

Long-Term Contracting with Markovian Consumers

By MARCO BATTAGLINI*

To study how a firm can capitalize on a long-term customer relationship, we characterize the optimal contract between a monopolist and a consumer whose preferences follow a Markov process. The optimal contract is nonstationary and has infinite memory, but is described by a simple state variable. Under general conditions, supply converges to the efficient level for any degree of persistence of the types and along any history, though convergence is history-dependent. In contrast, as with constant types, the optimal contract can be renegotiation-proof, even with highly persistent types. These properties provide insights into the optimal ownership structure of the production technology. (JEL D23, D42, D82)

Advances in information processing and new management strategies have made long-term, nonanonymous relations between buyers and sellers feasible in an increasing number of markets. Many retailers can now store large databases on consumers' choices and utilize them for pricing decisions at a very low cost. In part because of these new technologies, recent managerial schools have stressed the importance of capitalizing on long-term relations with customers (see, e.g., Louis V. Gerstner, Jr., 2002, and Jack Welch, 2001). When a long-term relationship is nonanonymous and types are persistent, the seller can mitigate the problem of asymmetric information by using consumers' choices to forecast future behavior. As a result, however, buyers are more reluctant to reveal private information that affects their consumption decisions: their strategic reaction may limit or even

eliminate the benefits for the seller. The existing literature has studied this problem, focusing on those cases in which the consumer's type is constant over time.¹ Here, it is well known that the seller finds it optimal to offer the optimal static contract period after period. In a sense, the seller commits not to use the information gathered from the consumer's choices.

A model of long-term contracting that assumes constant types, however, clearly misses an important dimension of the problem. Consider the case of a monopolist selling to an entrepreneur whose type depends on the number of customers waiting for service. As is well known, under standard assumptions on the arrival rate of customers, the type of this entrepreneur follows a Markov process (see, e.g., Samuel Karlin and Howard M. Taylor, 1975). Or, to give another example, consider the case of a company selling cellular telephones. These contracts often last for years and it would not be reasonable to assume that the telephone company, or the customer, does not take into account the likely, but uncertain, evolution of preferences (see, e.g., Eugenio J. Miravete, 2003, for evidence). In all these situations, the assumption that the consumer's type is constant is clearly not realistic. Even if types are very persistent, it is reasonable to assume that they may vary over time and follow a stochastic process.

In this paper, we characterize the optimal contract offered in an infinitely repeated setting

* Department of Economics, Princeton University, 001 Fisher Hall, Princeton, NJ 08544 (e-mail: mbattagl@princeton.edu). I gratefully acknowledge financial support from the National Science Foundation (Grant No. SES-0418150) and the hospitality of the Economics Department at the Massachusetts Institute of Technology for the academic year 2002–2003. I am grateful for helpful comments to Pierpaolo Battigalli, Douglas Bernheim, Stephen Coate, Eddy Dekel, Avinash Dixit, Glenn Ellison, Jeff Ely, Bengt Holmström, John Kennan, Alessandro Lizzeri, Rohini Pande, Nicola Persico, William Rogerson, Ariel Rubinstein, Jean Tirole, Asher Wolinsky, seminar participants at Bocconi University, Cornell University, Harvard University, MIT, Northwestern University, Princeton University, the Stanford Institute of Theoretical Economics, UCLA, University College London, and University of Wisconsin-Madison, as well as to three referees. Marek Picia provided valuable research assistance.

¹ The related literature is discussed in Section I.

by a monopolist to a consumer whose preferences evolve following a Markov process. In this case, even if types are highly persistent, the contract is very different from the contract with constant types because the seller finds it optimal to use information acquired along the interaction in a truly dynamic way. For this reason, the characterization of the optimal contract when there is heterogeneity within and across periods allows a new understanding of important aspects of a dynamic principal-agent relationship that previous models could not capture—particularly, with regard to the memory and complexity of the contract, its efficiency, and its robustness to renegotiation. Perhaps surprisingly, it also provides insights into the optimal ownership structure of the production technology.

As noted, when types are constant, the contract has no memory and the inefficiency of the optimal static contract is repeated period after period. With persistent but stochastic types, even in a simple stationary environment with one period memory (i.e., a Markov process), the contract is nonstationary and has *infinite memory*; despite this, however, it can be represented in a very economical way by a simple state variable. Even if types are arbitrarily highly correlated and the discount factor is arbitrarily small, the seller's optimal offer converges over time to the efficient supply schedule along *all* possible histories. The speed of convergence, however, is state contingent and occurs in a particular way, which extends a well-known property of the static model. On the one hand, in fact, we have a *generalized no distortion at the top* (GNDT) principle: after any history, if the agent reveals himself to have the highest possible marginal valuation for the good, supply is set efficiently from that date onward in any infinite history that may follow. On the other hand, and more importantly, we have a novel *vanishing distortion at the bottom* (VDB) principle: even in the history in which the agent always reveals to have the lowest marginal valuation for the good, the contract converges to the efficient menu offer. One immediate implication of this result is that in the “steady state,” or even after a few periods, the monopolist's supply schedule may be empirically indistinguishable from the outcome of an efficient competitive market; moreover, since higher efficiency is associated with a higher consumer

rent, it explains why “old” customers should be treated more favorably than “new” customers.²

In a stochastic environment, the incentives for renegotiation are also very different. As shown in the received literature (see discussion below) when types are constant over time, the monopolist benefits from the ability to commit to not renegotiating the contract, because the optimal contract is never time-consistent. With variable types, in contrast, this is not the case: indeed, even when types are highly correlated, a simple and easily satisfied condition guarantees renegotiation-proofness. Interestingly, when types are constant the optimal renegotiation-proof contract always requires the agent to use sophisticated mixed strategies: with correlated but stochastic types, the optimal renegotiation-proof contract has an equilibrium in pure strategies and simply requires the agent to report his type.

There is an intuitive argument which explains the dynamics of the distortions in the optimal contract and the efficiency result mentioned earlier. Assume that the agent's type can take two values: high and low marginal valuation for the good (respectively, θ_H and θ_L). Consider Figure 1, which shows the impact on profits at time zero of a marginal increase Δq in the quantity $q(h_t)$ offered to the consumer after a history h_t . On the one hand, this change increases the surplus that can potentially be appropriated by the seller if history h_t is realized (which is represented by the “thick arrow” on the left-hand panel in Figure 1).³ However, as in a static model, this increase in supply increases the rent that the principal must leave to the agent to satisfy incentive compatibility. In every period, optimal supply is determined by this marginal cost–marginal benefit trade-off, and the dynamic properties of the contract are driven by its evolution. To determine optimal supply, therefore, it is important to understand the impact over expected rents of this change at time t .

To this goal, consider the right panel of Figure 1 and assume that the rent of the high type

² See Igal Hendel and Alessandro Lizzeri (2003) and Georges Dionne and Neil A. Doherty (1994) for evidence of this phenomenon.

³ Because consumption is always distorted below its first best level, starting from the seller's optimum, a marginal increase in supply after history h_t corresponds to an increase in efficiency of the contract at that node.

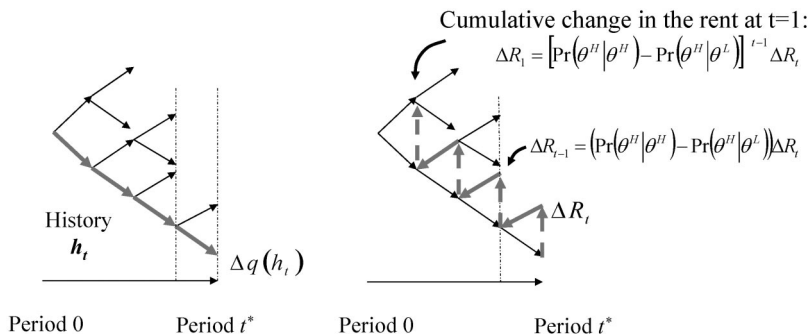


FIGURE 1. MARGINAL COST AND BENEFIT OF A CHANGE IN THE QUANTITY OFFERED AFTER HISTORY h_t

Notes: The arrows represent the history tree: an arrow pointing up (respectively, down) represents a high (respectively, low) type realization. The horizontal axis is the time line.

increases by ΔR_t at time t . At time $t - 1$ the expected utility of the agent in the history immediately preceding h_t increases as well because, although the agent is a low type at $t - 1$, he can become a high type in the following period, and then benefit from the increase in rent. Part of this extra expected rent can be extracted by the seller at $t - 1$, but not all, since incentive compatibility must be satisfied at that time as well. At time $t - 1$, the high type cannot receive less than what he would receive if he chose the option designed for the low type. Even if the seller extracts all the expected increase in consumption of the low type with an increase in price $\Delta p_{t-1} = \Pr(\theta_H|\theta_L)\Delta R_t$ at $t - 1$, the change in rent of the high type at $t - 1$, ΔR_{t-1} , would be equal to $\Pr(\theta_H|\theta_H)\Delta R_t - \Delta p_{t-1}$, that is, $(\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L))\Delta R_t$,⁴ which is positive if types are positively correlated. If the seller tries to extract this extra rent at $t - 1$, then, repeating the same argument, she still must provide an increase in rent to the high type at time $t - 2$ equal to $\Delta R_{t-2} = (\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L))\Delta R_{t-1}$, which can be written as $(\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L))^2\Delta R_t$. Proceeding backward, we arrive at an increase in the rent left to the consumer at time 1 proportional to $(\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L))^{t-1}$ (see the dashed arrows in the right panel of Figure 1).

While the marginal impact of the change in supply on expected surplus evaluated at time zero is proportional to the probability of the history h_t (i.e., $\mu_L\Pr(\theta_L|\theta_L)^{t-1}$), the impact on

the agent's expected rent is proportional to the "cumulative effect" of the difference in expectations of the types: $[\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)]^{t-1}$.⁵ Accordingly, the marginal cost–marginal benefit ratio at time t is proportional to

$$(1) \quad \left[\frac{\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)} \right]^{t-1}.$$

The dynamics of the optimal contract depends on the evolution of this cost-benefit ratio. When types are constant, the term in parentheses is *exactly* equal to one, so (1) is independent of t and the distortion is constant: this explains why, with constant types, it is optimal to offer the static contract repeatedly. When types are positively but imperfectly correlated, even if types are highly persistent, optimal supply converges to an efficient level along all histories as $t \rightarrow \infty$ because (1) converges to zero.

In general, any change in the contract at a time t has *cascade effects* on the expected rents in the previous periods. These effects depend not only on the transition probabilities, but also on the structure of the constraints that are binding at the optimum. As time passes, these cascade effects become increasingly complicated because the number of histories grows exponentially. A methodological contribution of this paper is in a novel characterization of the binding constraints by an inductive argument which

⁴ The probability that a type i in period t becomes a type j in $t + 1$ is denoted $\Pr(\theta_j|\theta_i)$.

⁵ The expected change in the agent's rent at time one is $\mu_H[\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)]^{t-1}\Delta R_t$, where μ_H is the probability that the agent is a high type in the first period. The constant μ_H , however, is irrelevant for our argument.

allows a substantial simplification of the problem.

The particular features of the optimal contract described above also have implications for the optimal ownership structure of the monopolist's business. It is indeed interesting to ask why the monopolist keeps control of the production technology: after all, only the consumer benefits directly from it and has information for its efficient use. We show that the optimal contract can be interpreted as offering the high-type consumer a call option to buy out the technology used by the monopolist. The sale of the technology, however, is state contingent and the monopolist tends to retain control more often than what would be socially optimal: by keeping the ownership rights, the monopolist can control future rents of the high types and this improves surplus extraction because types have different expectations for the future. This insight seems relevant to understand the ownership structure of a new technology. The initial owner of a new technology generally has monopoly power on its use thanks to a patent and must decide if it is more convenient to use the technology directly selling its products, or to sell the patent.

The paper is organized as follows. Section I surveys the related literature. In Section II we describe the model. In Section III we characterize the optimal contract and discuss its efficiency properties. Section IV discusses the theory of property rights that follows from the characterization. Section V discusses the properties of the monetary payments in the optimal contract. Section VI studies renegotiation-proofness. Section VII presents concluding comments.

I. Related Literature

As mentioned above, in dynamic models of price discrimination it is generally assumed that the agent's type is constant over time.⁶ In this case we have a "false dynamics" in which the monopolist finds it optimal to commit to a contract in which past information is ignored and

the optimal static menu is repeated in every period (see, e.g., Laffont and Tirole, 1993). With constant types the dynamics becomes interesting only when other constraints are binding, in particular when a renegotiation-proofness constraint must be satisfied. Seminal papers in this literature are Dewatripont (1989), Oliver Hart and Tirole (1988), and Laffont and Tirole (1990).⁷ In contrast to our findings with variable types, a common result in this literature with constant types is that the *ex ante* optimal contract is never renegotiation-proof.

Kevin Roberts (1982) and Robert M. Townsend (1982) are the first to present repeated principal-agent models with stochastic types. In these frameworks, however, types are serially independent realizations, and therefore incentives for present and future actions can easily be separated. Indeed, in this case, except for the first period, there is no asymmetric information between the principal and the agent because both share the same expectation for the future.⁸ David Baron and David Besanko (1984) and Laffont and Tirole (1996) extend this research, presenting two period procurement models in which the type in the second period is stochastic and correlated with the type in the first period. Because these models have only two periods, however, they cannot capture such important aspects of the dynamics of the optimal contract as its memory and complexity after long histories, or its convergence to efficiency. Aldo Rustichini and Asher Wolinsky (1995) characterize optimal pricing in a model with infinite horizon and Markovian types as ours. However, in their model consumers are not strategic and ignore that future prices depend on their current actions; demand, moreover, can assume two values, zero or one. None

⁷ These papers study the optimal renegotiation-proof contract with constant types under different assumptions. Hart and Tirole (1988) and Dewatripont (1989) present models with many periods: the first paper assumes that supply can have two values, zero or one; the second focuses on pure strategies and assumes some simplifications in the nature of the contractual agreement. Laffont and Tirole (1990) solve a model in which supply can assume more than two values, assuming two periods.

⁸ Because Townsend (1982) is specifically interested in modelling risk sharing, he assumes that the principal is less risk averse than the agent. In this case, even with i.i.d. types, the contract depends on the cumulated wealth of the agent.

⁶ For excellent overviews of the literature on dynamic contracting, see Patrick Bolton and Mathias Dewatripont (2005) and Jean-Jacques Laffont and Jean Tirole (1993).

of these papers with variable types considers renegotiation-proofness.⁹

II. The Model

We consider a model with two parties, a buyer and a seller. The buyer repeatedly buys a nondurable good from the seller. He enjoys a per-period utility $\theta_i q - p$ for q units of the good bought at a price p . In every period, the seller produces the good with a cost function $c(q) = \frac{1}{2} q^2$. The marginal benefit θ_i evolves over time according to a Markov process. To focus on the dynamics of the contract, we consider the simplest case in which each period the agent can assume one of two types, θ_L, θ_H with $\Delta\theta = \theta_H - \theta_L > 0$. The probability that state l is reached if the agent is in state k is denoted $\Pr(\theta_l|\theta_k) \in (0, 1)$; the distribution of types conditional on being a high (low) type is denoted $\alpha_H = (\Pr(\theta_H|\theta_H), \Pr(\theta_L|\theta_H))$ ($\alpha_L = (\Pr(\theta_H|\theta_L), \Pr(\theta_L|\theta_L))$). We assume that types are positively correlated, i.e., $\Pr(\theta_H|\theta_H) \geq \Pr(\theta_H|\theta_L)$. However, we do not make assumptions on the degree of correlation: indeed, an environment with constant types can be seen as a limiting case of our model in which the probability that a type does not change converges to one. In each period the consumer observes the realization of his own type; the seller, in contrast, cannot see it. At date 0 the seller has a prior $\mu = (\mu_H, \mu_L)$ on the agent's type.¹⁰ For future reference, note that the efficient level of output is equal to $q^e(\theta_i) = \theta_i$ in all periods and after any history of realizations of the types.

⁹ Dynamic environments with adverse selection and stochastic types have recently been used to study models of leasing, insurance, and other applications. See Pascal Courty and Hao Li (2000), and Hendel and Lizzeri (1999, 2003). John Kennan (2001) has studied a model with variable types, but in which only short-term contracts with one period length can be offered. Battaglini and Stephen Coate (2003) apply the techniques of the present paper to characterize the Pareto optimal frontier of taxation with correlated types.

¹⁰ The fact that the agent's type follows a Markov process can be modelled in many natural ways. The agent may be a firm whose type depends on its list of customers waiting for services, which according to the "inventory model" follows a Markov process (see Karlin and Taylor, 1975, §2.2.d). Or the agent's type may depend on his investment opportunities: if these follow a branching process, then they are described by a Markov process (see Karlin and Taylor, 1975, §2.2.f).

We assume that the relationship between the buyer and the seller is infinitely repeated and the discount factor is $\delta \in (0, 1)$. In period 1 the seller offers a supply contract to the buyer. The buyer can reject the offer or accept it; in the latter case the buyer can walk away from the relationship at any time $t \geq 1$ if the expected continuation utility offered by the contract falls below the reservation value $\underline{u} = 0$. In line with the standard model of price discrimination, the monopolist commits to the contract that is offered: in Section VI we relax this assumption, allowing the parties to renegotiate the contract.

It is easy to show that in the environment that we will study a form of the revelation principle is valid and allows us to consider without loss of generality only contracts that in each period t depend on the revealed type at time t and on the history of previous type revelations. In this case the contract $\langle \mathbf{p}, \mathbf{q} \rangle$ can be written as $\langle \mathbf{p}, \mathbf{q} \rangle = (p_t(\hat{\theta}_t|h_t), q_t(\hat{\theta}_t|h_t))_{t=1}^\infty$, where h_t and $\hat{\theta}_t$ are, respectively, the public history and the type revealed at time t , and $q_t(\cdot)$ and $p_t(\cdot)$ are the quantities and prices conditional on the declaration and the history.¹¹ In general, h_t can be defined recursively as $h_t := \{\hat{\theta}_{t-1}, h_{t-1}\}$, $h_1 := \emptyset$ where $\hat{\theta}_{t-1}$ is the type revealed in period $t-1$. The set of possible histories at time t is denoted H_t ; the set of histories at time j following a history h_i ($t \leq j$) is denoted $H_j(h_i)$. A strategy for a seller consists of offering a direct mechanism $\langle \mathbf{p}, \mathbf{q} \rangle$ as described above. The strategy of a consumer is, at least potentially, contingent on a richer history $h_t^C := \{\hat{\theta}_{t-1}, \theta_t, h_{t-1}^C\}$, $h_1^C := \theta_1$ because the agent always knows his own type. For a given contract, a strategy for the consumer, then, is simply a function that maps a history h_t^C into a revealed type: $h_t^C \mapsto b(h_t^C)$.

In the study of static models it is often assumed that all types are served, i.e., each type is offered a positive quantity, which is guaranteed by the assumption that $\Delta\theta$ is not too large. The same condition that guarantees this property in the static model also guarantees it in our dynamic model; therefore, to simplify notation, we assume that this condition is verified in our model.¹² This assumption can easily be relaxed,

¹¹ Note, therefore, that $p(\theta|h)$ is not the per-unit price paid after history $\{\theta, h\}$, but the total monetary transfer at that history.

¹² The condition that guarantees that all types are served is $\Delta\theta \leq (\mu_L/\mu_H)\theta_L$. As we will see, the distortion introduced

but this would complicate notation with no gain in insight.

In the first part of the analysis we focus on the case with unilateral commitment in which the monopolist can commit, but the consumer can leave the relationship anytime. This assumption seems the most appropriate in many markets.¹³ On the other hand, there are many situations in which renegotiation is an important component of the problem: in Section VI, we show that under general conditions the optimal contract is renegotiation-proof and therefore it can be applied to these environments too.

III. The Optimal Contract

The monopolist's optimal choice of contract maximizes profits under the constraint that after any history the consumer receives (at least) his reservation utility and, also after any history, there is no incentive to report a false type:

$$\begin{aligned}
 (\mathcal{P}_I) \quad & \max_{(p, \mu)} \mu_H [p(\theta_H|h_1) - q^2(\theta_H|h_1)/2] \\
 & + \delta E[\Pi(\theta|h_1, \theta_H)|\theta_t = \theta_H] \\
 & + \mu_L [p(\theta_L|h_1) - q^2(\theta_L|h_1)/2] \\
 & + \delta E[\Pi(\theta|h_1, \theta_L)|\theta_t = \theta_L] \\
 \text{s.t. } & IC_{h_i}(\theta_H), IC_{h_i}(\theta_L), IR_{h_i}(\theta_H), \\
 & IR_{h_i}(\theta_L) \quad \forall h_i
 \end{aligned}$$

where $E[\Pi(\theta|h_1, \theta_i)|\theta_t = \theta_i]$ $i = H, L$ is the expected value function of the monopolist after history $\{h_1, \theta_i\}$. The incentive constraints $IC_{h_i}(\theta_i)$ for $i = H, L$ are described by:

$$\begin{aligned}
 (IC_{h_i}(\theta_i)) \quad & q(\theta_i|h_t)\theta_i - p(\theta_i|h_t) \\
 & + \delta E[U(\theta|h_t, \theta_i)|\theta_t = \theta_i] \\
 & \geq q(\theta_j|h_t)\theta_i - p(\theta_j|h_t) \\
 & + \delta E[U(\theta|h_t, \theta_j)|\theta_t = \theta_i]
 \end{aligned}$$

$\forall i \neq j, i, j \in H, L$, where $U(\theta|h_t, \theta_i)$ is the value function of a type θ after a history $\{h_t, \theta_i\}$. These constraints guarantee that type i does not want to imitate type j after any history h_t . And the individual rationality constraint $IR_{h_i}(\theta_i)$ simply requires that the agent wants to participate in the relationship each period: $U(\theta|h_t, \theta_i) \geq 0$ for any i and h_t .

The classic approach to characterize the solution to this problem in a static environment is in two steps. First, a simplified program, in which the participation constraints of the high type and the incentive compatibility constraints of the low type are ignored, is considered (the "relaxed problem"). Then it is shown that there is no loss of generality in restricting attention to this case. In a static model, the remaining constraints of the relaxed problem are necessarily binding at the optimal solution: this simplifies the analysis because it allows us to substitute them directly in the objective function.

It is easy, however, to see that *in a dynamic model this cannot be true*. Given an optimal contract, we can always add a "borrowing" agreement in which the monopolist receives a payment at time t and pays it back in the following periods. If the net present value of this transaction is zero, then neither the monopolist's profit changes, nor any constraint, would be violated, so the contract would remain optimal: but the individual rationality constraints need not remain binding after some histories. More importantly, the incentive compatibility constraints may also not be binding. In order to provide incentives to the high type to reveal his private information, the monopolist may find it useful to use future payoffs instead of present payoffs to screen the agent's types. If this were the case, there would be a history after which the contract leaves to the high type more surplus than what a binding incentive compatibility constraint would imply.

The following result generalizes the "binding constraints" result of the static model, showing that in a dynamic setting, although constraints

by the monopolist is declining over time in all histories and, in the first period, it is equal to the distortion of the static model. Therefore if the monopolist serves all customers in the static model, then she serves all customers after all histories in our dynamic model too.

¹³ Discussing the life insurance market, Hendel and Lizzeri observe that the term value contracts in the insurance market which account for 37 percent of ordinary life insurance, "... are unilateral: the insurance companies must respect the terms of the contract for the duration, but the buyer can look for better deals at any time. [...] These features fit a model of unilateral commitment." (Hendel and Lizzeri, 2003, p. 302). Moreover, there is evidence that firms seem aware that the possibility to commit is important to win exclusive long-term contracts.

need not bind in every optimal scheme, there is no loss of generality in assuming that constraints in the relaxed problem are satisfied as equalities. Let us define $\mathcal{P}_H$ as the program in which expected profits are maximized, assuming that the incentive constraints of the high type and the participation constraints of the low type hold as equalities after any history, and no other constraint is assumed. We say that a supply schedule $q_t^*(\theta|h_t)$ is a solution of a given program if there exists a payment schedule $p_t^*(\theta|h_t)$ such that the menu $\{q_t^*(\theta|h_t), p_t^*(\theta|h_t)\}$ is a solution of the program.

LEMMA 1: *The supply schedule $q_t^*(\theta|h_t)$ solves $\mathcal{P}_I$ if and only if it solves $\mathcal{P}_H$.*

The result that the constraints may be assumed to hold as equalities without loss may be intuitively explained in a two-period version of the model (the complete argument, presented in the Appendix, is by induction on t). Assume that at time $t = 2$ the incentive compatibility constraint of the high type is not binding after a history $h_2 = \theta_L$. Consider this change in the contract: reduce the extra rent at $t = 2$ and reduce the price paid by the low type at $t = 1$ so that his participation constraint is satisfied as an equality after the change. The rent of the high type at time 1 depends on his outside option (the utility obtained by reporting himself untruthfully to be a low type), so it is affected by both these changes. Even if the net change in payments has a neutral effect on the low type's expected utility, however, it will reduce the outside option of the high type: because the high type is more optimistic about the future realization of his type, the reduction in future rents if he reports his type untruthfully will be larger than the

increase in payments at time t . If the value of the high type's outside option goes down, then his equilibrium rent goes down as well. Expected profits, therefore, would be larger after the change in the contract and all constraints would be respected: but this is not possible if the contract is optimal, so we have a contradiction. After a history $h_2 = \theta_H$ we proceed in a similar way: in this case profits remain constant after the change in prices, so the constraint needs not necessarily be binding at the optimum, but it can be reduced to an equality without loss. The argument for the participation constraints is analogous.

It is important to point out that Lemma 1 does not claim that any solution $\langle \mathbf{p}, \mathbf{q} \rangle$ of a relaxed problem in which the incentive constraint of the low type and the participation constraint of the high type are ignored is a solution of $\mathcal{P}_I$. In $\mathcal{P}_H$ we assume that the constraints are satisfied as equalities, so it is not just a relaxed version of $\mathcal{P}_I$. Indeed, such a claim would not be true: some solutions of the relaxed problem would imply future rents for the high type that would violate the incentive compatibility constraint of the low type after some histories. However, if $\langle \mathbf{p}, \mathbf{q} \rangle$ solves the relaxed problem, then there exists a $\mathbf{p}'$ such that $\langle \mathbf{p}', \mathbf{q} \rangle$ solves $\mathcal{P}_I$; and if $\langle \mathbf{p}, \mathbf{q} \rangle$ solves $\mathcal{P}_I$ then there exists a $\mathbf{p}''$ such that $\langle \mathbf{p}'', \mathbf{q} \rangle$ solves $\mathcal{P}_H$ and, *because of this*, solves the relaxed problem as well.

We can now focus the simpler problem with equality constraints $\mathcal{P}_H$; from the first-order conditions, we obtain:

PROPOSITION 1: *At any time t , the optimal contract is characterized by the supply function:*

$$(2) \quad q_t^*(\theta|h_t) = \begin{cases} \theta_H & \text{if } \theta = \theta_H \\ \theta_L - \Delta\theta \frac{\mu_H}{\mu_L} \left[\frac{\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)} \right]^{t-1} & \text{if } \theta = \theta_L \text{ and } h_t = h_t^L \\ \theta_L & \text{if } \theta = \theta_L \text{ and } h_t \in H_t \setminus h_t^L \end{cases}$$

where $h_t^L := \{\theta_L, \theta_L, \dots, \theta_L\}$, the history along which the agent always reports himself to be a low type in the first $t - 1$ periods.

From (2) we can see that the optimal contract is nonstationary and has unbounded memory: for any $T > 0$, we can always find two histories that are identical for the last T periods but that

induce different menus in the optimal mechanism.¹⁴ This fact, however, does not imply that

¹⁴ Consider two histories that differ only in the first realization of types, the first being high, the second being low, and which have low realizations in any period following date two. If these histories are longer than a positive parameter T , say they have $T + 1$ length, then they coincide

the contract has a complicated structure. From Proposition 1 we can see that the only thing that matters for the contract is whether we are on the lower branch or not. Since this depends only on the current type, and if in the previous periods the agent reported himself to be a high type, the state can be described by a simple 0-1 variable which can be defined recursively:

$$(3) \quad X_t = X(\theta_t, X_{t-1}) \\ = \begin{cases} 1 & \text{if } X_{t-1} = 1 \text{ and } \theta_t = \theta_L \\ 0 & \text{else} \end{cases}$$

for $t \geq 1$, and $X_0 = 1$. This variable starts with value one and remains one if the agent persists in reporting a low type; once the agent has reported himself to be a high type, the state switches to zero and remains constant forever. Let us define $\Lambda = [\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)/\Pr(\theta_L|\theta_L)]$; we have:

PROPOSITION 2: *The optimal solution is a function of time and the 0-1 state variable described by (3): $q_t^*(\theta_t, X_{t-1}) = \theta_t - \Delta\theta(\mu_H/\mu_L)X(\theta_t, X_{t-1})\Lambda^{t-1}$.*

The length of the memory of the optimal contract is a central issue in the literature on dynamic moral hazard (see William P. Rogerson, 1985), but it has not been studied in adverse selection models, because when the agent's type is perfectly constant we know that the contract is also constant over time and independent of past histories, so it has no memory. In the moral hazard literature, the memory of the contract is a direct consequence of the agent's risk aversion. With risk aversion, it is optimal not only to smooth consumption over states of the world, as in the static moral hazard framework, but also to smooth consumption over periods. To this end, the contract must keep track of the past realizations of the agent's income. In the model presented above, however, the agent is risk neutral; the persistence of

the distortion, therefore, does not depend on consumption smoothing, but it is a *necessary* feature of dynamic price discrimination. In a dynamic environment, the principal has more freedom to redistribute distortions over time and states in order to screen the agent's types. Propositions 1 and 2 characterize the optimal way to redistribute the distortion, proving that the principal finds it optimal to introduce distortions even in the far future, potentially for an unbounded number of periods. This is perhaps surprising since the agent's taste follows a Markov process and therefore the relevant economic environment has a memory of only one period.¹⁵

We now turn to the particular pattern in which distortions are introduced. In Sections III A and III B we discuss the dynamics of distortions and the asymptotic properties of the contract as $\delta \rightarrow 1$. In Section III C, we discuss the key assumptions of the model.

A. Efficiency: The GNDT and VDB Principles

In order to interpret (2), it is useful to compare it with the benchmark with constant types. In this case, there are only two possible histories: either the agent is always a high type, in which case the contract is efficient; or the agent is always a low type, and the contract is distorted below the efficient level in all periods by a constant $\Delta\theta(\mu_H/\mu_L)$.

When types follow a Markov process, the contract instantly becomes efficient as soon as the agent reports himself to be a high type: but now efficiency "invades" the histories in which the agent subsequently reports himself to be a low type. This is the *generalized no distortion at the top* (GNDT) principle. Its intuition is the following. Distortions are introduced only to extract more surplus from higher types; therefore there is no reason to distort the quantity offered to the highest type. After any history h_t the rent that must be paid to a high type to reveal himself is independent of the quantities that follow this history: since the incentive compatibility constraint for the high type is binding,

for at least the last T periods. At time 1 the monopolist offers an efficient contract in the first history, i.e., regardless of the realizations in the following periods, the quantity offered is efficient in any period following the first. Not so for the second history. Therefore, even if there is no intrinsic economic reason in the environment to offer different menus at date $T + 1$, the contracts are different.

¹⁵ As we discuss in greater detail in Section IV, the fact that the seller wants to distort supply for a unlimited and state-contingent number of periods has important implications for the allocation of the property rights of the production technology.

he receives the same utility as if he falsely reported himself to be a low type; therefore, only the quantities that follow such a history affect his rents. This implies that the monopolist is *residual claimant* on the surplus generated on histories after a high type report, and therefore the quantities that follow such histories are chosen efficiently. In our dynamic framework, this simple principle has strong implications because it forces the contract to be efficient not only in the first period in which the agent truthfully reveals himself to be a high type, but also in all the following periods.

A distortion persists on the lowest branch of the history tree (i.e., when the agent always declares to be a low type). By a simple manipulation of the formula in Proposition 2, the distortion can be written as $\theta_L - q^*(\theta_L|h_t) = \Delta\theta(\mu_H/\mu_L)X(\theta_t, X_{t-1})\{1 - [\Pr(\theta_L|\theta_H)/\Pr(\theta_L|\theta_L)]\}^{t-1}$, since the efficient level of output with a low type is θ_L . Given that types are positively correlated, we have $[\Pr(\theta_L|\theta_H)/\Pr(\theta_L|\theta_L)] \in (0, 1)$ and it follows that $\lim_{t \rightarrow \infty} q^*(\theta|h_t) = \theta = q^e(\theta)$, which proves:

PROPOSITION 3: *For any discount factor $\delta \in (0, 1)$, the optimal contract converges over time to an efficient contract along any possible history.*

This is the *vanishing distortion at the bottom* (VDB) principle. The monopolist introduces a distortion along the “lowest” history because this minimizes the cost of screening the agent’s types: however, even this distortion converges to zero as $t \rightarrow \infty$.

The optimal distortion simply equalizes the marginal cost of a decrease in supply (in terms of reduced surplus generated in the relationship) and its marginal benefit (in terms of reduced rent to be paid to the high type). With constant types, after any history h_t in which the agent declares to be a low type, the marginal benefit of increasing surplus with a higher $q(h_t)$ is independent of the length of the history: it is proportional only to μ_L , because once the type is low in the first period then it is low forever. Similarly, the marginal cost of an increase in $q(h_t)$ is proportional to μ_H , the probability that the high type receives the increase in the rent. Since, therefore, the marginal cost/marginal benefit ratio is time-independent, it is not surprising that the optimal distortion is also con-

stant over time. Indeed, the case with constant types is not asymptotically efficient because $\{1 - [\Pr(\theta_L|\theta_H)/\Pr(\theta_L|\theta_L)]\}^{t-1}$ is *exactly* one, and therefore, *independent* from t .

Clearly, as the persistence of types converges to one, we have that $[\Pr(\theta_L|\theta_H)/\Pr(\theta_L|\theta_L)] \rightarrow 0$. Not surprisingly, this implies that *ceteribus paribus* the contract converges in every period to the optimal static contract as $\Pr(\theta_H|\theta_H)$ and $\Pr(\theta_L|\theta_L)$ converge to one. There are, however, two important observations. First, convergence to efficiency appears to be relatively fast even if types are highly correlated.¹⁶ Second, as we discuss below, the results with fixed and stochastic types are very different when $\delta \rightarrow 1$, regardless of the level of persistence of the types.

B. Distribution of Surplus with Large Discount Factors

All the results presented above are valid for any $\delta \in (0, 1)$; if we assume that $\delta \rightarrow 1$, even stronger results emerge. In this case we can easily bound the inefficiency and determine the distribution of surplus between the seller and the buyer.¹⁷

With constant types, the average utility of the consumer is bounded away from zero and independent of δ ; the average payoff of the monopolist is equal to the profit that would be achieved in a static model and independent of δ as well. However, even an arbitrarily small reduction in the persistence of the types has a very high impact on surplus and payoffs when the discount factor is high.

¹⁶ Assume, for example, that types are ex ante equally likely, and the types are very much correlated (for example, the type is persistent 80 percent of the time). Then the expected inefficiency of the contract after ten periods will be $0.03779\Delta\theta$; the expected inefficiency after 50 periods will be $5.0517 \times 10^{-12}\Delta\theta$.

¹⁷ In this comparative statics exercise we change the discount factor, keeping the transition probabilities constant. Another interesting exercise would be to modify the level of persistence of the types, or the frequency of their changes. Increasing the frequency of changes would reinforce the effects of an increase in δ . But when we simultaneously increase types persistence and the discount factor, the result depends on which of the two (persistence and δ) converges faster to one. The case considered in the paper, in which only $\delta \rightarrow 1$, corresponds to the case in which the discount factor converges more quickly than persistence.

PROPOSITION 4: *When types are imperfectly correlated, even if correlation is not positive, then as $\delta \rightarrow 1$ the average profit of the monopolist converges to the first-best level of surplus and the average utility of the consumer converges to zero, regardless of the renegotiation-proofness constraint.*

When the discount factor is high, it does not matter what happens in the first T periods, for T finite. However, because we are working with a Markov process, in the long run the distribution of types converges to a stationary distribution which is independent from the initial value. This implies that at time one the ability of the consumer to predict his own type realizations in the far future is almost as good as the seller's ability. For this reason, when δ is high the monopolist can separate the agents paying only a minuscule rent to the higher type.¹⁸

C. Discussion

Before presenting further results, we now discuss the assumptions of the model, emphasizing the issues that are still open for future research. In particular, we focus on the stochastic process, the utility function, cash constraints, and the time horizon.

As noted, any change of the contract at time t has a "cascade" effect on expected utilities in the previous histories. These effects depend both on the structure of the constraints that are binding at the optimum, and on the transition probabilities, which determine the conditional expectation of the consumer at each history node. This is the reason the properties of the stochastic process are important in the characterization. A key assumption of the model is

that types are positively correlated.¹⁹ When this is the case, a "high" type has not only a higher marginal valuation for the good today, but also a higher expected valuation for a contract in the future. Without this assumption, a type would be "high" or "low" depending on which of these two components of utility prevails. Along with the rest of the literature on dynamic contracting,²⁰ we also assume that at any point in time the type θ_t can assume one of two values. When there are n possible values, the conditional distribution of future realization of θ_t is a $n - 1$ dimensional vector, so the characteristics of each agent are $n - 1$ dimensional. In this case, besides the problem of dynamic screening, we would have an additional problem of multidimensional screening. As is well known, in this case types are not "naturally" ordered, and the set of constraints that are binding can be more complicated. The environment studied above has the advantage of separating the study of the dynamics of the contract from the study of the multidimensionality of the types, which is a conceptually distinct problem, and therefore provides a better understanding of the dynamics.²¹

A related issue regards the transition probabilities in the stochastic process. Clearly, many different assumptions can be made regarding these probabilities. In this paper, we have considered the case in which the transition probabilities do not change over time. This, however, is not essential for the characterization: indeed even if the degree of positive correlation changes over time (but remains positive), we would be able to perform the same simplification of the incentive constraints as in Lemma 1.²² Another assumption of the model is that the transition probabilities between types

¹⁹ This assumption is used in the characterization of the optimal contract, but not in Proposition 4.

²⁰ See, e.g., Hart and Tirole (1988), Laffont and Tirole (1990), and Rustichini and Wolinsky (1995).

²¹ At the cost of higher complexities, the model can be extended to multidimensional types. Indeed, even in a multidimensional environment, types can be "endogenously" ordered to simplify the set of incentive constraints (see, for example, Jean-Charles Rochet, 1987).

²² In this case the optimal contract would not depend on the likelihood ratio Λ raised to the t as in Proposition 2, but on the multiplication of the changing likelihood ratios along the lowest history. For this extension, we would also need to continue assuming that a high type remains more likely to be high in the future.

¹⁸ It is worthwhile to point out the differences between this result and the results in Proposition 1 and 2 because the logic of their proofs is different. The proof of Proposition 4 does not require the assumption that types are positively correlated. For this reason, Proposition 4 is stronger than the result that would have followed from taking the limit in the formula of Proposition 1 as $\delta \rightarrow 1$. However, while Proposition 1 characterizes the optimal contract for any δ , Proposition 4 is only a limit result. Even in the limit case in which $\delta \rightarrow 1$, Proposition 4 shows that the contract converges to an efficient contract *in probability*, but it is silent on the behavior of the contract in any single history.

are all positive, although they may be small: this precludes a case in which there is a type i which will never become some other type j .²³ An interesting extension of the model could be to consider a process with more than two types and a form of long-term heterogeneity in which different transition probabilities correspond to each initial type. A systematic analysis of the properties of the stochastic process and the extension to the case of dynamic screening with multidimensional types is left for future research.

Regarding the utility and the cost function, the results can easily be extended to the case in which the cost function is a generic convex function and utility is a generic function $u(\theta, q)$, provided that the usual single crossing condition is assumed.²⁴ A relevant assumption, however, is that the utility is quasilinear (as generally assumed in the literature on nonlinear pricing). When the utility function is not quasilinear, we have an additional issue of consumption smoothing over time. In this case, too, the analysis of the quasilinear case allows us to separate the dynamic screening problem from the conceptually different problem of consumption shooting.²⁵

As standard in the literature, we do not impose cash constraints on the consumer's choice.²⁶ At the cost of additional complications, it would be a simple exercise to incorporate these constraints in our model. Under plausible assumptions, however, these constraints would be irrelevant for the analysis. In the model, monetary transfers can always be bounded above in all periods by a finite upper

bound which is generally very small. For example, the upper bound on monetary transfers depends (among other variables) on the persistence of agents' types: as persistence converges to one, the per-period payments converge to the same payments as in the static model. If we assume that cash constraints are satisfied in the standard static version of the model, then when types are sufficiently persistent (as perhaps reasonable to assume when consumption is frequent), cash constraints would not be binding in our dynamic model either.

Finally, we turn to the time horizon. Besides a direct theoretical interest, the analysis of a stationary model with infinite periods is useful for two reasons. First, with this assumption we can study long-term behavior and convergence of the contract, which would be impossible in a two-period model. It is also instrumental, however, in the study of price dynamics. For example, we will show that the transfer price of the monopolist's technology is declining over time. Since the model is stationary, the true value of the technology is constant and identical in any period. Therefore, this decline in price arises purely for strategic reasons: in a nonstationary model with finite periods we would not be able to separate the strategic effect from the natural decline in value due to the shorter horizon. It is, however, easy to show that our characterization would be valid even in a model with T periods.

IV. Property Rights

Before presenting results on the monetary transfers, it is useful to discuss property rights, since their allocation typically (although not necessarily) influences the flow of monetary transfers. Up to this point, we have assumed that the monopolist has the right to decide the quantity supplied in every period. Instead of selling output on a period-by-period basis, however, the monopolist may decide to sell the property rights of her exclusive technology to the consumer. Only the consumer benefits directly from the technology and has information for its efficient use. It is therefore natural to expect that the property rights are ultimately acquired by the agent who has a superior valuation of its future use. The decision to transfer property rights, however, depends on the history of the agent's types:

²³ This environment, however, can be approximated since transition probabilities can be arbitrarily small.

²⁴ This requires the cross derivative $u_{\theta q}$ to be positive. When the utility and cost functions are generic functions, however, the first-order conditions do not necessarily yield closed form solutions. All the results, however, continue to hold in this more general environment (see Battaglini and Coate, 2003, for details).

²⁵ The results, however, are robust to changes in the degree of risk aversion. In Battaglini and Coate (2003) we show that when risk aversion is below a critical value, the characterization with risk aversion is the same as the characterization without risk aversion and a small change in risk aversion would imply only a small change in the contract.

²⁶ Cash constraints limit the maximum amount of per-period monetary transfers between the principal and the agent. Clearly, such constraints can be incorporated in both a dynamic model and a static model.

PROPOSITION 5: *Without loss of generality, the optimal contract offers a call option to buy out the firm to the agent as soon as he reveals himself to be a high type. However, the monopolist never finds it optimal to sell the firm to an agent who has always revealed himself to be a low type.*

The first part of this result should not be surprising. After the agent reveals himself to be a high type, there is no residual asymmetric information. At this stage, and before the realization of the type in the following period, we should expect no reason for the monopolist to keep the ownership of the technology.²⁷ The interesting observation, however, is in the second part of the proposition: after a history in which the agent has never revealed he is a high type, the monopolist finds it strictly suboptimal to sell the firm and prefers to introduce a distortion in the value of the firm, not only in the period in which the type is revealed, but also in the subsequent periods.²⁸ Indeed, as we discussed in Section III A, the distortion is introduced to extract surplus from the high type. This suggests that it is natural to observe a distortion in the period in which the agent reveals his type. But this does not explain why the monopolist still wants to introduce a distortion in the following periods: given that the agent has revealed his low type, there is no asymmetric information anymore in this case either. This characteristic of the optimal contract depends on the dynamic nature of the incentive constraint and it is instructive to see why it is true.

Consider a simple two-period example. Assume that after the declaration in period 1 the monopolist sells the firm to the agent irrespective of the type. In the second period the agent would receive all the surplus, i.e., $\frac{1}{2} \theta_H^2$ if he is a high type and $\frac{1}{2} \theta_L^2$ if he is a low type. This implies that the high type receives a rent, i.e., an extra payoff with respect to the lower type,

equal to $R_{\text{own}} = \frac{1}{2} \Delta \theta (\theta_H + \theta_L)$. This rent is higher than the minimal rent that would guarantee truthful revelation: the incentive compatibility constraint requires only a rent equal to $R_{IC} = \Delta \theta \theta_L < R_{\text{own}}$. Imagine now that the monopolist, after the agent reveals himself to be a low type, keeps the ownership in order to reduce the rent of the high type at $t = 2$, instead of selling the firm. Assume, in particular, that instead of selling the good at cost in the second period, she sells to the high type θ_H units at price $\frac{1}{2} \theta_H^2 + \varepsilon$, i.e., she reduces the extra rent of the high type by ε in case in period 1 the agent declares to be a low type. For ε small, the contract remains incentive compatible in the second period. In order to satisfy the constraints at $t = 1$, suppose that the monopolist reduces the price paid by the low type by $\delta \Pr(\theta_H | \theta_L) \varepsilon$ dollars. The low type's incentives in period 1 are unchanged: if he reports himself to be a low type, he receives $\delta \Pr(\theta_H | \theta_L) \varepsilon$ dollars more in $t = 1$ and he expects to receive $\delta \Pr(\theta_H | \theta_L) \varepsilon$ less at $t = 2$; moreover, the contract does not change if the agent chooses to report himself to be a high type. Consider now the impact of this change on the incentive compatibility constraint of the high type at $t = 1$. If the high type deviates and reports himself to be a low type, he receives $\delta \Pr(\theta_H | \theta_L) \varepsilon$ more, the same as the low type since this is paid "in cash" at $t = 1$ with a reduced price. However, the expected loss for the high type is $\delta \Pr(\theta_H | \theta_H) \varepsilon$ because he is more optimistic than the low type about the future. Since $\delta [\Pr(\theta_H | \theta_L) - \Pr(\theta_H | \theta_H)] \varepsilon$ is negative, this implies that the outside option of the high type, i.e., the utility of reporting untruthfully, has a lower value and the monopolist can induce truthful revelation by leaving a lower rent to the high type at $t = 1$. The monopolist, therefore, prefers to keep strict ownership of the firm: *ownership enables control of the rent of the agent in the second period, and this control is important to extract surplus in the sale of the technology to the high type in the first period.* The characterization of the optimal contract in (2) goes beyond this observation. In our infinite and stationary environment, in fact, the monopolist finds it optimal to reduce the efficiency of the firm for potentially infinite periods, until she hears a "high-type" report. Moreover, as we will prove in Section V, the dynamics of the transfer price of the technology will be dictated by the dynamics of the optimal inefficiency in supply.

²⁷ Note that this result is different from the classical results by Jeremy Bulow (1982) concerning the trade-off between the sale and the rental of a durable good. In this literature, in fact, if a durable good is sold, then the quantity remains constant in the following periods; in our framework, instead, the firm is selling the *technology* to produce the good, and the future quantities depend on the realized type.

²⁸ I am grateful to Bengt Holmström and Asher Wolinsky who have independently suggested this point.

V. The Dynamics of Monetary Payments

As mentioned above, two payment schedules with the same present value can give the same incentives to an agent; therefore the prices charged in the optimal contract are not uniquely identified. Indeed, although it is true that we can assume without loss of generality that the optimal contract keeps the lowest type at his reservation utility in any period, we can construct equilibrium contracts that do not have this feature: an example is the contract in which the monopolist sells the technology to the agent. In general, when we have many periods, we can find optimal contracts in which the monopolist receives a large payment at some date t and she commits to pay it back in installments. The installments can, in principle, follow any time pattern. In this section, we focus on two types of optimal monetary transfers that seem more interesting from a theoretical and empirical point of view.

For any contract in which the monopolist borrows money and repays it in an arbitrary time pattern, we can distinguish two parts: a *supply contract* in which the relevant $IC_{h_t}(\theta_H)$ and $IR_{h_t}(\theta_L)$ constraints are satisfied as equalities, and a residual *lending contract*, in which the monopolist borrows some amount of money and pays it back over time to the agent. A reason why the supply contract is more interesting than other contracts is that if we assume that the monopolist is even slightly more patient than the agent, she would never find it optimal to ask the agent to anticipate payments for future supply (as occurs when the technology is sold to the agent), and therefore all the constraints would be binding after all histories.²⁹ The lending contract can take (almost) any form because the monopolist can commit to repay it according to any time pattern. The supply contract, however, is uniquely determined by the

incentive structure of the model. We can therefore ask what is the dynamics of prices and, more importantly, the dynamics of the consumer's utility in the optimal supply contract.

There is one particular case in which the monopolist receives anticipated payments from the consumer, which has special significance from an empirical and theoretical reason: the contract discussed in Section IV, in which the monopolist, as soon as compatible with profit maximization, sells the firm to the consumer who reports himself to be a high type. We call this arrangement the *sale-of-the-firm contract*. In this case, too, the monopolist can add on top of a sale-of-the-firm contract a lending contract, as defined above, in which she borrows more money than the value of the firm and repays the extra amount over time. Since we are not interested in this case, we assume without loss of generality that the $IR_{h_t}(\theta_L)$ constraint is satisfied as equality in all periods. Again, if this condition is satisfied, the sale-of-the-firm contract is uniquely determined. The interesting question in this case is the dynamics of the strike price of the call option on the technology.³⁰

Regarding the supply contract, we have:

PROPOSITION 6: *In the optimal supply contract, the average per-period utility of the agent starting from any date t is nondecreasing in t in all possible histories and strictly increasing in some history; therefore, the expectation at time zero of the average rent of the agent from date t is strictly increasing in t .*

Recent empirical work has highlighted that in some important markets long-term contracts are front-loaded: prices are initially high and decline over time.³¹ A consequence of this effect, therefore, is that the expected utility of a consumer from continuing to remain a monopolist's customer increases over time. Dionne and Doherty (1994) explain this phenomenon as a consequence of the possibility to renegotiate contracts over time. Building on Milton Harris and Holmström (1982), Hendel and Lizzeri

²⁹ If the monopolist is more patient than the agent, then the incentive compatibility for the low type would be binding in all periods and the monopolist would not find it optimal to lend money to the agent because the agent would not be able to commit to repay the debt. Note that the monopolist's objective function is continuous in her discount factor and the constraints do not depend on it: therefore, an infinitesimal reduction in the monopolist's discount factor would have only an infinitesimal impact on the optimal quantities $q_A(\theta_{h_t})$, but would eliminate equilibria in which the monopolist borrows money.

³⁰ I am grateful to William Rogerson who suggested this point.

³¹ This effect has been shown with California auto insurance data by Dionne and Doherty (1994), and more recently by Hendel and Lizzeri (2003) in the life insurance market.

(2000) present a model with no asymmetric information, but in which both the principal and the agent learn over time from a public signal the type of the agent: front-loading is therefore a consequence of reclassification risk. Our model suggests a new explanation for this phenomenon in which front-loading is precisely a consequence of the commitment power of the seller.³² Indeed, in the optimal contract, even without imposing renegotiation constraints, she finds it optimal to promise an efficient contract to the agent if he reports himself to be a high type, or to provide a contract with decreasing inefficiency. Because of this, she must commit to pay a rent to the high type that increases over time because the higher the efficiency of the contract, the more expensive it is to separate the agents' types.

We now turn to sale-of-the-firm contracts. As we discussed in Section IV, the monopolist finds it optimal to sell her technology only if the agent reveals himself to be a high type. It is therefore natural to look at the evolution of the call price of the option to buy the technology along the history in which it can be exercised (i.e., when the agent always reveals himself to be a low type). This question is interesting because, given the stationary structure of the model, the present value of the firm along this history is constant. Remember that the model has infinite periods, and because the preferences of the consumer follow a Markov process, the value of the firm depends only on the current state of the consumer's type.³³ This fact may suggest that the price of the firm is constant over time. We have, however:

PROPOSITION 7: *In the optimal sale-of-the-firm contract, the strike price of the call option to buy out the technology is strictly declining over time.*

What really matters in the determination of the transfer price of the firm is the outside

option of the high type (i.e., the value of reporting untruthfully). This outside option changes over time because the contract becomes increasingly efficient along the "lowest" history, and the improvements in the contract benefit the high type more than the low type. The higher efficiency of the contract, in fact, increases the agent's utility in the event in which he turns into a high type, and an agent who is a high type today has a higher probability of being a high type tomorrow. For this reason, the price for the service that the low type is willing to pay increases more slowly than the increase in utility of a deviation for a high type. For this reason, the outside option of the high type increases over time. This implies that the only way for the monopolist to induce a truthful revelation is to reduce the strike price of the call option on the property rights of the firm.

VI. Renegotiation-Proofness

So far we have assumed that the monopolist can commit to a contractual offer. We discussed this point above, arguing that this is the most appropriate assumption in many environments, in particular when the monopolist is serving many consumers and is interested in maintaining her reputation, or when renegotiation costs are larger than the benefits. There are situations, however, in which the seller cannot commit not to renegotiating the contract after some histories. The received literature has shown that if types are constant, the optimal contract is never renegotiation-proof. Perhaps surprisingly, given a condition that is easily satisfied, this is no longer true when the agent's type follows a stochastic process.³⁴ We say that a contract is renegotiation-proof if after no history h_t there is a new contract starting in period t that the consumer would accept in exchange for the original contract, and that is strictly superior for the monopolist. This definition is standard in the

³² In our model, we assume that the seller has monopoly power. Clearly, this assumption should be taken into consideration when using the model to interpret evidence in markets with more competitive environments.

³³ In an equilibrium of a direct mechanism, the consumer would reveal his type truthfully; therefore the firm would be sold only to high types. This implies that whenever the firm is sold, the expected present value of the firm is constant, irrespective of the period in which it is sold.

³⁴ This result should not be confused with the findings in Patrick Rey and Bernard Salanie (1990, 1996). Beside the fact that they consider a different model with moral hazard and constant types, these authors show that a renegotiation-proof contract can be implemented by a chain of short-term contracts (two-period contracts in which the principal can commit). They do not characterize the renegotiation-proof contract and they do not prove that the ex ante optimal contract is renegotiation-proof (which indeed would not be true in their models with constant types).

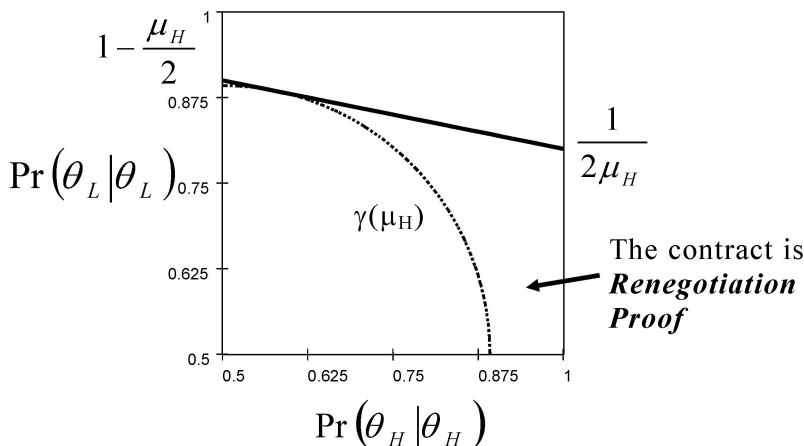


FIGURE 2. THE RENEGOTIATION-PROOFNESS CONDITION

literature and natural: when a contract is renegotiation-proof, either the monopolist or the consumer would reject a revision of the initial agreement.

PROPOSITION 8: *The optimal contract is renegotiation-proof if $\Pr(\theta_L|\theta_L) \leq 1 - \mu_H \Pr(\theta_H|\theta_H)$. Moreover, if this condition is not satisfied, there exists a $t < \infty$ such that the contract is not renegotiation-proof only in the first t periods.*

Figure 2 represents the condition of Proposition 7 in an example in which there is a 30-percent initial probability that the agent is a high type: the contract is renegotiation-proof for any point below the straight “thick line.” As can be seen from the figure, the set of parameters for which the contract is renegotiation-proof covers most of the set of feasible parameters. (Because types are positively correlated, $\Pr(\theta_L|\theta_L)$ and $\Pr(\theta_H|\theta_H)$ are both larger than $1/2$.)³⁵

The intuition of this result can be seen from Figure 3.³⁶ Assume, for the sake of illustration, that at time t the monopolist is contem-

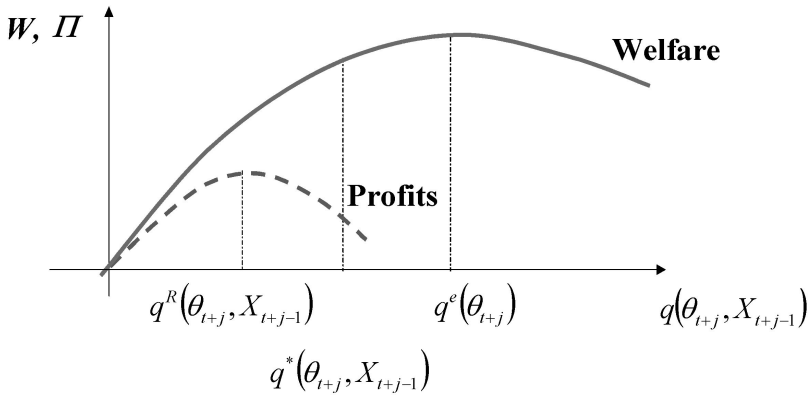
plating a change only in the quantity offered after the agent reports a type θ_{t+j} after a history X_{t+j-1} , keeping constant all other quantities. The complete argument clearly needs to consider a change on the entire sequence of contingent menus: this is a more sophisticated problem (solved in the Appendix), but this example provides a useful intuition. First, note that a contract is renegotiated only by a contract that is Pareto superior; otherwise either the seller or the buyer would not accept the change. Since welfare is strictly concave in the quantity supplied, and the ex ante optimal quantity at time t (i.e., $q^*(\theta_{t+j}, X_{t+j-1})$) is not larger than the efficient output $q^e(\theta_{t+j})$, it must be that welfare is strictly increasing in the interval $[0, q^*(\theta_{t+j}, X_{t+j-1})]$ (see Figure 3). This implies that the new output $q'(\theta_{t+j}, X_{t+j-1})$ prescribed by a renegotiated contract must be strictly larger than $q^*(\theta_{t+j}, X_{t+j-1})$.

Consider now a history h_t . If the agent has previously reported himself to be a high type, then the contract is efficient and not renegotiable; assume therefore that the agent has always reported himself to be a low type. If after h_t the monopolist could choose any other continuation contract, the quantity supplied after a history h_{t+j} following or equal to h_t in the new contract would be $q^R(\theta_{t+j},$

³⁵ If we measure the persistence of types by $\gamma = \max\{\Pr(\theta_L|\theta_L), \Pr(\theta_H|\theta_H)\}$, an immediate implication of Proposition 8 is that, given any initial prior, there is an upper bound $\gamma(\mu_H)$ on persistence such that for $\gamma \leq \gamma(\mu_H)$ the optimal contract is renegotiation-proof (area below the semicircle in Figure 2).

³⁶ The two concave functions in the figure represent the profits and welfare generated in period $t+j$ after history x_{t+j-1} : $q^*(\theta_{t+j}, X_{t+j-1})$ is the optimal contract, $q^e(\theta_{t+j})$ is

the efficient contract, and $q^R(\theta_{t+j}, X_{t+j-1})$ is the contract that is ex post optimal after history h_t .

FIGURE 3. THE EX POST MAXIMIZATION PROBLEM AFTER A HISTORY h_t

$X_{t+j-1}) = \theta_{t+j} - \Delta\theta[\Pr(\theta_H|\theta_L)/\Pr(\theta_L|\theta_L)]X(\theta_{t+j}, X_{t+j-1})\Lambda^{j-1}$. This formula is identical to the formula of the ex ante optimal contract in Proposition 2, except that instead of the prior μ_H , we use the appropriate posterior after h_t , $\Pr(\theta_H|\theta_L)$, and the state X_{t+j} is started afresh at time t . Comparing $q^R(\theta_{t+j}, X_{t+j-1})$ with the contract in Proposition 2, it is easy to verify that the ex ante optimal $q^*(\theta_{t+j}, X_{t+j-1})$ is larger than the ex post optimal level $q^R(\theta_{t+j}, X_{t+j-1})$ if

$$(4) \quad \frac{\mu_H}{\mu_L} \Lambda^{t+j-1} \leq \frac{\Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)} \Lambda^{j-1},$$

which is always satisfied if $\Pr(\theta_L|\theta_L) \leq 1 - \mu_H\Pr(\theta_H|\theta_H)$ since $t > 1$. (Obviously, the contract can be renegotiated only starting from the second period.) Because the profit function is also strictly concave, this implies that when (4) holds, any quantity larger than $q^*(\theta_{t+j}, X_{t+j-1})$ reduces expected profits at h_t . But then any change that would be accepted at h_t by the customer would necessarily reduce profits, implying that the quantity $q^*(\theta_{t+j}, X_{t+j-1})$ would not be renegotiated at any time t .

When types are constant, the optimal renegotiation-proof contract requires the consumer to play sophisticated mixed strategies, and this may appear unrealistic. These strategies are necessary to guarantee that, after any possible history, the monopolist's posterior beliefs are such that there are no ex post Pareto superior contracts. The result presented above, however, shows that when types are correlated

but follow a stochastic process, even if the correlation level is very high (as in Figure 2), consumers do not need to use mixed strategies in equilibrium, but simply truthfully report their type. The conflict between optimality and renegotiation-proofness, and the sophistication of equilibrium strategies necessary to guarantee the latter property, therefore, are implications of the assumption that types are constant or very highly correlated.³⁷

VII. Conclusion

This paper shows that a long-term contractual relationship in which the type of the buyer is constant over time is qualitatively different from a contractual relationship in which the type follows a Markov process, even if the types are highly persistent. While in the first case the contract is constant, in the second the contract is truly dynamic and converges to the efficient contract. Even if the environment has only one-period memory and risk-neutral agents, the optimal contract is not stationary and has unbounded memory. The structure of the optimal contract, however, is remarkably simple. In analogy with the static model, we have a stronger version of the *generalized no distortion at the top* principle, which implies that the entire state-contingent contract becomes forever efficient as soon as the agent reports himself to be

³⁷ A complete analysis of the mixed strategy equilibrium in the optimal renegotiation-proof contract with variable types when (4) is not satisfied is presented in Battaglini (2005).

a high type. In our dynamic setting, however, we also have a novel *vanishing distortion at the bottom* principle which clearly could not be appreciated in a static model.

With constant types there always is a conflict between optimality and renegotiation-proofness, and the latter property is guaranteed only if consumers use sophisticated mixed strategies. With stochastic types, in contrast, even if there is high persistence, the optimal contract is renegotiation-proof for natural specifications of the parameters. Consumers, moreover, adopt simple pure strategies.

The dynamic theory of contracting presented in the paper also provides insights into the ownership structure of the monopolist's exclusive technology and contributes to explaining some empirical findings. The monopolist may find it optimal to keep the ownership of the technology even when it would be inefficient in order to control the agents' future rents and therefore maximize rent extraction. This inefficient retention of property rights may potentially last for an arbitrarily large number of periods, but the allocation of property rights will be efficient with probability one in the long term.

APPENDIX

PROOF OF LEMMA 1:

For a generic maximization program $\mathcal{P}$, we define $\mathcal{V}(\mathcal{P})$, the value of the objective function at the optimum. Let us also define $\mathcal{P}_I^R$ the program in which expected profit is maximized only under the incentive compatibility constraints of the high type and the individual rationality constraints of the low type: $IC_{h_t}(\theta_H)$, $IR_{h_t}(\theta_L) \forall h_t$. We proceed in two steps.

CLAIM 1:

If $\langle \mathbf{p}, \mathbf{q} \rangle$ solves $\mathcal{P}_I^R$, then there exists a $\mathbf{p}'$ such that $\langle \mathbf{p}', \mathbf{q} \rangle$ satisfies $IR_{h_t}(\theta_L)$ and $IC_{h_t}(\theta_H)$ as equality $\forall h_t$, and achieves the same value as $\langle \mathbf{p}, \mathbf{q} \rangle$: $\mathcal{V}(\mathcal{P}_I^R) = \mathcal{V}(\mathcal{P}_H)$.

PROOF:

Given a solution $\langle \mathbf{p}, \mathbf{q} \rangle$, we first show that the price vector can be modified to guarantee that all the incentive constraints are satisfied as equalities without reducing the value of the program; then we show that the resulting contract with incentive constraints satisfied as equalities can be modified to make the individual rationality con-

straints satisfied as equalities as well, while leaving the incentive constraints untouched and keeping the value of the program constant.

Step 1: We show the result by induction. Let t be a finite integer. Assume that for any solution $\langle \mathbf{p}, \mathbf{q} \rangle$ of $\mathcal{P}_I^R$ and any $t' \leq t$, there is a $\mathbf{p}'$ such that $\langle \mathbf{p}', \mathbf{q} \rangle$ is also a solution of $\mathcal{P}_I^R$; and all incentive compatibility constraints are satisfied as equalities up to period t' , the value of the objective function is unchanged, and $\mathbf{p}'$ is identical to $\mathbf{p}$ in any period $j > t'$: $p'(\theta; h_j) = p(\theta; h_j)$ for $j > t'$. This step is clearly satisfied for $t = 1$, since in this period the incentive compatibility constraint is necessarily binding at the optimum. We now show that if $\mathbf{p}'$ exists $\forall t' \leq t$, then there must be a price vector $\mathbf{p}^{t+1}$ such that all the incentive compatibility constraints are satisfied as equalities up to period $t + 1$, $\mathbf{p}^{t+1}$ is identical to $\mathbf{p}$ in any period $j > t + 1$, and the value of the objective function is unchanged. Given the induction step, assume without loss of generality that the incentive constraints are satisfied as equalities for any $j \leq t$. There are two cases to consider. Assume first that at period $t + 1$, after a history $h_{t+1} = \{h_t, H_L\}$, the high type receives an expected continuation utility equal to the utility he would receive if he declares to be a low type plus a constant $\varepsilon > 0$. Modify the contract so that the new prices after histories $\{h_{t+1}, \theta_H\}$ and $\{h_{t+1}, \theta_L\}$ are respectively $p^{t+1}(\theta_H; h_{t+1}) = p(\theta_H; h_{t+1}) + \varepsilon$ and $p(\theta_L; h_{t+1}) = p(\theta_L; h_{t+1})$; ³⁸ simultaneously, reduce the price after history $\{h_t, \theta_L\}$ so that $p^{t+1}(\theta_L; h_t) = p(\theta_L; h_t) - \delta \Pr(\theta_H | \theta_L) \varepsilon$. We call this new price vector $\mathbf{p}^{t+1}$. This change would not reduce the monopolist's expected profit, it would not violate $IR_{h_t}(\theta_L)$ in any period, it would not violate the incentive compatibility constraints for histories following h_t , and it would satisfy $IC_{h_{t+1}}(\theta_H)$ as equality. Consider now the $IC_{h_t}(\theta_H)$ constraint at h_t . The utility of a high type that is truthful $U(\theta_H; h_t)$ is unchanged; if the high type reports himself to be a low type, however, he would receive:

$$(A1) \quad U(\theta_H; h_{t-1}) + \delta(\Pr(\theta_H | \theta_L) - \Pr(\theta_H | \theta_H)) \varepsilon \leq U(\theta_H; h_{t-1})$$

³⁸ Remember that $p(\theta; h_t)$ is the price charged after history h_t if the agent declared to be a type θ , so it is the price charged after a history $\{h_t, \theta\}$.

where the inequality follows from the fact that types are positively correlated. It follows that $IC_{h_j}(\theta_H)$ are satisfied for any $j \leq t$, and $\langle \hat{\mathbf{p}}^{t+1}, \mathbf{q} \rangle$ is a solution of $\mathcal{P}_I^R$. By the induction step we can find a new price vector $\mathbf{p}^{t+1}$ which is such that the incentive compatibility constraints are satisfied as equalities in all periods $j \leq t$, and that is identical to $\hat{\mathbf{p}}^{t+1}$ for periods $j > t$ (which is identical to $\mathbf{p}$ in periods $l > t + 1$). Since this change does not affect prices at $t + 1$ and following periods, in correspondence to $\mathbf{p}^{t+1}$ the incentive compatibility is satisfied as equality $\forall t \leq t + 1$, prices are equal to the prices in $\mathbf{p}$ for periods $j > t + 1$, and the value function is equal to the original.

Assume now that at period $t + 1$, after some history $h_t = \{h_{t-1}, \theta_H\}$, the high type receives a utility equal to the utility he would receive if he declares to be a low type plus a constant $\varepsilon > 0$. Modify the contract so that the new prices after histories $\{h_t, \theta_H\}$ and $\{h_t, \theta_L\}$ are respectively $p^{t+1}(\theta_H; h_t) = p(\theta_H; h_t) + \varepsilon$ and $p^{t+1}(\theta_L; h_t) = p(\theta_L; h_t)$; simultaneously, reduce prices after history $\{h_t, \theta_H\}$ so that $p^{t+1}(\theta_H; h_{t-1}) = p(\theta_H; h_{t-1}) - \delta \Pr(\theta_H | \theta_H) \varepsilon$. This new contract $\mathbf{p}^{t+1}$ would leave all the constraints of the relaxed problem satisfied $\forall h_t$ and the incentive constraint satisfied as equalities in the first $t + 1$ periods. And it would not reduce profits.

Step 2: By the previous step we can assume without loss of generality that all the incentive compatibility constraints are satisfied as equalities. We now show that the individual rationality constraints can also be reduced to equalities. It can be verified that the individual rationality constraint must be binding at $t = 1$. Again, we prove the result for the remaining periods by induction. Assume that in all periods $j \leq t$, $IR_{h_t}(\theta_L)$ holds as equality and that the expected utility of a low type agent after history h_{t+1} is $\kappa > 0$. Consider an increase by κ of the prices charged in the period $t + 1$, $p^{t+1}(\theta; h_t, \theta_i) = p(\theta; h_t, \theta_i) + \kappa \forall \theta$; and a reduction of the price at time t so that $p^{t+1}(\theta; h_{t-1}) = p(\theta; h_{t-1}) - \delta \kappa \forall \theta$. Clearly, this change would not violate the constraints of $\mathcal{P}_I^R$, would leave the incentive compatibility constraints untouched, and would satisfy all the individual rationality constraint as equality up to period $t + 1$. Profit would remain unchanged as well.

We now prove:

CLAIM 2:

Any solution of $\mathcal{P}_{II}$ satisfies all the constraints of $\mathcal{P}_I$, and $\mathcal{V}(\mathcal{P}_I^R) = \mathcal{V}(\mathcal{P}_I)$.

PROOF:

Since $\mathcal{V}(\mathcal{P}_I^R) = \mathcal{V}(\mathcal{P}_{II})$, we need only show that, in correspondence to the solution of $\mathcal{P}_{II}$, after any history h_t the low type does not want to imitate the high type (i.e., the $IC_{h_t}(\theta_L)$ constraint is satisfied) and the high type receives at least his reservation value (i.e., the $IR_{h_t}(\theta_H)$ constraint is satisfied). This guarantees that $\mathcal{V}(\mathcal{P}_I) \geq \mathcal{V}(\mathcal{P}_I^R)$ and hence the result.

Step 1: the $IC_{h_t}(\theta_L)$ constraints. Note that by $IC_{h_t}(\theta_H)$ and $IR_{h_t}(\theta_L)$, after any history h_t :

$$\begin{aligned} p(\theta_H; h_t) - p(\theta_L; h_t) = & \\ & (q(\theta_H; h_t) - q(\theta_L; h_t))\theta_L \\ & + \delta \Pr(\theta_H | \theta_L) \Delta U(\theta_H, h_t) \\ & + (q(\theta_H; h_t) - q(\theta_L; h_t))\Delta \theta \\ & + \delta (\Pr(\theta_H | \theta_H) - \Pr(\theta_H | \theta_L)) \Delta U(\theta_H, h_t) \end{aligned}$$

where $p(\theta; h_t) i = H, L$ is the price charged after the agent declares to be a type i and $\Delta U(\theta_H, h_t) = U(\theta_H; h_t, \theta_H) - U(\theta_H; h_t, \theta_L)$, the difference between the rent of a high type after a θ_H and a θ_L declaration (the continuation value of a low type is zero in $\mathcal{P}_{II}$). As can be seen from (2), in correspondence to the solution of $\mathcal{P}_{II}$,³⁹ after an agent declares to be a high type an efficient contract is offered in the optimal solution of the relaxed problem; so using $IC_{h_t}(\theta_H)$ and $IR_{h_t}(\theta_L)$ we can write:

$$\begin{aligned} U(\theta_H; h_t, \theta_H) = \Delta \theta \sum_{j=0}^{\infty} \delta^j (\Pr(\theta_H | \theta_H) \\ - \Pr(\theta_L | \theta_H)) q^e(\theta_L) \end{aligned}$$

where, remember, $q^e(\theta_L)$ is the efficient quantity when the type is θ_L . If the agent reports himself to be a low type, on the contrary, he will receive an inefficient quantity $q^*(\theta_L | h)$ that is never strictly higher than the efficient level $q^e(\theta_L)$: therefore his continuation value is not higher than $U(\theta_H; h_t, \theta_H)$.

³⁹ The formal derivation of (2) is in the proof of Proposition 1 below.

So $U(\theta_H; h_r, \theta_H) - U(\theta_H; h_r, \theta_L) \geq 0$ for any h_r ; and, since types are positively correlated, $(\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L))\Delta U(\theta_H, h_r) \geq 0$. It follows that:

$$p(\theta_H; h_t) - p(\theta_L; h_t) \geq (q(\theta_H; h_t) - q(\theta_L; h_t))\theta_L + \delta \Pr(\theta_H|\theta_L)\Delta U(\theta_H, h_t)$$

which implies that $IC_{h_t}(\theta_L)$ is satisfied at h_t .

Step 2: the $IR_{h_t}(\theta_H)$ constraints. By $IC_{h_t}(\theta_H)$ and $IR_{h_t}(\theta_L)$ we have:

$$\begin{aligned} q(\theta_H; h_t)\theta_H - p(\theta_H; h_t) \\ + \delta \Pr(\theta_H|\theta_H)U(\theta_H; h_t, \theta_H) &\geq \delta(\Pr(\theta_H|\theta_H) \\ - \Pr(\theta_L|\theta_L))U(\theta_H; h_t, \theta_L) &> 0 \end{aligned}$$

and therefore $IR_{h_t}(\theta_H)$ is satisfied as well.

We can now prove Lemma 1. Assume that $\langle \mathbf{p}, \mathbf{q} \rangle$ solves $\mathcal{P}_H$; then, by Claim 1 and 2 it must also solve $\mathcal{P}_L$. Assume that $\langle \mathbf{p}, \mathbf{q} \rangle$ solves $\mathcal{P}_L$; then, by Claim 2 it must also solve $\mathcal{P}_H^R$, since $\mathcal{V}(\mathcal{P}_H^R) = \mathcal{V}(\mathcal{P}_L)$. By Claim 1 there exists a $\mathbf{p}'$ such that $\langle \mathbf{p}', \mathbf{q} \rangle$ solves $\mathcal{P}_H$ and achieves the same value as $\mathcal{P}_L$. We conclude that $\mathbf{q}$ solves $\mathcal{P}_L$ if and only if it solves $\mathcal{P}_H$.

PROOF OF PROPOSITION 1:

Let us define $h_t^L := \{\theta_L, \theta_L, \dots, \theta_L\}$, the history along which the agent always reports himself to be a low type for $t - 1$ periods. Using the fact that $IC_{h_t}(\theta_H)$ and $IR_{h_t}(\theta_L)$ are equalities, we can formulate the utility of the high type at time 1 as:

$$\begin{aligned} U(\theta_H; h_1) &= \Delta\theta q_1^*(\theta_L|h_1) + \delta(\alpha_H - \alpha_L) \\ &\times \left(\Delta\theta q_2^*(\theta_L|h_2^L) + \delta(\alpha_H - \alpha_L) \begin{pmatrix} \Delta\theta q_3^*(\theta_L|h_3^L) + \dots \\ 0 \end{pmatrix} \right). \end{aligned}$$

This formula can be written as

$$\begin{aligned} (B2) \quad U(\theta_H; h_1) &= \Delta\theta \sum_{j=0}^{\infty} \delta^j (\Pr(\theta_H|\theta_H) \\ &- \Pr(\theta_H|\theta_L))^j q_{j+1}^*(\theta_L|h_{j+1}^L). \end{aligned}$$

Using the equality $IR_{h_1}(\theta_L)$ we know that the low type receives zero at time one. It follows that $\mathcal{P}_H$ can be represented as:

$$\begin{aligned} (B3) \quad E\Pi(\theta|h_1) &= \sum_{i=H,L} \mu_i \\ &\times \left[q(\theta_i)\theta_i - \frac{q(\theta_i)^2}{2} + \delta\alpha_i \left(\frac{W(\theta_H; \theta_i)}{W(\theta_L; \theta_i)} \right) \right] \\ &- \mu_H \Delta\theta \sum_{j=0}^{\infty} \delta^j (\Pr(\theta_H|\theta_H) \\ &- \Pr(\theta_H|\theta_L))^j q_{j+1}^*(\theta_L|h_{j+1}^L) \end{aligned}$$

where the first summation is the expected surplus generated by the supply contract $(W(\theta_i; \theta_j))$ is the expected social welfare generated by the contract from period 2 if the realization in period 1 is θ_j and the realization in period 2 is θ_i .

We have two possible cases:

Case 1: $h_t = \{h_{t-1}, \theta\} \in H, h_t^L \forall t \geq 1$. The first-order condition implies $q_t^*(\theta|h_t) = \theta$, and the contract is efficient.

Case 2: $h_t = h_t^L \forall t \geq 1$. The first-order condition with respect to a generic quantity offered along the lowest branch $q_t^*(\theta_L|h_t^L)$ implies that:

$$\begin{aligned} (B4) \quad q_t^*(\theta_L|h_t^L) &= \theta_L \\ &- \Delta\theta \frac{\mu_H}{\mu_L} \left[\frac{\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)} \right]^{t-1}, \end{aligned}$$

which completes the characterization of the optimal contract.

PROOF OF PROPOSITION 4:

Starting in period t from any history $\{h_t, \theta\}$, the expected first best surplus from time t onward is independent from t and equal to $W_t^*(\theta) = \sum_{j \geq t} \delta^{j-t} E_t[1/2 \theta_j^2 | \theta_t = \theta]$. Consider a contract $\mathbf{c}$ in which a fixed fee $F = W_1^*(\theta_L)$ is charged in period 1 and then an efficient menu plan in which $q_\tau(\theta) = \theta$, $p_\tau(\theta) = 1/2 \theta^2$ for any $\tau \geq 1$ is offered. This contract is clearly incentive compatible and individually rational for any $\tau \geq 1$; moreover, it is renegotiation-proof since it is efficient. Therefore it is a feasible option in the

monopolist's program, even if the renegotiation-proofness constraint must be satisfied, and must yield an average profit not larger than the profit of the optimal contract Π^* . This implies

$$(C5) \quad (1 - \delta)\Pi^* -$$

$$(1 - \delta)E\left[\sum_{\tau=1}^{\infty} \delta^{\tau-1} w^*(\theta_{\tau})\right] \geq$$

$$(1 - \delta)E\left[\sum_{\tau=1}^{\infty} \delta^{\tau-1} w^*(\theta_{\tau}) \middle| \theta_1 = \theta_L\right] -$$

$$(1 - \delta)E\left[\sum_{\tau=1}^{\infty} \delta^{\tau-1} w^*(\theta_{\tau})\right]$$

where $w^*(\theta)$ is the per-period Marshallian surplus when the type is θ . As $\delta \rightarrow 1$, the right-hand side can be written as

$$\Omega(\theta; \hat{t}) = \lim_{\delta \rightarrow 1} (1 - \delta)$$

$$\times \left\{ E\left[\sum_{\tau \geq \hat{t}} \delta^{\tau-1} w^*(\theta_{\tau}) \middle| \theta_1 = \theta_L\right] \right.$$

$$\left. - E\left[\sum_{\tau \geq \hat{t}} \delta^{\tau-1} w^*(\theta_{\tau})\right] \right\}$$

where $\hat{t}$ is a finite integer. Since (C5) must hold for any $\hat{t} \geq 1$ and $\lim_{\hat{t} \rightarrow \infty} \Omega(\theta; \hat{t}) = 0$ (because the process converges to a stationary distribution), we have that $\lim_{\delta \rightarrow 1} (1 - \delta)\Pi^*$ must be equal to $\lim_{\delta \rightarrow 1} (1 - \delta)E[\sum_{\tau=1}^{\infty} \delta^{\tau-1} w^*(\theta_{\tau})]$. This also implies that the agent's average payoff is zero.

PROOF OF PROPOSITION 5:

Since the optimal contract is efficient after the agent reveals himself to be a high type, the monopolist finds it optimal to offer to the consumer the same quantities that the consumer himself would choose if he could directly control supply. The first part of the result then follows from the fact that all players have quasilinear utilities and therefore they are indifferent

between paying or receiving a positive amount every period, or a large amount equal to the future expected payments at some period and zero afterward. The second part follows from the fact that from (2) we know that along the "lowest history," supply is distorted in the future with positive probability, and the monopolist cannot achieve these distortions without control of the technology.

PROOF OF PROPOSITION 6:

When $IR_{h_t}(\theta_L)$ is satisfied as equality for all h_t , the low type receives zero expected utility in all periods. Therefore we need only show that the average utility of the high type is nondecreasing in time. Using (2) and the incentive compatibility constraint of the high type we can write:

$$U(\theta_H; t, X_{t-1})$$

$$= \Delta \theta \sum_{j=0}^{\infty} \delta^j (\Pr(\theta_H | \theta_H) - \Pr(\theta_H | \theta_L))^j$$

$$\cdot \left[\theta_L - \Delta \theta X(\theta_L, X_{t+j-1}) \frac{\mu_H}{\mu_L} \right.$$

$$\left. \times \left(1 - \frac{\Pr(\theta_H | \theta_L)}{\Pr(\theta_L | \theta_L)} \right)^{t+j-1} \right]$$

where $U(\theta_H; t, X_{t-1})$ is the expected utility of a high type at time t given the state X_{t-1} . Consider now two periods: t and $t' < t$. It is easy to show that $U(\theta_H; t, X_{t-1}) - U(\theta_H; t', X_{t'-1})$ is proportional to $X(\theta_L, X_{t'-1}) - X(\theta_L, X_{t-1})[1 - (\Pr(\theta_H | \theta_L)/\Pr(\theta_L | \theta_L))]^{(t-t')}$, which is nonnegative because $X_{t-1} \leq X_{t'-1}$ and strictly positive if $X(\theta_L, X_{t'-1}) = 1$. Therefore, the average rent of the agent is nondecreasing in any history and strictly increasing in a nonempty subset of histories. It follows that, at time zero, the expected average rent starting from period t is strictly increasing in t .

PROOF OF PROPOSITION 7:

Since the monopolist's technology is sold as soon as compatible with profit maximization, its sale can occur only along a history in which the agent has always reported himself to be a low type. Consider any such history h_t . The price $P(h_t)$ paid for the technology by the high type is determined by the equation

$$(F6) \quad W^*(\theta_H) - P(h_t) = U(\theta_H, \theta_L; h_t)$$

where $U(\theta_i, \theta_j; h_t)$ is the utility of a type θ_i from declaring to be a type θ_j after a history h_t ; and $W^*(\theta_i)$ is the expected first best surplus from time t if the type at t is θ_i .⁴⁰ Since $W^*(\theta_H)$ is clearly history independent, the result follows by the fact that supply is increasing over time and therefore $U(\theta_H, \theta_L; h_t)$ is increasing (see [B2]).

PROOF OF PROPOSITION 8:

Consider the problem of ex post maximization faced by the monopolist after a history h_t with $t > 1$ in which the agent has never reported himself to be a high type. At this stage, expected profits can be written as:

$$(G7) \quad E[\Pi(\theta|h_t)|h_t] \\ = \Pr(\theta_H|\theta_L)[W(\theta_H, \mathbf{q}_H) - R(\mathbf{q}_L)] \\ + \Pr(\theta_L|\theta_L)W(\theta_L, \mathbf{q}_L)$$

where $\mathbf{q}_i$, $i = H, L$ is the sequence of quantities in the menus offered if the agent reports himself to be a type i at t ; $W(\theta_i, \mathbf{q}_i)$ is the expected surplus generated in the contract if the agent is of type i and $\mathbf{q}_i$ is offered; and $R(\mathbf{q}_L)$ is the expected rent of the high type starting from h_t which guarantees incentive compatibility. (By Lemma 1 it depends only on $\mathbf{q}_L$ as in [B2], and the rent of the low type is zero.) Indeed (G7) is a compact way to write (B3) when the posterior probability that the type is high starting from h_t is $\Pr(\theta_H|\theta_L)$. The monopolist's ex post problem ($\mathcal{P}_{ex\ post}^*$) consists of maximizing (G7) under the additional constraint that the expected rents of the agent are at least as high as the expected rents starting from h_t obtained keeping the original, ex ante optimal contract. It is, however, useful to consider the program ($\mathcal{P}_{ex\ post}^*$) in which (G7) is maximized under the additional constraint that *expected welfare* is at least as high as the level achieved with the original ex

ante optimal contract, which, after h_t , is a constant that we denote $W^*: \sum_{i=H,L} \Pr(\theta_i|\theta_L)W(\theta_i, \mathbf{q}_i) \geq W^*$. We denote this constraint (G8). If we show that the ex ante optimal quantities solve $\mathcal{P}_{ex\ post}^*$, then they must also solve $\mathcal{P}_{ex\ post}$ and be renegotiation-proof. The Lagrangian of $\mathcal{P}_{ex\ post}^*$ is:

$$(G9) \quad \mathcal{L}_P = (1 + \tau)[\lambda_L W(\theta_H, \mathbf{q}_H) \\ + W(\theta_L, \mathbf{q}_L)] - \lambda_L R(\mathbf{q}_L)$$

where λ_L is the ex post likelihood ratio $[\Pr(\theta_H|\theta_L)/\Pr(\theta_L|\theta_L)]$, and τ is the Lagrangian multiplier associated with (G8). We proceed in three simple steps:

Step 1: $\tau > 0$. Given that $\Pr(\theta_L|\theta_L) \leq 1 - \mu_H \Pr(\theta_H|\theta_H)$ implies (4) $\forall_j \geq 0$, if $\tau = 0$ then the solution of (G9) implies that all the quantities in $\mathbf{q}_H$ are set efficiently and the quantities in $\mathbf{q}_L$ are distorted downward more than the solutions of the ex ante optimal problem (the argument is the same as in Section VI). But then the welfare constraint (G8) must be violated, a contradiction.

Step 2: Let λ° be the ex ante likelihood ratio (μ_H/μ_L) . We can show that $[\lambda_L/(1 + \tau)] = \lambda^\circ \Lambda^{t-1}$. Indeed, it can be verified that the quantities following history h_t in the optimal solution of the ex ante problem maximize:

$$(G10) \quad \mathcal{L}_A = \lambda^\circ W(\theta_H, \mathbf{q}_H) \\ + W(\theta_L, \mathbf{q}_L) - \lambda^\circ \Lambda^{t-1} R(\mathbf{q}_L)$$

If $[\lambda_L/(1 + \tau)] < \lambda^\circ \Lambda^{t-1}$, then the solution of (G9) would be less distorted than the solution of (G10), implying that the welfare constraint (G8) is not binding and so $\tau = 0$, a contradiction. Similarly we can prove that the reverse inequality is not possible. From $[\lambda_L/(1 + \tau)] = \lambda^\circ \Lambda^{t-1}$ we conclude that the solution of (G9) and (G10) coincide and the optimal ex ante contract is renegotiation-proof.

Step 3. Finally, it is easy to see that if $\Pr(\theta_L|\theta_L) > 1 - \mu_H \Pr(\theta_H|\theta_H)$, then there must be a finite $\tilde{t}$ such that for $t > \tilde{t}$, then (4) is satisfied, and the argument in steps 1 and 2 is valid for any $t > \tilde{t}$.

⁴⁰ To avoid confusions in what follows, it is worth emphasizing that $U(\theta_j, \theta_i; h_t)$ and $U(\theta_j; h_t, \theta_i)$ are different objects: the first represents the case in which a type j reports untruthfully to be a type i after a history h_t ; the second represents the case in which a type j truthfully reports his type after a history $h_{t+1} = \{h_t, \theta_i\}$. Indeed the second expression is equivalent to $U(\theta_j, \theta_j; h_t, \theta_i)$.

REFERENCES

- Baron, David P. and Besanko, David.** "Regulation and Information in a Continuing Relationship." *Information Economics and Policy*, 1984, 1(3), pp. 267–302.
- Battaglini, Marco and Coate, Stephen.** "Pareto Efficient Income Taxation with Stochastic Abilities." National Bureau of Economic Research, Inc., NBER Working Papers: No. 10119, 2003.
- Battaglini, Marco.** "Optimality and Renegotiation in Dynamic Contracting." Centre for Economic Policy Research, CEPR Discussion Paper: No. 5014, 2005.
- Bolton, Patrick and Dewatripont, Mathias.** *Contract theory*. Cambridge, MA: MIT Press, 2005.
- Bulow, Jeremy I.** "Durable-Goods Monopolists." *Journal of Political Economy*, 1982, 90(2), pp. 314–32.
- Courty, Pascal and Li, Hao.** "Sequential Screening." *Review of Economic Studies*, 2000, 67(4), pp. 697–717.
- Dewatripont, Mathias.** "Renegotiation and Information Revelation over Time: The Case of Optimal Labor Contracts." *Quarterly Journal of Economics*, 1989, 104(3), pp. 589–619.
- Dionne, Georges and Doherty, Neil A.** "Adverse Selection, Commitment, and Renegotiation: Extension to and Evidence from Insurance Markets." *Journal of Political Economy*, 1994, 102(2), pp. 209–35.
- Gerstner, Louis V., Jr.** *Who says elephants can't dance*. New York: Harper Collins, 2002.
- Harris, Milton and Holmström, Bengt.** "A Theory of Wage Dynamics." *Review of Economic Studies*, 1982, 49(3), pp. 315–33.
- Hart, Oliver D. and Tirole, Jean.** "Contract Renegotiation and Coasian Dynamics." *Review of Economic Studies*, 1988, 55(4), pp. 509–40.
- Hendel, Igal and Lizzeri, Alessandro.** "Adverse Selection in Durable Goods Markets." *American Economic Review*, 1999, 89(5), pp. 1097–1115.
- Hendel, Igal and Lizzeri, Alessandro.** "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance." *Quarterly Journal of Economics*, 2003, 118(1), pp. 299–327.
- Karlin, Samuel and Taylor, Howard M.** *A first course in stochastic process*. San Diego: Academic Press, 1975.
- Kennan, John.** "Repeated Bargaining with Persistent Private Information." *Review of Economic Studies*, 2001, 68(4), pp. 719–55.
- Laffont, Jean-Jacques and Tirole, Jean.** "Adverse Selection and Renegotiation in Procurement." *Review of Economic Studies*, 1990, 57(4), pp. 597–625.
- Laffont, Jean-Jacques and Tirole, Jean.** *A theory of incentives in procurement and regulation*. Cambridge, MA: MIT Press, 1993.
- Laffont, Jean-Jacques and Tirole, Jean.** "Pollution Permits and Compliance Strategies." *Journal of Public Economics*, 1996, 62(1–2), pp. 85–125.
- Miravete, Eugenio J.** "Choosing the Wrong Calling Plan? Ignorance and Learning." *American Economic Review*, 2003, 93(1), pp. 297–310.
- Rey, Patrick and Salanie, Bernard.** "Long-Term, Short-Term and Renegotiation: On the Value of Commitment in Contracting." *Econometrica*, 1990, 58(3), pp. 597–619.
- Rey, Patrick and Salanie, Bernard.** "On the Value of Commitment with Asymmetric Information." *Econometrica*, 1996, 64(6), pp. 1395–1414.
- Roberts, Kevin.** "Long-Term Contracts." Unpublished Paper, 1982.
- Rochet, Jean-Charles.** "A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context." *Journal of Mathematical Economics*, 1987, 16(2), pp. 191–200.
- Rogerson, William P.** "Repeated Moral Hazard." *Econometrica*, 1985, 53(1), pp. 69–76.
- Rustichini, Aldo and Wolinsky, Asher.** "Learning about Variable Demand in the Long Run." *Journal of Economic Dynamics and Control*, 1995, 19(5–7), pp. 1283–92.
- Townsend, Robert M.** "Optimal Multiperiod Contracts and the Gain from Enduring Relationships under Private Information." *Journal of Political Economy*, 1982, 90(6), pp. 1166–86.
- Welch, Jack.** *Jack: Straight from the gut*. New York: Warner Books, Inc., 2001.