Long-Term Contracting with Markovian Consumers

Marco Battaglini [1]

This version: August 2003[2]

[1] Princeton University and CEPR. I gratefully acknowledge the hospitality of the economics department at the Massachusetts Institute of Technology for the academic year 2002-2003. E-mail: mbattagl@princeton.edu.

[2] I am grateful for helpful comments to Pierpaolo Battigalli, Stephen Coate, Eddy Dekel, Glenn Ellison, Jeff Ely, Bengt Holmstrom, John Kennan, Alessandro Lizzeri, Rohini Pande, Nicola Persico, William Rogerson, Ariel Rubinstein, Jean Tirole and Asher Wolinsky, as well as to seminar participants at Cornell, Harvard, MIT, Northwestern, Princeton and U.C.L.A. Marek Picia provided valuable research assistance.

Abstract

To study how a firm can capitalize on a long-term customer relationship, we characterize the optimal contract offered by a monopolist to a consumer whose preferences evolve following a Markov process. Even in the simplest case with two types the optimal contract is non-stationary and has infinite memory, but it can be represented by a simple state variable in a very economical way. A weak sufficient condition guarantees that the optimal contract converges to the efficient contract for any degree of types' persistence and along any history, though the speed of convergence is history dependent. In contrast with the case with constant types, under general conditions there is no conflict between profit maximization and renegotiation-proofness. These properties provide insights into the optimal distribution of ownership rights of the technology controlled by the monopolist and contribute to explaining some empirical findings.

Keywords: contract theory, price discrimination, adverse selection, ownership rights.

JEL Classification: D42, D82, D23.

I Introduction

Advances in information processing and new management strategies have made long-term, non-anonymous 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.[1] When a long-term relation is non-anonymous and types are persistent, the seller can mitigate the problem of asymmetric information by using consumers' choices to forecast future behavior. However, as a result, 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: 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 that are waiting for service. As it is well known, under standard assumptions on the arrival rate of customers, the type of this entrepreneur follows a Markov process.[2] Or, to give another example, take the case of a company selling cellular phones. 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.[3] 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 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, such a characterization allows a better understanding of some aspects of a dynamic principal-agent relationship that previous models could not capture: in particular, regarding the mem-

[1] See, for example, Gertsner [2002] or Welch [2002].

[2] A more detailed description of such a model can be found in Section II.

[3] See Miravete [2002] for recent convincing evidence that the preferences of customers of telephone services evolve stochastically over time; and Section II for further discussion.

ory and complexity of the optimal contract, its efficiency, and its renegotiation-proofness. Perhaps surprisingly, it also provides insights into the optimal ownership structure of the production technology.

As we said, when types are constant, the contract has no memory and the inefficiency of the optimal static contract is repeated period after period. With stochastic types, even in a simple Markovian environment with one period memory, risk neutral agents and two types, the contract is non-stationary and has infinite memory; however, despite this, 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 optimal menu 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* principle: after any history, if the agent reports to be a high type, 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* principle: even in the lowest history in which the agent always reports to be a low type, the contract converges to the efficient menu. 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 higher consumer's rents, it explains why, as observed in the empirical literature, "old" customers are treated more favorably than "new" customers.[4]

In a stochastic environment, the incentives for renegotiation are also very different. As shown in seminal papers by Dewatripont [1989], Hart and Tirole [1988] and Laffont and Tirole [1990], 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. In our model, on the contrary, this is not generally the case, even when types are highly correlated. Indeed, we show a condition that is easily satisfied and guarantees that the optimal contract is renegotiation-proof. 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", though incorrect, argument for the efficiency result mentioned

[4] Indeed Hendel and Lizzeri [2000] and Dionne and Doherty [1994] find evidence that contracts are front-loaded: prices are high in initial periods and decline over time. See below for a more detailed discussion.

above. At each point in time the consumer has an informational advantage with respect to the monopolist: knowing his own state the consumer has privileged information about the likely evolution of the type too. Since the distribution of types converges to a stationary distribution, however, at date 1 the difference in conditional expectations between the lowest and the highest type about the likely realization at time t, for t large, is not substantial. This implies that *at date 1* it is "cheap" to sell the expected surplus generated by the relationship from date t onward: indeed the monopolist can separate the consumer's types by only paying the small difference in their conditional expectations. This may suggest that the monopolist finds it optimal to sell up-front to the consumer as much expected surplus as can be generated after any history h_t: i.e. the efficient surplus or close to it.

This argument, however, is incorrect. Indeed, the menu offered at period t has little *direct* effect on the rents that must be left to the agent in period 1 to induce him to reveal his type truthfully. However it may have a substantial effect on the rents that must be left in period t-1: remember that the agent receives new private information in any period and in a direct mechanism the monopolist has also to extract this information in any period. A more efficient contract at period t, then, will substantially increase the rents of any type immediately below t. Therefore, through the change in the cost of extracting surplus at time t-1, a more efficient contract has an *indirect* impact on the cost of extracting surplus at time t-2 and so on up to period 1. Because of these indirect effects, a small change at time t can significantly impact the consumer's utility at time zero.

Consider Figure 1 which shows the impact on profits at time zero of a change Δq in the quantity $q(h_t)$ offered to the consumer after a history h_t.[5] Assume that $q(h_t)$ is lower than the efficient level and that the principal increases it by Δq. On the one hand, the change in $q(h_t)$ affects surplus (and profits) if history h_t is realized (represented by the "thick arrow" on the left hand panel). The right hand panel of Figure 1, however, shows that the change also affects the rent that must be paid to the high type: specifically, by the incentive compatibility constraint,[6] it increases by $\Delta\theta\Delta q$ at time t; again by the incentive compatibility constraint, it increases at time t-1 by the difference in the conditional expectations $[\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)]\,\Delta\theta\Delta q$..., and so on up to period zero. While the marginal impact of the change in supply on expected surplus is proportional to the probability of the history h_t, $\Pr(\theta_L|\theta_L)^{t-1}$, the impact on the agent's rents at time zero is proportional to the "cumulative effect" of the difference in expectations of

[5] In Figure 2, an arrow that points up means that the type is high; an arrow that points down means that the type is low.

[6] This is the constraint that guarantees that the high type does not want to imitate the low type.

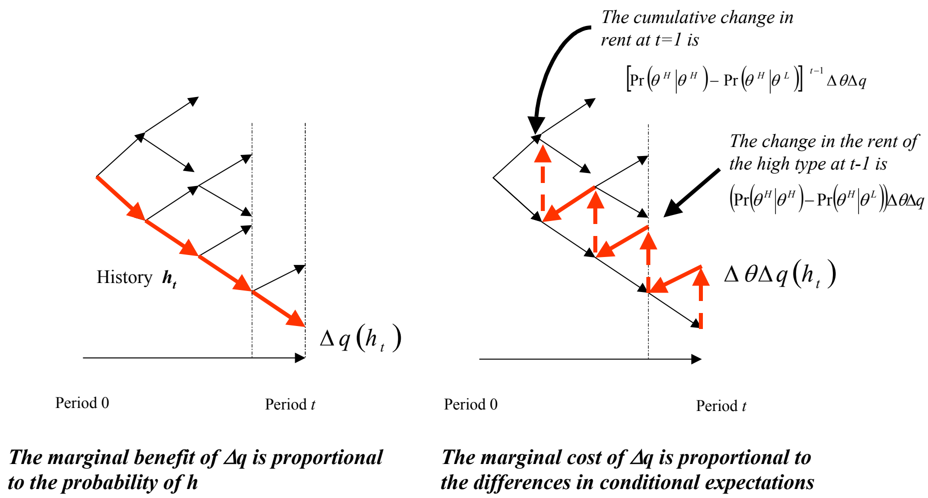

Figure 1: Marginal cost and marginal benefit of a change in the quantity offered after history h_t. An arrow that points up means that the type is high; an arrow that points down means that the type is low. History h_t corresponds to the case in which the type is low for t periods.

the future change $[\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)]^{t-1}$. Therefore the marginal cost and the marginal benefit for a change in quantity at time t after history h_t reach time zero from two different "paths," as it can be seen from the thick arrows in the right and left panel of Figure 1.

Even if the marginal cost of a more efficient contract at time t converges to zero, and indeed it does, the marginal benefit of a change in a quantity after an history at time t converges to zero too because, even if the discount rate is one, the probability of any particular history of length t converges to zero. Therefore the quantity offered in the contract converges to the efficient level for this history only if the cost converges faster than the benefit:

$$\frac{[\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)]^t}{\Pr(\theta_L|\theta_L)^t} \underset{t\to\infty}{\longrightarrow} 0 \tag{1}$$

In the following, we show that, independently of the discount factor or the degree of correlation of types, the assumption that types are positively correlated is sufficient to obtain a simple characterization of the optimal contract and to prove that (1) is true along all histories.

These results have also implications for the optimal ownership structure of the monopolist's business. In fact, it is interesting to ask why the monopolist keeps control of the production technology: after all, only the consumer directly benefits from it and has information for its efficient use. Indeed, we show that the optimal contract can be *inter-*

preted as offering to the consumer with a high type 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 of the future. This insight seems relevant to understand the ownership structure of a new technology. The initial owner of a new technology has generally monopoly power on its use thanks to a patent and has to decide if it is more convenient to use the technology directly selling its products, or to sell the patent.[7]

All the results discussed above hold for any discount factor. If we assume that the discount factor is high, stronger results hold true. Indeed, we show that as the discount factor converges to one, the monopolist appropriates all the expected surplus of the relationship, which, in turn, converges to the efficient surplus.

The paper is organized as follows. Section II describes the model. Section III characterizes the optimal contract. Section IV discusses the dynamic properties of the optimal supply function. Section V discusses the results on the property rights. Section VI discusses the properties of the monetary payments in the optimal contract. Section VII studies renegotiation-proofness. Section VIII presents the asymptotic results for $\delta \to 1$. Section IX concludes. The following subsection discusses some related literature.

A Related literature

As mentioned above, dynamic models of price discrimination generally assume that the agent's type is constant over time. We have already said that, 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.[8] With constant types the dynamics becomes interesting only when other constraints are binding, in particular when a renegotiation-proofness constraint must be satisfied. This problem has been solved only under some assumptions. Hart and Tirole [1988] and Dewatripont [1989] present models with many periods and constant types: the first paper, however, assumes that supply can have two values, 0 or 1; the second focuses on pure strategies and assumes some simplifications in the nature of the contractual agreement. Laffont and Tirole [1990] have solved a model in which supply can assume more than two values,

[7] The original author of the operating system software *DOS* sold it to *Microsoft*; *Microsoft*, on the contrary, sold licenses of single copies of the software to *IBM.* Forgetting the well known history of these two transactions, it might seem that, at least from a theoretical point of view, the two decisions are equivalent in terms of expected revenues; our model helps to understand why they are not.

[8] The proof of this result can be found in Laffont and Tirole [1993], §1.10, pp. 103.

but have considered only two periods.[9]

Roberts [1982] and 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 be easily separated: indeed in this case, except for period 1, there is no asymmetric information between the principal and the agent because both share the same expectation for the future.[10] Baron and Besanko [1984] and Laffont and Tirole [1996] extend this research presenting two periods procurement models in which the type in the second period is stochastic and correlated with the type in period 1. Courty and Li [2000] enrich the analysis of these two papers considering different assumptions on the distribution of types in a model with two periods and unitary demand. Because all these models have only two periods, however, they can not capture important aspects of the dynamics of the optimal contract like its memory and complexity after long histories or its convergence to efficiency; moreover they do not consider renegotiation-proofness.

Dynamic environments with adverse selection and stochastic types have been recently applied to the study of models of leasing and insurance.[11] Hendel and Lizzeri [1999] consider a model in which a durable good stochastically depreciates generating a dynamic adverse selection framework similar to the case studied in this paper in which the type of the consumer changes over time. They characterize the equilibrium when the life of the good is two periods and study the interaction between the new and used markets of the good.[12] Biehl [2001] studies a similar problem in the two periods framework of Laffont and Tirole [1996] and compares the profitability of leasing versus selling a durable good with unitary demand.

Hendel and Lizzeri [2000] study, both from a theoretical and empirical point of view, a model of dynamic insurance inspired by Harris and Holmstrom [1982] in which the

[9] See Laffont and Tirole [1993] for a survey of dynamic models with renegotiation.

[10] 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 in his i.i.d. types, the contract depends on the cumulated wealth of the agent. Because the assumption of players with different risk aversion is less justified in a buyer-seller environment, following the tradition in price discrimination theory, we assume players with equal risk aversion. This allows us to focus on the effect of serial correlation in the types' realizations on dynamic incentives that this literature has not studied.

[11] Dynamic models of adverse selection are also studied in models of taxation and risk sharing following Townsend [1982]. This literature focused on the case of stochastic but serially independent types. See, for example, Green [1987] and Atkeson and Lucas [1992]. Battaglini and Coate [2003] use some of the techniques of the present paper to characterize in closed form the Pareto optimal frontier of taxation with correlated types.

[12] This model is extended by Hendel and Lizzeri [2002a and 2002b] to analyze leasing contracts, showing how these contracts help to screen consumers in a way that matches empirical evidence. Hendel, Lizzeri and Siniscalchi [2002] extend the analysis to the case with than two periods but when the discount factor converges to 1.

expected type of the agent is stochastic and changes over time. They characterize the optimal contract with symmetric information under the assumption that the seller can commit to it but the agent can not. Despite the fact that we consider an environment with asymmetric information, the optimal contract that we derive is consistent with the empirical evidence that they uncover.

All these papers are interested in modelling optimal long-term contracts, some of them assuming full commitment and some of them requiring renegotiation-proofness. The related literature on the "Ratchet effect", on the contrary, is characterized by the assumption that the principal can only offer short term contracts with one-period length. Hart and Tirole [1988] study this problem too, and Kennan [2001] extends their model to the case in which types may change over time, maintaining the assumption of a 0-1 supply function. Laffont and Tirole [1987,1988] characterize the optimal contract with a general supply function, but in two periods. As it is well known, the fact that in our paper and in the work cited before the seller can use long-term contracts makes the results very different from the prediction of this literature, even when only a weak form of commitment consistent with renegotiation-proofness can be used.[13]

II The model

A Setup

We suppose two parties, a buyer and a seller. In each period the buyer has a reservation value $\underline{u} = 0$ and enjoys a utility $\theta_t q - p$ for q units of a non-durable good bought at a price p. The monopolist's cost function is $c(q) = \frac{1}{2}q^2$. The marginal benefit θ_t evolves over time according to a Markov process. To focus on the dynamics of the contract and simplify notation, we consider the simplest case with 2 types $\Theta = \{\theta_i\}_{i=H,L}$ and, without loss of generality, $\Delta\theta = \theta_H - \theta_L > 0$. The probability that state l is reached if the agent is in state k, i.e. the (l,k) component of the transition matrix Λ, is denoted $\Pr(\theta_l | \theta_k) \in (0,1)$; the distribution of types conditional on being a high (low) type, i.e. the first (second) column of Λ, is denoted $\boldsymbol{\alpha}_H$ ($\boldsymbol{\alpha}_L$).[14] We order the types assuming that $\boldsymbol{\alpha}_H$ first order stochastically dominates $\boldsymbol{\alpha}_L$. In each period the consumer observes the realization of his own type; the seller, on the contrary, can not see it. At date 0 the seller has a prior $\boldsymbol{\mu} = (\mu_H, \mu_L)$ on the agent's type.

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

[13] See Section VI for more discussion; and Laffont and Tirole [1993] for an excellent survey.

[14] The corresponding row vectors are respectively α_H^T and α_L^T.

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 utility offered by the contract is lower than $\underline{u} = 0$. In line with the standard model of price discrimination, the monopolist commits to the contract that is offered: in Section VII we relax this assumption allowing the parties to renegotiate the contract. It is easy to show that in this environment 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 $\langle \mathbf{p}, \mathbf{q} \rangle = \left(p_t \left(\widehat{\theta} \,| h_t \right), q_t \left(\widehat{\theta} \,| h_t \right) \right)_{t=1}^{\infty}$, where h_t and $\widehat{\theta}$ 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. In general, h_t can be defined recursively as $h_t := \left\{ \widehat{\theta}_{t-1}, h_{t-1} \right\}$, $h_1 := \emptyset$ where $\widehat{\theta}_{t-1}$ 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 an history h_t ($t \leq j$) is denoted $H_j(h_t)$. A strategy for a seller consists in 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 $h_t^C := \left\{ \widehat{\theta}_{t-1}, \theta_t, h_{t-1}^C \right\}$, $h_1^C := \theta_1$ because the agent always knows his own type. For a given contract, a strategy for the consumer, then, is simply a function that maps an 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.[15] The same condition that guarantees this property in the static model also guarantees it in our dynamic model; therefore, to simplify notation, we also assume it.[16] This assumption can easily be relaxed, but this would complicate notation with no gain in insight.

Both the seller and the buyer maximize their expected utility in the contractual relationship. For future reference, the first best is easily calculated in this model: in each period the efficient quantity to produce is $q^E(\theta) = \theta$ and, therefore, $w^*(\theta) = \frac{1}{2}\theta^2$ is the per period welfare.

B Discussion

We now briefly comment the main assumptions of the model presented above.

Stochastic types. The key aspect of our framework is the assumption of stochastic and

[15] The condition that guarantees that all types are served is $\Delta\theta \leq \frac{\mu_L}{\mu_H}\theta_L$.

[16] As we will see, the distortion introduced 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 customer after all histories in our dynamic model too.

variable types. Although only few empirical works have studied evolution of consumers' preferences, existing literature seems to confirm this assumption. Miravete [2002], in particular, using rich data from a tariff experiment run by South Central Bell in 1986, presents convincing evidence that consumers, at least in telephone services, are uncertain about their future demand. From a more theoretical point of view, the assumption that the process follows a Markov chain is quite general and flexible because, except for stochastic dominance, we are making no assumptions on the (positive) degree of types' correlation. In particular as the coefficients in the diagonal of Λ converge to 1, we can have an arbitrarily high level of correlation among types. Indeed, the case with constant types is a special example of the model presented above in which the matrix Λ is diagonal and therefore non-generic. In many situations, moreover, this assumption can be explicitly microfounded. Given the generality of the agency problem that we are analyzing, we present only two examples: the flexibility of the Markovian model allows us to find many others.[17]

1. The "list of customers" model. Assume that the agent is a firm buying a factor of production and its marginal value depends on the length of the queue of its customers waiting for service: intuitively, the longer the queue, the higher is the price that the agent can charge for the final product. At the beginning of each period when the agent buys the factor of production for the day, therefore, the type is the number of customers that have not been served in the previous period. During the day, however, new customers arrive with probability $a_k = \Pr\{k \text{ new consumers arrive}\} > 0$ with $k \in \{0,1,2\}$ and $\sum_{k=0}^{2} a_k = 1$. Assume that if two consumers are already in the list then a new consumer would not sign up so that the (non-negative) length of the queue l_t is at most 2. At the end of the period one customer is served with probability one. It is easy to verify that the type of the firm at the beginning of any period t is $\max\{\min\{l_{t-1} + \xi_t, 2\} - 1, 0\}$ and follows a Markov process with two states $\{0,1\}$ and transition matrix:

$$\Lambda = \begin{bmatrix} \sum_{i=1}^{2} a_i & a_2 \\ a_0 & \sum_{i=0}^{1} a_i \end{bmatrix}$$

One can also immediately see that the first column first order stochastically dominates the second in a strict sense and therefore this model is a particular version of the framework presented above.

2. The "venture capitalist's" model. A technological company has $x < n$-1 research programs. Each of them can generate a number $\xi = 0, 1, ..b$ offspring projects with probability $\Pr\{\xi = k\} = a_k > 0$. Call j_t the number of projects at time t and assume

[17] Any standard text on Markov processes can be consulted for other applications that can be used for microfundation. See for example Karlin and Taylor [1975].

that j_t can be at most n-1 because if more projects are available, the firm can only develop n-1 of them. A standard argument shows that j_t follows a Markov chain with $n \times n$ transition matrix and that the stochastic dominance condition is verified. Again, if we assume that the company asks a venture capitalist for money and managerial advice and the marginal value of each of them is increasing in the number of projects, then we fall in the case of the general model presented above.

Infinite horizon. Besides a direct theoretical interest, the analysis of a stationary model with infinite periods is useful for two reasons. First we can study long term behavior and convergence of the contract which is clearly impossible in a two period model. However, it is also important to study the dynamics of prices. 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. Indeed, in a non-stationary 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. Note however that it is easy to show that our characterization would be valid even in a model with T finite periods.

Unilateral commitment. In the first part of the analysis we focus an 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. Discussing the life insurance market, Handel and Lizzeri observe that the Term value contracts in the insurance market, which account for 37% 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. [...] This features fit a model of unilateral commitment."[18] It is moreover true that firms seem aware that the possibility to commit is important to win exclusive long term contracts.[19] On the other hand there are many situations in which renegotiation is an important component of the problem. As we anticipated, in Section VII we show that under general conditions the optimal contract is renegotiation-proof and therefore it can be applied to these environments too.

Discount factor. We are making no assumptions on the discount factor. Although it reasonable to assume that the monopolist is fairly forward looking, we allow the monopolist to be very much myopic. The model can also be easily extended to other types of time preferences as, for example, hyperbolic discounting.

[18]Handel and Lizzeri [2002], p.4.

[19]Using the words of the CEO of IBM Lou Gerstner: "You can't get into that kind of business without making the commitment to carry the infrastructure and loss until a contract that could extend over five or ten years becomes profitable. There is no such a thing as a toe in the water. When you take this plunge it's a full-body immersion." Gertsner [2002], p.133.

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.

$$\begin{array}{ll} \max_{\langle \mathbf{p},\mathbf{q}\rangle} & \mu \left[p\left(\theta_H \left| h_1 \right.\right) - q^2\left(\theta_H \left| h_1 \right.\right)/2 + \delta E\left[\Pi\left(\theta \left| h_1, \theta_H \right.\right) \left|\theta_t = \theta_H \right.\right]\right] + \\ & (1-\mu)\left[p\left(\theta_L \left| h_1 \right.\right) - q^2\left(\theta_L \left| h_1 \right.\right)/2 + \delta E\left[\Pi\left(\theta \left| h_1, \theta_L \right.\right) \left|\theta_t = \theta_L \right.\right]\right] \\ s.t. & IC_{h_t}\left(\theta_H\right),\ IC_{h_t}\left(\theta_L\right),\ IR_{h_t}\left(\theta_H\right),\ IR_{h_t}\left(\theta_L\right) \end{array} \qquad (\mathcal{P}_I)$$

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

$$\begin{array}{rl} & q\left(\theta_i \left| h_t \right.\right)\theta_i - p\left(\theta_i \left| h_t \right.\right) + \delta E\left[U\left(\theta \left| h_t, \theta_i \right.\right) \left|\theta_t = \theta_i \right.\right] \\ \geq & q\left(\theta_j \left| h_t \right.\right)\theta_i - p\left(\theta_j \left| h_t \right.\right) + \delta E\left[U\left(\theta \left| h_t, \theta_j \right.\right) \left|\theta_t = \theta_i \right.\right] \end{array} \qquad (IC_{h_t}(\theta_i))$$

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

The classic approach to characterize the solution of 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 however easy to see that *in a dynamic model this cannot be true.* Given an optimal contract we can always add a "borrowing" contract 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 not be binding as well. 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 it is *not necessarily* true, there is no loss of generality in assuming that constraints are binding. Lets define $\mathcal{P}_{II}$ 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 are binding. We say that a supply schedule $q_t^*(\theta\,|h_t)$ is a solution of a program if there exist 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 $\mathcal{P}_{II}$.*

This result may be intuitively explained in a two period model (the compete argument is by induction on t). Assume that at time $t = 2$ the incentive compatibility constraint of the high type is not binding after an 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 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 rent of the high type: because the high type is more optimistic about the future realization of his type, the reduction in future rents will be larger than the increase in payments at time t. Expected profits, therefore, are larger after the change in the contract and all constraints are respected: a contradiction. After an history $h_2 = \theta_L$ we proceed in a similar way: in this case profits remain constant after the change, so the constraint needs not be necessarily binding at the optimum, but it can be made binding 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 $\mathcal{P}_I$.[20] Indeed this would not be true:[21] 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 $\mathcal{P}_{II}$ and, *because of this*, solves the relaxed problem as well.

We can now solve the simpler problem with binding constraints and obtain:

Proposition 1 *At any time t, the optimal contract is characterized by the supply func-*

[20] In $\mathcal{P}_{II}$ we *assume* that the constraints are binding, so it is not just a relaxed version of $\mathcal{P}_I$.

[21] 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.

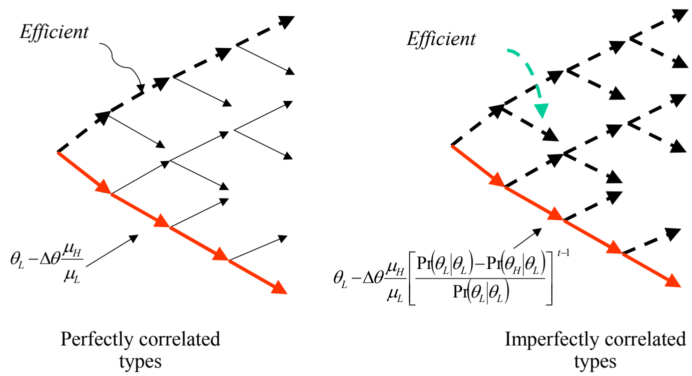

Figure 2: Comparison of the optimal contracts with constant and Markovian types

tion:

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

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

Given (2), the optimal contract has a very simple structure which is represented in Figure 2. The left panel of this Figure shows the benchmark case in which types are constant. This contract is efficient if the agent is a high type and inefficient if he is a low type: in the latter case the inefficiency is constant over time and equal to $\Delta\theta\frac{\mu_H}{\mu_L}$; all the remaining histories have zero probability. When types follow a Markov process, similarly, the contract instantly becomes efficient as soon as the agent reports to be a high type (dashed arrows in the left panel of Figure 2): but now efficiency "invades" also the histories in which the agent subsequently reports to be a low type. We can call this result a *Generalized No Distortion at the Top* principle because its logic is not dissimilar to the logic of the "No Distortion at the Top" result in static price discrimination.

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 the agent to reveal his high type is independent from the quantities that follow this history: since the incentive compatibility constraint for the high type is binding, only the quantities that follow a history in which the agent falsely reports to be a low type 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 to be a high type, but also in all the following periods.

A distortion, however, persists on the lowest branch of the history tree. The denominator of the distortion -the square parenthesis in the middle expression of (2)- is the probability $\mu_L \Pr(\theta_L | \theta_L)^{t-1}$ that the agent reaches the node in which the distorted supply is actually offered to the agent. Not surprisingly the higher is the probability that the increase in supply increases surplus, and therefore the profits of the monopolist who is residual claimant, the smaller is the optimal downward distortion.

On the other hand the numerator reflects the impact of a change in supply after t periods on the rent of the high type. The higher is $[\Pr(\theta_H | \theta_H) - \Pr(\theta_H | \theta_L)]$ the more profitable it is for a high type to imitate a low type, and therefore the higher is the rent that he receives to preserve incentive compatibility. This term is raised to the t-1 because, as represented in the left panel of Figure 1, we are interested in the marginal effect on the rent *at time zero*, so we are evaluating the expected change of the expected change ... of the expected change for t-1 times. Again, not surprisingly when this term increases, the monopolist is willing to accept a higher distortion in order to reduce the high type's rent. Although the inefficiency along this history persists indefinitely, however, as we discuss more fully below, it converges to zero, yielding the *Vanishing Distortion at the Bottom* result mentioned in the Introduction.

Proposition 1 uniquely determines the optimal quantities supplied after any history. In the following section we discuss in detail the evolution over time of this supply function and its history dependence. The result presented above, however, does not uniquely characterize the payments in the optimal contract. Section VI explores the possible payment schedules and their evolution.

IV The dynamics of supply: the "Vanishing Distortion at the Bottom" principle

This section focuses on two properties of the optimal state contingent supply function: first, we discuss its memory and the complexity of its structure; then, its efficiency.

The length of the memory of the optimal contract is a central issue in the literature on dynamic moral hazard, but it has not been studied in adverse selection models:[22] this because when the agent's type is perfectly constant we know that the contract is also

[22] A classic reference is Rogerson [1985] who proves that the optimal dynamic contract with moral hazard has necessarily a positive memory. Chiappori et al. [1994] extend some of these results and survey this literature.

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. This implies that it is not only optimal to smooth consumption over states of the world, as in the static moral hazard framework, but also that it is optimal to smooth consumption over periods; to do this the contract has to keep track of the past realization of the agent's income. In the model presented above, however, the agent is risk neutral: therefore, if the intuition were similar to the moral hazard case, then the contract should have no memory in our model too. Moreover, since the agent's taste follows a Markov process and therefore the relevant economic environment has a memory of only one period, it may seem natural that if the contract has any memory, this must be one.

From (2), however, it is clear that the memory of the optimal contract is unbounded: for any $K > 0$, we can always find two histories which are identical for the last K periods but induce different menus in the optimal mechanism. Consider, for example, two histories that differ only in the first realization of types, the first being high, the second being low, and that have low realizations in any period following date two. If these histories are longer than a positive parameter K, say they have $K+1$ length, then they coincide for at least the last K periods. As we have seen in Proposition 1, at time 1 the monopolist will offer 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 $K+1$, the contracts will be different.

This result is perhaps surprising in an environment as simple as the one studied here with only two types. When there are many types, the need to keep track of the past history is even higher so it is easy to predict that the need for a long memory would be higher too. The fact that the optimal contract has unbounded memory, however, does not imply that 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 declaration on the previous state and if in the previous state we were on the lower branch, the state can be described by a simple 0-1 variable which can be defined recursively

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

and $X_0 = 1$. This state variable therefore starts with value one and remains one if the agent persists in reporting to be a low type. Once the agent has reported to be a high type the state switches to zero and remains constant forever. We have:

Proposition 2 *The optimal solution is a function of time and the* 0-1 *state variable described by (3):*

$$q_t^*(\theta_t, X_{t-1}) = \theta_t - \Delta\theta\frac{\mu_H}{\mu_L}X(\theta_t, X_{t-1})\left(\frac{\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)}\right)^{t-1} \tag{4}$$

The result in (4), therefore, shows that at least in the two types case and despite the long memory, the problem has an extremely simple and economical structure which makes the result more realistic even in a world with limited computational resources.

As anticipated, the second main implication of Proposition 1 is that the contract converges to efficiency along all histories, even the 'lowest' history in which the agent always reports to be a low type. To understand the intuition of this result consider first the static model. As it is well known, in this case the quantity offered to the low type is lower than the efficient level. This distortion, however, is not necessary for incentive compatibility, since the monopolist could easily offer an incentive compatible contract that is also efficient. The monopolist introduces a distortion because this minimizes the rents that need to be left to the agent: the optimal distortion simply equalizes the marginal cost of an increase 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 constant over time.

In a dynamic setting with changing types, however, the inefficiency depends on time because the cost/benefit ratio is also time-dependent. On the one hand the marginal benefit of an increase in $q(h_t)$ converges to zero as t increases because it becomes more and more unlikely that the particular history h_t will be reached. But, on the other hand, the marginal cost $\mu_H[\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)]^{t-1}\Delta\theta$ is also converging to zero. A simple manipulation of (4), however, yields:

$$q_t^*(\theta_L|h_t) - \theta_L = \Delta\theta\frac{\mu_H}{\mu_L}X(\theta_t, X_{t-1})\left[1 - \frac{\Pr(\theta_L|\theta_H)}{\Pr(\theta_L|\theta_L)}\right]^{t-1}$$

Since by first order stochastic dominance we have $\frac{\Pr(\theta_L|\theta_H)}{\Pr(\theta_L|\theta_L)} \in (0,1)$, it follows that

$$\lim_{t\to\infty} q_t^*(\theta_L|h_t) = \theta_L = q^E(\theta_L)$$

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.*

Since the distortion of the optimal contract is confined to the "lowest" history (see Figure 2) and this converges to zero, Proposition 5 therefore shows that, in analogy with the classical "No Distortion at the Top" result, in the Markovian case we also have a novel *Vanishing Distortion at the Bottom* principle that clearly could not be appreciated in static models or in models with constant types.

It is interesting to note that the case with constant types does not converge to efficiency because the term in the square parenthesis in (4) is *exactly* one and so *it is independent from t*. Any perturbation of the parameters[23] would imply convergence to efficiency. In this sense the result with constant types is not robust. On the contrary, as the persistence of types converges to one, we have that $\frac{\Pr(\theta_L|\theta_H)}{\Pr(\theta_L|\theta_L)} \to 0$. Not surprisingly, this implies that, *ceteribus paribus*, the contract converges in every period to the optimal static contract (though, obviously, the entire sequence of menus would never converge). There are, however, two important observations. The first regards the rate of convergence. Assume, for example, that types are ex ante equally likely ($\mu_H = \mu_L = \frac{1}{2}$), and the types are very much correlated (for example, the type is persistent 80% of the time). Then the expected inefficiency of the contract after 10 periods will be $0.03779\Delta\theta$; the expected inefficiency after 50 periods is $5.0517 \times 10^{-12}\Delta\theta$.[24] Therefore, even after a few periods, we should expect to observe efficiency, or a small inefficiency.

The second observation is that, as we will show below, for any fixed degree of correlation, as $\delta \to 1$ the result with fixed and stochastic types are substantially different. Indeed when $\delta \to 1$ there is a "discontinuity" in the results with imperfect and perfect correlation: in the former case, average surplus converges to the efficient level; in the latter, surplus is independent of δ and equal to the static inefficient level.

V 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 on her exclusive

[23] Remember that by stochastic dominance $\frac{\alpha_{HH}-\alpha_{HL}}{\alpha_{LL}}$ is never larger then one.

[24] The probability that the contract is inefficient after 10 periods is 0.067109 and after 50 periods it is 8.9203×10^{-6}. In the event that the contract is inefficient, the inefficiency is equal to 0.056314Δ after 10 periods and $5.6632 \times 10^{-7}\Delta$ after 50 periods.

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:

Proposition 4 *Without loss of generality, the optimal contract offers a call option to buy out the firm to the agent as soon as he reveals to be a high type. However, the monopolist never finds it optimal to sell the firm to an agent who has always revealed to be a low type.*

The first part of this result should not be surprising. After the agent reveals he is 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.[25]

The interesting observation, however, is in the second part of the proposition: in 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.[26] As in the static model, and as we discussed in Section IV, the distortion is introduced to extract surplus from the high type: this would suggest 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 too.

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.[27] Consider Figure 3 which, for simplicity, represents an history tree of a simplified two periods model. 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 $R_{OWN} = \frac{1}{2}\Delta\theta\left(\theta_H + \theta_L\right)$. This rent is higher than the minimal rent that would guarantee truthful revelation: the incentive compatibility constraint only requires a rent equal to $R_{IC} = \Delta\theta\theta_L < R_{OWN}$.

[25] Note that this result is different from the classical results by 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 over time; in our framework, instead, the firm is selling the *technology* to produce the good, and the future quantities depend on the realized type.

[26] I am grateful to Bengt Holmstrom and Asher Wolinsky who have independently suggested this point.

[27] Indeed, the example presented below is also useful to appreciate the intuition behind Lemma 1.

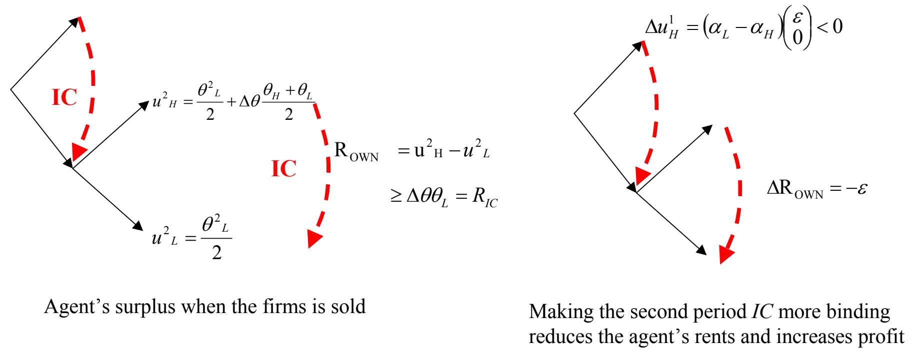

Figure 3: A two period example. If the monopolist sells the firm the incentive compatibility constraint is not tight in the second period and this is not optimal.

Assume now that the monopolist, after the agent reveals to be a low type, keeps the ownership in order to reduce the rent of the high type in $t = 2$ 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 he is 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 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 he is 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 $\delta(\alpha_{LH} - \alpha_{HH}) \varepsilon$ 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 paying a lower rent. The monopolist, therefore, strictly prefers to keep the ownership structure of the firm: ownership allows to control the rents 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 an "high-type" report. Moreover, as we will prove in Section VI studying in detail monetary payments, the dynamics of the transfer price of the technology will be dictated by the dynamics of the inefficient supply schedule.

VI The dynamics of monetary payments

As mentioned above, two payment schedules with the same present value may 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 by 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.

The supply contract. For any contract in which the monopolist borrows money and repays it in an arbitrary time pattern, we can always distinguish two parts: a *supply contract* in which the relevant $IC_{h_t}(\theta_H)$ and $IR_{h_t}(\theta_L)$ constraints are binding and a *lending contract*, in which the monopolist borrows some amount of money and pays it back over time to the agent. One 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, then she would never find it optimal to borrow money from the agent.[28] 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.

The "sale of the firm" contract. There is one particular case in which the monopolist receives anticipated payments from the consumer that has special significance from an

[28] If the monopolist is more patient then 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 $q_t(\theta\,|h_t)$, but would eliminate equilibria in which the monopolist borrows money.

empirical and theoretical reason: the contract discussed Section V in which the monopolist, as soon as compatible with profit maximization, sells the firm to the consumer who reports to be a high type. 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 in generality that the $IR_{h_t}(\theta_L)$ constraint is binding 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 firm.[29]

In the next two subsections we discuss the properties of these contracts.

A Front loading in supply contracts

Only few works have studied dynamic adverse selection problems from an empirical point of view.[30] However a result that seems robust and confirmed by more than one researcher is a front-loading effect: the observation of high initial prices which decline over time. This effect has been shown with California auto insurance data by Dionne and Doherty [1994] and more recently by Hendel and Lizzeri [2000] in the life insurance market. 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.

These two papers have provided two different explanations that certainly capture the phenomenon in many real environments. Dionne and Doherty [1994] see it as a consequence of the possibility to renegotiate contracts over time. Hendel and Lizzeri, on the contrary, 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.

As we said, both stories capture important aspects of the problem; however some issues remain open. Regarding the first explanation, it is not clear to what extent contracts are really renegotiable over time, especially for the case of auto insurance: contracts may be state contingent, but they are often in the form of a take it or leave it offer. Indeed it is reasonable to assume that the seller of the insurance has often commitment power with respect to many customers and would find it too expensive to renegotiate frequently the contract signed with each consumer.[31] Regarding the second explanation, although it is

[29] I am grateful to William Rogerson who suggested to discuss this point.

[30] See Chiappori [2000] for an excellent review of models of insurance under asymmetric information.

[31] Moreover it is reasonable to expect that even if there is some renegotiation between a consumer and a local branch of, say, an insurance company, this renegotiation is constrained by guidelines imposed by the parent company, which are necessarily not contingent on the needs of a single consumer.

true that in many markets asymmetric information may not play a major role,[32] in many other important markets there is a significant degree of asymmetric information.[33]

Our model shows that front-loading is a characteristic of the optimal contract even with asymmetric information and no renegotiation, proving that, on average, 'old' customers receive a better treatment than 'new' customers:

Proposition 5 *In the optimal supply contract the average per period utility of the agent starting from any date t is non-decreasing 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.*

Contrary to previous work, in our model front-loading is precisely a consequence of the commitment power of the monopolist. Since, as we discussed above, she finds it optimal to promise an efficient contract to the agent if he reports to be a high type, or to provide a contract with smaller and smaller inefficiency, she must commit to leave a higher rent because the higher is the efficiency of the contract, the more expensive it is to separate the agents' types.

We have not attempted an empirical investigation to verify which theory fits reality best, but it is certainly possible to test them empirically and verify if front-loading is a consequence of dynamic optimization in the presence of asymmetric information, as our theory predicts; or it is a consequence of some form of lack of commitment power, as previous theories predict.[34] Most likely the answer to this question depends on the characteristics of the market under analysis.

B The declining value of the monopolist's technology

As we discussed in Section V, the monopolist never finds it optimal to sell her technology except if the agent reveals himself to be a high type. It is therefore natural to look at the evolution of the price of the technology along the history in which the agent reveals to be a low type, because after any such history the agent has the opportunity to buy it out by reporting, truthfully or untruthfully, that he is a high 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

[32] For example this may be the case of the life insurance contracts studied by Hendel and Lizzeri [2000] for which detailed medical check-ups are required.

[33] This seems to be true, for example, for term life insurance in which there is evidence of adverse selection. See for example Brugiavini [1990] and Friedman and Warshawski [1990].

[34] For these results it is actually necessary that the monopolist cannot commit and that types are very highly persistent (previous papers assume constant types), since, as Proposition 7 below proves, if this is not the case then the optimal contract is renegotiation proof.

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. 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. This fact may suggest that the price of the firm is constant over time. However we have:

Proposition 6 *In the optimal "Sell the Firm" contract, the strike price of the call option to buy out the firm 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 is changing over time because the contract is becoming more and more efficient along this history and the improvements in the contract along the lowest history 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 is a high type in the future and an agent who is a high type today has a higher probability to be a high type tomorrow. This means that the price for the service that the low type is willing to pay increases slower than the increase in utility of a deviation for a high type: this is the reason why the outside option of the high type increases over time. It follows 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 for the high type.

VII 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 can not commit to not renegotiating the contract after some histories. In seminal papers, Dewatripont [1989], Hart and Tirole [1988] and Laffont and Tirole [1990], have shown that if types are constant, the optimal contract is never renegotiation-proof.[35] Perhaps surprisingly, when the agent's type follows a stochastic process this is not true: given a condition which is easily satisfied for realistic parametric specifications, the optimal contract is indeed renegotiation-proof.

[35] See Laffont and Tirole for a discussion of the literature [1993].

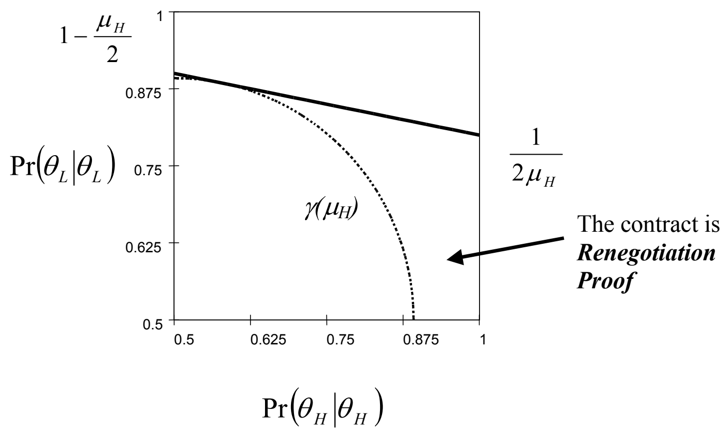

Figure 4: The renegotiation proofness condition.

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

Proposition 7 *If* $\Pr(\theta_L|\theta_L) \leq 1-\mu_H \Pr(\theta_H|\theta_H)$, *then the optimal contract is renegotiation-proof. Moreover, if this condition is not satisfied, then there exists a* $t < \infty$ *such that the contract is not renegotiation-proof only in the first* t *periods.*

Figure 4 represents the condition of Proposition 7 in an example in which there is a 30% initial probability that the agent is a high type: the contract is renegotiation-proof for any point below the straight 'thick line'. As it 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 (remember that types are correlated, so $\Pr(\theta_L|\theta_L)$ and $\Pr(\theta_H|\theta_H)$ are both larger than 1/2). In particular, 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 7 is that, given any initial prior, there is an upper-bound $\gamma(\mu_H) > 0$ on persistence such that for $\gamma < \gamma(\mu_H)$ the optimal contract is renegotiation-proof (area below the semicircle in Figure 4). In the example $\gamma(.3)$ would be larger than 87% (see Figure 4.A); but even assuming that $\mu_H = .5$, $\gamma(.5)$ would remain higher than 71%.

The intuition of this result can be seen from Figure 5. Assume, for the sake of illustration, that at time t the monopolist is contemplating a potential change only in the quantity offered after the agent reports a type θ_{t+j} after an history X_{t+j-1}, keeping constant all other quantities. The complete argument, clearly needs to consider a change

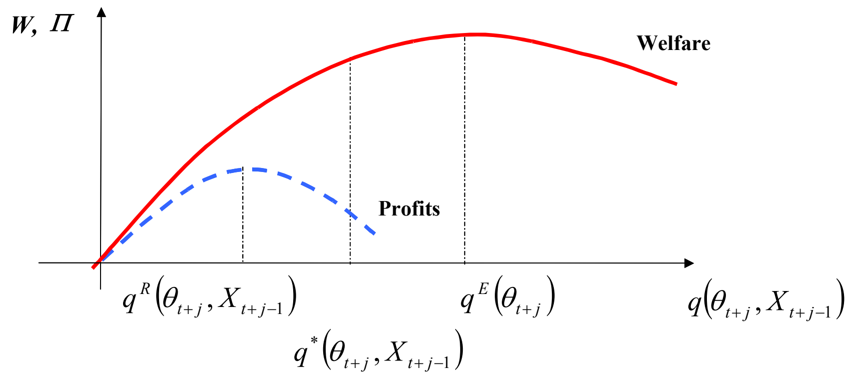

Figure 5: The ex-post maximization problem after an history h_t. The two concave functions represents the profits and welfare generated in period $t+j$ after history h_{t+j}: $q^*(\theta_{t+j}, X_j)$ is the optimal contract, $q^E(\theta_{t+j})$ is the efficient contract and $q^R(\theta_{t+j}, X_j)$ is the contract that is ex post optimal after history h_t.

on the entire sequence of contingent menus: this is a more sophisticate 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, $q^*_{t+j}(\theta_{t+j}, X_{t+j-1})$, is not larger than the efficient output $q^E(\theta_{t+j})$, it must be that welfare is strictly increasing in the interval $\left[0, q^*_{t+j}(\theta_{t+j}, X_{t+j-1})\right]$ (see Figure 5). This implies that the new output $q'_t(\theta_{t+j}, X_{t+j-1})$ prescribed by a renegotiated contract must be strictly larger than $q^*_{t+j}(\theta_{t+j}, X_{t+j-1})$.

Consider now a history h_t. If the agent has previously reported to be a high type, then the contract is efficient and not renegotiable; assume therefore that the agent has always reported to be a low type. If after h_t the monopolist could choose *any other contract*, the quantity supplied after an history h_{t+j} following or equal to h_t in the new contract would satisfy the first order condition:

$$q^R_t(\theta_{t+j}, X_{t+j-1}) = \theta_{t+j} - \Delta\theta \frac{\Pr(\theta_H | \theta_L)}{\Pr(\theta_L | \theta_L)} X(\theta_{t+j}, X_j) \left(\frac{\Pr(\theta_H | \theta_H) - \Pr(\theta_H | \theta_L)}{\Pr(\theta_L | \theta_L)} \right)^j \quad (5)$$

This formula is identical to the formula of the ex ante optimal contract (4), 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 (5) with (4) it is easy to verify that the ex ante optimal $q^*_{t+j}(\theta_{t+j}, X_{t+j-1})$ is larger than the ex-post optimal level $q^R_{t+j}(\theta_{t+j}, X_{t+j-1})$ if

$$\frac{\mu_H}{\mu_L} \left(\frac{\Pr(\theta_H | \theta_H) - \Pr(\theta_H | \theta_L)}{\Pr(\theta_L | \theta_L)} \right)^{t+j-1} \leq \frac{\Pr(\theta_H | \theta_L)}{\Pr(\theta_L | \theta_L)} \left(\frac{\Pr(\theta_H | \theta_H) - \Pr(\theta_H | \theta_L)}{\Pr(\theta_L | \theta_L)} \right)^j \quad (6)$$

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 (6) holds, any quantity larger than $q^*_{t+j}(\theta_{t+j}, X_{t+j-1})$ 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 $q^*_{t+j}(\theta_{t+j}, X_{t+j-1})$ would not be renegotiated at any time t.

The result that the optimal contract is never renegotiated should be interpreted in the light of the previous literature. As we mentioned, when types are constant the contract is never renegotiation-proof. The previous literature has characterized the equilibria of these models under some assumptions.[36] A common feature of all these characterizations is that the optimal contract requires the consumer to play sophisticated mixed strategies, that 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 shows that when types are correlated but follow a stochastic process, even if the correlation level is very high (as in Figure 4) there is not necessarily a contradiction between optimality and renegotiation-proofness. Moreover, this proves that 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 that is necessary to guarantee the latter property, therefore, are implications of the assumption that types are constant or very highly correlated.

On the other hand, Proposition 7 shows how even a weak form of commitment (consistent with renegotiation-proofness) may induce outcomes that are very different from the benchmark case in which the seller is forced to offer one-period contracts (as in the Ratchet Effect literature). Kennan [2001] has extended the model by Hart and Tirole [1988] to the case when types may change over time, showing that if the seller can offer only one-period contracts, then there is an equilibrium in which supply has a cyclical behavior. Although this result depends on the assumption that supply may assume only two values, and so can not be applied directly to our framework, it is easy to see that our characterization and the convergence result are true even with 0-1 supply, implying that the optimal renegotiation-proof contract would not be cyclical, but converge along all histories to an efficient menu.

[36] Dewatripont [1989] consider some restrictions on the space of contracts that can be offered; Hart and Tirole [1988] consider contracts with 0-1 supply; Laffont and Tirole [1990] consider two periods. All these papers assume constant types.

VIII Large discount factors

All the results presented above are valid for any $\delta \in (0,1)$. If we assume that $\delta \to 1$, however, we have much stronger results: indeed, in this case we can easily bound the inefficiency and determine the distribution of surplus. As a benchmark case, observe that with constant types the average utility of the consumer is bounded away from zero for any δ, even for $\delta \to 1$, and the average payoff of the monopolist is also independent of the discount factor and equal to the optimal static profit. The following proposition shows that this property is not robust since an arbitrarily small perturbation of the evolution of types has a very high impact on surplus and payoffs for high discount factors.

Proposition 8 *For any generic Λ, even if $\boldsymbol{\alpha}_H$ does not stochastically dominate $\boldsymbol{\alpha}_L$, then as $\delta \to 1$ 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.*

The intuition of Proposition 8 is immediate because the assumption that $\delta \to 1$ considerably simplifies the analysis. When the discount factor is high, it does not matter what happens in the first K periods, for K 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 the consumer, at time one, does not really have much private information regarding the steady state. Note that the steady state is going to be characterized by a distribution and its realizations will be known only to the agent; but at time one the expected distribution of any type and of the monopolist almost coincide. For this reason the monopolist can separate the agents paying only a minuscule rent to the higher type.

It is worth to point out the differences between this result and the result in Proposition 1 because the logic of their proofs is different. The proof of Proposition 8, in fact, does not require the assumption that the distribution of types conditional on being a high type stochastically dominates the distribution conditional on being a low type. This implies that the result in Proposition 8 is stronger than the result that would have followed from a mere application of Proposition 1 to the case in which $\delta \to 1$. However, while Proposition 1 characterizes the optimal contract for any δ, Proposition 8 is only a limit result. This implies that if $\delta < 1$ the contract may converge as $t \to \infty$ to a contract with a small but strictly positive inefficiency. More importantly, even in the limit case in which $\delta \to 1$, Proposition 8 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, which may well not converge to zero even as $\delta \to 1$.

Secondly, note that the result holds even if the contract must be renegotiation-proof, proving that the monopolist does not gain very much from commitment power when the discount factor is high, even if the condition in Proposition 7 does not hold.

IX 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 evolves following a Markov process, even if the types are arbitrarily highly persistent. While in the first case the contract is constant, in the latter the contract is truly dynamic and converges to the efficient contract. Moreover in the former case the monopolist's profits and the surplus are independent of the discount factor and always lower than the efficient level; instead in the latter case surplus converges to the optimal level and the monopolist extracts all rents as the discount factor converges to one.

The characterization of the optimal contract has uncovered some interesting features of dynamic contracting. Even if the environment has only one-period memory and risk neutral agents, the optimal contract is not stationary and has unbounded memory. This, however, does not imply that the structure of the contract is complex: indeed it can be represented by a simple state variable. In analogy with the static model, we have a stronger version of the *No Distortion at the Top* principle, which implies that the entire state-contingent contract becomes forever efficient as soon as the agent reports to be 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. Finally, with constant types there is always a conflict between optimality and renegotiation-proofness, and the latter property is guaranteed only if consumers use sophisticated mixed strategies. With stochastic types, on the contrary, even if there is high persistence, the optimal contract is renegotiation proof for reasonable parameters' specifications and consumers 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. We have shown that the monopolist may inefficiently find it optimal to keep the ownership of the technology in order to control the agents' future rents and therefore maximize rent extraction. This inefficient retention of the property rights may potentially last for infinite periods, although the allocation of property rights will be efficient with probability one in the long term. Moreover the contract displays an important property highlighted by empirical research: the average continuation utility of the agent increases over time and 'old' customers receive a better treatment than new customers.

It is clear that this theory of dynamic contracting can be applied to problems that are more general than price discrimination. Battaglini and Coate [2003], for example, extend these techniques to characterize the Pareto optimal frontier of taxation in a dynamic model with correlated types. Some questions remain open, as the characterization of the optimal renegotiation-proof contract when the correlation of types is near to one. We are not aware of any other model that studies an infinite contractual relationship with stochastic types with (or without) renegotiation.[37] We conjecture that the results of this paper would be true in that environment too. We leave such an extension for future research.

[37] See Laffont and Tirole [1993] for a survey of the literature.

Appendix

A. Proof of Lemma 1

For a generic program $\mathcal{P}$, we define $\mathcal{V}(\mathcal{P})$, the value achieved by the objective function at the optimum. Lets also define $\mathcal{P}_I^R$ the program in which expected profits are 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 $IR_{h_t}(\theta_H)$ and $IC_{h_t}(\theta_L)$ as equality, and achieves the same value as $\langle \mathbf{p}, \mathbf{q} \rangle$: $\mathcal{V}\left(\mathcal{P}_I^R\right) = \mathcal{V}\left(\mathcal{P}_I\right)$*

Proof. We first show that the price vector can be modified to guarantee that all the incentive constraints are binding; then we show that it can be modified to make the individual rationality constraints binding as well.

Step 1 We show the result by induction. Assume that for any solution $\langle \mathbf{p}, \mathbf{q} \rangle$ of $\mathcal{P}_I^R$ and any $t < \infty$, there is a $\mathbf{p^t}$ such that $\langle \mathbf{p^t}, \mathbf{q} \rangle$ is also a solution of $\mathcal{P}_I^R$ and: all incentive compatibility constraints are binding up to period t; and $\mathbf{p}^t$ is identical to $\mathbf{p}$ in any period $j > t$: $p^t(\theta; h_j) = p(\theta; h_j)$ for $j > t$. This step is clearly satisfied at $t = 1$, since the incentive compatibility constraint is necessarily binding at the optimum. We now show that for any solution $\langle \mathbf{p}, \mathbf{q} \rangle$ there must be a price vector $\mathbf{p^{t+1}}$ such that all incentive compatibility constraints are binding up to period t+1 and $\mathbf{p^{t+1}}$ is identical to $\mathbf{p}$ in any period $j > t+1$. Given the induction step, assume without loss of generality that the incentive constraints are binding for any $j \leq t$. There are two cases to consider. Assume first that at period $t+1$, after some history $h_t = \{h_{t-1}, \theta_L\}$, 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 $\{h_t, \theta_H\}$ and $\{h_t, \theta_L\}$ are respectively $p^{t+1}(\theta_H; h_t) = p(\theta_H; h_t) + \varepsilon$ and $p(\theta_L; h_t) = p(\theta_L; h_t)$;[38] simultaneously, reduce the price after history $\{h_{t-1}, \theta_L\}$ so that:

$$p^{t+1}(\theta_L; h_{t-1}) = p(\theta_L; h_{t-1}) - \delta \alpha_{LH} \varepsilon$$

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

[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 an history $\{h_t, \theta\}$.

is truthful $U(\theta_H; h_{t-1})$ is unchanged; if the high type, however, reports to be a low type he would receive:

$$U(\theta_H; h_{t-1}) + \delta(\alpha_{LH} - \alpha_{HH})\varepsilon \leq U(\theta_H; h_{t-1}) \tag{A.1}$$

where the inequality follows by first order stochastic dominance. It follows that $IC_{h_j}(\theta_H)$ are satisfied for any $j \leq t$ and $\langle \mathbf{p^{t+1}}, \mathbf{q} \rangle$ is a solution of $\mathcal{P}_I^R$. By the induction step we can find a new price vector $\mathbf{p}^t$ which is such that the incentive compatibility constraints are binding in all periods $j < t$. Since this price vectors is vector is identical to p^{t+1} for periods $j > t$, the incentive compatibility continues to be binding at t+1 as well.

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: $p^{t+1}(\theta_H; h_{t-1}) = p(\theta_H; h_{t-1}) - \delta\alpha_{HH}\varepsilon$. This new contract would leave all the constraints of the relaxed problem satisfied with the incentive constraint binding in the first $\widehat{t} + 1$ 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 binding. We now show that the individual rationality constraints can be made binding too. The individual rationality constraint must be binding at stage 1. Again, we prove the result for the remaining periods by induction. Assume that in all periods $j \leq t$ $IR_{h_j}(\theta_L)$ is tight 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 $\mathcal{P}_I^R$, it would leave the incentive compatibility constraints binding, and satisfy all the individual rationality constraint as equality up to period $t+1$. Profit would remain unchanged as well. ■

We now prove:

Claim 2 $\mathcal{V}(\mathcal{P}_I^R) = \mathcal{V}(\mathcal{P}_I)$

Proof. Since $\mathcal{V}(\mathcal{P}_I^R) = \mathcal{V}(\mathcal{P}_{II})$, we only need to 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) and the high type receives at least his reservation value ($IR_{h_t}(\theta_H)$ constraint) in the optimal contract of the relaxed problem. 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:

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

where $p(\theta_i; h_t)$ $i = H, L$ is the price charged after the agent declares to be a type i and 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 it 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:

$$U(\theta_H; h_t, \theta_H) = \Delta\theta \sum_{j=0}^{\infty} \delta^j \left(\Pr(\theta_H | \theta_H) - \Pr(\theta_L | \theta_H)\right)^j q^E(\theta_L)$$

where, remember, $q^E(\theta_L)$ is the efficient quantity when the type is θ_L. If the agent reports to be a low type 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 $U(\theta_H; h_t, \theta_H)$. So we have $U(\theta_H; h_t, \theta_H) - U(\theta_H; h_t, \theta_L) \geq 0$ for any h_t and, since types are correlated: $(\alpha_{HH} - \alpha_{LH})\,\Delta U(\theta_H, h_t) \geq 0$. It follows that:

$$p(\theta_H; h_t) - p(\theta_L; h_t) \geq (q(\theta_H; h_t) - q(\theta_L; h_t))\,\theta_L + \delta\alpha_{LH}\Delta U(\theta_H, h_t)$$

and $IC_{h_t}(\theta_L)$ is satisfied.

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

$$q(\theta_H; h_t)\,\theta_H - p(\theta_H; h_t) + \delta\alpha_{HH} U(\theta_H; h_t, \theta_H) \geq \delta(\alpha_{HH} - \alpha_{LL})\,U(\theta_H; h_t, \theta_L) > 0$$

and therefore $IR_{h_t}(\theta_H)$ is satisfied too. ■

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

B. Proof of Proposition 1

Lets define $h_t^L := \{\theta_L, \theta_L, ...\theta_L\}$ the history along which the agent always report to be a low type for t-1 periods. Using the binding $IC_{h_t}(\theta_H)$ and $IR_{h_t}(\theta_L)$ we can formulate

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

the utility of the high type at time 1 as:

$$\begin{aligned} U(\theta_H; h_1) &= \Delta\theta q_1^*(\theta_L | h_1) \\ &\quad +\delta(\boldsymbol{\alpha}_H - \boldsymbol{\alpha}_L)^T \left(\begin{array}{c} \Delta\theta q_2^*\left(\theta_L \left| h_2^L \right.\right) + \delta(\boldsymbol{\alpha}_H - \boldsymbol{\alpha}_L)^T \left(\begin{array}{c} \Delta\theta q_3^*\left(\theta_L \left| h_3^L \right.\right) + ... \\ 0 \end{array} \right) \\ 0 \end{array} \right) \end{aligned}$$

which can be written as

$$U(\theta_H; h_1) = \Delta\theta \sum_{j=0}^{\infty} \delta^j \left(\Pr(\theta_H | \theta_H) - \Pr(\theta_H | \theta_L)\right)^j q_{j+1}^*\left(\theta_L \left| h_{j+1}^L \right.\right) \tag{B.2}$$

By the binding $IR_{h_1}(\theta_L)$ the low type receives zero at time one. It follows that $\mathcal{P}_{II}$ can be represented as:

$$\begin{aligned} E\Pi(\theta | h_1) &= \sum_{i=H,L} \mu_i \left[q(\theta_i)\theta_i - \frac{q(\theta_i)^2}{2} + \delta\boldsymbol{\alpha}_i^T \left(\begin{array}{c} W(\theta_H; \theta_i) \\ W(\theta_L; \theta_i) \end{array} \right) \right] \\ &\quad -\mu_H \Delta\theta \sum_{j=0}^{\infty} \delta^j \left(\Pr(\theta_H | \theta_H) - \Pr(\theta_H | \theta_L)\right)^j q_{j+1}^*\left(\theta_L \left| h_{j+1}^L \right.\right) \end{aligned} \tag{B.3}$$

Two cases:

Case 1: $h_t = \{h_{t-1}, \theta\} \in H_t \setminus h_t^L \ \forall t \geq 1$. 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 $q_t^*\left(\theta_L \left| h_t^L \right.\right)$ implies that:

$$q_t^*\left(\theta_L \left| h_t^L \right.\right) = \theta_L - \Delta\theta \frac{\mu_H}{\mu_L} \left[\frac{\Pr(\theta_H | \theta_H) - \Pr(\theta_H | \theta_L)}{\Pr(\theta_L | \theta_L)} \right]^{t-1} \tag{B.4}$$

which completes the characterization of the optimal contract.

C. Proof of Proposition 4

Since after the agent reports to be a high type the optimal contract is efficient, 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 quasi-linear 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 afterwards. 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 can not achieve these distortions without control of the technology. ■

D. Proof of Proposition 5

When $IR_{h_t}(\theta_L)$ is binding, the low type receives zero utility in any period. Therefore we only need to show that the average utility of the high type is non-decreasing in time. Using (2) and the incentive compatibility constraint of the high type we can write:

$$\begin{aligned} U(\theta_H; t, X_{t-1}) &= \Delta\theta \sum_{j=0}^{\infty} \delta^j \left(\Pr(\theta_H|\theta_H) - \Pr(\theta_H|\theta_L)\right)^j \cdot \\ & \left[\begin{array}{c} \theta_L - \\ \Delta\theta X(\theta_L, X_{t-1}) \frac{\mu_H}{\mu_L} \left(1 - \frac{\Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)}\right)^{t+j-1} \end{array} \right] \end{aligned}$$

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

$$X(\theta_L, X_{t'-1}) - X(\theta_L, X_{t-1}) \left(1 - \frac{\Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)}\right)^{(t-t')}$$

which is non-negative 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 non-decreasing in any history and strictly increasing in a non-empty subset of histories. It follows that, at time zero, the expected average rent starting from period t is strictly increasing in t. ■

E. Proof of Proposition 6

Since the monopolist's technology is sold as soon as compatible with profit maximization, its sale can occurs only along an history in which the agent has always reported 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

$$W^*(\theta_H) - P(h_t) = U(\theta_H, \theta_L; h_t) \tag{E.5}$$

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 an history h_t; and $W^*(\theta_i)$ expected first best surplus from time t if the type at t is θ_i.[40] Since $W^*(\theta_L)$ 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 (B.2)). ■

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

F. Proof of Proposition 7

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

$$E\left[\Pi\left(\theta\left|h_t\right.\right)\left|h_t\right.\right] = \Pr\left(\theta_H\left|\theta_L\right.\right)\left[W\left(\theta_H,\mathbf{q}_H\right) - R\left(\mathbf{q}_L\right)\right] + \Pr\left(\theta_L\left|\theta_L\right.\right)W\left(\theta_L,\mathbf{q}_L\right) \tag{F.6}$$

where $\mathbf{q}_i$ $\mathbf{i} = H, L$ is the sequence of quantities in the menus offered if the agents reports to be a type i at t; $W\left(\theta_i,\mathbf{q}_i\right)$ is the expected surplus generated in the contract if the agent is of type i and $\mathbf{q}_i$ is offered; and $R\left(\mathbf{q}_L\right)$ 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 (B.2), and the rent of the low type is zero). Indeed (F.6) is a compact form to write (B.3) when the posterior probability that the type is high is $\Pr\left(\theta_H\left|\theta_L\right.\right)$ starting from h_t. The monopolist's ex post problem ($\mathcal{P}_{ex\ post}$) consists in maximizing (F.6) 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 (F.6) 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^*:

$$\sum_{i=H,L}\Pr\left(\theta_i\left|\theta_L\right.\right)W\left(\theta_i,\mathbf{q}_i\right) \geq W^* \tag{F.7}$$

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:

$$\mathcal{L}_P = (1+\tau)\left[\widehat{\lambda}W\left(\theta_H,\mathbf{q}_H\right) + W\left(\theta_L,\mathbf{q}_L\right)\right] - \widehat{\lambda}R\left(\mathbf{q}_L\right) \tag{F.8}$$

where $\widehat{\lambda}$ is the ex post likelihood ratio $\frac{\Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)}$ and τ is the Lagrangian multiplier associated with (F.7). We proceed in three simple steps:

Step 1: $\tau > 0$. Given that $\Pr\left(\theta_L\left|\theta_L\right.\right) \leq 1 - \mu_H\Pr\left(\theta_H\left|\theta_H\right.\right)$ implies (6), if $\tau = 0$ then the solution of (F.8) implies that all the quantities in $\mathbf{q}_H$ are set efficiently and the quantities in $\mathbf{q}_H$ are distorted downward more than the solutions of the ex ante optimal problem (the argument is the same as in Section VII). But then (F.7) must be violated, and this is a contradiction.

Step 2. Denote λ° as the ex ante likelihood ratio $\frac{\mu_H}{\mu_L}$ and $A = \left[\frac{\Pr(\theta_H|\theta_H)-\Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)}\right]$. Then we have that $\frac{\widehat{\lambda}}{(1+\tau)} = \lambda^\circ A^{t-1}$. Indeed, it is easy to verify that the quantities following history h_t in the optimal solution of the ex ante problem maximize:

$$\mathcal{L}_A = \lambda^\circ W\left(\theta_H,\mathbf{q}_H\right) + W\left(\theta_L,\mathbf{q}_L\right) - \lambda^\circ A^{t-1}R\left(\mathbf{q}_L\right) \tag{F.9}$$

If $\frac{\widehat{\lambda}}{(1+\tau)} < \lambda^{\circ}\left[\frac{\Pr(\theta_H|\theta_H)-\Pr(\theta_H|\theta_L)}{\Pr(\theta_L|\theta_L)}\right]^{t-1}$ then the solution of (F.8) would be less distorted than the solution of (F.9), implying that (F.7) is not binding and so $\tau = 0$, contradiction. Similarly we can prove that the reverse inequality is not possible. Therefore we conclude that the solution of (F.8) and (F.9) 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 $\widetilde{t}$ such that for $t > \widetilde{t}$, then (6) is satisfied and the argument in steps 1 and 2 is valid for any $t > \widetilde{t}$. ■

G. Proof of Proposition 8

Starting in period t from any history $\left\{h_t, \theta_t = \widetilde{\theta}\right\}$, the expected first best surplus from time t is independent from t and equal to $W^*\left(\widetilde{\theta}\right) = \sum_{j\geq t} \delta^j E_t\left[\frac{1}{2}\theta_j^2 \middle| \theta_t = \widetilde{\theta}\right]$. Consider a contract c^+ in which a fixed fee $F = W^*(\theta_L)$ is charged in period 1 and then an efficient menu plan offered in which:

$$q_t^+(\theta) = \theta, \ p_t^+(\theta) = \frac{1}{2}\theta^2 \text{ for any } t \geq 1$$

This contract is clearly incentive compatible and individually rational for any $t \geq 1$; 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:

$$\begin{aligned}(1-\delta)\Pi^* - (1-\delta)E\left[\sum_{t=0}^{\infty}\delta^t w^*(\theta)\right] \ \geq \ & (1-\delta)E\left[\sum_{t=0}^{\infty}\delta^t w^*(\theta)|\theta_1 = \theta_L\right] \\ & -(1-\delta)E\left[\sum_{t=0}^{\infty}\delta^t w^*(\theta)\right]\end{aligned} \quad \text{(G.10)}$$

As $\delta \to 1$, the right hand side can be written as:

$$\Omega\left(\theta;\widehat{t}\right) = \lim_{\delta\to 1}(1-\delta)\left\{E\left[\sum_{t\geq\widehat{t}}\delta^t w^*(\theta)|\theta_1 = \theta_L\right] - E\left[\sum_{t\geq\widehat{t}}\delta^t w^*(\theta)\right]\right\}$$

Since (G.10) must holds for any $\widehat{t} \geq 1$ and $\lim_{\widehat{t}\to\infty}\Omega\left(\theta;\widehat{t}\right) = 0$ (because the process converges to a stationary distribution) we have that

$$\lim_{\delta\to 1}(1-\delta)\Pi^* = \lim_{\delta\to 1}(1-\delta)E\left[\sum_{t=0}^{\infty}\delta^t w^*(\theta)\right]$$

Which also implies that the agent's average payoff is zero. ■

References

[1] Atkeson A and R. Lucas (1992): “On Efficient Distribution With Private Information,” *Review of Economic Studies*, 59:427-453.

[2] Battaglini M. and S. Coate (2003): “Pareto Optimal Dynamic Taxation,” Cornell and Princeton, mimeo.

[3] Baron, D. and D. Besanko (1984): “Regulation and Information in a Continuing Relationship,” *Information Economics and Policy*, 1(3):267-302.

[4] Biehl A. (2001): Durable-goods Monopoly with Stochastic Values,” *RAND Journal of Economics*, 32(3):565-577.

[5] Brugiavini A. (1990): “Longevity Risk and the Life cycle.” Ph. D. Dissertation, LSE, London.

[6] Bulow, J.(1982): “Durable Good Monopolists,” *Journal of Political Economy*, 90:314-332.

[7] Chiappori, P. (2000): “Econometric Models Under Asymmetric Information,” *Handbook of Insurance*, eds. G. Dionne.

[8] Chiappori, P., I. Macho, P. Rey and B. Salanie (1994): “Repeated Moral Hazard: Memory, Commitment, and the Access to Credit Markets,” *European Economic Review*, 1527-1553.

[9] Courty, P. and H. Li (2000): “Sequential Screening,” *Review of Economic Studies*, 67(4):697-718.

[10] Dewatripont, M. (1989): “Renegotiation and Information Revelation Over Time: The case of Optimal labor Contracts,” *Quarterly Journal of Economics*, 104(3):589-619.

[11] Dionne G. and N. Doherty (1994): “Adverse Selection, Commitment and Renegotiation: Extension to and Evidence from insurance Markets,” *Journal of Political Economy*, 102:209-235.

[12] Friedman, B. and M. Warshawsky (1990): “The Cost of Annuities: Implications for Savings Behavior and Bequests.” *Quarterly Journal of Economics*, CII(1):135-154.

[13] Gertsner, L. Jr. V. (2002), *Who Says Elephant Can’t Dance*, Harper Collins. New York.

[14] Green E. (1987): "Lending and the Smoothing of Uninsurable income," in E. Prescott and N. Wallace (eds.), *Contractual Arrangements for Intertemporal Trade*. Minneapolis, University of Minnesota Press, pp. 3-25.

[15] Hart O. and J. Tirole (1988): "Contract Renegotiation and Coasian Dynamics," *Review of Economic Studies*, 55:509-540.

[16] Hendel I. and A. Lizzeri (2000): "Adverse Selection in Durable Goods Markets," *American Economic Review*, 89(5):1097-115.

[17] Hendel I. and A. Lizzeri (2002a): "The Role of Leasing Under Adverse Selection," *Journal of Political Economy*, 101(1):113-143.

[18] Hendel I. and A. Lizzeri (2002b): "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," NBER working paper n. 7470. Forthcoming in the *Quarterly Journal of Economics*.

[19] Hendel I., A. Lizzeri and M. Siniscalchi (2002): "Efficient Sorting in a Dynamic Adverse Selection Model," mimeo.

[20] Harris M.; B. Holmstrom (1982), "A Theory of Wage Dynamics," *The Review of Economic Studies*, 49(3):315-333.

[21] Karlin S. and H. Taylor (1975): *A First Course in Stochastic Processes*, Academic Press, San Diego.

[22] Kennan, J. (2001): "Repeated Bargaining wit Persistent Private Information," *Review of Economic Studies*, 68: 719-755.

[23] Laffont J.J. and J. Tirole (1987): "Comparative Statics of the Optimal Dynamic Incentive Contract," *European Economic Review*, 21:127-148.

[24] Laffont J.J. and J. Tirole (1988): "The Dynamics of Incentive Contracts," *Econometrica*, 56(5):1153-1175.

[25] Laffont J.J. and J. Tirole (1990): "Adverse Selection and Renegotiation in Procurement," *The Review of Economic Studies*, 57(4):597-625.

[26] Laffont J.J. and J. Tirole (1993): *A Theory of Incentives in Procurement and Regulation*, Cambridge, MIT Press.

[27] Laffont J.J. and J. Tirole (1996): "Pollution Permits and Compliance Strategies," *Journal of Public Economics*, 62:85-125.

[28] Miravete, E. (2002): "Choosing the Wrong Calling Plan? Ignorance and Learning," *American Economic Review*, forthcoming.

[29] Osborne, M. and A. Rubinstein (1994): *A Course in Game Theory*, Cambridge, MIT Press.

[30] Roberts, K. (1982): "Long-term Contracts," mimeo.

[31] Rogerson, W. (1985): "Repeated Moral Hazard," *Econometrica*, 53:69-76.

[32] Townsend, R. (1982): "Optimal Multiperiod Contracts and the Gain from Enduring Relationship under Private Information," *Journal of Political Economy*, 90:1166-1186.