Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Games and Economic Behavior 60 (2007) 213–246
GAMES and Economic Behavior
www.elsevier.com/locate/geb

Optimality and renegotiation in dynamic contracting

Marco Battaglini

Department of Economics, Princeton University, 001 Fisher Hall, Princeton, NJ 08544, USA

Received 18 April 2005

Available online 17 January 2007

Abstract

We characterize the optimal renegotiation-proof contract in a dynamic principal–agent model in which the type of the agent may change stochastically over time. We show that, under general conditions, the optimal contract with commitment is renegotiation proof even when type realizations are serially correlated. When the renegotiation-proofness constraint is binding, it is always optimal to partially screen the types by offering a menu of choices to the agent; and the distortion induced by the renegotiation-proofness constraint is non-decreasing in the persistence of types.

JEL classification: D42; L51

Keywords: Contract theory; Dynamic contracts; Renegotiation

1. Introduction

It is well known that in principal–agent models the optimal contract generally prescribes allocative inefficiencies. These distortions increase the principal's utility because they make it easier to screen the agent's types. For this reason, a contract that is ex ante optimal may not be time-consistent: ex post, after some history, both the principal and the agent would agree on renegotiating these inefficiencies. A contract that can not be Pareto improved after any history and therefore that does not provide such renegotiation opportunities is said to be *renegotiation proof*. Though different assumptions can be made on how a contract can be renegotiated in practice, renegotiation proofness is a minimal requirement of time consistency for a dynamic contract. How do renegotiation proof contracts differ from standard contracts with commitment? What do they look like?

E-mail address: mbattagl@princeton.edu.

0899-8256/$ – see front matter
doi:10.1016/j.geb.2006.10.007

In seminal papers, Hart and Tirole (1988), Dewatripont (1989), and Laffont and Tirole (1990) have provided a first answer to these questions by characterizing optimal renegotiation proof contracts assuming that the type of the agent is constant throughout time. All these papers have come to the conclusion that the ex ante optimal contract with commitment is never renegotiation proof: renegotiation proofness always weakens the principal's ability to screen the agent's types.

In this paper we characterize the optimal renegotiation proof contract in a principal–agent model in which agent types may change stochastically across periods. This generalization has a direct interest from an applied perspective since it is natural to assume that the type of an agent, though serially correlated, may change. But it also highlights some features of dynamic contracting that could not otherwise be observed under the assumption of constant types. We show that the cost for the principal of respecting the renegotiation constraint is non-decreasing in the degree of serial correlation of the types. More importantly, we show that under general conditions there is no conflict between renegotiation proofness and optimality, even if types are highly correlated.

We study a two period principal–agent model of procurement in which the agent can assume two types: low or high cost of production. The key parameter of the model is the probability α that the agent's type remains unchanged between period 1 and period 2. When $\alpha = 1$, we have the model with constant types studied in the previous literature as a special case, which we use as benchmark. Figure 1 summarizes the main features of the optimal renegotiation proof contract as a function of α. The optimal contract is characterized by two thresholds α_1 and α_2 with $\alpha_2 > \alpha_1$, which identify three possible cases.

When $\alpha \leqslant \alpha_1$ (area I) *the optimal contract with commitment coincides with the optimal renegotiation proof contract*. This result is perhaps surprising because, as mentioned, it never holds with constant types. As seen in Fig. 1, this threshold is a function of the likelihood ratio $\Gamma_0 = \frac{v}{1-v}$ between the prior probability that a type is low-cost, denoted v, and the probability that it is high-cost $(1 - v)$. As the fraction of high-cost types is increased, α_1 converges to one; but, for plausible parameters' values, α_1 is very high: when, for example, there is a 20% prior probability of low-

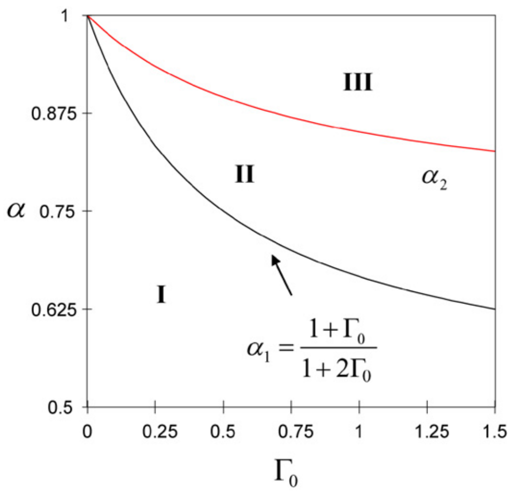

Fig. 1. The optimal contract as a function of types' persistence (α) and Γ_0, the likelihood ratio between the prior probability of an efficient and inefficient firm.

cost types, then the contract is renegotiation proof even if types are persistent more than 80% of the time.

When $\alpha > \alpha_1$, the optimal contract with commitment is no longer renegotiation-proof. The optimal contract with commitment is fully separating in the first period and, in this region, it prescribes an allocation that in the second period is more distorted than what would be optimal at $t = 2$ given the information revealed at $t = 1$. These distortions would not be credible when $\alpha > \alpha_1$. Two cases are possible.

In $\alpha \in (\alpha_1, \alpha_2]$, the optimal renegotiation proof contract is still fully separating in the first period. The principal offers a menu with two options at $t = 1$: one option attracts only high-cost types, the other attracts only low-cost types. In the second period the principal offers again a menu with two options that is contingent on the choice in period one. As in the case with commitment, the principal (she) will not find it optimal to offer an efficient menu; however she is forced to reduce the inefficiency at $t = 2$ to satisfy the renegotiation proofness constraint. This feature of the optimal renegotiation proof contract is also novel with respect to previous characterizations. When types are constant the optimal contract is not fully separating in the first period, except when the discount factor sufficiently small. On the contrary, the set $(\alpha_1, \alpha_2]$ is always non-empty, for any discount factor δ, and for any size of the payoffs in the second period.

If $\alpha > \alpha_2$ the principal finds it optimal to have a pooling equilibrium in which types do not fully separate in the first period. The equilibrium level of pooling, and with it the cost for the principal to satisfy the renegotiation proofness constraint, are monotonically increasing in persistence; however, *full pooling of types is never optimal* for any level of persistence: both in the first and second period the principal always prefers to offer the agent a menu of choices.

To understand why the ex ante optimal contract can be renegotiation proof even with high correlation, and the other results mentioned above, consider the extreme example in which types are stochastic, but serially uncorrelated. Here, it is well known that the principal finds it optimal to offer a contract that is fully efficient in the second period (see Roberts, 1982; Baron and Besanko, 1984). Indeed, with uncorrelated types the agent has no informational advantage in the first period with respect to the continuation of the game: the principal finds it optimal to maximize the expected surplus generated in period two because he can extract all of it in the first period with a simple take it or leave it offer. Efficiency, however, implies that the contract is also Pareto optimal and therefore renegotiation proof: the principal will never be able to offer a different contract that maintains the agent's utility and increases social welfare as well.

When we modify this benchmark case introducing some degree of correlation of types over time, the principal no longer finds it optimal to offer a contract that is efficient in the second period, even if correlation is very small. Renegotiation proofness, however, is guaranteed in this case by the residual uncertainty on the future realizations of the types. Even if in the first period types are perfectly screened, there is still uncertainty on the type in the second period. This implies that at $t = 2$ the second period optimal contract for the principal *still prescribes an allocative inefficiency* in order to screen this residual uncertainty away. The optimal contract with commitment is renegotiation proof when it prescribes an inefficiency in the second period that is smaller than the inefficiency that would be optimal in the second period if types perfectly separated in the first period. When this is the case, it is credible that the inefficiency in the second period will be at least as large as promised ex ante, because the principal would impose an even higher degree of inefficiency in the second period if she could. On the other hand, the fact that the principal has to guarantee the same utility promised in the ex ante contract assures that the inefficiency will not be larger than promised.

As correlation increases, the ex ante optimal contract is no longer renegotiation proof, but it still depends on the level of correlation because this determines the discrepancy between the ex ante optimal level of the allocative inefficiency and the second-period optimal level. Contrary to the case with constant types, we show that the optimal contract is not conditionally optimal for the principal in the second period.

This paper connects two distinct lines of research. On the one hand we have the literature that has studied renegotiation of contracts in dynamic settings when the agent's type is constant across periods. Dewatripont (1989), Hart and Tirole (1988) and Laffont and Tirole (1990), characterized the optimal renegotiation proof contract under different assumptions of the economic environment.[1] Our paper departs from this literature by assuming that types may change stochastically from period to period. For this generalization, we adopt the standard principal–agent framework introduced in Laffont and Tirole (1990), and we extend it to the case with variable types: this allows us to have a clear benchmark case to evaluate the impact of types variability across periods and compare it to the case with constant types.

The other literature to which our paper relates is the one on dynamic adverse selection with stochastic types, in which renegotiation is not a concern. In early contributions, Roberts (1982) and Baron and Besanko (1984) have studied the optimal contract with variable types and commitment. Laffont and Tirole (1996) have studied a two-period model of regulation in which the agent's types are variable, but the principal can commit. Dynamic models of pricing in which consumer's types change stochastically are considered, among others, by Battaglini (2005), Biehl (2001), Courty and Li (2000), Hendel and Lizzeri (2002), Hendel et al. (2005), Kennan (2001), and Rustichini and Wolinsky (1995). These papers, however, do not characterize the optimal renegotiation proof contract because either they focus on the case with full commitment, or on the case in which long-term contracts are ruled out and only spot contracts are possible.

The connection among these two literatures has previously been partially explored in two papers. Battaglini (2005) has characterized a sufficient condition for renegotiation proofness of the ex ante optimal contract, but has not fully characterized the optimal renegotiation proof contract. Blume (1998) studied renegotiation in a model in which a durable good monopolist serves a buyer with time varying valuations. In this model, supply is restricted to take two possible values, and stronger assumptions than in the present paper are made on the transition probabilities. Although this model is insightful in comparing the rental versus the sale of a durable good, it cannot be used to study the relationship between optimality and renegotiation proofness since optimal supply does not vary continuously with the distribution of types. In our model, moreover, production takes place in every period and the good is not durable.

The paper is organized as follows. Section 2 presents the model. As a benchmark, Section 3 characterizes the ex ante optimal contract with full commitment. Section 4 studies how screening is affected by the renegotiation constraint. Section 5 characterizes the optimal renegotiation proof contract with variable types. Section 6 concludes.

[1] Dewatripont (1989) characterizes the optimal renegotiation-proof contract in a general framework, but with restrictions on the contract space and assuming pure strategies. Hart and Tirole (1988) consider a model with T periods and two types in which supply could assume two values. Laffont and Tirole (1990) fully characterize the optimal renegotiation-proof contract in a model with two periods and two types with a general supply function.

2. Model

We consider a dynamic principal–agent model with two periods. To have a clear benchmark, the model is a direct generalization of the classic model introduced by Laffont and Tirole (1990) to the case with variable types. We interpret it as a model of cost based procurement, but the setting is general and can be used to describe many other standard principal–agent interactions.[2]

In each period a principal needs to purchase one unit of a public good from an agent. The principal is a government agency which is entrusted with the task of providing a public service, and the agent is a private firm which can technically produce it. The public good provides social surplus S in each period, and its cost in period t depends on two components: on the type of the agent at time t, θ^t; and on the level of effort that the agent chooses at time t, e_t. The cost is: $c^t = \theta^t - e_t$. We assume that θ^t may take two values θ_L and θ_H with $\theta_H > \theta_L$ and $\Delta\theta = \theta_H - \theta_L$. The variable e_t is the level of effort that the firm can exert to reduce the project's cost, and takes values in $\mathbb{R}^+$. The intertemporal discount factor is δ.

The principal designs a procurement contract on the basis of the observable variables. The principal observes the realized cost, but not the level of effort or the firm's type. Without loss of generality, the principal reimburses the cost c^t of realizing the project and pays an incentive fee s_t. The compensation s_t, therefore, cannot directly depend on effort and the firm's type, but only on the realized cost. In each period the firm's manager has utility $s_t - \varphi(e_t)$ where $\varphi(e_t)$ is a convex cost function. To obtain simple closed forms, in the remainder of the paper we assume $\varphi(e_t) = \frac{\eta}{2}e_t^2$ with $\eta > 0$. As it will be evident, this assumption is not essential for most of the results presented below. The only results in which this assumption will be used is Proposition 5 in which we show that if it is optimal to have pooling, the optimal level of pooling is unique.

The agent's type changes stochastically over time: the probability that a type persists is $\Pr(\theta^{t+1} = \theta_i \mid \theta^t = \theta_i) = \alpha$ for $i, j = H, L$ and $t \geqslant 0$. We assume that types are positively correlated ($1/2 \leqslant \alpha \leqslant 1$), but we do not make assumptions on the degree of correlation. The initial prior is that a fraction $v \in (0, 1)$ of firms has a low cost parameter. It is useful to define the ex ante likelihood ratio $\Gamma_0 = \frac{v}{1-v}$ and the second-period ratios after an low-cost type realization, $\Gamma_L = \frac{\alpha}{1-\alpha}$, and an high-cost type realization, $\Gamma_H = \frac{1-\alpha}{\alpha}$. One of the leading examples of a procurement problem are military contracts (see Rogerson, 1994). These contracts often involve research projects that last many years (as the design of a new plane) and are obviously plagued by asymmetric information problems. As in our model, the characteristics of the projects (cost of carrying out research, probability of obtaining results, etc.) are most likely to be variable over time, though persistent.

The per-period welfare level when the project is realized (net of the agent's utility) is $S - (1 + \lambda)(s_t + c^t)$, where $\lambda > 0$ is the shadow cost of public funds. As in Laffont and Tirole, we assume λ to be exogenous for simplicity, but it can be easily derived assuming that the principal can raise resources to finance monetary transfers only with distortionary taxes.[3] As we change λ, we can capture different possible scenarios. The higher is λ the more expensive it is to make transfers. When λ is positive but bounded, the principal, when maximizing total welfare (which includes the agent's utility), takes into account the benefits of making a transfer to the private firms producing the good; in the limit case in which $\lambda \to \infty$, however, the principal simply wishes to minimize the cost of the project. The appropriate choice of λ depends on the

[2] See Section 5.4 for a discussion in greater detail.

[3] For a screenining model that formally derives the shadow cost of public funds, see Battaglini and Coate (2006).

characteristics of the economy under analysis: the more taxation is distortionary, the higher λ should be assumed.[4]

A *contract* C consists of a sequence of message spaces $(\mathcal{M}^t)_{t=1,2}$ that can be sent by the agent to the principal in period t, a sequence of functions $c(m^t)$ and $s(m^t)$ assigning, respectively, a cost level and a monetary transfer (positive or negative) from the principal to the agent for each history of messages $m_1 \in \mathcal{M}^1$ in period 1 and $m^2 = (m_1, m_2) \in \mathcal{M}^1 \times \mathcal{M}^2$ in period 2.[5] For any message m_1, a contract C defines an associated 1-period *truncation contract* $C^2(m_1) = \{c(m_1, \cdot), s(m_1, \cdot)\}$. A contract C defines a (trivial) game in which the agent is the only player. For any associated Perfect Bayesian Equilibrium E of this game we can define a posterior probability that the agent is a low cost type at $t = 2$ $\mu^E(m_1)$ at any node m_1. In the remainder of the paper, whenever clear, we will omit the superscript E referring to the associated equilibrium.

The goal of the principal is to design a contract that maximizes the expected discounted social welfare. We will focus on the case in which S is large enough so that it always optimal to provide the public good. Define $e^* = \frac{1}{\eta}$: this is the level of effort that the firm of type θ would choose if it fully internalizes the cost and benefit of the project. Let $S(\theta) = S - (1+\lambda)(\theta - e^*) - \varphi(e^*)$ be level of welfare produced by type θ if effort is e^*, and define $v^* = \max\{\frac{(1+\delta)v}{1+\delta v}, \frac{1}{2}\}$. As shown in Appendix A, a sufficient condition for the public good being always produced in the optimal renegotiation proof contract is:

Assumption 1. $v^* S(\underline{\theta}) - S(\bar{\theta}) < 0$.

As it can be easily verified, this condition is satisfied if S is sufficiently large.

In this model the *first best* is achieved in the hypothetical command economy in which the government agency can observe effort and types, and impose by decree any level of effort. In this case, firm θ would be required to exert effort e^*, which corresponds to a cost $c^*(\theta) = \theta - e^*$, and would receive no monetary transfers.

Consider now the *second best* when the agency does not observe the type and the effort and firms are free to chose the level of effort that is most profitable. If the principal can commit to a contract, the revelation principle allows us to use a direct mechanism in which $\mathcal{M}^1 \equiv \mathcal{M}^2 \equiv \{\theta_H, \theta_L\}$. In this case, the problem consists of maximizing expected discounted welfare in which the cost c^t and the incentive fee s_t are contingent on the realized type, under the incentive compatibility constraint that the agent desires to report his type truthfully, and the individual rationality constraint that each type receives a reservation utility $\underline{u}$, which we normalize to zero (see program $\mathcal{P}_I$ in Section 3). The solution to this problem is the *ex ante optimal contract*.

In this work, we require the contract to satisfy an additional constraint, *renegotiation proofness*.

Definition 1. A contract C with an associated equilibrium E is renegotiation proof if, for any message $m^1 \in \mathcal{M}^1$, at the beginning of the second period the principal cannot replace its associated truncation $C^2(m^1)$ with a new one period contract which strictly increases social expected welfare given the posterior $\mu^E(m^1)$ and that guarantees at least the same rent to each type of the agent.

[4] Clearly, the problem in which $\lambda = 0$ is not interesting: in this case the principal does not care about the size of monetary transfers and the first best can always be achieved.

[5] As it can easily verified, we can ignore without loss of generality the possibility that messages can be sent by the sender as well.

A contract is said to be *renegotiation proof* if there is a Perfect Bayesian Equilibrium E such that C with E is renegotiation proof. This definition is standard in the literature (see Dewatripont, 1989; Hart and Tirole, 1988; Laffont and Tirole, 1990). When a contract is not renegotiation proof, then both the principal and the regulated firm would agree on renegotiating the initial contract. When the contract is renegotiation proof, on the contrary, any attempt to renege the initial contract by the principal would face the opposition of the regulated firm.

3. The benchmark case with commitment

As a benchmark, in this section we discuss the optimal regulatory framework when the principal can commit to a long term contract.

With commitment, the optimal contract specifies a cost c_h and a transfer s_h for any possible history h in period 1 ($\{H, L\}$) and in period 2 ($\{HH, LH, LL, HL\}$).[6] As mentioned, the revelation principle guarantees that we can restrict attention to a direct mechanism in which the firm truthfully reports her type at any possible history node. The principal's problem with commitment is:

$$\min_{\langle \mathbf{c},\mathbf{s}\rangle} \sum_{i=H,L} \Pr(\theta_i) \left\{ \begin{array}{c} (1+\lambda)\big(c_i+\varphi(\theta_i-c_i)\big)+\lambda\big(s_i-\varphi(\theta_i-c_i)\big) \\ +\delta \displaystyle\sum_{j=H,L} \Pr(\theta_j \mid \theta_i) \left[\begin{array}{c} (1+\lambda)(c_{ij}+\varphi(\theta_j-c_{ij})) \\ +\lambda(s_{ij}-\varphi(\theta_j-c_{ij})) \end{array} \right] \end{array} \right\} \qquad (\mathcal{P}_I)$$

under the usual constraints:

$$\begin{array}{ll} IC_H, IC_L, IC_i(H), IC_i(L) & \text{for } i=H,L, \\ IR_H, IR_L, IR_i(H), IR_i(L) & \text{for } i=H,L \end{array} \qquad (1)$$

where IR_H, IR_L, $IR_i(H)$, $IR_i(L)$ are the participation constraints which guarantee that after any possible history the firm receives at least her reservation value; and IC_H, IC_L and $IC_i(H)$, $IC_i(L)$ for $i = H, L$ are the incentive constraints which guarantee that the regulated agent is willing to report its type truthfully, in the first period and in the second period after a history i.[7]

As it is formally proven in the appendix, in period 1 and in period 2 after any history h only the incentive compatibility constraint of the low-cost type and the participation constraint of the high-cost type can be binding; and these constraints can be assumed to be satisfied as equalities without loss of generality. Given this, we can write the profit π_{iL} enjoyed by a firm who in the second period is low-cost after it has declared to be a type $i = H, L$ in period 1 as:

$$\pi_{iL} = s_{iL} - \varphi(\theta_L - c_{iL}) = \pi_{iH} + \Phi(c_{iH}) \qquad (2)$$

where $\Phi(c_{iH}) = \varphi(\theta_H - c_{iH}) - \varphi(\theta_L - c_{iH})$, and the second equality follows from the binding participation constraint for the high-cost type.

Consider now the incentive constraint in the first period. We have:

$$\begin{aligned} \pi_L &= \big[s_L - \varphi(\theta_L - c_L)\big] + \delta\big[\alpha\pi_{LL} + (1-\alpha)\pi_{LH}\big] \\ &= \big[s_H - \varphi(\theta_L - c_H)\big] + \delta\big[\alpha\pi_{HL} + (1-\alpha)\pi_{HH}\big] \\ &= \pi_H + \Phi(c_H) + \delta(2\alpha - 1)\big[\Phi(c_{HH}) - \pi_{HH}\big] \\ &= \Phi(c_H) + \delta(2\alpha - 1)\Phi(c_{HH}), \end{aligned} \qquad (3)$$

[6] For simplicity henceforth we denote a history θ_i at $t = 1$ (respectively $\{\theta_i, \theta_j\}$ at $t = 2$) with the relevant subscript i (respectively ij).

[7] A detailed description of these constraints is found in Appendix B, see Section B.1.

where the last equality follows from the fact that the participation constraint of the high-cost type is binding. We can therefore write the objective function as:

$$\min_{\mathbf{c_h}\ \forall h}\left\{\begin{array}{l}(1+\lambda)\sum_{i=H,L}\Pr(\theta_i)\left\{\begin{array}{l}c_i+\varphi(\theta_i-c_i)\\ \quad+\delta\sum_{j=H,L}\Pr(\theta_j\mid\theta_i)\big[c_{ij}+\varphi(\theta_j-c_{ij})\big]\end{array}\right\}\\ \quad+\lambda\nu\big[\Phi(c_H)+\delta(2\alpha-1)\Phi(c_{HH})\big]\end{array}\right\}.$$

This is an unconstrained minimization problem with a convex objective function; from the first order necessary and sufficient conditions with respect to c_h $\forall h$ we obtain the optimal levels of effort:

$$e_h^{\mathcal{C}}=\begin{cases}\frac{1}{\eta} & h=L,LL,LH,HL,\\ \frac{1-\eta\Gamma_0\frac{\lambda}{1+\lambda}\Delta\theta}{\eta} & h=H,\\ \frac{1-\eta(\frac{2\alpha-1}{\alpha})\Gamma_0\frac{\lambda}{1+\lambda}\Delta\theta}{\eta} & h=HH.\end{cases}\tag{4}$$

The solution (4) shows the time-inconsistency problem faced by the principal in this stochastic environment.[8] In the optimal regulatory framework, the principal gives up some efficiency in the second period in order to reduce the rents left to the agent in the first period: this reduces the shadow cost of transfers at $t=1$ and increases social welfare. If the principal could rewrite the contract in the second period after history HH, for example, she would choose the second-period optimal level $e^P(HH)=\frac{1-\eta\Gamma_H\frac{\lambda}{1+\lambda}\Delta\theta}{\eta}$ and not the (generically different) level $e_{HH}^{\mathcal{C}}=\frac{1-\eta(\frac{2\alpha-1}{\alpha})\Gamma_0\frac{\lambda}{1+\lambda}\Delta\theta}{\eta}$ prescribed by (4). The optimal contract, however, is efficient after histories L, LL, LH and HL. Because the incentive compatibility constraint of the low-cost type is satisfied as equality in equilibrium, the rent obtained by the low-cost type depends only on the utility that he receives if in the first period he chooses to report that he is a high-cost. The principal pays this rent and extracts all the remaining expected surplus; the principal is residual claimant in the following nodes, and finds it optimal to choose efficient cost levels.

The case with changing types, however, is different from the canonical case with constant types.[9] Even if the history is economically irrelevant, the principal treats the agent differently according to it at $t=2$. A firm with a low cost history is treated better than a firm with a high cost history: in the second case, the firm is required to exert less effort than the efficient level. Since the equilibrium level of the firm's rent is increasing in the level of effort, a firm with a LL history receives a higher rent in the second period than a firm with a history HL: $e_{HH}^{\mathcal{C}}<e_{LH}^{\mathcal{C}}\Rightarrow\pi_{HL}<\pi_{LL}$.

[8] The optimal contract is fully characterized by (4). Indeed, given (4) the monetary transfers s_t can be found immediately using the binding constraints (2), (3) and the binding individual rationality constraint for the high-cost firm. Because of this, and because it would not add to the results, here and in the following propositions we omit the explicit solution of s_t for simplicity.

[9] Because the focus of the paper is in the optimal renegotiation-proof contract, we do not present here a full narrative discussion of the intuition behind the properties of the optimal contract with commitment, but we highlight only the features that are useful for the discussion of renegotiation-proofness. See Battaglini (2005) for a complete discussion of the properties of optimal contacts with commitment when types are variable.

This feature of the optimal contract is not present when types are constant over time. The benchmark case with constant types follows immediately from (4). If we let $\alpha \to 1$, we have that with constant types the principal offers the optimal static contract in every period:

$$e_h^{\mathcal{C}} = \begin{cases} \frac{1}{\eta} & h = L, LL, LH, HL, \\ \frac{1-\eta\Gamma_0\frac{\lambda}{1+\lambda}\Delta\theta}{\eta} & h = H, HH. \end{cases} \tag{5}$$

As it can be seen from (5), and also well known in the literature, the optimal contract with constant types ignores past history and is a simple repetition of the optimal static contract.[10]

Because contracts (4) and (5) are ex post inefficient, they are not credible: both the principal and the agent may agree on renegotiating them. In the following sections, therefore, we will characterize the optimal contract that is never renegotiated.

4. Information revelation and renegotiation

When the contract can be renegotiated the revelation principle does not hold any longer and we can not assume full separation of types in period one without loss of generality. How much information is revealed by the agents' choices in period one? How do types separate? In this section we answer these questions and characterize the basic structure of the optimal renegotiation proof contract.

When the contract must be renegotiation proof, the principal solves program $\mathcal{P}_I$ with an additional constraint. For any history h, the contract must remain conditionally optimal in the second period given the level of utility promised to the firm at $t = 1$:

$$\{c_{h,j}^*, s_{h,j}^*\}_{j=H,L} \in \arg\min \sum_{j=H,L} \Pr(\theta_j \mid h) \begin{bmatrix} (1+\lambda)(c_{hj} + \varphi(\theta_j - c_{hj})) \\ +\lambda(s_{hj} - \varphi(\theta_j - c_{hj})) \end{bmatrix} \tag{$\mathcal{R}$}$$
$$\text{s.t.} \quad IC_h(j) \text{ and } \pi_{h,j} \geqslant \pi_{h,j}^* \text{ for } j = H, L$$

where $\Pr(\theta_j \mid h)$ are the posterior probabilities at the beginning of period 2 given a history h, and $\pi_{h,j}^*$ is the rent promised by the initial contract to a firm of type θ_j after period 1 history h. Observe that there is no loss of generality in assuming that the optimal contract specifies only two options at $t = 2$: indeed, given any posterior, at $t = 2$ the problem of the principal is a standard static principal–agent problem in which the utility promised to the types at $t = 1$ is the reservation level at $t = 2$.

Constraint ($\mathcal{R}$) can be usefully simplified. Let $\tilde{c}(\mu)$ be the cost level prescribed for the high-cost type in the optimal contract at $t = 2$ when the posterior at the beginning of the period is μ[11]:

$$\tilde{c}(\mu) = \theta_H - e^* + \frac{\mu}{1-\mu}\frac{\lambda}{1+\lambda}\Delta\theta.$$

Let $\mu(h)$ be the posterior probability that the firm is low-cost after history h, we have:

Lemma 1. *The renegotiation proofness constraint* ($\mathcal{R}$) *is satisfied after history h if and only if*:

$$\theta_H - e^* \leqslant c_{h,H} \leqslant \tilde{c}\big(\mu(h)\big) \tag{6}$$

[10] With constant types histories HL and LH have zero probability. Obviously, however, the optimal contract with constant types must specify the contractual terms after all histories (deviations are possible out-of-equilibrium). In this sense, the optimal contract can be seen as a repetition of the optimal static contract.

[11] This is simply the cost that would be prescribed for the inefficient type in an optimal static contract if the prior is μ.

where $c_{h,H}$ is the cost implemented in the second period after a history $\{h, \theta_H\}$.

The intuition for this result is the following. Both the surplus produced by the project (not considering the cost of incentive transfers) and the principal's objective functions (which incorporates incentive transfers) at node h are concave in the cost chosen for the high cost firm $c_{h,H}$. If $c_{h,H} > \tilde{c}(\mu(h))$, then a marginal reduction in $c_{h,H}$ would be a Pareto improvement, and it would also increase the principal's objective function: so it would be renegotiated. If $c_{h,H} < \theta_H - e^*$, then a marginal increase in $c_{h,H}$ would also be a Pareto improvement and increase the principal's objective function as well: again, it would be renegotiated. When (6) is satisfied the only way to have a Pareto improvement is to reduce $c_{h,H}$: but such a change would be vetoed by the principal since it reduces her objective function.

At this stage we cannot be very specific regarding what a history h at $t = 2$ can be: this depends on the message space $\mathcal{M}^1$. Since the revelation principle does not hold in this environment, we cannot assume a direct mechanism. The next two lemmas characterize the structure of the optimal renegotiation proof contract. They allow us to restrict attention to contracts in which only two menus are chosen: one chosen only by the low-cost types, and the other by the high-cost type or both types.

Let us assume that the optimal contract specifies a menu A^o for period 1 (in which each option is a price-quantity pair associated to a message $m^1 \in \mathcal{M}^1$). Given this, the possible histories are $H^1 = \{i\}_{i \in A^o}$ in period 1 after the agent's choice, and $H^2 = \{iH, iL\}_{i \in A^o}$ in period two after the second period realization of the agent's type. We proceed in two steps. In the first step, we rule out double randomizations:

Lemma 2. *In the optimal renegotiation proof contract no pair of options i or j are chosen by both types with positive probability.*

This lemma implies that for any couple of options i, j offered by the principal, only three cases are possible. We can have full separation, in which case each type selects a different contract; or, if we have a pooling equilibrium, we can have two cases: either one of the two options is chosen only by the low-cost type, and the other contract is chosen by both types with positive probability; or one contract is chosen only by the high-cost type, and the other contract is chosen by both types with positive probability.[12]

The second step further restricts the possibilities:

Lemma 3. *Without loss of generality, in the optimal renegotiation proof contract there is no pair of options i and j such that with positive probability one is chosen by both types and the other only by the high-cost type.*

From Lemmata 2 and 3 we know that at the optimum we can only have two cases. In the first case, at $t = 1$ the principal offers a menu in which options are chosen exclusively by each type of agent. Since the optimal contract is fully separating, all the options chosen by one type can be merged into one option without loss of generality. In the second case, at $t = 1$ the principal offers

[12] Note that Lemma 2 holds *for any couple i, j* of options offered by the principal at $t = 1$: it does not rule out the possibility that more than two options are offered in equilibrium. For example, it does not rule out the possibility that three options are offered, two separating and a third that is chosen with positive probability by both types. This and other potential cases are ruled out by Lemma 3.

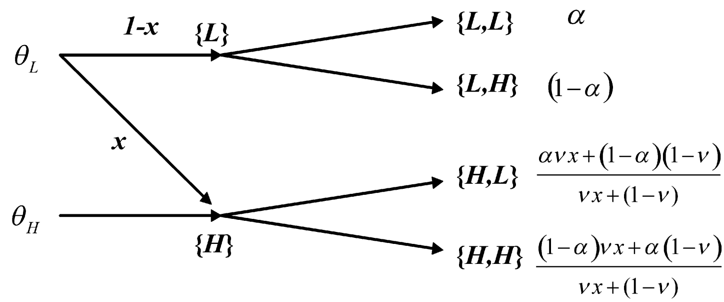

Fig. 2. The equilibrium distribution of types at $t = 2$: x is the probability that the efficient type θ_L pools with the inefficient type θ_H.

a menu in which some options are chosen only by the low-cost firm, and only one option that is chosen by both types. Again, we can merge all the options chosen only by the low-cost type into a single option. We therefore conclude:

Proposition 1. *Without loss of generality, in the optimal renegotiation proof contract the principal can offer the agent a choice between two options*; *one will only be chosen by the low-cost type, and the other will be chosen by either the high-cost type or both types.*

Figure 2 describes the transition probabilities and the posterior distribution of types at $t = 2$ in the optimal contract. Only the low-cost firm may randomize in equilibrium. Option $\{L\}$ is chosen by the low-cost firm only. With probability $x \in [0, 1]$, however, this firm may choose the "inefficient" option (history $\{H\}$), which is also chosen by high types. This choice of the low-cost firm affects the posterior distribution of types in the second period after $\{H\}$. To simplify the comparison with the case with commitment we refer to $\{L\}$ as the low-cost option, and $\{H\}$ as the high-cost option.

Let $C(\theta, c, s) = (1 + \lambda)(c + \varphi(\theta - c)) + \lambda(s - \varphi(\theta - c))$ be the net (per period) welfare cost of the project if the type is θ, the transfer is s, and the realized cost of the project is c; and let x be the probability that the low-cost type mimics the high-cost type. Using Proposition 1, the optimal renegotiation proof contract can be expressed as the solution of the program:

$$\min_{\langle x,\mathbf{c},\mathbf{s}\rangle} \left\{ \begin{array}{l} v(1-x)C(\theta_L, c_L, s_L) + vxC(\theta_L, c_H, s_H) + (1-v)C(\theta_H, c_H, s_H) \\ \quad + \delta \left[\begin{array}{l} \sum_{j=H,L} \Pr(\theta_j \mid \theta_L)[v(1-x)C(\theta_j, c_{Lj}, s_{Lj}) + vxC(\theta_j, c_{Hj}, s_{Hj})] \\ \quad \times (1-v) \sum_{j=H,L} \Pr(\theta_j \mid \theta_H) C(\theta_j, c_{Hj}, s_{Hj}) \end{array} \right] \end{array} \right\} \qquad (\mathcal{P}_{II})$$

under constraints sets (1) and (6), and such that $x = 0$ if IC_L is strictly satisfied and $x \in [0, 1]$ if it is satisfied with equality.

5. The optimal renegotiation proof contract

We can now characterize the optimal contract by solving $\mathcal{P}_{II}$. In this section we show that, as anticipated by Fig. 1 in the Introduction, we may have three relevant cases, depending on the persistence of the types. In the first case, the optimal renegotiation proof contract coincides with the ex ante optimal contract with commitment (Area I in Fig. 1). In the second case, it is not ex ante optimal, but is still fully separating (Area II). Finally, in the third case, $x \in (0, 1)$ and we have pooling of types in the first period (Area III). These three cases are studied in order in the next three subsections.

5.1. When is the renegotiation proofness constraint irrelevant?

In this section we characterize the threshold on persistence α_1 that guarantees that there is no conflict between ex ante optimally and renegotiation proofness.

Both with constant and with variable types, the principal finds it optimal to require the high-cost firm to choose a cost level higher than what would be efficient to choose with no asymmetric information. As can be seen from (5), when types are constant, the distortions on costs are actually equal in the first and second period—both are $\Gamma_0 \frac{\lambda}{1+\lambda} \Delta\theta$. On the contrary, when there is commitment but types are stochastic, the distortion in the second period is

$$\left(\frac{2\alpha - 1}{\alpha}\right) \Gamma_0 \frac{\lambda}{1+\lambda} \Delta\theta$$

which is strictly smaller than the distortion in the first period if the types are positively correlated ($\alpha > 1/2$). In the limit case when types are uncorrelated ($\alpha \to \frac{1}{2}$), the distortion in the second period is actually zero. Because of this, the principal is promising a contract that becomes more efficient in the second period than in the first when the agent's types become less correlated.

Assume that α is such that the principal at $t = 1$ finds it optimal to promise a contract for $t = 2$ in which the required cost is less than or equal to the cost that would be optimal to choose in the second period with no commitment. This occurs when persistence is below a threshold, $\alpha < \alpha_1 = \frac{1+\Gamma_0}{1+2\Gamma_0}$. In this case, the contract generates more social surplus at $t = 2$ than the contract that would have been chosen with no commitment at that stage. If the principal tries to renegotiate it with a contract that has higher costs, it is impossible for both the firm and the principal to be (weakly) better off. If, on the contrary, the principal proposes a contract that has a lower cost level, then we would have a more efficient contract: but the increase in the rent of the low-cost type would be so high that the principal would be worse off after the change. We have:

Proposition 2. *If $\alpha \leqslant \frac{1+\Gamma_0}{1+2\Gamma_0} \equiv \alpha_1$, there is no conflict between renegotiation proofness and ex ante optimality, and the optimal renegotiation proof contract e_h^L is given by* (4).

The reason why the optimal contract with commitment is never renegotiation proof when types are constant can be easily explained in light of the previous discussion. When types are constant, the optimal distortion in the second period is zero, since after the first period there is no residual uncertainty left (in a fully separating equilibrium). Therefore, it is always smaller than the ex ante optimal distortion. This implies that ex post the principal always finds it optimal to reduce the distortion, and the agent would agree with the change.

Proposition 2 suggests that in many environments the problem of time-inconsistency of contracts may be less important than what is suggested by the literature on renegotiation with constant types. It is natural to interpret the assumption that types are constant as an approximation to the assumption that types are highly correlated. However Proposition 2 shows that this interpretation is not entirely valid, since the optimal contract with commitment may be time-consistent even if types are very persistent.

5.2. When is the contract separating?

When $\alpha > \alpha_1$ the ex ante optimal contract is not renegotiation proof. However, as we prove in this section, the optimal renegotiation proof contract remains separating and in pure strategies while α remains below a second threshold α_2 (larger than α_1).

Proposition 1 guarantees that in the first period only the low-cost firm may randomize between the options offered in the optimal contracts. This does not necessarily imply that the low-cost firm is randomizing (and therefore IC_L is satisfied with equality), nor that we can ignore the incentive constraint of the high-cost firm. To complete the characterization of the optimal renegotiation proof contract let us define the auxiliary problem $\mathcal{P}_{III}$:

$$\min_{\langle x,\mathbf{c},\mathbf{s}\rangle}\left\{\begin{array}{l} v(1-x)C(\theta_L,c_L,s_L)+vxC(\theta_L,c_H,s_H)+(1-v)C(\theta_H,c_H,s_H) \\ \quad +\,\delta\left[\begin{array}{l}\displaystyle\sum_{j=H,L}\Pr(\theta_j\mid\theta_L)[v(1-x)C(\theta_j,c_{Lj},s_{Lj})+vxC(\theta_j,c_{Hj},s_{Hj})] \\ \displaystyle\quad\times(1-v)\sum_{j=H,L}\Pr(\theta_j\mid\theta_H)C(\theta_j,c_{Hj},s_{Hj})\end{array}\right]\end{array}\right\} \qquad (\mathcal{P}_{III})$$

under the constraint that:

$IC_L, IC_i(L)$ and IR_H $IR_i(H)$ $i = L, H$ are satisfied as equalities and (6) is satisfied.

We say that a cost schedule c_h is *renegotiation proof optimal* if, for any h, there is a transfer schedule s_h such that $\{c_h, s_h\}$ solves $\mathcal{P}_{II}$. The following lemma characterizes the optimal renegotiation proof cost schedule.

Lemma 4. *A cost schedule c_h is renegotiation proof optimal if and only if it solves $\mathcal{P}_{III}$.*

A few remarks are necessary to interpret this result. The usual approach to solve principal–agent models is to solve a "relaxed problem" in which some constraints are ignored, find a solution, and then verify that the solution to this problem solves the general problem too. This is not what we are doing with Lemma 4. Indeed, if we only considered the solution of the "usual" relaxed program that ignores the incentive constraints of the high-cost firm and the participation constraints of the low-cost firm without *imposing* that the remaining constraints are satisfied as equalities, we would certainly have solutions that do not solve the general program. While in a static setting, the incentive constraint of the low-cost type is necessarily binding, and this is sufficient to guarantee that the solution to the relaxed problem solves the general program as well, this is not generally true in a dynamic setting. There may exist a solution to the relaxed problem in which the incentive compatibility constraint is not binding after some history at $t = 2$, and that would violate some incentive compatibility constraint of the high-cost firm in $\mathcal{P}_{II}$. Similarly there are solutions $\{c_h, s_h\}$ to $\mathcal{P}_{II}$ in which the incentive compatibility of the low-cost type in the second period is not binding. Lemma 4, however, proves that we can use $\mathcal{P}_{III}$ as an auxiliary program to characterize the set of optimal solutions $\mathbf{c} = \{c_h\}_{\forall h}$. While the relaxed problem is essentially identical to the full problem in a static model because the omitted constraints are superfluous at the optimum, program $\mathcal{P}_{III}$ is not identical to the full program, it is merely a "tool" to characterize the cost schedule.

We can now find the optimal cost schedule c_h $\forall h$, by solving $\mathcal{P}_{III}$. It is easy to verify that c_L, c_{LH}, c_{iL} for $i = 1, 2$ are set at the efficient level in $\mathcal{P}_{III}$: $c_L = c_{iL} = \theta_L - e^*$ for $i = L, H$ and $c_{LH} = \theta_H - e^*$.[13] If we define

$$\mu(\alpha, x) = \frac{vx\alpha + (1-v)(1-\alpha)}{vx + (1-v)}$$

as the posterior probability that the firm is efficient after a history H, and use some simple algebra to rewrite $\mathcal{P}_{III}$, we can characterize the optimal cost schedule solving the program:

$$\begin{aligned} &\min_{\{c_{HH}, c_H, x\}} \quad A(x, c_H) + \delta B(x, c_{HH}) \\ &\text{s.t.} \quad \theta_H - e^* \leqslant c_{HH} \leqslant \tilde{c}\big(\mu(\alpha, x)\big) \end{aligned} \tag{7}$$

where

$$\begin{aligned} A(x, c_H) = (1+\lambda) &\left\{ \begin{array}{l} v(1-x)\big(\varphi(e^*) + \theta_L - e^*\big) \\ \quad + vx\big(\varphi(\theta_L - c_H) + c_H\big) + (1-v)\big(\varphi(\theta_H - c_H) + c_H\big) \end{array} \right\} \\ &+ \lambda v\big[\Phi(c_H)\big] \end{aligned}$$

is the total cost of the project in the first period; and

$$\begin{aligned} B(x, c_{HH}) = (1+\lambda) &\left[\begin{array}{l} [v\alpha + (1-\alpha)(1-v)](\varphi(e^*) + \theta_L - e^*) \\ \quad + v(1-\alpha)(1-x)(\varphi(e^*) + \theta_H - e^*) \\ \quad + [\alpha(1-v) + xv(1-\alpha)](\varphi(\theta_H - c_{HH}) + c_{HH}) \end{array} \right] \\ &+ \lambda v(2\alpha - 1)\Phi(c_{HH}) \end{aligned}$$

is the total cost of the project for the principal in the second period.

The renegotiation constraint must now be binding at the optimum:

Lemma 5. *If $\alpha > \alpha_1$, then $c_{HH} = \tilde{c}(\mu(\alpha, x))$ in* (7).

This implies that the problem can be written as the minimization of

$$A(x, c_H) + \delta B\big(x, \tilde{c}\big(\mu(\alpha, x)\big)\big) \tag{8}$$

in which the control variables are x and c_H. This expression makes clear the trade-off in the determination of x, and therefore in the determination of the optimal degree of pooling in the equilibrium contract. An increase in x has a *direct* and an *indirect* effect on welfare. The direct effect always reduces welfare in both periods. In the first period, the higher x is, the less efficient the screening of types is. In the second period, moreover, a low-cost firm that chooses the "*L*" path receives an efficient contract even if her type changes; on the contrary, after "*H*," a firm that is low-cost in $t = 1$ but high-cost in $t = 2$ receives a distorted contract. The higher x is, therefore, the greater the expected distortion, and the lower the expected welfare.

On the other hand, an increase in x also has an indirect positive effect on social welfare by increasing $\tilde{c}(\mu(\alpha, x))$. An increase in x directly increases the probability that a firm is low-cost after history $\{H\}$ because more firms that were low-cost in the first period choose this path. Since types are correlated over time, this increases the fraction of low-cost firms in the second

[13] Once the equality constraints are incorporated in the objective function, the problem becomes a simple minimization of a convex function. The efficiency of c_L, c_{LH}, c_{iL} follows from the first order necessary and sufficient conditions.

period. The increase in the posterior, finally, induces a higher cost level c_{HH}, since by Lemma 5, $c_{HH} = \tilde{c}(\mu(\alpha, x))$. This is beneficial because it reduces the gap between the ex ante optimal cost level with commitment after history $\{HH\}$ and the optimal level of the cost that can be credibly enforced after $\{HH\}$.

It is precisely because the marginal benefit of x is "mediated" by the marginal change of $\tilde{c}(\mu(\alpha, x))$ on $B(x, \tilde{c}(\mu(\alpha, x)))$ that full separation of types in $t = 1$ remains optimal when α is not too much larger than α_1. As proven in Proposition 2, with $\alpha \leqslant \alpha_1$, the ex ante optimal cost level for the second period is already more efficient than the optimal level in the second period: therefore, having pooling in the first period would unequivocally reduce welfare. Similarly, when $\alpha = \alpha_1$ the marginal benefit of increasing the cost level in the second period through an increase in x is zero; and when $\alpha > \alpha_1$, but $\alpha - \alpha_1$ is not too large, the indirect marginal benefit of x is not sufficient to compensate for its direct marginal cost. We have:

Proposition 3. *There exists a threshold* $\alpha_2 \in (\alpha_1, 1]$ *such that the optimal renegotiation proof contract is fully separating in the first period if and only if* $\alpha \leqslant \alpha_2$. *If* $\alpha \in (\alpha_1, \alpha_2]$ *effort is efficient after a history L, and it is ex post optimal after history H, then given the equilibrium beliefs*:

$$e_h^L = \begin{cases} \frac{1}{\eta} & h = L, LL, HL, LH, \\ \frac{1 - \eta \Gamma_0 \frac{\lambda}{1+\lambda} \Delta\theta}{\eta} & h = H, \\ \frac{1 - \eta \Gamma_H \frac{\lambda}{1+\lambda} \Delta\theta}{\eta} & h = HH. \end{cases} \tag{9}$$

It is interesting to note how the optimal contract (9) differs from the ex ante optimal contract with commitment (and therefore the optimal renegotiation proof contract when $\alpha \leqslant \alpha_1$), and from the second period optimal contract.

With constant types, the optimal renegotiation proof contract is always conditionally optimal: given the equilibrium strategies and the implied posterior beliefs, the contract is optimal for the principal in the second period. On the contrary, when $\alpha \leqslant \alpha_2$, the optimal renegotiation proof contract is not second-period optimal. Indeed, after a history $\{L\}$ it would be optimal to introduce some distortion (since there is still uncertainty about the type's realization), but the optimal renegotiation proof contract prescribes no inefficiency. This shows that while in the second period the renegotiation proofness constraint is not really binding for the principal with constant types (she indeed takes a conditionally optimal action), the constraint really binds with changing types.

After a history $\{HL\}$, (4) and (9) are efficient and prescribe a level of effort that coincides with the second period optimal level and the level prescribed by the ex ante optimal contract with commitment. The optimal renegotiation proof contract after history $\{HH\}$ with $\alpha > \alpha_1$ (i.e., c_{HH}^L), however, differs from the level in the optimal contract with commitment, $c_{HH}^{\mathcal{C}}$:

$$c_{HH}^{\mathcal{C}} - c_{HH}^L = \left[\left(\frac{2\alpha - 1}{\alpha}\right)\Gamma_0 - \Gamma_H\right]\frac{\lambda}{1+\lambda}\Delta\theta > 0. \tag{10}$$

Interestingly the optimal renegotiation proof contract with $\alpha > \alpha_1$ is second-period optimal for the principal after history H, as in the case with constant types, but this cost level is strictly lower (more efficient) than what would be ex ante optimal with commitment. When $\alpha \leqslant \alpha_1$ the cost level is ex ante optimal but strictly larger (more inefficient) than the second period optimal level.

It is finally interesting to note that, as proven by Hart and Tirole (1988) and Laffont and Tirole (1990), the optimal renegotiation proof contract is in mixed strategies when types are constant except when δ is sufficiently small. This is necessary to generate the correct posterior beliefs in the second period that guarantee that there is no Pareto superior contract. These strategies may require considerable strategic sophistication. Except when the interaction is repeated many times, it may seem unrealistic that the play of the game follows the exact equilibrium prediction. This may be a problem in situations in which the principal–agent relationship is desultory. When types change over time and $\alpha \leqslant \alpha_2$, on the contrary, the equilibrium is in pure strategies and only requires the agent to report its type.

5.3. When do we have a pooling equilibrium?

In this section we study the conditions for the existence of pooling equilibria and their properties. Propositions 2 and 3 guarantee that no pooling equilibrium exists in $(\frac{1}{2}, \alpha_2)$. However, they do not prove that a pooling equilibrium exists at all since α_2 could be equal to one. Moreover, the objective function is non-linear in x and α, so we potentially may have multiple optimal levels of pooling.

The key to the existence of an equilibrium with pooling is the relationship between the payoffs in the first and second periods. When the payoffs in the first period are substantially more important than the payoffs in the second period (for example if δ is small), then the problem is essentially static and, not surprisingly, full separation is optimal so that $\alpha_2 = 1$. Is it true that it becomes optimal to have some pooling at $t = 1$ when the payoffs of the second period are sufficiently large compared to the payoffs in the first period?

The relationship between the payoffs in the first and second period is determined by δ. In this sense this variable not only reflects the discount factor, but, more in general, the relationship between the payoffs in the two periods. This is the interpretation given in Laffont and Tirole (1990), who assume δ to be in $(0, \infty)$. Accordingly, in this section we will assume $\delta \in (0, \infty)$.

Proposition 4. *There exists a threshold $\delta^* \in (0, \infty)$ such that $\alpha_2 < 1$ if $\delta > \delta^*$, and $\alpha_2 = 1$ otherwise.*

With constant types an increase in δ only increases the marginal benefit of pooling because the cost of pooling is concentrated in the first period.[14] However, this fails to be true when $\alpha < 1$. In this case, the cost of pooling spreads to the second period too. As we discussed before, if a low-cost agent declares itself to be high-cost, it would receive an inefficient cost assignment in the second period if its type changes. On the contrary, the cost assignment is always efficient if it truthfully reports its type at $t = 1$. So as we increase the relevance of the payoff in the second period (increasing, for example, δ) we not only increase the benefits of pooling, but also some of its costs. This cost of pooling in the second period is proportional to the probability that the high-cost agent changes type, $1 - \alpha$, so as $\alpha \to 1$, this cost fades away in the second period. For this reason it is *not true* that for any $\alpha < 1$ there is a δ large enough that guarantees pooling; but it is true that when δ is large enough the upper bound α_2 is lower than one.

Is the optimal level of pooling unique when $\delta > \delta^*$? The indirect "beneficial" effect of x at $t = 2$ is not monotonically increasing in x. On the one hand, an increase in x reduces the marginal

[14] In this case x affects welfare in the second period only indirectly through a change in $c(\mu(\alpha, x))$, therefore we do not have the direct effect described in Section 5.2 and an increase in x always (weakly) increases welfare at $t = 2$.

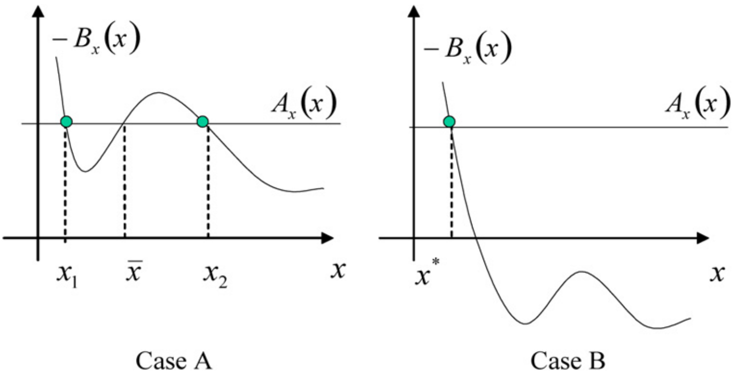

Fig. 3. Uniqueness of the solution: only case B is possible.

effect $\frac{\partial}{\partial c}B(x, \tilde{c}(\mu(\alpha, x)))$; on the other hand, it increases $\frac{\partial}{\partial x}\tilde{c}(\mu(\alpha, x))$. If case A in Fig. 3 were possible, then we might have multiple equilibria (x_1, x_2). Proposition 5 proves that only case B in Fig. 3 is possible.

Proposition 5. *For $\alpha > \alpha_2$ there is a unique optimal level of pooling $x^*(\alpha, \delta) \in (0, 1)$, and the optimal contract is characterized by*:

$$e_h^L = \begin{cases} \frac{1}{\eta} & h = L, LL, HL, LH, \\ \frac{1}{\eta} - \frac{v}{vx^*(\alpha,\delta)+(1-v)}\left(\frac{\lambda}{1+\lambda} - x^*(\alpha, \delta)\right)\Delta\theta & h = H, \\ \frac{1}{\eta} - \frac{vx^*(\alpha,\delta)\alpha+(1-v)(1-\alpha)}{vx^*(\alpha,\delta)(1-\alpha)+\alpha(1-v)}\frac{\lambda}{1+\lambda}\Delta\theta & h = HH. \end{cases}$$

Clearly, the equilibrium level of pooling, as measured by $x^*(\alpha, \delta)$, depends on all the parameters of the model. However, it is possible to bound its possible range and characterize its sensitivity to the persistence of types. Using the fact that we can rule out case A in Fig. 3 we can prove that $x^*(\alpha, \delta)$ is monotonic in α.

Proposition 6. *The optimal pooling level $x^*(\alpha, \delta)$ is non-decreasing in α and strictly increasing for $\alpha > \alpha_2$.*

Laffont and Tirole (1990) proved that when types are constant full-pooling is impossible. Proposition 6, therefore, has an immediate implication.[15]

Corollary 1. *With imperfect persistence pooling is never larger than with constant types, so full pooling is never optimal for any level of correlation α.*

This result extends Laffont and Tirole's (1990) finding that full pooling is suboptimal with constant types to the general case when types can have any degree of correlation over time. This finding has practical implications: at $t = 1$, it is always optimal to offer a menu of options that separates types in a significant way.

[15] The fact that full pooling is not optimal at $\alpha = 1$ can also be directly seen from (25).

5.4. Discussion

We conclude this section with a discussion of the model and its applications, and the questions that are open for future research.

Applications. In Section 2, we motivated the model by suggesting that it describes the problem of optimal design of a cost based procurement contract, but the model has many other potential applications. For example, it can be used to study the dynamic interaction between a buyer and a seller of a non-durable good.[16] Assume for simplicity $\eta = \frac{1}{2}$. As it can be easily verified, in the model of Section 2 the agent's preferences can written (up to a constant) as $s + u(\theta, c)$: the transfer s plus a function of the cost and the type, $u(\theta, c) = \theta c - \frac{1}{2}c^2$.[17] Given this, to obtain a model of price discrimination, we only need to reinterpret the choice variables. We can reinterpret the variable c as the quantity supplied by the seller q and, without loss of generality, s as the transfer price of the good net of production costs $C(q)$: $-s = p - C(q)$. If we assume that the utility of the agent for q units of the good is θq and the cost of producing the good is $\frac{1}{2}q^2$, then $s + u(\theta, q)$ is precisely the net utility of a consumer for q units at price p, $\theta q - p$. When λ is large the principal is interested primarily in minimizing the monetary payments to the agents, so the model is isomorphic to the problem of optimal consumer screening by a profit maximizing monopolist. The case with variable but persistent types seems well suited to model this type of long-term, non-anonymous market interactions between buyers and sellers, which is becoming more and more common in Internet transactions where retailers such as *Amazon.com* collect and store large amounts of information and may condition contractual terms (as price discounts) on past history. Our model therefore contributes to understanding these types of interactions when the seller does not have commitment because the contract can be renegotiated.[18]

Extensions and open questions. We now turn to the discussion of possible extensions of the basic framework, focusing in particular on the time horizon, the type space and the contractual environment. In our previous analysis we considered the simplest form of dynamics, a model with two periods. Since many environments span a longer time horizon, it would be interesting to extend the basic model to more than two periods. Battaglini (2005) characterizes the ex ante optimal contract with commitment in a model with multiple finite or infinite periods. A full characterization of the optimal renegotiation proof contract in a infinite horizon model is still a question open for future research. This extension presents no conceptual complications, except that the analysis becomes more tedious; such analysis, however, would allow to study the asymptotic properties of the model as $t \to \infty$. We conjecture that the optimal renegotiation proof contract would converge to an efficient contract, but we leave the analysis of this issue for future work.

As in Hart and Tirole (1988), Laffont and Tirole (1990), Blume (1998), we assumed that the state variable θ^t is binary. Only two papers considered the more complicated case with many (but constant across periods) types. Dewatripont (1989) considers the case with n finite types, but he has to assume that the equilibrium is in pure strategies, and to restrict the set of possible contracts. Laffont and Tirole (1988) consider a model of dynamic contracting with a continuum of types but do not present a complete characterization of the contract (the full characterization is

[16] For a related discussion see also Laffont and Tirole (1990).

[17] The agent's utility is $s - \frac{1}{2}(\theta - c)^2$. Setting $\alpha = s - \frac{1}{2}\theta^2$ it can be written as $\alpha + \theta c - \frac{1}{2}c^2$.

[18] An infinite horizon version of this model is studied by Battaglini (2005), who focuses on the case with commitment.

presented in Laffont and Tirole (1990), where two types are assumed): their main result is that full separation is not optimal in the first period. When types may change stochastically the analysis is even more complicated. In this case the type of an agent is given not only by θ^t, but also by the shape of the conditional distribution of future realizations. When there are n types the distribution is an $(n-1)$-dimensional object, and we are in a multidimensional screening problem. As it is well known, multidimensionality complicates the analysis of screening problems; our framework allows us to separate the dynamics of the principal–agent interaction, which is the focus of the paper, from the conceptually distinct problem of the multidimensionality of types in order to produce a sharper analysis of the dynamics.

A related question concerns the assumption of a Markovian stochastic process. It would be interesting to study the optimal contract under alternative assumptions. This analysis, however, would be complicated by two problems. First, if the transition probabilities do not depend only on the current type but also on the previous history, then the analysis is, again, multidimensional. Second, to fully capture the properties of these types of stochastic processes we would need to consider a temporal horizon with more than two periods.[19]

Regarding the contractual environment, in this work we have focused the analysis on long-term contracts with renegotiation. Different assumptions are also possible. One case is that the principal can fully commit to a contract, and renegotiation is not an issue. We have presented this case, which is plausible in many environments, in Section 3 as a benchmark. At the other extreme is the case in which the principal cannot make any intertemporal commitment at all and can only offer spot contracts. This is certainly an interesting case in some environments, but unrealistic in most long-term principal agent interactions. Indeed, although contract can be renegotiated, it is rarely the case that a non-anonymous principal–agent interaction is independent from past performance. It is therefore important to understand the limits imposed by renegotiation proofness on long-term agreements.

As proven by Laffont and Tirole (1986, 1988), the optimal contract with no intertemporal commitment can be very different from the optimal renegotiation proof contract. Since a renegotiation can be vetoed by the agent, the principal can credibly commit to leave a minimal level of utility to the agent by offering a long term renegotiable contract. When only spot contracts are possible, the lack of even this type of commitment generates the *Ratchet effect*: because the agent fears that any information that is revealed can be used against him/her, little information is revealed in the early stages of the game. Unfortunately, even when types are constant, these contracts are very intractable since the set of binding incentive compatibility constraints generally depends on the parameters of the model. A study of the optimal contract in a model very similar to the model described in Section 2 (but with constant types) is studied by Laffont and Tirole (1987). Because of the complications generated by the Ratchet effect, however, the optimal contract with no commitment can be characterized only using numerical methods.[20]

[19] A non-Markovian assumption on the distribution of types is made in Blume (1998). In this paper, he assumes that there are two types: one type has a constant high valuation, the other type has valuations that change according to an i.i.d. process. In our model the expected utility of a type depends on his current realization; in Blume's model, on the contrary, types are permanently different. Because his paper studies a different problem (how to sell a durable good) and restricts demand to be 0 or 1, however, it is difficult to compare its results with the present work. Battaglini (2005) considers an intermediate case in which types are Markovian, but may have different persistence levels. Although in that paper the optimal renegotiation proof contract is presented only for a given parameter range (the paper focuses on the case with commitment), the results are consistent with the more comprehensive characterization presented in the present work.

[20] The relationship between short-term contracts and long-term contracts with commitment is also studied by Rey and Salanie (1996). These authors, however, do not characterize the optimal contract with renegotiation, or the equilibrium

In an important contribution Kennan (2001) has studied a seller–buyer model with variable types in which the principal can only offer a sequence of spot contracts. He shows that the equilibrium may involve a cycle of screening offers (in which types choose different alternatives) and pooling offers (in which types behave in the same way). When the principal offers a screening contract, the posterior probability fluctuates according to the agent's choices. When the posterior drops low enough, a pooling contract becomes optimal. In this case, although types do not separate, the posterior does not remain constant: when types follow a Markov process, it converges back to its stationary level. When this level is high enough, a screening contract may become optimal, and a new cycle starts again. A key assumption in the characterization of this model is that supply can take only 0 or 1 values, which substantially restricts the ability of the seller to screen the types of the agent. Because of this it is difficult to say what the contract would look like when supply can take a continuum of values as in our model.

6. Conclusion

Since Ramsey's (1927) seminal paper on optimal policy in a dynamic environment, the literature has struggled with a time-inconsistency issue. Indeed the ex ante optimal policy is rarely time-consistent because it can often be substituted in subsequent periods by a Pareto superior policy. In the context of principal–agent problems, time inconsistency of the ex ante optimal contract always reduces welfare when the agent's type is constant over time (as typically postulated in the existing literature) because it makes it more difficult to provide credible incentives that induce the agent to reveal his type.

In this paper we have characterized the optimal renegotiation proof contract in a principal–agent model in which the type of the agent can vary across periods. This case is important both from a strictly theoretical point of view, and from a more applied perspective, since an agent's types are variable and stochastic in many real world environments. We have shown that, in this case, even when types are very persistent, the standard equilibrium with commitment may be renegotiation proof. This is good news for applied work because, by entitling to use the revelation principle in these environments, it may substantially simplify the study of contractual relations when it would not be plausible to rule out the possibility of renegotiation by assumption.

Appendix A

In this appendix we show that Assumption 1 implies that the public project is always realized in the optimal renegotiation proof contract.

We first show that the project is always realized in the first period. Assume by contradiction that this is not true. In this case at $t = 1$ the principal offers a menu in which the low type strictly prefers not to produce the public good (otherwise it would be optimal to require the high cost type to realize the project as well). The expected utility for the principal can not be larger than:

$$vS(\theta_L) + \delta\big[vE\Phi(\alpha) + (1-v)E\Phi(1-\alpha)\big] \tag{11}$$

with spot contracts: they show that an optimal long-term renegotiation proof contract (more than 2 periods) can be approximated (under some conditions) by a sequence of 2-period contracts with commitment.

where $E\Phi(\mu)$ is the expected utility of the principal if the optimal one period contract is chosen at $t = 2$ and the probability that the type is low-cost is μ.[21] Since $E\Phi(\alpha) < S(\theta_L)$, we can bound (11) from above by: $v(1 + \delta)S(\theta_L) + \delta(1 - v)E\Phi(1 - \alpha)$. If in the first period the project is realized by both types, then a lower bound of the expected utility for the principal is:

$$S(\theta_H) + \delta E\Phi\big(v\alpha + (1 - v)(1 - \alpha)\big). \tag{12}$$

The first term in (12) is the expected utility if there is no screening in the first period and both types receive a cost refund equal to the cost of the high-cost type; the second term reflects the fact that in the second period the principal can offer an optimal one-period contract. Equation (12) can be bounded below by $(1 + \delta v)S(\theta_H) + \delta(1 - v)E\Phi(1 - \alpha)$ because $E\Phi(\mu)$ is increasing in μ. Given this, we conclude that (11) is strictly smaller than (12) if $\frac{(1+\delta)v}{1+\delta v}S(\theta_L) - S(\theta_H) < 0$, which is guaranteed by Assumption 1. This implies that the original contract can not be optimal, a contradiction.

Now consider period 2. Assume by contradiction that there is a history h after which the optimal renegotiation proof contract requires only the low-cost type to produce and in which with positive probability production does not take place. Since this contract is inefficient at $t = 2$, it must be that at this node the principal strictly prefers it to a contract in which both agents are producing (otherwise the contract would be renegotiated). This is possible only if $\mu > v^*$, where μ is the posterior probability that the type is low-cost after history h, otherwise we would have $\mu S(\theta_L) - S(\theta_H) < 0$, a contradiction. It is easy to verify that it is not possible that the posterior probability of a low-type is larger than v^* (which is larger than the prior v) after all nodes. Let h' be a node in which the posterior probability of a low-cost type is $\mu' \leqslant v^*$, so that both types are providing the public good. Since after h, by assumption, the project is not produced with positive probability, it must be that a positive mass of high-cost types is choosing h, so the high-cost type is indifferent between nodes h and h'. Conditional on being at either node h or h', let x be the conditional probability that a low-cost type chooses node h and let y be the conditional probability that the same node is chosen by a high-cost type. Probabilities x and y can not be both interior in $(0, 1)$.[22] Since by assumption the mass of high-cost types is positive after h, it must be that $x = 1$, $y \in (0, 1)$. But this is impossible. Consider a marginal decrease in y. This leaves the contract after h renegotiation proof. After history h', however, there are only high-cost types, so the posterior probability that a type is low-cost at $t = 2$ is $1 - \alpha$, independent of y. We conclude that after the marginal change in y, the contract remains renegotiation proof at history h' as well. However, the change strictly increases the principal's payoff because $1 - \alpha < \frac{1}{2} \leqslant v^*$ so $(1 - \alpha)S(\theta_L) - S(\theta_H) < 0$. From this contradiction we conclude that the projects is realized with probability one after all histories.

[21] At $t = 1$ only the low-cost type produces, so in the second period the principal can potentially condition the contract on this information.

[22] Since after node h the planner strictly prefers to exclude the high-cost types, after a marginal change in x and y the contract remains renegotiation proof at node h. The renegotiation constraint after node h' is satisfied if after the change if the posterior probability kept larger than μ'. This implies that x, y must be the solution of a maximization problem with a linear objective function and a linear constraint, which must have a corner solution 0 or 1, a contradiction.

Appendix B

B.1. Characterization of the optimal contract with commitment

In this section we complete the argument presented in Section 3 for the characterization of the optimal contract with commitment.

The set of constraints of $\mathcal{P}_I$ is described by the following inequalities. The incentive compatibility constraints are:

$$\begin{aligned}&\big[s_H-\varphi(\theta_H-c_H)\big]+\delta\big[\alpha\big(s_{HH}-\varphi(\theta_H-c_{HH})\big)+(1-\alpha)\big(s_{HL}-\varphi(\theta_L-c_{HL})\big)\big]\\&\quad\geqslant\big[s_L-\varphi(\theta_H-c_L)\big]+\delta\big[\alpha\big(s_{LH}-\varphi(\theta_H-c_{LH})\big)+(1-\alpha)\big(s_{LL}-\varphi(\theta_L-c_{LL})\big)\big],\end{aligned}\tag{IC_H}$$

$$\begin{aligned}&\big[s_L-\varphi(\theta_L-c_L)\big]+\delta\big[(1-\alpha)\big(s_{LH}-\varphi(\theta_H-c_{LH})\big)+\alpha\big(s_{LL}-\varphi(\theta_L-c_{LL})\big)\big]\\&\quad\geqslant\big[s_H-\varphi(\theta_L-c_H)\big]+\delta\big[\alpha\big(s_{HL}-\varphi(\theta_L-c_{HL})\big)+(1-\alpha)\big(s_{HH}-\varphi(\theta_H-c_{HH})\big)\big],\end{aligned}\tag{IC_L}$$

$$s_{iH}-\varphi(\theta_H-c_{iH})\geqslant s_{iL}-\varphi(\theta_H-c_{iL})\quad\big(IC_i(H)\ i=H,L\big),$$
$$s_{iL}-\varphi(\theta_L-c_{iL})\geqslant s_{iH}-\varphi(\theta_L-c_{iH})\quad\big(IC_i(L)\ i=H,L\big),$$

and the participation constraints are:

$$\big[s_H-\varphi(\theta_H-c_H)\big]+\delta\big[\alpha\big(s_{HH}-\varphi(\theta_H-c_{HH})\big)+(1-\alpha)\big(s_{HL}-\varphi(\theta_L-c_{HL})\big)\big]\geqslant 0,\tag{IR_H}$$

$$\big[s_L-\varphi(\theta_L-c_L)\big]+\delta\big[(1-\alpha)\big(s_{LH}-\varphi(\theta_H-c_{LH})\big)+\alpha\big(s_{LL}-\varphi(\theta_L-c_{LL})\big)\big]\geqslant 0,\tag{IR_L}$$

$$s_{iH}-\varphi(\theta_H-c_{iH})\geqslant 0,\tag{$IR_i(H)$}$$

$$s_{iL}-\varphi(\theta_L-c_{iL})\geqslant 0.\tag{$IR_i(L)$}$$

To complete the argument in Section 3, we only need to show that in the relaxed problem, in which we ignore the incentive constraint of the high-cost type and the participation constraints of the low-cost type, we can assume the incentive constraint of the low-cost type and the participation constraints of the high-cost type to be satisfied with equality without loss of generality; and that the solution of this relaxed problem is a solution of the full problem as well. The strict concavity of the problem guarantees that cost function c_h found with this procedure must be the unique solution of $\mathcal{P}_I$.

It can be easily verified that the incentive compatibility constraints of the low-cost type and the participation constraint of the high-cost type are binding in the first period. It is also simple to see that the participation constraint of the high-cost type must be binding in the second period. Assume that this were not true after a history h and $\pi_{hH}=\kappa_1>0$, then we can uniformly reduce s_{hH} and s_{hL} by κ_1; and contextually increase s_h by $\delta\kappa_1$. The modified contract continues to be a solution of $\mathcal{P}_I$ because it still satisfies the low-cost type's incentive constraints in both periods; and it satisfies the participation constraints of the high-cost type with equality.

Consider now the incentive compatibility constraint of the low-cost type after a history $h=\{H\}$ and assume, by contradiction, that: $s_{HL}-\varphi(\theta_L-c_{HL})=\pi_{HL}+\Phi(c_{HH})+\kappa_2$, with $\kappa_2>0$. Modify the contract in the following way. First reduce s_{HL} by κ_2; contextually increase s_H by

$\delta(1-\alpha)\kappa_2$. After the change, welfare is constant and the incentive compatibility and participation constraints are satisfied in period 2. In period 1 the participation constraint of the high-cost type is also satisfied; consider the incentive constraint. If the low-cost agent mimics the high-cost, the payoff world be:

$$\pi_L - \delta(2\alpha - 1)\kappa_2 < \pi_L.$$

So this constraint remains satisfied too. We can now reduce s_L by $\delta(2\alpha - 1)\kappa_2$: the resulting contract satisfies all the constraints with equality and yields higher welfare. So the initial contract was not optimal, a contradiction. We conclude that $IC_H(L)$ is binding. The case for $IC_L(L)$ is analogous.

We now prove that the solution of the relaxed problem is also a solution of the full problem. To see that the second period incentive compatibility constraints of the high-cost type are satisfied after any history h note that by using the equalities implied by the $IC_h(L)$ and $IR_h(H)$ constraints we have:

$$\begin{aligned} s_{hL} - s_{hH} &= \varphi(\theta_{hL} - c_{hL}) - \varphi(\theta_{hL} - c_{hH}) \\ &= \varphi(\theta_{hH} - c_{hL}) - \varphi(\theta_{hH} - c_{hH}) + \big[\Phi(c_{hH}) - \Phi(c_{hL})\big] \\ &\Rightarrow s_{hH} - \varphi(\theta_{hH} - c_{hH}) \geqslant s_{hL} - \varphi(\theta_{hH} - c_{hL}) \end{aligned}$$

where the last passage follows from the fact that by (4) $c_{hH} \geqslant c_{hL}$ for any h and $\Phi(c)$ is decreasing in c. Similarly, for period one we have:

$$\begin{aligned} s_L - s_H &= \varphi(\theta_H - c_L) - \varphi(\theta_H - c_H) + \delta\big[(1-\alpha)(\pi_{HL} - \pi_{LL})\big] \\ &\quad + \big[\Phi(c_H) - \Phi(c_L)\big] + (2\alpha - 1)(\pi_{HL} - \pi_{LL}) \\ &\Rightarrow s_L - \varphi(\theta_L - c_L) + \delta(1-\alpha)\pi_{LH} \geqslant s_H - \varphi(\theta_L - c_H) \\ &\quad + \delta(1-\alpha)\pi_{HH} \end{aligned}$$

where the last expression follows by the fact that $\Phi(c_H) - \Phi(c_L) \leqslant 0$ (since $c_H \geqslant c_L$ and $\Phi(c)$ is decreasing) and $\pi_{LL} \geqslant \pi_{HL}$ (since $c_{LL} \leqslant c_{HL}$). The fact that the participation constraint of the low-cost type is satisfied follows from (2) and (3). □

B.2. Proof of Lemma 1

Consider the renegotiation constraint ($\mathcal{R}$) after a history h. Total output (i.e. the value of the project net of its cost) is non-increasing with respect to c_{hH} in the interval $[\theta_H - e^*, \infty)$ and the principal objective function (i.e. welfare, output net of the cost λ of leaving a rent to the agent) is non-decreasing with respect to c_{hH} in the interval $(-\infty, \tilde{c}(\mu(h))]$ (respectively, strictly decreasing and strictly increasing in the interior of the interval). Therefore, when $c_{hH} \in [\theta_H - e^*, \tilde{c}(\mu(h))]$, renegotiating the contract in favor of a cost $c > c_{hH}$ would imply lower output, making it impossible to obtain a Pareto improvement. On the contrary, replacing the cost prescribed in the original contract with a new cost $c < c_{hH}$ would imply lower welfare. We therefore conclude that when $c_{hH} \in [\theta_H - e^*, \tilde{c}(\mu(h))]$, c_{hH} cannot be profitably renegotiated. Similarly, we can see that the condition is necessary: when c_{hH} does not belong to $[\theta_H - e^*, \tilde{c}(\mu(\alpha, x))]$, the sign of the derivative of output and welfare is the same, so there is a marginal change in c which would improve output and welfare at the same time. □

B.3. Proofs of Lemmata 2 and 3

Without loss of generality, the optimal contract specifies only two options at $t = 2$. It can be verified that given a posterior and a level of utility promised by the contract ex ante, the optimal renegotiation proof contract in the second period can assume only three possible forms: it can be conditionally optimal given the posterior; it can be a sell-out contract, in which the cost is chosen efficiently in the second period; or it can be "rent constrained."[23] In a rent constrained contract the incentive compatibility constraint is binding; but, in order to guarantee to the low-cost type the utility promised at $t = 1$, the high-cost type is assigned a lower cost level than what it would be ex post optimal.[24] Except for the sell-out contract, the incentive compatibility constraint of the low-cost agent is always binding at $t = 2$. As usual in static principal–agent problems, in all these cases the incentive compatibility constraint of the high-cost types is not binding at $t = 2$. It is also easy to prove that the participation constraint of the high-cost type is binding at $t = 1$. We now prove Lemma 2.

Lemma 2. *Without loss of generality, in the optimal renegotiation proof contract no pair of options i or j are chosen by both types with positive probability.*

Proof. Assume by contradiction that, say, contracts 1 and 2 are chosen by both types with positive probability: conditionally on choosing contract 1 or 2, the low-cost agent chooses contract 2 with probability $x \in (0, 1)$, and the high-cost types chooses contract 1 with probability $y \in (0, 1)$. Without loss of generality we can label the histories so that the cost in the first period after contract 1 is chosen (history $\{1\}$) is smaller than after 2 is chosen (history $\{2\}$): $c_1 \leqslant c_2$. We call the contract after history $\{1\}$ (respectively, $\{2\}$) "the low-cost (respectively, high-cost) contract." For future reference it is also useful to write the objective function with this double randomization as a function of x and y. We will keep the contracts $j \neq 1, 2$ and the probability that they are chosen constant throughout the proof so total cost can be written as

$$(1 - p_H)K_H + (1 - p_L)K_L + C(x, y)$$

[23] Given any posterior, at $t = 2$ the problem of the regulator is a standard static principal–agent problem in which the utilities promised in the renegotiation proof contract to the low-cost type (U_L^*) and to the high-cost type (U_H^*) are the reservation levels at $t = 2$ for the respective types. Without loss of generality, we can assume $U_H^* = 0$, so the key determinant of the contract at time 2 is U_L^*. Let μ be the posterior at $t = 2$ and let $U(\mu)$ be the level of utility left to the low-cost type in the optimal static contract when the posterior is μ, and U^e the minimal rent that needs to be left to the low-cost type to guarantee that an efficient contract in which both types exert e^* is incentive compatible. Depending on U_H^*, four cases are possible in an optimal contract. If $U_H^* \leqslant U(\mu)$, the at $t = 2$ the principal can implement the optimal static contract because the participation constraint is not binding. If $U(\mu) < U_H^* \leqslant U^e$, the participation constraint is binding and we have a rent constrained contract. The last two cases occur when $U_H^* > U^e$. In this case the utility promised to the low-cost type is high enough that even if the effort level of the high-cost type is chosen to be efficient, the incentive compatibility constraint of the low-cost type is not binding. Two cases are possible. If the incentive compatibility constraint of the high-cost type is not binding at the efficient solution in which both types exert e^*: in this case the contract is an efficient sell-out. If the incentive compatibility of the high-cost type is binding, then the low-cost type's effort would be inefficient. This case, however, can be ignored. Indeed it is easy to verify that it would never be optimal to promise such a high U_L^* at $t = 1$: it is possible to uniformly reduce the transfers to both types without affecting the incentive compatibility constraints.

[24] As usual, when the incentive constraint of the efficient firm is binding, her rent is decreasing in the cost chosen by the inefficient firm.

where $p_i > 0\ \forall i = H, L$ is the probability that type i chooses contracts 1 or 2, K_i is the expected cost associated with type i choosing contracts $j \neq 1, 2$, and $C(x, y)$ is the expected welfare cost associated with types H and L choosing contracts 1 or 2:

$$C(x,y) = \left\{ \begin{array}{l} p_L v(1-x)C(\theta_L, c_1, s_1) + p_L vxC(\theta_L, c_2, s_2) \\ \quad + p_H(1-v)(1-y)C(\theta_H, c_2, s_2) + p_H(1-v)yC(\theta_H, c_1, s_1) \\ \quad + \delta \left[\begin{array}{l} p_L v \sum\limits_{j=H,L} \Pr(\theta_j \mid \theta_L)[(1-x)C(\theta_j, c_{1j}, s_{1j}) + xC(\theta_j, c_{2j}, s_{2j})] \\ \quad + p_H(1-v) \sum\limits_{j=H,L} \Pr(\theta_j \mid \theta_H)[(1-y)C(\theta_j, c_{2j}, s_{2j}) + yC(\theta_j, c_{1j}, s_{1j})] \end{array} \right] \end{array} \right\} \tag{13}$$

where $C(\theta, c, s) = (1+\lambda)(c + \varphi(\theta - c)) + \lambda(s - \varphi(\theta - c))$. We proceed in three steps.

Step 1. We first show that if x, y are interior in $(0, 1)$, the incentive compatibility of the low-cost agent must be binding in the second period after any history $i = 1, 2$. To see this assume, for example, that it is not true after history $\{1\}$. In this case we have a sell-out contract after $\{1\}$: since it is efficient, the renegotiation proofness constraint is not binding. So the optimal levels of x, y are determined by maximizing welfare under the constraint that the contract is renegotiation proof after only history $\{2\}$. Assume the posterior after this history is μ_2. We can have two cases. If $c_{2H} \in [\theta_H - e^*, \tilde{c}(\mu_2))$, then, by Lemma 1, renegotiation constraint is not binding after this history as well. Since the objective function is linear in x, y, at least one of them must be a corner solution 0 or 1, a contradiction. If $c_{2H} = \tilde{c}(\mu)$, then, since $\tilde{c}(\mu)$ is increasing is μ, the renegotiation constraint is satisfied after a change in x, y if and only if the posterior is kept larger or equal to μ_2. The solution x, y, therefore, must minimize (13) under the constraint that

$$\alpha v p_L x + (1-\alpha)(1-v)p_H y \geqslant \mu_2\big[v p_L x + (1-v)p_H y\big]. \tag{14}$$

But this is still a linear problem with a linear constraint, so it implies a corner solution 0 or 1, a contradiction.

Step 2. We now show that if x, y are interior, then conditional on a agent choosing contract $\{1\}$, the probability that the agent is low-cost (γ_1) at $t = 1$ is higher than the probability conditional on choosing contract $\{2\}$ (γ_2). For this purpose, it is useful to re-write the agent's profits in terms of the cost levels. If y is interior, the high-cost type must be indifferent between the two contracts; therefore the profit if she chooses contract 2, $\pi_H(2)$, must be equal to the profits if she chooses contract 1, $\pi_H(1)$:

$$\begin{aligned} \pi_H(2) = \pi_H(1) &= s_1 - \varphi(\theta_H - c_1) + \delta \left[\begin{array}{l} \alpha(s_{1H} - \varphi(\theta_H - c_{1H})) \\ \quad + (1-\alpha)(s_{1L} - \varphi(\theta_L - c_{1L})) \end{array} \right] \\ &= \pi_L(1) - \big[\Phi(c_1) + \delta(2\alpha - 1)\Phi(c_{1H})\big] \end{aligned}$$

where the last equality follows from the fact that by Step 1, the incentive compatibility of the low-cost type is satisfied as equality at $t = 2$ and therefore $s_{1L} - \varphi(\theta_L - c_{1L}) = \Phi(c_{1H})$; and since the participation constraint the high-cost type is binding at $t = 2$, $s_{1H} - \varphi(\theta_H - c_{1H}) = 0$. But since the participation constraint of the high-cost agent is binding at $t = 1$, we have $\pi_H(2) = 0$ and:

$$\pi_L(1) = \big[\Phi(c_1) + \delta(2\alpha - 1)\Phi(c_{1H})\big].$$

A similar argument yields:

$$\pi_L(2) = \big[\Phi(c_2) + \delta(2\alpha - 1)\Phi(c_{2H})\big].$$

So by the indifference condition of the low-cost agent, x interior implies:

$$\big[\Phi(c_1) + \delta(2\alpha - 1)\Phi(c_{1H})\big] = \big[\Phi(c_2) + \delta(2\alpha - 1)\Phi(c_{2H})\big]. \tag{15}$$

The indifference condition of the high-cost agent is automatically guaranteed by an appropriate choice of the monetary transfers: since it is independent of the choice of the costs, we can ignore it. We now consider two cases.

Case 2.1. Assume that $c_{2H} < \tilde{c}(\mu_2)$ and $c_{1H} < \tilde{c}(\mu_1)$.[25] It is easy to verify that in the optimal solution it cannot be that $c_{iH} < \theta_H - e^* \ \forall i \in 1, 2$: by Lemma 1, we conclude that the renegotiation proofness constraint is not binding after histories 1 and 2. The optimal contract after these histories, therefore, is a solution of the minimization of $C(x, y)$ without the renegotiation proofness constraint. However this problem is linear in x and y, so they cannot be interior as assumed, a contradiction.

Case 2.2. We can therefore restrict the analysis to only two cases: $c_{2H} = \tilde{c}(\mu_2)$ and $c_{1H} \leqslant \tilde{c}(\mu_1)$ (Case 2.2.1) or $c_{2H} < \tilde{c}(\mu_2)$ and $c_{1H} = \tilde{c}(\mu_1)$ (Case 2.2.2).

Case 2.2.1. Note that:

$$\gamma_i = \frac{\mu_i - (1 - \alpha)}{2\alpha - 1} \quad \forall i = 1, 2$$

where μ_i is the posterior probability at $t = 2$ that the agent is low-cost after history $i = 1, 2$. Since $\alpha \geqslant \frac{1}{2}$, $\gamma_1 \geqslant \gamma_2$ if and only if $\mu_1 \geqslant \mu_2$. If $x \in (0, 1)$, the indifference condition for the low-cost agent implies:

$$\big[\Phi(c_{1H}) - \Phi(c_{2H})\big] = -\frac{\Phi(c_1) - \Phi(c_2)}{\delta(2\alpha - 1)} \leqslant 0 \tag{16}$$

so $c_{1H} \geqslant c_{2H}$ since $\Phi(\cdot)$ is decreasing. But then, since $\tilde{c}(\cdot)$ is increasing, we have that $\tilde{c}(\mu_2) = c_{2H} \leqslant c_{1H} \leqslant \tilde{c}(\mu_1)$ implies $\mu_1 \geqslant \mu_2$, and $\gamma_1 \geqslant \gamma_2$.

Case 2.2.2. Let us define $c_{ND}(\gamma)$ the optimal cost that the principal would choose if she knew that the probability that a agents is low-cost is γ and she could not discriminate between the two types (ND stands for no-discrimination). Assume that $c_1 < c_{ND}(\gamma_1)$. Consider a marginal increase in c_1 and a reduction in c_{1H} such that the incentives for the low-cost agent to mix are preserved (condition (16)). This change does not violate the renegotiation constraint. Since the principal's objective function is concave in c_1, the change in c_1 increase welfare. However since c_{1H} is conditionally optimal, its change has only a second order effect. It follows that the total change increases welfare, a contradiction. It must be that $c_1 \geqslant c_{ND}(\gamma_1)$. Assume now that $c_2 \geqslant c_{ND}(\gamma_2)$. Consider a marginal increase in c_{2H} and a reduction in c_2 so that the indifference condition for the high-cost agent is preserved. This change does not violate the renegotiation constraint since $c_{2H} < \tilde{c}(\mu_2)$. By concavity, both changes weakly increase welfare and at least one strictly increases it, a contradiction: therefore $c_2 < c_{ND}(\gamma_2)$. We conclude that $c_{ND}(\gamma_1) \leqslant c_1 \leqslant c_2 < c_{ND}(\gamma_2)$, which implies $\gamma_1 \geqslant \gamma_2$.

Step 3. It can be verified that $c_2 \geqslant c_{ND}(\gamma_2)$ if and only if $c_1 \leqslant c_{ND}(\gamma_1)$. Indeed if this were not the case, it would be possible to simultaneously increase (or decrease) c_1 and c_2 keeping all incentives constraints satisfied in the first period. This would imply that after any history the

[25] As proven in Proposition 2, this case occurs if $\alpha < \alpha_1$. See Section 5.1 for details.

cost would be nearer to the first best and therefore welfare (which is concave) would be higher in the first period and nothing would change in the second period, a contradiction. Therefore we can focus on two possible cases: $c_{ND}(\gamma_1) \leqslant c_1 \leqslant c_2 \leqslant c_{ND}(\gamma_2)$ (Case 3.1) and $c_1 < c_{ND}(\gamma_1) \leqslant c_{ND}(\gamma_2) < c_2$ (Case 3.2). We claim that if both x and y are interior, we can reduce at least one of them without loss of generality. Note that, compared to the case with constant types, here the analysis is complicated by the fact that the sign of the marginal effect of x on the cost function is ambiguous. With constant types a marginal effect of x affects only the first period: now it affects both. Given that $c_1 \leqslant c_2$ implies that $c_{1H} \geqslant c_{2H}$ (and therefore if contract 1 is more efficient in the first period than it must be less efficient in the second) shifting types from branch {2} to branch {1} may actually reduce welfare.

Case 3.1. Differentiating (13) we have ("$a \propto b$" stands for "a it is proportional to b"):

$$\frac{\partial C(x,y)}{\partial y} \propto \left\{ \begin{array}{l} \left[C(\theta_H, c_1, s_1) - C(\theta_H, c_2, s_2)\right] \\ \quad + \delta \sum_{j=H,L} \Pr(\theta_j \mid \theta_H)\left[C(\theta_j, c_{1j}, s_{1j}) - C(\theta_j, c_{2j}, s_{2j})\right] \end{array} \right\}.$$

Since the expected discounted profit of the high-cost agent is the same in the two contracts:

$$\frac{\partial C(x,y)}{\partial y} \propto \left\{ \begin{array}{l} \left[c_1 + \varphi(\theta_H - c_1) - c_2 - \varphi(\theta_H - c_2)\right] \\ \quad + \delta \sum_{j=H,L} \Pr(\theta_j \mid \theta_H)\left[c_{1j} + \varphi(\theta_j - c_{1j}) - c_{2j} - \varphi(\theta_j - c_{2j})\right] \end{array} \right\}. \tag{17}$$

We have that:

$$\left[c_1 + \varphi(\theta_H - c_1) - c_2 - \varphi(\theta_H - c_2)\right] \geqslant 0, \tag{18}$$

since $c_1 \leqslant c_2 \leqslant c_{ND}(\gamma_2) \leqslant \theta_H - e^*$. Similarly since $c_{1H} \geqslant c_{2H} \geqslant \theta_H - e^*$ (see Case 2.2.1) and $c_{1L} = c_{2L} = \theta_L - e^*$, we have that the second term of (17) is also non-negative. Note moreover that by (16) if $c_1 = c_2$ then $c_{1H} = c_{2H} \geqslant \theta_H - e^*$: since $c_{1L} = c_{2L} = \theta_L - e^*$, then the two contracts are identical and can be merged without loss. So we can assume without loss of generality that $c_1 < c_2$, and (18) is strict. Therefore $\frac{\partial C(x,y)}{\partial y} > 0$.

We can now consider two cases.

Case 3.1.1. Assume first that $\frac{\partial C(x,y)}{\partial x} \leqslant 0$. Consider a marginal increase in x and a contextual marginal decrease in y such that the posterior probability after history {1} is unchanged. The changes in x, y must induce a strict reduction in total welfare cost. If $c_{2H} \in [\theta_H - e^*, \tilde{c}(\mu))$ or if the posterior after history {2} is increased, then, by Lemma 1, a marginal change in x, y would not affect the renegotiation constraint: so the contract would remain renegotiation proof. If $c_{2H} = \tilde{c}(\mu)$ and the posterior decreases after the change, then we marginally reduce c_{2H} so that the contract remains renegotiation proof after history {2} too. In order to preserve the indifference condition at $t = 1$ between the two contracts for the high-cost agent, we can marginally increase c_2. Since c_{2H} is conditionally optimal, its change would have only a second order effect on welfare. Since, on the contrary $c_2 \leqslant c_{ND}(\gamma_2)$, we would (weakly) increase in welfare at $t = 1$. The total effect, therefore, strictly increases welfare, a contradiction.

Case 3.1.2. Assume now $\frac{\partial C(x,y)}{\partial x} > 0$. Consider a marginal decrease in x. By Lemma 1 the contract remains renegotiation proof after history {1}. If $c_{2H} \in [\theta_H - e^*, \tilde{c}(\mu))$, then, by Lemma 1, a marginal change in x, y would not affect the renegotiation constraint: so the contract would remain renegotiation proof. If $c_{2H} = \tilde{c}(\mu)$, then we marginally reduce c_{2H} so that the contract remains renegotiation proof after history {2} too. In order to preserve the indifference condition at $t = 1$ between the two contracts for the high-cost agent, we can marginally increase c_2. Since c_{2H} is conditionally optimal, its change would have only a second order effect on welfare. Since $c_2 \leqslant c_{ND}(\gamma_2) < \theta_L - e^*$, the change in c_2 would weakly increase welfare at $t = 1$.

Case 3.2. Assume that $c_{1H} = \tilde{c}(\mu_1)$, where μ_1 is the posterior probability of a low-cost agent at $t = 2$ after history $\{1\}$. We can marginally reduce c_{1H} and simultaneously marginally increase c_1 so that the indifference condition between the two contracts is preserved for the low-cost agent. The change in c_{1H} has zero marginal effect, since $\tilde{c}(\mu_1)$ is conditionally optimal; on the contrary the increase in c_1 strictly increases welfare since $c_1 < c_{ND}(\gamma_1)$. So welfare strictly increases, a contradiction: it must be that $c_{1H} < \tilde{c}(\mu_1)$. The principal's solution for x, y, therefore minimizes (13), which is linear in x, y under the linear constraint (14), which guarantees that the posterior in the second period after history $\{2\}$ is at least as large as μ_2. This program cannot have both x and y interior, so a double randomization cannot be optimal in this case either. □

We now prove Lemma 3.

Lemma 3. *Without loss of generality, in the optimal renegotiation proof contract there is no pair of options i and j such that with positive probability one is chosen by both types and the other only by the high-cost type.*

Proof. Assume by contradiction that, conditional on choosing contracts 1 or 2, the low-cost type chooses contract 1 with probability $1 - x = 1$ and the other type chooses contract 1 with probability $y \in (0, 1)$ and contract 2 with probability $1 - y$. We consider two cases.

Case 1. Assume first that the low-cost agent strictly prefers contract 1 to contract 2.

Step 1.1. Contract 2 must be low-cost: i.e. both in period 1 and in period 2 after any possible history it prescribes an efficient cost level. Clearly the contract is efficient if at time 2 the agent's type is L. Assume it is not efficient at time 2 if the agent's type is H. In this case effort must be distorted below the efficient level, otherwise it is easy to verify that an increase in e would unambiguously increase welfare and respect all the constraints. Consider therefore a new contract in which we marginally reduce cost after history $\{2, H\}$: $c^*_{2,H} = c_{2,H} - \varepsilon$. This (weakly) increases the rent that the low-cost type receives at $t = 2$ after contract 2. Let us reduce the monetary payment at $t = 1$ if contract 2 is chosen so that the high-cost agent's profit is constant. This new contract is renegotiation proof, and does not affect incentives at $t = 1$ (since the low-cost agent's incentive compatibility was not binding), and it is more efficient than the original: so welfare must be higher, a contradiction. The fact that the contract is efficient at $t = 1$ is proven in a similar way.

Step 1.2. If we marginally reduce y, contract 1 remains renegotiation proof by Lemma 1 because the posterior increases after history $\{1\}$; and contract 2 is efficient, so renegotiation proof. The agent's rents are unchanged after all histories and welfare is higher, since an efficient contract is chosen with higher probability. So the original contract could not be optimal, a contradiction.

Case 2. Assume now that the low-cost agent does not randomize, but she is also indifferent between the two contracts. Assume that $C(0, y)$ is non-decreasing in y. Then if we marginally reduce y we do not reduce welfare. Contract 1 remains renegotiation proof since the posterior that the agent is low-cost is increasing. Contract 2 is chosen only by high-cost agents, so the posterior that the agent is low-cost at $t = 2$ is equal to $1 - \alpha$, independent of y: therefore the change does not affect the renegotiation proofness constraint either. Since $C(0, y)$ is linear in y, it must be globally non-decreasing in y, so we can reduce y to zero without reducing welfare and respecting the renegotiation proofness constraint. Assume therefore that $C(0, y)$ is strictly decreasing at y. If we marginally increase y, contract 2 remains renegotiation proof. If $c_{1H} \in [\theta_H - e^*, \tilde{c}(\mu_1))$, then a marginal change in y would not affect the renegotiation constraint: so the change strictly increases welfare without violating the renegotiation proofness constraint, a contradiction. If $c_{1H} = \tilde{c}(\mu_1)$, then we marginally reduce c_{1H} so that contract 1 remains renegotiation proof too. This change has only a second order effect on welfare, while the decrease in y strictly increases welfare. Therefore the new contract is strictly better after the increase, again a contradiction. □

B.4. *Proof of Proposition 2*

Given Lemma 1 we only need to show that the cost function implied in the ex ante optimal contract is in $[\theta_H - e^*, \tilde{c}(\mu(\alpha, x))]$. If $h = L$, then in the ex ante optimal contract $c^{\mathcal{C}}_{LH} = \theta_H - e^*$ and the constraint is satisfied. After $h = H$, since $c^{\mathcal{C}}_{HH}$ is always larger than $\theta_H - e^*$, the ex ante contract is renegotiation proof if $c^{\mathcal{C}}_{HH} \leqslant \tilde{c}(\mu(\alpha, x))$, which is true if and only if $\frac{2\alpha-1}{\alpha}\Gamma_0 \leqslant \Gamma_H$, which can be written as the condition in the statement of the proposition. □

B.5. *Proof of Lemma 4*

We first present the "If" part of statement, proving that any solution $c(h)$ of P_{III} solves the general problem. Then we present the "only if" part, that any solution $c(h)$ of $\mathcal{P}_{II}$ solves P_{III}.

"**If**" We first prove that the value of program P_{II} in which constraints IC_H, $IC_i(H)$ and $IR_i(L)$ for $i = H, L$ are ignored is identical to the value of program P_{III}: i.e. if there is a solution $\{c(h), s(h)\}$ of the first problem, then there exists a solution $\{c(h), s'(h)\}$ of the second problem which yields the same value to the principal. Then we show that the solution of P_{III} satisfies all the constraints of P_{II}. These two steps imply that the solution of P_{III} solves P_{II}.

It is easy to verify that in the relaxed problem IC_L, IR_H, $IR_H(H)$ and $IR_L(H)$ are satisfied as equalities following the same steps as in Proposition 1. Therefore we focus on $IC_i(L)$ for $i = H, L$. Consider first $IC_H(L)$. Assume that this constraint is not satisfied as an equality. As in Proposition 1, we now show that we can modify the contract respecting all the constraints and increase (at least weakly) welfare. Differently from Proposition 1, however, we need to respect the incentive compatibility constraint in the first period which guarantees that the low-cost agent is, if necessary, willing to mix between the two options. Modify the contract as follows. Reduce s_{HL} by $\varepsilon > 0$ and increase s_H by $\delta(1-\alpha)\varepsilon$. Let us also reduce s_L by $\delta(2\alpha - 1)\varepsilon$. All the other terms of the contract remain constant, moreover the probability x that contract $\{L\}$ is chosen in the first period by the low-cost agent is also unchanged. This modification of the initial contract does not affect any incentive or participation constraint in the first or second period. In the first period, moreover, the low-cost agent remains indifferent (if it was before the change) between the two contracts and therefore is still willing to mix with probability x. Since $\mu(\alpha, x)$ and c_{HH} are unchanged, the renegotiation constraint (6) continues to be satisfied as well. However in the new contract welfare is higher because the present value of expected payments is reduced by $\delta v(2\alpha - 1)\varepsilon > 0$. The case of $IC_L(L)$ is analogous, however now the new contract with constraints satisfied as equalities is only weakly superior for the principal.

Given that IC_L and $IC_H(L)$ are binding, in $\mathcal{P}_{III}$ the low-cost agent receives a rent:

$$\Phi(\theta_H - c_H) + (2\alpha - 1)\Phi(\theta_H - c_{HH}).$$

And given that the participation constraints are binding the high-cost agent receives zero. It follows that the optimal solution of $\mathcal{P}_{III}$ is characterized by (7). In the proof of Proposition 1, where we showed that the solution of the relaxed problem with constraints satisfied as equalities solves all the constraints of $\mathcal{P}_I$, we only used the fact that constraints of this problem are satisfied as equalities and that $c_L \leqslant c_H$ and $c_{LL} \leqslant c_{HL}$, both of these two properties are still satisfied in (7). Therefore all the steps in Proposition 1 can be replicated to prove that the solution of the relaxed problem with constraints satisfied as equalities solves all the constraints of $\mathcal{P}_{II}$.

"**Only if**" By the "If" part, the value of program $\mathcal{P}_{II}$ must be identical to the value of $\mathcal{P}_{III}$, and the value of $\mathcal{P}_{III}$ is identical to the value of $\mathcal{P}_{II}$ in which only the constraints IC_L, $IC_i(L)$ and $IR_i(H)$ for $i = H, L$ and (6) are considered. This implies that any solution $\{c(h), s(h)\}$ of $\mathcal{P}_{II}$ is also a solution of this relaxed problem. But then we know that there must be a solution $\{c(h), s'(h)\}$ which solves $\mathcal{P}_{III}$.

B.6. Proof of Lemma 5

Assume that $c_{HH} < \tilde{c}(\mu(\alpha, x))$. Then, since $x \in [0, 1]$, program (7) can be represented by the Kuhn–Tucker Lagrangian:

$$\mathcal{L} = A(x, c_H) + \delta B(x, c_{HH}) + \xi\big[c_{HH} - (\theta_H - e^*)\big]$$

where $\xi \leqslant 0$ is the Kuhn–Tucker multiplier of the constraint $c_{HH} \geqslant (\theta_H - e^*)$ and $x \in [0, 1]$. The optimal solution must satisfy the first order condition for c_{HH}: $\delta \frac{\partial}{\partial c_{HH}} B(x, c_{HH}) + \xi \geqslant 0$, which implies

$$\frac{\partial}{\partial c_{HH}} B(x, c_{HH}) \geqslant 0. \tag{19}$$

We must also satisfy the Kuhn–Tucker condition for x: $x \cdot \frac{\partial}{\partial x}[A(x, c_H) + \delta B(x, c_{HH})] = 0$, which implies that $x = 0$ since

$$\begin{aligned}\frac{\partial}{\partial x}\big[A(x, c_H) + \delta B(x, c_{HH})\big] &= v\big[\big(\varphi(\theta_L - c_H) + c_H\big) - \big(\varphi(e^*) + \theta_L - e^*\big)\big] \\ &\quad + \delta v(1-\alpha)\big[\big(\varphi(\theta_H - c_{HH}) + c_{HH}\big) - \big(\varphi(e^*) + \theta_H - e^*\big)\big] > 0.\end{aligned}$$

As an immediate consequence, we have that $\frac{\mu(\alpha,x)}{1-\mu(\alpha,x)} = \frac{(1-\alpha)}{\alpha}$. The threshold α_1 is defined by $\frac{(1-\alpha_1)}{\alpha_1} = \Gamma_0(\frac{2\alpha_1-1}{\alpha_1})$; therefore $\alpha > \alpha_1$ implies $\frac{\mu(\alpha,x)}{1-\mu(\alpha,x)} < \Gamma_0(\frac{2\alpha-1}{\alpha})$, and

$$\begin{aligned}\tilde{c}\big(\mu(\alpha, x)\big) &= e^* - \frac{\lambda}{1+\lambda}\frac{\mu(\alpha, x)}{1-\mu(\alpha, x)}\Delta\theta \\ &> e^* - \Gamma_0\left(\frac{2\alpha-1}{\alpha}\right)\Delta\theta = c^C_{HH}\end{aligned}$$

that is: $\tilde{c}(\mu(\alpha, x))$ is larger than the ex ante optimal solution of the unconstrained problem c^C_{HH}. Since $B(x, c_{HH})$ is strictly convex in c_{HH}, this implies that $\frac{\partial}{\partial c_{HH}} B(x, c_{HH}) < 0$: but this contradicts (19). We therefore conclude that $c_{HH} = \tilde{c}(\mu(\alpha, x))$. □

B.7. Proof of Proposition 3

Consider the derivative of $A(x, c_H) + \delta B(x, \tilde{c}(\mu(\alpha, x)))$ evaluated at $\alpha = \alpha_1$. Regarding the first term, we have $\frac{\partial}{\partial x} A(x, c_H) = v[(\varphi(\theta_L - c_H) + c_H) - (\varphi(e^*) + \theta_L - e^*)] > 0$, and for the second term:

$$\frac{\partial}{\partial x} B\big(x, \tilde{c}\big(\mu(\alpha, x)\big)\big) \tag{20}$$

$$= \left[\frac{\partial}{\partial \tilde{c}(\mu(\alpha, x))} B\big(x, \tilde{c}\big(\mu(\alpha, x)\big)\big) \frac{\partial}{\partial x} \tilde{c}\big(\mu(\alpha, x)\big) + \frac{\partial}{\partial x} B\big(x, \tilde{c}\big(\mu(\alpha, x)\big)\big) \right] \tag{21}$$

$$= (1+\lambda) \left\{ \begin{array}{l} [\alpha(1-v) + xv(1-\alpha)] \\ \quad \times \left[\begin{array}{l} (1 - \varphi'(\theta_H - \tilde{c}(\mu(\alpha, x)))) \\ \quad + \frac{\lambda}{1+\lambda} \Gamma_0(\frac{2\alpha-1}{\alpha(1-v)+xv(1-\alpha)}) \Phi'(\tilde{c}(\mu(\alpha, x))) \end{array} \right] \cdot \frac{\partial}{\partial x} \tilde{c}\big(\mu(\alpha, x)\big) \\ \quad + v(1-\alpha)\big[\big(\varphi(\theta_H - \tilde{c}\big(\mu(\alpha, x)\big)\big) + \tilde{c}\big(\mu(\alpha, x)\big)\big) - \big(\varphi(e^*) + \theta_H - e^*\big)\big] \end{array} \right\}$$

where $\varphi'(\cdot)$ and $\Phi'(\cdot)$ are the first order derivatives. Lets now evaluate this expression at $\alpha = \alpha_1$. By definition of α_1, $\frac{(1-\alpha_1)}{\alpha_1} = \Gamma_0(\frac{2\alpha_1 - 1}{\alpha_1})$; therefore:

$$\frac{\partial}{\partial x} B\big(x, \tilde{c}\big(\mu(\alpha_1, x)\big)\big) = (1+\lambda)$$

$$\times \left\{ \begin{array}{l} \big[\alpha(1-v) + xv(1-\alpha)\big] \\ \quad \times \left[\begin{array}{l} \cdot 1 - \varphi'(\theta_H - \tilde{c}(\mu(\alpha, x))) \\ \quad + \frac{\lambda(1-\alpha)}{\alpha(1+\lambda)} \frac{\alpha}{\alpha(1-v)+xv(1-\alpha)} \Phi'(\tilde{c}(\mu(\alpha, x))) \end{array} \right] \frac{\partial}{\partial x} \tilde{c}\big(\mu(\alpha, x)\big) \\ \quad + v(1-\alpha)\big[\big(\varphi(\theta_H - \tilde{c}\big(\mu(\alpha, x)\big)\big) + \tilde{c}\big(\mu\big(\alpha, x)\big)\big) - \big(\varphi(e^*) + \theta_H - e^*\big)\big] \end{array} \right\} \Bigg|_{\alpha = \alpha_1}.$$

But we have that

$$1 - \varphi'\big(\theta_H - \tilde{c}\big(\mu(\alpha, x)\big)\big) + \frac{\lambda(1-\alpha)}{\alpha(1+\lambda)} \frac{\alpha}{\alpha(1-v) + xv(1-\alpha)} \Phi'\big(\tilde{c}\big(\mu(\alpha, x)\big)\big)$$

$$\geqslant 1 - \varphi'\big(\theta_H - \tilde{c}\big(\mu(\alpha, x)\big)\big) + \frac{(1-\alpha)}{\alpha} \frac{\lambda}{(1+\lambda)} \Phi'\big(\tilde{c}\big(\mu(\alpha, x)\big)\big) = 0, \quad \forall x \in [0, 1].$$

where the first inequality follows from the fact that $\frac{\alpha}{\alpha(1-v)+xv(1-\alpha)} < 1$ and $\Phi'(\tilde{c}(\mu(\alpha, x))) < 0$; and the second equality follows from the fact that, by the definition of $\tilde{c}(\mu)$, the first order condition of the second-period optimal contract given that the posterior is Γ_H is zero at $\tilde{c}(\mu(\alpha, x))$. It follows that $\frac{\partial}{\partial x}[A(x, c_H) + \delta B(\tilde{c}(\mu(\alpha, x)))] > 0$ for any $x \in [0, 1]$ at $\alpha = \alpha_1$. Since the objective function is continuous with continuous derivative, there must be a

$$\alpha_2 \equiv \inf\left\{ \alpha \in (\alpha_1, 1] \;\Big|\; \frac{\partial}{\partial x}\big[A(0, c_H) + \delta B\big(0, \tilde{c}\big(\mu(\alpha, 0)\big)\big)\big] < 0 \right\}$$

as required such that $x = 0$ is a corner solution in $(\alpha_1, \alpha_2]$.

The optimal cost levels are derived from the first order necessary and sufficient conditions. From these levels, the optimal effort function (9) follows immediately since $c_t = \theta^t - e_t$. □

B.8. Proof of Proposition 4

Since $\alpha > \alpha_1$, $\frac{\partial}{\partial \tilde{c}(\mu(\alpha,x))} B(x, \tilde{c}(\mu(\alpha, x))) < 0$ at $x = 0$. As it can be seen from, (25) as $\alpha \to 1$, $\frac{\partial}{\partial x} B(x, \tilde{c}(\mu(\alpha, x))) < 0$ at $x = 0$. It follows that there is an $\hat{\alpha} < 1$ and a δ^* such that $\alpha > \hat{\alpha}$ and $\delta > \delta^*$ implies $\frac{\partial}{\partial x}[A(0, c_H) + \delta B(\tilde{c}(0))] < 0$, and the optimal x must be positive.

B.9. Proof of Proposition 5

It is convenient to represent the marginal effect of x on second period costs as:

$$B_x(x,\alpha)=\frac{\partial}{\partial x}B\big(x,\tilde{c}\big(\mu(\alpha,x)\big)\big)=D(x,\alpha)E(x,\alpha)+F(x,\alpha) \tag{22a}$$

where (see Eq. (25)):

$$D(x,\alpha)=\frac{\eta\Delta\theta^2\lambda^2}{1+\lambda}\big[(1-\alpha)-(1-x)v\alpha\big], \tag{23}$$

$$E(x,\alpha)=\frac{v(1-v)(2\alpha-1)}{[vx(1-\alpha)+\alpha(1-v)]^2},$$

$$F(x,\alpha)=v(1+\lambda)(1-\alpha)\begin{bmatrix}(\varphi(\theta_H-\tilde{c}(\mu(\alpha,x)))+\tilde{c}(\mu(\alpha,x)))\\ -(\varphi(e^*)+\theta_H-e^*)\end{bmatrix}$$

$$=\frac{\eta v(1+\lambda)(1-\alpha)}{2}\left(\frac{vx\alpha+(1-v)(1-\alpha)}{vx(1-\alpha)+\alpha(1-v)}\frac{\lambda}{1+\lambda}\Delta\theta\right)^2. \tag{24}$$

Assume now that there are (at least) two solution x_1 and x_2. Since $\frac{\partial}{\partial x}A(x,c_H)$ is a positive constant independent of x, by the first order condition

$$\frac{\partial}{\partial x}\big[A(x_i,c_H)+\delta B\big(x_i,\tilde{c}\big(\mu(\alpha,x_i)\big)\big)\big]=0\quad \forall i=1,2$$

we have that $B_x(x_i,\alpha)<0$ for $i=1,2$. Moreover by the second order condition $B_{xx}(x_i,\alpha)\geqslant 0$ $\forall i=1,2$. By continuity, therefore, there exists a $\bar{x}\in(x_1,x_2)$ such that $B_x(\bar{x},\alpha)<0$ and $B_{xx}(\bar{x},\alpha)\leqslant 0$ (see for example Fig. 3. Case A, note that $\bar{x}$ may even coincide with x_2). However

$$B_x(\bar{x},\alpha)=D(\bar{x},\alpha)E(\bar{x},\alpha)+F(\bar{x},\alpha)<0\quad\Rightarrow\quad D(\bar{x},\alpha)<0$$

since both $E(\bar{x},\alpha)$ and $F(x,\alpha)$ are strictly positive. But then:

$$B_{xx}(\bar{x},\alpha)=D_x(\bar{x},\alpha)E(\bar{x},\alpha)+D(\bar{x},\alpha)E_x(\bar{x},\alpha)+F_x(x,\alpha)>0$$

since $E_x(x,\alpha)$ is obviously negative, $D_x(x,\alpha)>0$ and

$$F_x(x,\alpha)=\frac{\partial}{\partial\tilde{c}}\big[\varphi\big(\theta_H-\tilde{c}\big(\mu(\alpha,x)\big)\big)+\tilde{c}\big(\mu(\alpha,x)\big)\big]\frac{\partial}{\partial x}\tilde{c}\big(\mu(\alpha,x_i)\big)>0$$

since $\tilde{c}(\mu(\alpha,x))>\theta_H-e^*$, φ is convex and $\frac{\partial}{\partial x}\tilde{c}(\mu(\alpha,x_i))>0$. But this is a contradiction: we conclude that there can be only a unique solution. □

B.10. Proof of Proposition 6

To understand how the "degree" of pooling is affected by a change in α, we need to study the impact of α on the first order condition

$$\frac{\partial}{\partial x}\big[A\big(x^*(\alpha,\delta),c_H(\alpha,\delta)\big)+\delta B\big(x^*(\alpha,\delta),\tilde{c}\big(\mu\big(\alpha,x^*(\alpha,\delta)\big)\big)\big)\big]=0.$$

The first term $\frac{\partial}{\partial x\partial\alpha}A(x,c_H)$ depends on α only indirectly, through $c_H(\alpha,\delta)$, and $x^*(\alpha,\delta)$: but because of the Envelope Theorem, this is irrelevant for the impact of α on the first order condition. The key variable, therefore, is the impact of correlation on $\frac{\partial}{\partial x}B(x,\mu(\alpha,x),\alpha)$:[26]

[26] In the expression below we make explicit that α also affects $B(x,\mu(\alpha,x))$ directly by writing $B(x,\mu(\alpha,x),\alpha)$.

$$\frac{\partial}{\partial x}B\big(x,\mu(\alpha,x),\alpha\big)=\frac{\eta\Delta\theta^2\lambda^2}{1+\lambda}\big[(1-\alpha)-(1-x)\alpha v\big] \times\frac{v(1-v)(2\alpha-1)}{[vx(1-\alpha)+\alpha(1-v)]^2} +\frac{\eta v(1+\lambda)(1-\alpha)}{2}\left(\frac{vx\alpha+(1-v)(1-\alpha)}{vx(1-\alpha)+\alpha(1-v)}\frac{\lambda}{1+\lambda}\Delta\theta\right)^2. \tag{25}$$

This expression is generally not monotonic in α. It is however possible prove the monotonicity of the solution by exploiting the equilibrium conditions. If $B_x(x,\alpha)>0$, then we have a corner solution with $x^*(\alpha)=0$, which is obviously non decreasing in α. Therefore it is sufficient to prove that $B_x(x,\alpha)\leqslant 0$ implies $B_x(x,\alpha)=\frac{\partial}{\partial x}B(x,\tilde{c}(\mu(\alpha,x)))$ is monotonically non-increasing in α. It is useful to distinguish two cases.

Case 1. $vx\leqslant 1-v$. As in Proposition 7 we can represent the marginal effect of x on second period costs as:

$$B_x(x,\alpha)=\frac{\partial}{\partial x}B\big(x,\tilde{c}\big(\mu(\alpha,x)\big)\big)=D(x,\alpha)E(x,\alpha)+F(x,\alpha) \tag{26a}$$

which are defined in (23). Therefore we have that

$$B_{x\alpha}(x,\alpha)=\frac{\partial^2}{\partial x\partial\alpha}B\big(x,\tilde{c}\big(\mu(\alpha,x)\big)\big)=D(x,\alpha)E_\alpha(x,\alpha)+D_\alpha(x,\alpha)E(x,\alpha)+F_\alpha(x,\alpha). \tag{27a}$$

Observe that $E_\alpha(x,\alpha)\propto vx\alpha+(1-\alpha)(1-v)>0$. We have

$$F_\alpha(x,\alpha)\propto-\frac{vx\alpha+(1-v)(1-\alpha)}{(vx(1-\alpha)+\alpha(1-v))^2}\left\{\begin{array}{l}\big[vx\alpha+(1-v)(1-\alpha)\big]\\ \quad+2\frac{(1-\alpha)[(1-v)^2-(vx)^2]}{vx(1-\alpha)+\alpha(1-v)}\end{array}\right\} \tag{28}$$

$$\propto-\left\{\begin{array}{l}\big[vx\alpha+(1-v)(1-\alpha)\big]\big[vx(1-\alpha)+\alpha(1-v)\big]\\ \quad+2(1-\alpha)\big[(1-v)^2-(vx)^2\big]\end{array}\right\}.$$

and therefore it is negative. The first term in the parenthesis is always positive, therefore the result follows from the fact that if $vx<1-v$, then the last term is positive too. Since $D(x,\alpha)$ and $D_\alpha(x,\alpha)$ are negative, it follows that $B_{x\alpha}(x,\alpha)<0$.

Case 2. $vx>1-v$. This time it is convenient to decompose $B_x(x,\alpha)$ as (see (25)):

$$B_x(x,\alpha)=\frac{\eta\Delta\theta^2\lambda^2}{1+\lambda}vG(\alpha)\cdot H(\alpha)$$

where:

$$G(\alpha)=\big[vx(1-\alpha)+\alpha(1-v)\big]^{-2};\quad\text{and}$$
$$H(\alpha)=\big(1-\alpha-(1-x)\alpha v\big)(1-v)(2\alpha-1)+\frac{1-\alpha}{2}\big[xv\alpha+(1-v)(1-\alpha)\big]^2.$$

First note that if $H(\alpha)>0$ then $B_x(x,\alpha)>0$, and we have a corner solution in which $x(\alpha,\delta)=0$, so $\frac{\partial x(\alpha,\delta)}{\partial\alpha}=0$. Assume therefore that $H(\alpha)<0$. We can write the second order derivative of $H(\alpha)$ as:

$$H_{\alpha\alpha}(\alpha) = -4(1-v)\big[1+(1-x)v\big] - \big[xv\alpha + (1-v)(1-\alpha)\big]\big[xv-(1-v)\big] \\ - \big[xv-(1-v)\big]\big[xv(2\alpha-1)+2(1-\alpha)(1-v)\big].$$

which is clearly non-positive given $vx > 1-v$. Observe that

$$H\left(\frac{1}{2}\right) = \frac{1}{4}\big[xv+(1-v)\big]^2 > 0.$$

If $H(\alpha) > 0$ for any $\alpha \in [\frac{1}{2}, 1]$, then we have a corner solution $x(\alpha,\delta) = 0\ \forall\alpha$, and the result is proven. If there exists a $\alpha' \in (\frac{1}{2}, 1]$ such that $H(\alpha') < 0$, then there must be a α'' such that $H(\alpha'') > 0$ and $H_\alpha(\alpha'') < 0$. Strict concavity of $H(\alpha)$ however implies that $H_\alpha(\alpha) < 0$ for any $\alpha > \alpha''$. So $H(\alpha) < 0$ implies $B_{x\alpha}(x,\alpha) < 0$. □

Acknowledgments

I am grateful for helpful comments to Thomas Wiseman, two anonymous referees and the participants to the 4th annual Duke/Northwestern/Texas I.O. Theory Conference. Alice Hsiaw and Gabor Virag provided outstanding research assistance. I gratefully acknowledge support from the Alfred P. Sloan foundation and from NSF grants SES-0418150 and SES-0547748.

References

Battaglini, M., 2005. Long-term contracting with Markovian consumers. Amer. Econ. Rev. 95, 637–658.

Battaglini, M., Coate, S., 2006. Pareto efficient income taxation with stochastic abilities. Mimeo.

Baron, D., Besanko, D., 1984. Regulation and information in a continuing relationship. Info. Econ. Pol. 1 (3), 267–302.

Biehl, A., 2001. Durable-goods monopoly with stochastic values. RAND J. Econ. 32, 565–577.

Blume, A., 1998. Contract Renegotiation with Time Varying Valuations. J. Econ. Manage. Strategy 7, 397–433.

Courty, P., Li, H., 2000. Sequential screening. Rev. Econ. Stud. 67, 697–718.

Dewatripont, M., 1989. Renegotiation and information revelation over time: The case of optimal labor contracts. Quart. J. Econ. 104, 589–619.

Hart, O., Tirole, J., 1988. Contract renegotiation and Coasian dynamics. Rev. Econ. Stud. 55, 509–540.

Hendel, I., Lizzeri, A., 2002. The role of leasing under adverse selection. J. Polit. Economy 101, 113–143.

Hendel, I., Lizzeri, A., Siniscalchi, M., 2005. Efficient sorting in a dynamic adverse-selection model. Rev. Econ. Stud. 72 (2), 467–497.

Kennan, J., 2001. Repeated bargaining with persistent private information. Rev. Econ. Stud. 68, 719–755.

Laffont, J.J., Tirole, J., 1986. Using cost observation to regulate firms. J. Polit. Economy 94, 614–641.

Laffont, J.J., Tirole, J., 1987. Comparative statics of the optimal incentive contract. Europ. Econ. Rev. 31, 901–926.

Laffont, J.J., Tirole, J., 1988. The dynamics of incentive contracts. Econometrica 56, 1153–1175.

Laffont, J.J., Tirole, J., 1990. Adverse selection and renegotiation in procurement. Rev. Econ. Stud. 57, 597–625.

Laffont, J.J., Tirole, J., 1996. Pollution permits and compliance strategies. J. Public Econ. 62, 85–125.

Ramsey, F., 1927. A contribution to the theory of taxation. Econ. J. 37, 47–61.

Rey, P., Salanie, B., 1996. On the value of commitment with asymmetric information. Econometrica 64, 1395–1414.

Roberts, K., 1982. Long-term contracts. Mimeo.

Rogerson, W., 1994. Economic incentives and the defense procurement process. J. Econ. Perspect. 8, 65–90.

Rustichini, A., Wolinsky, A., 1995. Learning about variable demand in the long run. J. Econ. Dynam. Control 19, 1283–1292.