}^T \sum_{y^t} \beta^{t-1} u(c(y^t)) \Pr(y^t | e_L) - e_L. \end{aligned} \tag{IC}$$

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 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:

$$\begin{aligned} & \min_{\{c(y^t)\}_{t=1}^T} K(T) \\ & \text{s.t. PC and IC} \end{aligned} \tag{CM}$$

2.1 Alternative interpretation: Sorting Types

With a simple relabeling of terms, our model applies to adverse selection problems. In these situations, the productivity of the agent, i.e. the probability distribution over output that he induces by working at the firm, is randomly assigned by nature. The agent knows his productivity, but the firm cannot observe it. Productivity may be high or low: $\theta \in \{\theta_H, \theta_L\}$. Denote the probability of a given history of outcomes, conditional on productivity type, as $\Pr(y^t | \theta_H)$. Assume that an agent with high productivity has an outside utility of \tilde{U}_H . The low productivity worker, instead, has an outside utility of $\tilde{U}_L < \tilde{U}_H$.

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

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

Relabeling $\underline{U} + e_H = \tilde{U}_H$, and setting $\Pr(y^t | \theta_H) = \Pr(y^t | e_H)$, 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:

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

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

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

which, substituting \underline{U} from the PC, is equivalent to

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

Setting $\Pr(y^t | \theta_L) = \Pr(y^t | e_L)$, this last equation is equivalent to our original IC. This condition 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 to the design of optimal contracts when firms face potential workers who have private information about their own abilities. All results presented in the paper using the persistent effort moral hazard framework apply to this adverse selection framework as well.

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

The optimal contract can be characterized from 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 $LR(y^t)$, is defined as the ratio of the probability of observing y^t under a deviation, to the probability under the recommended level of effort:

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

Proposition 1 *The optimal sequence $\{c(y^\tau)\}_{\tau=1}^T$ 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 $\tilde{y}^{t'}$ of (possibly) different lengths t and t' ,*

$$c(y^t) > c(\tilde{y}^{t'}) \Leftrightarrow LR(y^t) < LR(\tilde{y}^{t'})$$

This simple characterization is due to the fact that, in spite of its dynamic structure, this problem can be reduced to a standard static moral hazard case. We clarify this before presenting the proof for the proposition. It is of course key that the agent chooses effort only once. This implies that, although incentives are optimally smoothed over time, they are evaluated only once by the agent, at the moment of choosing his action. Define $\gamma_T \equiv \frac{1-\beta}{1-\beta^T}$. It is easy to see that the principal is indifferent between minimizing the total cost of the contract as in problem CM and solving the following normalized problem:

$$\begin{aligned} & \min_{\{c(y^t)\}_{t=1}^T} \gamma_T \sum_{t=1}^T \sum_{y^t} \beta^{t-1} \{c(y^t)\} \Pr(y^t|e_H) \\ \text{s.t. } & (\gamma_T)\underline{U} \leq \gamma_T \left\{ \sum_{t=1}^T \sum_{y^t} \beta^{t-1} \{u(c(y^t))\} \Pr(y^t|e_H) \right\} - (\gamma_T)e_H \quad (\text{NCM}) \\ & \gamma_T \left\{ \sum_{t=1}^T \sum_{y^t} \beta^{t-1} \{u(c(y^t))\} \Pr(y^t|e_H) \right\} - (\gamma_T)e_H \\ & \geq \gamma_T \left\{ \sum_{t=1}^T \sum_{y^t} \beta^{t-1} \{u(c(y^t))\} \Pr(y^t|e_L) \right\} - (\gamma_T)e_L \end{aligned}$$

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 multiplied by the averaging term γ_T ; together, they define a weight for time t : $\omega_t^T \equiv \gamma_T \beta^t$, where $\sum_t \omega_t^T = 1$, and the superindex indicates the dependence of this value on the length of the contract, T . We can rename a history y^t of arbitrary length as $h_i \in H_T$, $i = 1, \dots, I(T)$, where $H_T \equiv \cup_{t=1}^T Y^t$ is the set of all possible histories in a T -period problem and $I(T) = \sum_{t=1}^T 2^t$. History y^t corresponding to h_i happens with “weighted” probability

$$P_i(e) \equiv \omega_t \Pr(y^t|e), \quad (1)$$

and we have $\sum_i P_i = 1$. 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 in the averaged problem, to the per period value that, discounted, sums up to the original utility amount. These normalized constraints are equivalent to the dynamic ones.

In a static moral hazard problem, the information structure 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. As we just argued, the states in the dynamic case are all histories in H_T . Each $h_i \in H_T$ happens with probability $P_i(e)$. 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.

Proof of Proposition 1. As we just argued, our problem is formally equivalent to a static moral hazard problem. Our assumptions on the utility of the agent imply that the objective function is continuous, differentiable and strictly convex. As is standard in the literature, we can write the problem with contingent utility levels, $u(y^t)$, as choice variables (as opposed to consumption levels). The domain of consumption translates into a domain constraint (DC):

$$u(y^t) \in [u(c_{\min}), u(c_{\max})]. \quad (\text{DC})$$

The change of variables makes the constraints linear in the choice variables, so compactness and convexity of the domain follow easily. Hence, a solution exists and is unique. Since utility is separable in consumption and effort, the standard argument applies to show that the PC is binding. From the FOC's,

$$c(y^t) : \frac{1}{u'(c(y^t))} = \lambda + \mu [1 - LR(y^t)] \quad \forall y^t, \quad (2)$$

where λ and μ are the multipliers associated with the PC and the IC respectively. For the IC to be satisfied it is necessary that $\mu > 0$. 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, i.e., to those that have a high likelihood ratio. Unlike in the static problem, however, in our dynamic model the distribution of output over time determines the timing of the most informative histories, and hence the timing of punishments. For example, if output is i.i.d., informative histories will be histories in the latest periods of the contract, meaning incentives will be delayed. This case is studied in detail in the next section.

It is worth noting that our framework falls into the class of problems of repeated asymmetric information studied in Golosov, Kocherlakota and Tsyvinski (2003). Hence, it is of no surprise that equation 2 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 (RMH) problem, generalized by Golosov, Kocherlakota and Tsyvinski (2003):

$$\frac{1}{u'(c(y^t))} = \sum_{y_{t+1}} \frac{1}{u'(c(y^t, y_{t+1}))} \Pr(y_{t+1} | e_H, y^t). \quad (3)$$

This property implies that the agent, if allowed, would like to save part of his consumption transfer every period in order to smooth his consumption over time. As shown in Rogerson (1985), it also

implies that the expected consumption of the agent decreases with time whenever $\frac{1}{u'(c)}$ is convex, increases if it is concave, and is constant whenever utility is logarithmic. We show in the next section, however, that the implications for variability of consumption over the contract in our model are very different from the repeated actions and no persistence model in Rogerson (1985).

4 Outcomes Independently and Identically Distributed

In this section, we study a particular specification of the probability distribution of output: we assume it is independently and identically distributed (i.i.d.) across periods. For the rest of the paper, we analyze the two outcomes case, $Y_t = \{y_L, y_H\}$, and we assume

$$\begin{aligned}\Pr(y_t = y_H|e_H) &= \pi \\ \Pr(y_t = y_H|e_L) &= \hat{\pi} \quad \forall t = 1, \dots, T.\end{aligned}$$

with $\pi > \hat{\pi}$. This assumption puts additional structure on the probability distribution of histories, and allows for the optimal contract to be further characterized.

Note that

$$\begin{aligned}\frac{\Pr(y_t = y_H|e_L)}{\Pr(y_t = y_H|e_H)} &= \frac{\hat{\pi}}{\pi} < 1, \\ \frac{\Pr(y_t = y_L|e_L)}{\Pr(y_t = y_L|e_H)} &= \frac{1 - \hat{\pi}}{1 - \pi} > 1.\end{aligned}$$

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

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

Hence, in the two outcome setup there is perfect substitutability of output realizations across time. The tuple $\{t, x(y^t)\}$ 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 $x(y^{t-1})$ to determine current consumption. Simply put, the number of high outputs in his work history, together with his tenure in the contract are sufficient to determine the agent's current consumption.

Denote \bar{y}^t the history at t containing $y_t = y_H$ for all t , that is, the history of length t that satisfies $x(\bar{y}^t) = t$. Similarly, let \underline{y}^t denote the history with $x(\underline{y}^t) = 0$. The following corollary summarizes the direct implications of Proposition 1 in the two output i.i.d. framework.

Corollary 1 *Assume output can only take two values, $\{y_L, y_H\}$, and it is i.i.d. over time. The following properties hold:*

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

2. For any two histories of the same length y^t and \tilde{y}^t , $c(y^t) \geq c(\tilde{y}^t)$ if and only if $x(y^t) \geq x(\tilde{y}^t)$, regardless of the sequence in which the realizations occurred in each of the histories.
3. As t increases, $c(\bar{y}^t)$ increases and $c(\underline{y}^t)$ decreases. Hence, as t increases, $c(\bar{y}^t) - c(\underline{y}^t)$ increases.

Property 2 is not necessarily true in the optimal contract corresponding to a RMH problem (Rogerson, 1985). When effort is chosen every period, past realizations are less important than recent ones in determining compensation in a given period. In particular, the *spread* of utility in a given node (i.e., $u(y_H | y^{t-1}) - u(y_L | y^{t-1})$) is determined by effort disutility and the likelihood ratio of the outcome in that period only. However, the *level* of expected utility at that node (i.e., $E[u(y_t | y^{t-1})]$) depends on all past history, because the optimal contract spreads incentives for past efforts into all future consumptions. The combination may imply that two histories with the same $x(y^t)$, for example $y^t = (y_H, y_L)$ and $\tilde{y}^t = (y_L, y_H)$, may be assigned $c(y^t) < c(\tilde{y}^t)$.⁸ In our framework with persistence, however, the whole history is evaluated as a single signal about the initial effort, and hence the number of high realizations fully determines the ordering of consumption, regardless of its timing.

Property 3 illustrates an implication of i.i.d. output on the timing of incentives. As seen in 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, and it translates into a wider range of utilities in the contract as time goes by.

As it turns out, we can further characterize the stochastic properties of consumption over time. In order to do this, it is useful to characterize the distribution of the likelihood ratios over time, and use its moments. We now turn to that.

Probability Distribution over likelihood ratios

To each history y^t corresponds a likelihood ratio $LR(y^t)$. The probability of observing that particular likelihood ratio is the probability of the $x(y^t)$ that generates it. Clearly, $x(y^t)$ at t follows a binomial distribution with t trials and a probability of success π (or $\hat{\pi}$ if low effort is chosen). Hence, in equilibrium, under high effort choice,

$$\Pr(LR(y^t) | e_H) = \Pr(y^t | e_H) = \binom{x(y^t)}{t} \pi^{x(y^t)} (1 - \pi)^{t - x(y^t)},$$

where $\binom{x}{t}$ denotes the standard combinatorial function. Hence, we have a well defined distribution over likelihood ratios. 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:

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

⁸In the RMH numerical example presented later in Table 2, for example, we have $c(y_H, y_L) = 1.16$ and $c(y_L, y_H) = 2.26$.

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, $E [LR(y^t) | e_L]$. Let \widehat{E} denote this expectation at $t = 1$:

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

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

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

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

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

Lemma 1 *The variance of the likelihood ratios is an increasing and strictly convex function of t .*

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 $t = 1, \dots, T - 1$, the one-period increase in the variance equals

$$v_t - v_{t-1} = \widehat{E}^t - 1 - (\widehat{E}^{t-1} - 1) = (\widehat{E} - 1) \widehat{E}^{t-1},$$

which increases with t . ■

Leading example

Before moving on to analyze the properties of the optimal contract further, we provide a complete closed form solution of an example, assuming $u(c) = 2\sqrt{c}$. This corresponds to a CRRA utility function with coefficient of risk aversion equal to $\frac{1}{2}$. We use this specification as our leading example.

Without loss of generality, we set $c_{\min} = 0$. We limit our analysis to values of e , \underline{U} , π , $\widehat{\pi}$, τ and T that satisfy

$$\frac{U}{e} + 1 \geq \frac{1}{\bar{v}} \left[\left(\frac{1 - \widehat{\pi}}{1 - \pi} \right)^T - 1 \right], \quad (4)$$

where \bar{v} denotes the weighted average across periods of the variance of the likelihood ratios:

$$\bar{v} \equiv \sum_{t=1}^T \omega_t^T v_t.$$

Under condition 4, the constraint DC does not bind, (i.e., $\sqrt{c(y^t)} \geq 0$ for all y^t .)⁹ The explicit solutions for the multipliers are:

$$\lambda = \frac{\gamma_T (\underline{U} + e_H)}{2} \quad \text{and} \quad \mu = \frac{\gamma_T (e_H)}{2} \frac{1}{\bar{v}}.$$

⁹It is easy to find parameters that satisfy condition 4 (it holds in all numerical examples given in the paper.) See Section 4.1 for a discussion of the cases when this condition is not satisfied.

The solution for consumption is:

$$c(y^t) = [\lambda + \mu(1 - LR(y^t))]^2 \quad \forall t, \quad \forall y^t \in Y^t.$$

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

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

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

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

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. Hence, an increase in the variance of the likelihood ratios translates into cost savings.

Long term implications of unobserved persistent productivity

In this section we derive some further properties of the process for consumption in our setup and later we provide some comparisons with the repeated moral hazard case. We first analyze the properties of the inverse of the marginal utility. Later in this section we discuss the relationship of this measure to the variance of consumption.

Proposition 2 *The variance of $\frac{1}{w'(c(y^t))}$ is an increasing and strictly convex function of t . Moreover, the variance of $\frac{1}{w'(c(y^t))}$ conditional on y^{t-1} decreases with $x(y^{t-1})$.*

Proof. From eq. 3 and the formula for the variance we have

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

The result in the proposition follows from lemma 1. From the same first order conditions we have that

$$E\left(\frac{1}{w'(c(y^t))}|e_H, y^{t-1}\right) = \lambda + \mu(1 - LR(y^{t-1})).$$

and

$$Var\left(\frac{1}{w'(c(y^t))}|e_H, y^{t-1}\right) = \mu^2 [LR(y^{t-1})]^2 v_1,$$

where $[LR(y^{t-1})]^2 v_1$ is the variance of $LR(y^t)$ conditional on y^{t-1} . The result follows from the fact that $LR(y^{t-1})$ decreases with $x(y^{t-1})$. ■

We postpone the discussion of this proposition to state the following corollary for our leading example.

Corollary 2 *When the agent's utility is given by $u(c) = 2\sqrt{c}$, the variance of $u(c(y^t))$ is given by*

$$\text{Var}(u(y^t)) = 4\mu^2 v_t = \left(\frac{e\gamma_T}{\bar{v}}\right)^2 v_t.$$

This variance is an increasing and strictly convex function of t . Moreover, the variance of $u(c(y^t))$ conditional on y^{t-1} is given by

$$\text{Var}(u(y^t) | e_H, y^{t-1}) = 4\mu^2 [LR(y^t)]^2 v_1 = \left(\frac{e\gamma_T}{\bar{v}}\right)^2 [LR(y^t)]^2 v_1,$$

This variance decreases with $x(y^{t-1})$.

A measure of the information contained in histories of length t is the variance of the likelihood ratios at time t .¹⁰ In the i.i.d. case, the higher precision of information in the latter periods of the contract translates into an increase of v_t over time. Moreover, early luck determines how important the provision of incentives (i.e., conditional variance of utility) is in latter periods.

We now briefly discuss the relationship between the variance of the inverse of marginal utility with the variance of consumption. First, we note that, for a widely used specification of CRRA utility, $u(c) = \ln(c)$, Proposition 2 characterizes the variance of c_t . In general, however, a change in the variance of the inverse of marginal utility (or in the variance of utility) does not translate into a proportional (or even a same-sign) change in the variance of consumption. This is due to the range of potential curvatures of the utility function. For relatively high curvatures, corresponding to relatively low levels of consumption, an increase in the variance in utility may be achieved by a decrease in the variance of consumption, if it happens simultaneously with a decrease in the expected consumption. Hence, a general result for the variance of consumption cannot be stated. In Table 1 we illustrate this point with a numerical example. The parameters for the example are as follows:

T	\underline{U}	e	β	π	$\hat{\pi}$	α
4	8	0.5	0.95	0.6	0.5	1

Table 1.a shows that the variance of consumption increases with t , aligned with the variance of utility. Also, as predicted by corollary 2, the conditional variance of utility decreases with the number of good outcomes in a history, as we report in Table 1.b for $t = T$. However, in this example, $\text{Var}(c(y^4) | y_L, y_L, y_L) = 0.06 < \text{Var}(c(y^4) | y_L, y_L, y_H) = 0.10$, i.e., the conditional variance of *consumption* increases with the number of good outcomes, due to the difference in expected consumption (which is 0.35 for the (y_L, y_L, y_L) history and 0.94 for the (y_L, y_L, y_H) one). Low expected consumption makes marginal utility very high following history (y_L, y_L, y_L) . This same effect is causing the variance of consumption to be a concave function of t , in spite of the variance of the likelihood ratios and the variance of utility being a convex function of t .

¹⁰See Kim (1995) for a discussion of this measure, as well as two related measures: rankings over distributions of the likelihood ratios according to (i) mean preserving spread of cumulative distributions, and (ii) Blackwell sufficiency.

1.a) PERSISTENT Effort: Unconditional Moments

	$E[c(y^t)]$	$Var[c(y^t)]$	$E[u(y^t)]$	$Var[u(y^t)]$
$t = 1$	1.33	0.08	2.29	0.07
$t = 2$	1.35	0.16	2.29	0.14
$t = 3$	1.37	0.23	2.29	0.21
$t = 4$	1.38	0.29	2.29	0.29

1.b) PERSISTENT Effort: Conditional Variance at $t = 4$

	$E[c(y^4) y^3]$	$Var[c(y^4) y^3]$	$E[u(y^4) y^3]$	$Var[u(y^4) y^3]$
$x(y^3) = 0$	0.35	0.06	1.08	0.26
$x(y^3) = 1$	0.94	0.10	1.90	0.11
$x(y^3) = 2$	1.52	0.07	2.46	0.05
$x(y^3) = 3$	2.01	0.04	2.83	0.02

$$K = 5.0281, \quad K^* = 4.8688$$

Table 1. Numerical example with persistent effort.

Comparison with long term implications of repeated (non-persistent) hidden effort models

We concluded from the discussion in this section that our model of a contractual relationship with moral hazard and persistence has strong long-term implications on the evolution of the utility of the agent. It is interesting to compare these implications with those of the optimal contract in a RMH problem.¹¹ In a standard RMH problem, the conditional distribution of output at time t depends only on the effort chosen by the agent at time t , that is, productivity can vary from high to low across periods, depending on the effort choice.

For the comparison to be meaningful, we construct a RMH example in which the principal implements high effort every period. Whenever the agent chooses high effort in a given period, he implements a probability π of high output in the given period; this probability is $\hat{\pi}$ if effort is low. The objective function of the principal, hence, is equal to the one in eq. 2. We assume that effort is equally costly every period, and the discounted sum of effort disutility is equal to e ; that is, in the RMH problem, effort disutility in a given period, denoted by \tilde{e} , satisfies:

$$\tilde{e} = e\gamma_T.$$

¹¹See Rogerson (1985) for a formal analysis of this standard textbook model.

This implies that the participation constraint of this related problem is exactly as our PC equation in page 4, and that our IC equation is one of the incentive constraints of the problem: the constraint that choosing high effort in both periods should be preferred to choosing low in all of them. However, the RMH problem has 2^{t-1} extra incentive constraints at each t , which assure that, contingent on each of the possible 2^{t-1} histories in the previous period, the agent wants to choose high effort at t , with a disutility cost of \tilde{e} :

$$\begin{aligned} & \sum_{\tau=t}^T \sum_{y^t} \beta^{\tau-1} u \left(c \left(y^{\tau \setminus t} \right) \right) \Pr(y^\tau | e_H) - e_H \\ & \geq \sum_{\tau=t}^T \sum_{y^t} \beta^{\tau-1} u \left(c \left(y^{\tau \setminus t} \right) \right) \Pr(y^\tau | e_L) - e_L, \end{aligned} \quad (\text{IC})$$

where $y^{\tau \setminus t}$ is the history of length τ that coincides with y^t in the first t realizations.

We use the utility specification in our leading example to provide some specific comparison of the properties of the optimal contracts. We concentrate on the long term implications for consumption under the optimal contracts. In the RMH, the unconditional expected utility is equal to that of the model with persistence, $E[u(y^t)] = \gamma_T (\underline{U} + e)$, for all t . However, the expected consumption is not. This is due to different properties of the variance of utility. The conditional variance of utility is

$$\text{Var}(\tilde{u}(y^t) | e_H, y^{t-1}) = \frac{\tilde{e}^2}{v_1} \left(\frac{1}{\sum_{j=0}^{T-t} \beta^j} \right)^2 = \frac{(e\gamma_T)^2}{v_1} \left(\frac{1}{\sum_{j=0}^{T-t} \beta^j} \right)^2. \quad (6)$$

This variance grows with t as in our model with persistence, but is not conditional on past history. This implies that the unconditional variance is simply

$$\text{Var}(u(y^t) | e_H) = \frac{(e\gamma_T)^2}{v_1} \sum_{\tau=1}^t \left(\frac{1}{\sum_{j=0}^{T-\tau} \beta^j} \right)^2.$$

Table 2 reports the moments of the numerical solution to the related RMH problem corresponding to the example in Table 1. Three important differences with the persistent example stand out. First, the total cost of the contract, K , is higher in the RMH problem. Second, as seen in Table 2.a, both the variance of utility and of consumption are a very convex function of time. Comparing these values to those in Table 1.a we see that, even if this variance is lower at $t = 1$ in the RMH, it becomes much higher in later periods (see Fig. 1 for a comparison). Third, as implied by equation 6, the conditional variance of utility is independent of $x(y^t)$ in the RMH, implying that the conditional variance of consumption increases with $x(y^t)$. Table 2.b reports this variance, and shows that there is a positive correlation between the level of expected consumption and the variance of consumption. When effort is persistent, instead, this correlation between level and variance is (mostly) negative, as reported in Table 1.b (with the exception, discussed earlier, of nodes corresponding to histories with the lowest $x(y^t)$, for t close to T .)

2.a) REPEATED Effort: Unconditional Moments

	$E[c(y^t)]$	$Var[c(y^t)]$	$E[u(y^t)]$	$Var[u(y^t)]$
$t = 1$	1.32	0.04	2.29	0.03
$t = 2$	1.33	0.11	2.29	0.09
$t = 3$	1.36	0.25	2.29	0.20
$t = 4$	1.47	0.78	2.29	0.64

2.b) REPEATED Effort: Conditional Variance at $t = 4$

y^3	$E[c(y^4) y^3]$	$Var[c(y^4) y^3]$	$E[u(y^4) y^3]$	$Var[u(y^4) y^3]$
(y_L, y_L, y_L)	0.58	0.17	1.38	0.44
(y_H, y_L, y_L)	0.86	0.28	1.74	0.44
(y_L, y_H, y_L)	0.96	0.32	1.85	0.44
(y_L, y_L, y_H)	1.18	0.41	2.07	0.44
(y_H, y_H, y_L)	1.33	0.47	2.21	0.44
(y_H, y_L, y_H)	1.58	0.57	2.43	0.44
(y_L, y_H, y_H)	1.72	0.63	2.54	0.44
(y_H, y_H, y_H)	2.21	0.83	2.90	0.44

$$K = 5.0783 \quad K^* = 4.8688$$

Table 2. Numerical example with repeated effort (RMH).

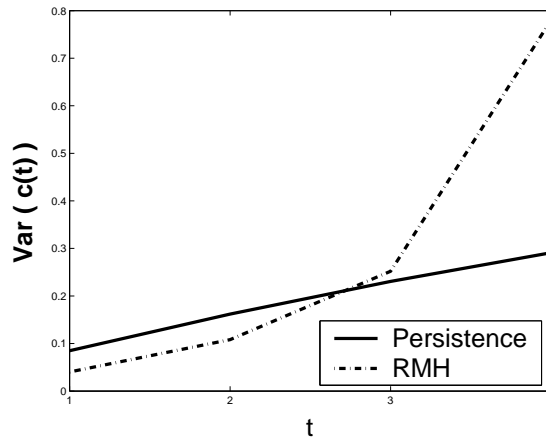


Fig. 1. Evolution over time of the variance of consumption.

4.1 Asymptotic Optimal Contract

If the principal and the agent can commit to an infinite contractual relationship ($T = \infty$) and utility is unbounded below, 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

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

and the First Best cost is

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

A solution to the problem CM may not exist when $T = \infty$. 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:

$$c(y^t) = \begin{cases} c_0 & \text{if } LR(y^t) \leq L \\ \underline{c} & \text{if } LR(y^t) > L \end{cases}.$$

Proposition 3 *Assume output is i.i.d. and the agent has a utility function that satisfies $\lim_{c \rightarrow c_{\min}} u(c) = -\infty$. 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. See Appendix. ■

As we increase the threshold L , the set of histories that have punishment c_p decreases, as does the probability of those histories in equilibrium. Hence, a lower \underline{c} is needed to guarantee that the contract is still incentive compatible. However, as the proof shows, for the new incentive compatible \underline{c} the expected punishment decreases, allowing us to decrease c_0 . The intuition for this result parallels that of Mirrlees (1974); in his static example, output realizations lie on a continuum $y \in [0, \infty)$ and are distributed according to a lognormal whose mean depends on unobservable effort. The corresponding likelihood ratio for continuous output tends to infinity as y approaches 0. In our framework, the binomial distribution of $x(y^t)$ converges to a normal distribution as t grows; the corresponding likelihood ratios with $x(y^t)$ lying in a continuum also tend to infinity as $x(y^t)$ approaches 0.

It should be noticed that the proof for this result depends on the unlimited punishment power

of the principal. When utility is bounded below, however, results are less clear.¹² For our leading example, when T is finite we can always find a $\underline{U}(T)$ such that the DC constraint does not bind. For these parameters, we can show that difference in the cost of the second best contract and the first best contract decreases as T increases. We can write the weighted average of the likelihood ratios as:

$$\bar{v}(T) = \hat{E} \gamma_T \frac{1 - (\beta \hat{E})^T}{1 - \beta \hat{E}} - 1.$$

We can show that

$$\frac{\partial \bar{v}(T)}{\partial T} > 0.$$

Then, using the expression for the per period cost of the optimal contract, we have:

$$k(T) - k^*(T) = \frac{1}{4} \gamma_T^2 \frac{e^2}{\bar{v}(T)}, \quad (7)$$

which is clearly decreasing in T . Notice that the limit of $\bar{v}(T)$ depends on $\beta \hat{E}$ as follows:

$$\begin{aligned} \lim_{T \rightarrow \infty} \bar{v}(T) &= \lim_{T \rightarrow \infty} \frac{1 - \beta}{1 - \beta^T} \sum_{t=1}^T \beta^{t-1} (\hat{E}^t - 1) \\ &= \lim_{T \rightarrow \infty} \frac{1 - \beta}{1 - \beta^T} \left(\hat{E} \sum_{t=1}^T (\beta \hat{E})^{t-1} \right) - 1 \\ &= \begin{cases} \infty & \text{if } \beta \hat{E} > 1 \\ \hat{E} \frac{1 - \beta}{1 - \beta \hat{E}} - 1 & \text{if } \beta \hat{E} \leq 1 \end{cases}, \end{aligned}$$

If the DC were not to bind as T approached infinity, $k(T)$ would converge to k^* only if $\beta \hat{E} > 1$, and would be bounded away from it by a positive number otherwise. However, the condition on the parameters that guarantees that the DC constraint is not binding (see eq. 4) is not satisfied when T approaches infinity. To see this, rewrite eq. 4 as

$$\left(\frac{U}{e} + 1 \right) \sum_{t=1}^T \beta^{t-1} \geq \left[\frac{\left(\frac{1 - \hat{\pi}}{1 - \pi} \right)^T - 1}{\sum_{t=1}^T \beta^{t-1} (\hat{E}^t - 1)} \right].$$

The left hand side of this condition converges to $\left(\frac{U}{e} + 1 \right) / (1 - \beta)$ as T approaches infinity. The limit of the right hand side can be found using L'Hopital's rule (twice):

$$\lim_{T \rightarrow \infty} \left[\frac{\left(\frac{1 - \hat{\pi}}{1 - \pi} \right)^T - 1}{\sum_{t=1}^T \beta^{t-1} (\hat{E}^t - 1)} \right] = - \left(\frac{1 - \hat{\pi}}{\beta \hat{E}} \right)^T \frac{\ln \left(\frac{1 - \hat{\pi}}{1 - \pi} \right) \ln \left(\frac{1}{\beta} \frac{1 - \hat{\pi}}{1 - \pi} \right)}{\frac{\hat{E} \ln(\beta \hat{E})}{1 - \beta \hat{E}}} = \infty$$

¹²We have recently become aware of an article by Jewitt, Kadan and Swinkles (2008) that studies optimal contracts in the presence of utility bounds. They provide a formal proof of existence and uniqueness for the case discussed here, for a general utility function specification and a continuum of output levels.

Note that this limit would be 0 (and hence DC would not be binding) if $\beta\widehat{E} > \frac{1-\widehat{\pi}}{1-\pi}$. However, recalling that $\widehat{E} = \widehat{\pi}\frac{\widehat{\pi}}{\pi} + (1-\widehat{\pi})\frac{1-\widehat{\pi}}{1-\pi}$, it is easy to see that $\beta\widehat{E} < \frac{1-\widehat{\pi}}{1-\pi}$ always. Since eq. 4 is not satisfied in the limit, the closed form solution for the contract derived earlier cannot be used to analyze the optimal contract for T arbitrarily large. The solution for the multipliers λ and μ when DC binds is modified: it depends on the set of histories for which DC binds, which, in turn, depends on the value of λ and μ . Moreover, the PC and the IC may not bind in the optimal contract. Closed forms are not available, but numerical examples that solve the fixed point problem for finite (but large) T can be computed. In our numerical examples we found that, for a given \widehat{E} , high values of β , for which $\beta\widehat{E} > 1$, implied that the cost of the second best contract got very close to that of the first best; lower values for β for which $\beta\widehat{E} < 1$, however, implied second best contracts whose cost seemed to quickly converge (in a small number of periods) to a number bounded away from k^* .

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 1 *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. We consider stochastic processes that contain informative outcomes up to period $\tau \geq 1$, and for any $t > \tau$ they satisfy:

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

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, $\bar{\pi}$, 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. 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 4 *The cost of a contract strictly decreases if the duration of persistence, τ , increases.*

The structure of the contract is easy to characterize. When $\tau < T$, the likelihood ratio of any uninformative history following y^τ remains constant and equal to $LR(y^\tau)$:

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

Hence, by the first order conditions, consumption is constant from τ until T . The result in Prop. 4 is illustrated in the following corollary for our leading example.

Corollary 3 *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 $t \leq \tau_1$.*

We can prove the result directly. Let subscript i denote variables corresponding to a contract of duration τ_i . We have

$$v_{ti} = \begin{cases} \widehat{E}^t - 1 & \text{for } t \leq \tau_i \\ 0 & \text{for } t > \tau_i. \end{cases}$$

It is easy to see that $\tau_1 < \tau_2$ implies $\bar{v}_1 < \bar{v}_2$. This, in turn, implies $\mu_1 > \mu_2$ and, by eq. (5), $k_2(T) < k_1(T)$. For the second part of the corollary, recall that we can express variance of utility as a function of \bar{v} using the solution for μ , as in Corollary 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.

When duration increases, the value of \bar{v} increases, which lowers μ . The individual variances v_t of the periods that had informative outcomes, however, remain unchanged. As a result, less incentives are allocated to the early periods. This is intuitive since, with higher duration, more informative realizations are available in late periods.

Note that, when duration τ is finite, the asymptotic result breaks down. In the proof of Prop. 3, we can no longer be sure to find a y^t such that $LR(y^t) > L$ for any arbitrarily large L . In our leading example, since the weights ω_t converge to $(1 - \beta)\beta^{t-1}$, the weighted variance converges to a finite number and the cost of the contract is bounded away from the first best.

PERSISTENT Effort: Unconditional Moments

	$E[c(y^t)]$	$Var[c(y^t)]$	$E[u(y^t)]$	$Var[u(y^t)]$
$t = 1$	1.33	0.10	2.29	0.08
$t = 2$	1.36	0.20	2.29	0.17
$t = 3$	1.38	0.28	2.29	0.26
$t = 4$	1.38	0.28	2.29	0.26

$$K = 5.0464 \quad K^* = 4.8688$$

Table 3. Numerical example with duration of persistence $\tau = 3$.

Table 3 presents a numerical example that illustrates the implications of changes in duration. In all periods but the last, we observe higher variance of consumption than in the example of Table 1 (which corresponds to $\tau = T$.) This translates into higher cost of the contract. The increase in

variance is spread evenly across periods (approximately 1.24 times the variance of consumption in Table 1, for the first three periods). We do not report the conditional variance in period 4, since it is zero for all histories, reflecting the fact that compensation stays constant after period 3.

Longer duration of persistence means more information. More information translates into a lower value of the multiplier of the IC constraint, μ , reflecting the fact that the IC is easier to satisfy. However, the availability of better quality information is materialized in more extreme values of the likelihood ratios. Rearranging the first order conditions of the Second Best we have that for any two histories y^t and \tilde{y}^t of any length,

$$\frac{1}{u'(c(y^t))} - \frac{1}{u'(c(\tilde{y}^t))} = \mu \left(\frac{\Pr(\tilde{y}^t|e_L)}{\Pr(\tilde{y}^t|e_H)} - \frac{\Pr(y^t|e_L)}{\Pr(y^t|e_H)} \right). \quad (8)$$

The patterns for the variability of compensation that we described in this section can be understood in terms of contemporaneous changes in the likelihood ratios and in μ . For the square root utility, eq. 8 means that the difference in utility is proportional to the difference in the likelihood ratios; for logarithmic utility, it is the difference in consumption levels. The factor of proportionality between differences in likelihood ratios and differences in compensation is μ . In the increases in the duration of persistence studied in this section, we have shown that the decrease in the multiplier (the sensitivity of compensation to the likelihood ratios) is large enough so that variability of compensation decreases.

6 Conclusion

We study a simple representation of the problem of a firm that hires from a pool of workers heterogeneous in their productivity. This productivity is private information of the worker, and is persistent: it affects the distribution of output of the firm in every period in which a worker is employed. 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, the larger the differences in unobserved productivity, the bigger the efficiency gains from postponing incentives, and the higher the level of insurance provided to the agent in early periods. In the special case of i.i.d. output and square root utility, the optimal contract implies a negative correlation of the level and the variance of consumption, in contrast with the optimal contract in a setting of repeated moral hazard.

Our model is only a partial approximation to the problem of compensation design in the presence of persistent hidden information. In its simplicity, it abstracts from an important feature in many interesting examples: the agents may be able, or required, to exert further unobservable efforts during their relationship with their employers – efforts that may or may not be persistent. Such examples include the design of unemployment schemes, CEO compensation, or optimal wage schemes in any industry in which the productivity of workers depends both on everyday effort and their persistent productivity.

Combining a repeated effort incentive problem with the persistence framework presented here, then, is a natural next step towards understanding the importance of persistence in many relevant

contracting environments. Some important contributions have been made in this direction, which may be classified into two groups. First, Fernandes and Phelan (2000) and Doepke and Townsend (2006) provided useful recursive formulations, which can be used to compute the solution to specific examples. Although they do not allow for a characterization of the optimal contract, these recursive formulations highlight one important difficulty in problems with persistence: there is no common knowledge of preferences at the beginning of each period. Hence, the principal needs to check the potential profitability of joint deviations of effort that would be easily ruled out if effort were not persistent. In our paper, we deal with this difficulty proposing a model that eliminates the joint deviations, while capturing the main features of an information system in the presence of persistence.

A second group of papers provides some interesting characterizations of the optimal contracts, for particular examples. Mukoyama and Sahin (2005) and Kwon (2006) study a similar problem with repeated persistent efforts. They restrict their analysis to cases in which the principal implements high, equally costly, effort every period. When early effort is assumed to have higher impact on later period's outcomes than later efforts, these papers show that perfect insurance may arise in the early periods, in sharp contrast with the repeated moral hazard predictions. Jarque (2008) studies a repeated moral hazard problem with a continuum of efforts under two main assumptions: linear disutility in effort and linear effect of past efforts on the persistent productive state that determines the probability distribution over output. She finds that the contract has the same properties as a repeated moral hazard problem, since it is optimal for the principal to regard the level of the productive state as if it were the level of effort in a standard problem without persistence. The assumptions in these three frameworks make a characterization possible, but at the expense of restricting the structure of persistence. Our model allows us to explore the implications of more general output processes.

In our paper, we compare the model with persistent productivity to one with repeated moral hazard. We believe that this comparison also provides some interesting insights on the results in the literature on moral hazard and persistence mentioned above. The persistent structure of incentives that we characterize in our framework with only one effort is particularly simple because only two distributions over output need to be statistically discriminated (i.e., there is only one incentive constraint.) If the agent has the possibility of deviating every period, as in the repeated moral hazard problem, an extra incentive constraint appears every period and for each possible past history of output. Two distributions over output need to be discriminated in every period, corresponding to high or low effort in the given period. If these efforts are persistent, as in the three papers above, the number of potential distributions driving output increases: in any given period, the equilibrium distribution has to be contrasted with the multiple different distributions resulting from every possible combination of past efforts. By imposing an ordering over these distributions that translates easily into an ordering of expected utility for the agent, the papers cited above are able to characterize the solution to the optimal contract.

A general characterization of the optimal contract in a problem with repeated persistent effort would need to evaluate the relative importance of each deviation, as opposed to assuming it. We conjecture from our analysis that this will depend on the correlation of the different distributions

under different effort deviations; weighted sums of likelihood ratios of histories are likely to be useful in the characterization. However, it is also true that more needs to be learnt about a second effect that arises in the presence of persistence: in general, the agent aims to smooth effort disutility over time. Our work abstracts completely from this problem. The literature on repeated moral hazard with hidden savings suggests that recursive formulations with added state variables, and hence involved computations, may be needed to establish the implications of this effect on the optimal contract.

7 Appendix

Proof of Proposition 3. 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:

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

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:

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

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

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

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

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

so we can write

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

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

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

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

$$\delta(L) = (e_H - e_L) \frac{\sum_t \beta^{t-1} (1 - F_t(L))}{\sum_t \beta^{t-1} (F_t(L) - \widehat{F}_t(L))} (1 - \beta). \quad (9)$$

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

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

The actual cost will be strictly lower than $\frac{c_0}{1 - \beta}$ since, with probability $\sum_t \beta^{t-1} (1 - F_t(L)) > 0$ 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, $(1 - F_t(L))$ decreases in all periods where it was positive. In those same periods, both $F_t(L)$ and $\widehat{F}_t(L)$ increase, but we have:

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

This implies

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

Substituting this inequality in expression (9),

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

We have that $\delta(L)$ is decreasing in L as long as $\sum_t \beta^{t-1} (F_t(L) - \widehat{F}_t(L)) > 0$ for L . From the above inequalities, this will hold whenever $1 - \widehat{F}_t(L) > 0$ for some t . For the discrete case, this holds if there exists a path y^t such that $L(y^t) > L$, which is guaranteed in the i.i.d. case. Hence, for any $\varepsilon > 0$ we can find an L low enough so that $K(\infty) < K^*(\infty) + \varepsilon$. ■

Proof of Proposition 4. Denote by $C_1 = \{c_1(y^t)\}_{t=1}^T$ 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 $C_2 = \{c_2(y^t)\}_{t=1}^T$. 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 $LR(y^{t-1}, y_L) \neq LR(y^{t-1}, y_H)$ 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 $LR(y^{t-1}, y_L) = LR(y^{t-1}, y_H)$ for all y^{t-1} . Since the solution for the optimal contract is unique, we conclude that 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 . ■

References

- [1] Autor, David H. “Why do Temporary Help Firms Provide Free General Skills Training?,” *Quarterly Journal of Economics*, Vol. 116, No. 4: 1409–1448 (2001).
- [2] Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994a. “The internal economics of the firm: Evidence from personnel data.” *Quarterly Journal of Economics* 109, no. 4:881–919.
- [3] ———. 1994b. “The wage policy of a firm.” *Quarterly Journal of Economics* 109, no. 4:921–55.
- [4] Blackwell, D., and M. A. Girshick. *Theory of Games and Statistical Decisions*. New York. John Wiley and Sons, Inc., 1954.
- [5] Doepke, Matthias, and R. Townsend. “Dynamic mechanism design with hidden income and hidden actions,” *Journal of Economic Theory*, vol. 126(1), pages 235-285, January (2006)
- [6] Fernandes, A. and C. Phelan. “A Recursive Formulation for Repeated Agency with History Dependence,” *Journal of Economic Theory*, **91** (2000): 223-247.
- [7] Fudenberg, D., B. Holmström, and P. Milgrom. “Short–Term Contracts and Long–Term Agency Relationships.” *Journal of Economic Theory*, 51, 1-31 (1990)
- [8] Gibbons, Robert, and Michael Waldman. “Careers in organizations: Theory and evidence.” In *Handbook of labor economics*, vol. 3, ed. Orley C. Ashenfelter and David Card. Amsterdam: North-Holland. (1999a)
- [9] ———. “A theory of wage and promotion dynamics inside firms.” *Quarterly Journal of Economics* 114, no. 4:1321–58. (1999b)
- [10] ———. “Enriching a Theory of Wage and Promotion Dynamics inside Firms.” *Journal of Labor Economics*, Vol. 24, no. 1 (20006)
- [11] Golosov, M., N. Kocherlakota and A. Tsyvisnki. “Optimal Indirect and Capital Taxation.” *The Review of Economic Studies* 70(3): 569-588 (2003)
- [12] Grochulski, B. and T. Piskorski. “Risky Human Capital and Deferred Capital Income Taxation.” Mimeo (2006)

- [13] Grossman, Sanford and Oliver D. Hart. “An Analysis of the Principal–Agent Problem.” *Econometrica* **51**, Issue 1 (Jan.,1983), 7-46.
- [14] Jewitt, I., “Justifying the First–Order Approach to Principal–Agent Problems,” *Econometrica*, **56**(5), 1177-1190 (1988).
- [15] Holmström, B. “Moral Hazard and Observability,” *Bell Journal of Economics*, Vol. **10** (1) pp. 74-91. (1979)
- [16] Holmström, B, and P. Milgrom. “Aggregation and Linearity in the Provision of Intertemporal Incentives”, *Econometrica*, Econometric Society, vol. 55(2), pages 303-28, March (1987)
- [17] Jarque, A. “Repeated Moral Hazard with effort Persistence.” Federal Reserve Bank of Richmond WP 08-4 (2008)
- [18] Jewitt, I., O. Kadan and J. Swinkels. “Moral Hazard with Bounded Payments.” *Journal of Economic Theory*, 143, 59-82 (2008).
- [19] Kim, S. K. “Efficiency of an Information System in an agency Model.” *Econometrica*, vol **63**(1), pages 89–102 (1995)
- [20] Kwon, I. “Incentives, Wages, and Promotions: Theory and Evidence.” *Rand Journal of Economics*, **37** (1), 100-120 (2006)
- [21] Lazear, E. “The power of Incentives.” *American Economic Review Papers and Proceedings*, Vol. 90 No. 2 (2000)
- [22] Miller, Nolan. “Moral Hazard with Persistence and Learning,” Mimeo (1999)
- [23] 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)
- [24] Mukoyama, T. and A. Sahin, “Repeated Moral Hazard with Persistence,” *Economic Theory*, vol. **25**(4), pages 831-854, 06 (2005)
- [25] Rogerson, William P. “Repeated Moral Hazard”, *Econometrica*, Vol. **53**, No. 1. (1985), pp. 69-76.
- [26] Rogerson, W. P. “The First–Order Approach to Principal–Agent Problems,” *Econometrica*, **53**(6), pp. pp. 1357-1367 (1985b).
- [27] Shavell, S. and L. Weiss: “The Optimal Payment of Unemployment Insurance Benefits over Time”, *Journal of Political Economy*, **87** (1979), 1347-1362.