Contracting over Time when Writing is Costly[*]

Pierpaolo Battigalli
Bocconi University
pierpaolo.battigalli@uni-bocconi.it

Giovanni Maggi
Princeton University
maggi@princeton.edu

December 2000

Abstract

In this paper we examine a model of contracting where parties interact repeatedly and can contract at any point in time, but writing enforceable contracts is costly. We argue that the costs of writing contracts can provide a theoretical explanation for two common observations: the fact that long-term contracts are often preferred to spot contracts, and the fact that relationships are often managed by a combination of formal (externally enforced) and informal (self-enforcing) contracts.

[*]Giovanni Maggi thanks the Department of Economics at NYU and the European University Institute (EUI), where he was visiting during part of this project. Pierpaolo Battigalli gratefully acknowledges …nancial support from the EUI. We thank the participants in the workshop "Contract Enforcement and Incompleteness" (Siena, July 3-5, 2000) and in particular our discussant K…r Eliaz for useful and constructive comments.

1. Introduction

It is often argued that writing detailed contracts can be very costly, and that this may be a cause of contractual incompleteness. The costs of writing contracts have particularly interesting implications for the structure of contracts in a repeated relationship, as they generate tradeoffs at (at least) two levels: the choice between long-term contingent contracts and short-term spot contracts, and the choice between formal (externally enforced) and informal (self-enforcing) contracts.[1] In this paper we attempt a rigorous examination of these tradeoffs. We will argue that the presence of writing costs can provide a theoretical explanation for two common observations: the fact that long-term contracts are often preferred to spot contracts, and the fact that relationships are often regulated by a combination of formal and informal contracts.

Several theoretical explanations have been proposed for the widespread use of long-term contracts. Most of these explanations share the idea that long-term contracts provide long-term commitment. Long-term commitment may be valuable to induce parties to make relationship-specific (long-term) investments, to facilitate intertemporal smoothing or insurance, or to provide incentives to reveal private information.[2] Another explanation of long-term contracts – which we consider complementary to the ones just mentioned – is the presence of transaction costs. This idea is expressed, for example, by Hart and Holmstrom (1987, p. 130): "if a relationship is repetitive, it may save on transaction costs to decide in advance what actions each party should take rather than to negotiate a succession of short term contracts." When transaction costs take the form of fixed contracting costs (for example due to bargaining costs), this argument makes intuitive sense. However, if transaction costs take the form of 'variable' writing costs (as they will in our model), and spot contracting is feasible (i.e. parties can contract after observing the state of nature and before actions are taken), it is no longer clear that they can explain the use of long-term contracts. If anything, intuition might suggest the opposite conclusion, because spot contracting bypasses the cost of describing contingencies. We

[1] We will use interchangeably the expressions "formal" and "externally enforced" contract; likewise for "informal" and "self-enforcing" contract. We refrain from using the terminology "explicit" vs. "implicit" contracts – which is common in the literature – because contracts may be quite explicit even though they cannot be enforced in court.

[2] Papers that highlight these benefits of long-term contracts include Townsend (1982), Lambert (1983), Allen (1985), Rogerson (1985), Harris and Holmstrom (1987), Crawford (1988), Malcomson and Spinnewyn (1988), Rey and Salanie (1990), Fudenberg et al. (1990).

attempt a rigorous examination of this issue by means of a simple model that makes explicit the language used to write contracts.

More specifically, in section 2 we consider a principal-agent model with verifiable contingencies and actions. The state is described by a set of dichotomous elementary events, and the agent can take a set of dichotomous elementary actions. There is a one-to-one correspondence between elementary events and actions, in the sense that it is efficient to take a given action if and only if the corresponding event occurs. The agent's interests are in conflict with those of the principal, so that in the absence of a contract the agent will always take the inefficient actions. The parties interact repeatedly and can write contracts at any point in time (which includes the possibility of spot contracts). The language consists of primitive sentences that describe elementary events and actions; this language can be used to describe state-dependent constraints on behavior. Each primitive sentence has a cost, and the total cost of writing a contract is the sum of the costs of its primitive sentences.

Absent writing costs, the model has little predictive power, as there is a plethora of optimal contracting plans (including a contingent contract, a sequence of spot contracts, and a host of intermediate solutions). But with an arbitrarily small writing cost, the model yields a unique optimum. Depending on parameters, the optimum is either a contingent contract or a noncontingent 'default' contract (i.e. a noncontingent contract that is subject to exceptions or amendments over time).

We emphasize that a contingent contract may be optimal even though the cost of describing contingencies can be avoided by writing spot contracts. This conclusion depends heavily on our language-based approach. In particular, things would be radically different if writing costs were modeled in a more conventional way, along the lines of Dye's (1985a) well-known model: we show that, with writing costs a' la Dye, if the number of possible states is large enough a contingent contract is dominated by a sequence of spot contracts.

If writing costs are not small, the optimal contracting plan may be incomplete. Contractual incompleteness can take the form of rigidity and/or discretion. In the optimal contracting plan, the set of actions is partitioned in three subsets: one subset of actions is regulated either by contingent rules or by 'default' rules; another group of actions is regulated by rigid rules; and some actions are left to the agent's discretion. Regardless of the degree and type of contractual incompleteness, however, the model yields a simple prediction: if a long-term contract is defined as a contract that is written once and for all, the model predicts that a long-term contract is optimal if the relationship is sufficiently durable and uncertainty is relatively high. This is true irrespective of the level of the writing cost, which only determines the extent of contract incompleteness.

In section 3 we introduce the possibility of self-enforcing contracts, i.e. contracts that are enforced by reputation mechanisms rather than by external courts. The advantage of a self-enforcing contract is that it can be communicated informally, rather than being written formally, because it need not be enforceable in court, and this saves on writing costs. On the other hand, the absence of an external enforcement mechanism may limit the effectiveness of a self-enforcing contract. Thus, there may be a trade-off between formal and informal contracting.

We find that formal and informal contracts tend to be used jointly, with some tasks regulated formally and others regulated informally. In principle there can be three types of contractual incompleteness in this setting: discretion, rigidity of the formal contract, and rigidity of the informal contract. We show however that rigidity in the informal contract is never optimal. In the optimal contract, low-cost tasks are regulated by informal rules, intermediate-cost tasks are regulated by formal (contingent or rigid) rules, and high-cost tasks are left to the agent's discretion.

The relative importance of formal versus informal contracting is captured by the ratio between the number of tasks regulated formally and the number of tasks regulated informally. We find that this ratio may increase with the writing cost; in particular, it is possible that an increase in the writing cost results in fewer tasks being regulated informally without changing the number of tasks regulated formally (but making the formal contract more rigid). Furthermore, as the writing cost approaches zero, the optimal contract need not be fully formal, and may even be fully informal. We also find that the relative importance of formal contracting may increase with the durability of the relationship, even though a more durable relationship makes informal contracts easier to sustain.

Next we discuss the related literature. There are a few papers in the macro/labor field where long-term contracts are motivated by the presence of fixed contracting costs (perhaps due to costs of bargaining), as opposed to our 'variable' writing costs (i.e. costs that increase with the length of the contract). Examples of this literature are Gray (1976, 1978) and Dye (1985b). Of course, if there are fixed contracting costs, long-term contracts may be preferred to short-term contracts. However, the implications of fixed contracting costs are quite different from those of the variable writing costs considered here. For example, if in our model we replaced our writing costs with a fixed contracting cost, it would always be optimal to write a complete contingent contract (or no contract at all). Fixed contracting costs cannot explain the presence of rigidity or discretion in a long-term contract, or the use of 'default' contracts. Moreover, if self-enforcing contracts are available, fixed contracting costs cannot explain the simultaneous use of formal and self-enforcing contracts.

Our results are also interesting in relation to Maskin and Tirole (1999). They argue that the presence of unforeseen contingencies (or, by an extension of their argument, the costs of describing contingencies) does not imply inefficiencies in contracting, provided that parties can design an appropriate message-based mechanism to be played after contingencies are observed and before actions are taken. Since in our setting parties have the option of observing the state before contracting, mechanisms à la Maskin-Tirole are redundant, hence none of our conclusions changes if such mechanisms are available. In section 2.4 we discuss in more detail the relevance of our results for the Maskin-Tirole critique.

The interaction between self-enforcing contracts and formal contracts is analyzed also in Baker et al. (1994) and Pearce and Stacchetti (1998).[3] These papers propose a different – and in many respects complementary – explanation for the combined use of formal and informal contracts. They consider a repeated principal-agent model where parties can write a formal contract based on verifiable signals of the agent's action, and/or an informal contract based on unverifiable signals. Both papers find that it may be optimal to offer a combination of a formal wage and an informal 'bonus'. There are important differences between these models and our model of section 3.2, both in the focus of the analysis and in the comparative-statics predictions. We will discuss these differences at length in section 3.2.1.

[3] There is also a vast literature on purely self-enforcing contracts. Bull (1987) and Mac Leod and Malcomson (1989) are two examples of this literature.

Our paper is related to the literature on complexity costs as a cause of contractual incompleteness, in particular Dye (1985a), Anderlini and Felli (1994, 1999), Krasa and Williams (1999), MacLeod (2000) and Battigalli and Maggi (2000). Unlike these papers, we examine the link between complexity costs and contractual incompleteness in the context of a long-term relationship.[4]

2. A model of formal contracting

We start by modeling the language used to write contracts.

$\Sigma^e = \{e_1; e_2; e_3; ...\}$ is a finite collection of primitive sentences, each of which describes an elementary event concerning the external environment. For example, e_1: "the passenger has a moustache", e_2: "the passenger's bag is red".

$\Sigma^a = \{a_1; a_2; a_3; ...\}$ is a finite collection of primitive sentences describing elementary actions (behavioral events, or tasks), for example, a_1 : "check the passenger's passport", a_2 : "search the passenger's bag".

With a slight abuse of terminology, we will use the notation e_k (resp. a_k) to indicate both an elementary event (resp. action) and the primitive sentence that describes it.

We assume that this language is the (only) common-knowledge language for the parties and the courts. This ensures that there are no problems of ambiguous interpretation of the contract.

A state is a complete description of the exogenous environment, represented by a valuation function $s : \Sigma^e \to \{0; 1\}$, where $s(e_k) = 1$ means that primitive sentence e_k is true at state s and $s(e_k) = 0$ means that primitive sentence e_k is false at state s.[5] In other words, $s(e_k)$ is a dummy variable that takes value one if elementary event e_k occurs and zero otherwise, and a

[4] We should also mention a paper by Bart Lipman (1997), which analyzes the implications of computation costs for the tradeoff between long-term and short-term contracting. He considers boundedly rational agents who trade repeatedly and can learn the payoff implications of future contingencies only by paying a 'computation cost'. Relatively high computation costs lead to short-term contracts. Low computation costs may lead to long-term contracts.

[5] To simplify the exposition we describe the basic notation omitting time subscripts. We will introduce time subscripts later in this section, when we describe the timing of the game.

state is a realization of the vector of dummy variables $(s(e_1); s(e_2); ....)$.

Similarly, a behavior is a complete description of all elementary actions, represented by a valuation function $b : \Omega^a \to \{0; 1\}$; where $b(a_k) = 1$ means that elementary action a_k is executed, and $b(a_k) = 0$ that a_k is not executed.

The efficient behavior depends on the state. We assume a simple one-to-one correspondence between elementary tasks and elementary events. The principal wants task k to be performed if and only if elementary event k occurs. In our airport example, the principal wants the agent to check the passenger's passport if and only if he has a moustache, and to search his bag if and only if the bag is red.

Principal and agent are risk neutral. The principal gets an incremental benefit of one from "matching" e_k with a_k, while he gets zero incremental benefit if there is a "mismatch". The principal's per-period utility is:

$$\pi(s; b; m) = \sum_{k=1}^{N} [s(e_k)b(a_k) + (1 - s(e_k))(1 - b(a_k))] - m \tag{2.1}$$

where m is the payment to the agent.[6]

The agent's interests are always in conflict with the principal's, in the sense that his preferred actions are always opposite the principal's preferred actions. This assumption is convenient but not essential to the qualitative results. The agent's one-period utility is:

$$U(s; b; m) = m - \sum_{k=1}^{N} \delta_k [s(e_k)b(a_k) + (1 - s(e_k))(1 - b(a_k))], \quad (0 < \delta_k < 1) \tag{2.2}$$

The parameter δ_k captures the disutility associated with task k for the agent. The agent's reservation utility is zero. Payoffs are common knowledge to the contracting parties, and the state and the parties' behavior are verifiable in court. Thus, there are no issues of moral hazard or adverse selection. We assume that preferences and realized payoff levels are not verifiable in

[6] The qualitative results would remain unchanged if we generalized the principal's utility function in the following way: $\pi = \sum_{k=1}^{N} [\alpha_1 s(e_k)b(a_k) + \alpha_2(1 - s(e_k))(1 - b(a_k)) - \alpha_3(1 - s(e_k))b(a_k) - \alpha_4 s(e_k)(1 - b(a_k))] - m$, where $\alpha_j > 0$. This function allows for different rewards according to the type of "match" (i.e., according to whether e_k is matched with a_k or $\neg e_k$ is matched with $\neg a_k$), and for damages in case of "mismatches".

court, and that the principal cannot "sell the activity" to the agent (i.e., the agent cannot be made the recipient of the gross payoff π).[7]

Next we define a contract and the costs of writing it. A contract is a pair $(g; m)$ where $g = (\gamma_k)_{k=1}^{N}$ is a set of N clauses and m is a transfer from the principal to the agent (wage).[8]

Each clause γ_k regulates a task. Given our simple matching structure between tasks and elementary events, we can restrict our attention to four types of clause: (i) a contingent clause, that constrains the agent to do a_k if and only if e_k occurs, $C_k : [a_k \leftrightarrow e_k]$; (ii) a noncontingent positive clause, constraining the agent to do a_k whatever happens, $R_k : [a_k]$; (iii) a noncontingent negative clause, constraining the agent to do not a_k whatever happens, $\bar{R}_k : [\neg a_k]$; (iv) the empty clause, D, that imposes no constraint on the agent (note that since we include the empty clause among the possible clauses, there is no loss of generality in assuming that the number of clauses in the contract is N). For example, if $N = 3$, the set of clauses $(R_1; D; C_3)$ constrains the agent to do a_1 whatever happens and to do a_3 if and only if elementary event e_3 occurs, leaving the agent free with regard to task 2. We denote G the collection of all possible sets of clauses (thus, for any contract $(g; m)$, $g \in G$).

A contract that specifies wage $m \geq 0$ constrains the principal to pay at least m to the agent; a contract that specifies $m \leq 0$ constrains the agent to pay at least $-m$ to the principal.

Describing a task or an elementary event is costly. To simplify, we assume that the cost of describing a task and the cost of describing an elementary event are both equal to c. It follows that writing a contingent clause C_k costs $2c$, and writing a noncontingent clause (R_k or $\bar{R}_k$) costs c. We also assume that specifying the wage in the contract is costless, thus the cost of writing a set of clauses $\gamma \in G$ is $Cost(\gamma) = 2cN_C^{\gamma} + cN_R^{\gamma}$, where N_C^{γ} is the number of contingent clauses and N_R^{γ} is the number of noncontingent clauses. The writing cost is borne entirely by the principal.[9]

[7] If preferences were verifiable, the first-best outcome could trivially be achieved by a contract of the form "The agent's behavior b must maximize the sum of the parties' utilities:" On the other hand, if realized payoff levels were verifiable, the first-best outcome could be achieved by offering the agent a transfer that increases one-for-one with the principal's realized payoff level. And selling the activity to the agent would be equivalent to specifying a contingent transfer as in the previous point.

[8] We can ignore the possibility of making payments contingent on the state or on the agent's actions, because the agent is risk neutral and his actions are verifiable.

[9] On the "strategic" effects of transaction costs paid by both parties, see Anderlini and Felli (1997).

Next we describe the timing of the game (and introduce time subscripts in the notation). The parties interact for infinitely many periods and have common discount factor $\delta \in (0,1)$. The parameter δ can also be interpreted as capturing the stability of the relationship.[10] In each period $t = 1, 2, \ldots$ the timing is the following: the state of nature $s_t \in S$ is observed, then the principal offers a contract (g_t, m_t) to the agent, where $g_t = (\gamma_{kt})_{k=1}^{N}$, $\gamma_{kt} \in \{C_k, R_k, \bar{R}_k, D\}$. (Note, offering a contract $((D)_{k=1}^{N}, 0)$ can be interpreted as offering no contract, as this contract imposes no constraints on players.) The principal pays the cost of drafting contract (g_t, m_t). If the contract is accepted, the principal makes the payment m_t and then the agent acts, both players being constrained by the contract.[11] If the contract is rejected, the agent gets his reservation utility (zero).

We have implicitly assumed that the contract proposed in period t can specify only a wage and a job description for time t. This assumption is without loss of generality. We could allow the contract at time t to specify wages and tasks for future dates, but there would be no gains from doing so.[12]

In the markovian equilibrium analyzed in this section, the wage m_t will be set at the minimum level that induces the agent to accept the proposed contract. Since the determination of the wage is a trivial aspect of the analysis, we will focus on the set of clauses. With a slight abuse of our terminology, from now on we will refer to a "contract" simply as its set of clauses.

If at time t the principal wants to offer a different contract than the one at time $t-1$, he can save substantial writing costs by proposing a modification of the previous contract, rather than drafting a whole new contract. Contract modifications can take two forms: (i) amendments, that is, permanent modifications of the contract; or (ii) exceptions, that is, temporary modifications applied only for the current period. We allow the principal to modify the existing contract with any set of amendments and exceptions at any point in time.[13]

[10] The parameter δ can be interpreted as the composition of two parameters, $\delta = \theta\delta'$, where θ is the probability that the game will continue and δ' is the 'true' discount factor.

[11] The sequence between payment and actions is not essential here, but will matter for self-enforcing contracts; in section 3 we will come back to the issue of timing. Also, nothing would change in the model if we allowed the principal to choose whether to write the contract before or after the state s_t is observed, as there is no gain from writing the contract before the state is observed.

[12] Recall that there is no gain from long-term commitment and no fixed contracting costs in this model.

[13] Implicit in this formulation is the assumption that the contract for time $t-1$ can be used as default contract

To capture this idea, we need to distinguish between the effective contract (the contract actually enforced at time t) and the default contract. The effective contract at t is given by the default contract at t plus the exceptions at t, and the default contract at t is given by the default contract at $t-1$ plus the amendments at t. The default contract will be the key state variable of our problem, while the amendments and exceptions will be the control variables.

More formally, the default contract is a set of clauses

$$\bar{g}_t : (\nu_{k;t})_{k\in N},$$

where $\nu_{k;t} \in \{C_k; R_k; \hat{R}_k; D\}$. The default contract at time t is given by:

$$\bar{g}_t = \tilde{f}(\bar{g}_{t-1}; g_t^A) = \left((\nu_{k;t-1})_{k\in N\setminus K_t^A}; (\alpha_{k;t})_{k\in K_t^A} \right);$$

where $\alpha_{k;t} \in \{C_k; R_k; \hat{R}_k\}$ is the amendment for task k and $g_t^A = (\alpha_{k;t})_{k\in K_t^A}$ is the set of amendments. For each task k; the default clause at $t = 0$ is the empty clause: $\nu_{k;0} = D$.

The effective contract at time t is given by:

$$g_t = f(\bar{g}_t; g_t^E) = \left((\nu_{k;t})_{k\in N\setminus K_t^E}; (\varepsilon_{k;t})_{k\in K_t^E} \right);$$

where $\varepsilon_{k;t} \in \{C_k; R_k; \hat{R}_k\}$ is the exception for task k and $g_t^E = (\varepsilon_{k;t})_{k\in K_t^E}$ is the set of exceptions.

The writing cost paid in period t is $Cost(g_t^A) + Cost(g_t^E)$.[14]

We assume that the stochastic process governing the external environment is a Markov chain;[15] the transition probabilities are denoted by $\mu(s_{t+1}|s_t)$. We will first focus on the special

for time t, but contracts from earlier dates cannot. For example, we do not allow contract g_t to say "contract g_{t-2} applies with the following modifications...". A more general model would allow for richer 'recalling' possibilities, but we conjecture that, if there is a costs of recalling more remote contracts, the key insights of the analysis would not change.

[14] We assumed that the language described at the outset is the only common-knowledge language. In principle, the parties could construct a new language, for example by attaching a symbol to each state and to each behavior, and write a contract with the new language. Note that the parties would have to attach a vocabulary that translates the new language into the original one, in order for the courts to be able to interpret the contract. If the relationship is one-shot, the new language cannot be more efficient than the original one, because the cost of writing the vocabulary in the contract is at least as great as its benefits. In a repeated relationship, however, this approach might in principle be efficient. (We thank Leonardo Felli and Luca Anderlini for bringing this point to our attention.) A more general model would allow for this kind of recoding of the language, but we think that, if there is a cost of writing the new symbols in the contract, the results of the analysis may not change radically.

[15] A Markov chain is a Markovian process with one-period memory and stationary transition probabilities (see, e.g., Gallager (1996, p. 103)).

case of an i.i.d. process and then show how the results change with persistent shocks. To streamline the exposition, we also assume that in the first period the state is $(1, \ldots, 1)$, i.e., $s_1(e_k) = 1$ for all k. In the appendix we solve the model dropping this assumption.

In this section we focus on stationary Markov perfect equilibria, that is, subgame perfect equilibria in which currrent decisions depend only on the state variable, i.e. the current state of the environment and the default contract of the previous period [see Fudenberg and Tirole (1991, Ch. 13)].

The game may have also subgame perfect equilibria that support some cooperation without the aid of formal contracts. These are equilibria where current decisions depend on past behavior ('punishment' strategies). We think of these equilibria as "self-enforcing contracts". The reason we ignore these equilibria in this section is to focus more sharply on the role of formal contracts, and on the issue of long-term versus short-term contracts. We will consider self-enforcing contracts in the next section.

Given the simple structure of the interaction, solving for the stationary Markov perfect equilibria boils down to maximizing the expected discounted value of the surplus net of writing costs. To state the problem formally, we need to introduce the policy function $h : G \times S \to G \times G$, or in more explicit notation, $(g_t^A, g_t^E) = h(g_{t-1}, s_t)$. The policy function induces, at each t, a random value for the surplus net of writing costs, which we denote $NS_t^h : S^t \to \mathbb{R}$. The problem can then be stated as

$$\max_h E\left[\sum_{t=1}^{\infty} \delta^{t-1} NS_t^h\right] \tag{2.3}$$

An optimal contracting plan is a solution of problem (2.3).

2.1. Independent shocks

As a first step of the analysis, we consider the case in which elementary events are identically and independently distributed across t and k. In the next section we will consider the case of serially correlated events.

Let us assume that, for every $k = 1, ..., N$, $t = 2, 3, ...$ and $s_t, s_{t-1} \in S$,

$$\mu(s_t|s_{t-1}) = p^{\sum_k s_t(e_k)}(1-p)^{[N-\sum_k s_t(e_k)]}, \; p \in (\frac{1}{2}; 1) \tag{2.4}$$

(recall that $s_1(e_k) = 1$ for all k). The probability that an elementary event occurs is given by p, that is, $p \equiv \Pr\{s_t(e_k) = 1\}$ for all k and t. We can think of p as capturing the degree of uncertainty in the environment: the higher p, the lower the uncertainty (notice that the variance of dummy variable is decreasing in p).

It turns out that in an optimal contracting plan each task k is regulated in one of four ways:

1. A C_k clause written at time $t = 1$, with no subsequent modifications. We will refer to this as a contingent rule, denoted by C_k.

2. A default clause R_k followed by an exception every time $s_t(e_k) = 0$ (e_k does not occur). We will refer to this as a default rule cum exceptions, denoted DE_k.

3. A default clause R_k with no subsequent modifications. We will refer to this as a rigid rule, denoted R_k.

4. No clause at any t (discretion for task k), denoted by D_k.

This is the right juncture to discuss the notion of contract incompleteness in this dynamic setting. Contract incompleteness can take two basic forms: (a) rigidity, meaning that the contractual obligations do not discriminate sufficiently between states, and (b) discretion, in the sense that the contractual obligations do not completely specify the agent's behavior. In this setting, a simple measure of contractual rigidity is the number of tasks regulated by rigid rules, and a measure of discretion is the number of tasks that are left unregulated.

Importantly, the notion of contract incompleteness must be understood in a dynamic perspective. For example, the presence of noncontingent clauses in a contract does not imply that there is contractual rigidity, because the noncontingent clauses may be modified over time. In particular, note that a default rule cum exceptions implements the first best outcome for task k at all times, just as a contingent rule.

It is straigthforward to derive the net incremental value of these four rules:

$$
\begin{array}{ll}
k^{th} \text{ Rule} & \text{Incremental Net Value} \\
C_k & \frac{1-\delta_k}{1-d} - 2c \\
DE_k & \frac{1-\delta_k}{1-d} - \frac{(1-dp)c}{1-d} \\
R_k & 1 - \delta_k + \frac{dp(1-\delta_k)}{1-d} - c \\
D_k & 0
\end{array}
\qquad (2.5)
$$

In the next proposition, N_C, N_{DE}, N_R and N_D denote respectively the numbers of tasks regulated by rules 1, 2, 3 and 4 above. We refer to a complete contingent contract as a contracting plan where each task is handled by a contingent rule, and to a complete default contract cum exceptions as a contracting plan where each task is handled by a default rule cum exceptions. (Also, when we use the expression "increasing" or "decreasing" without further specification we mean it in the weak sense.)

Proposition 1. (i) If c is smaller than a critical level c^*, the optimum is either a complete contingent contract or a complete default contract cum exceptions. The former is preferred if d is higher than a critical level $\hat{d}(p)$, where $\hat{d}(\cdot)$ is a strictly increasing function.

(ii) In general, a set of low-δ_k tasks is regulated entirely by contingent rules or entirely by default rules cum exceptions; a set of intermediate-δ_k tasks is regulated by rigid rules; and a set of high-δ_k tasks is left to the agent's discretion(each of these sets may be empty). If $d > \hat{d}(p)$, the contract is written once and for all at $t = 1$.

(iii) N_C is increasing in d; N_{DE} and N_D are decreasing in d. If d is higher than some critical level d^* (function of other parameters), a complete contingent contract is optimal.

(iv) N_C and N_D are decreasing in p; N_{DE} and N_R are increasing in p.

Absent writing costs, the model has little predictive power, because there is a vast multiplicity of optimal contracting plans. Any contracting plan that implements the first best is optimal. These include a complete long-term contingent contract, a sequence of complete spot contracts, and a whole host of intermediate solutions. However, an arbitrarily small writing cost is sufficient to pin down a unique optimum. Point (i) states that, if the writing cost is small, the optimum is either a complete contingent contract or a complete default contract cum exceptions (note that both implement the first best outcome). The former tends to be optimal when the relationship is more durable and when there is more uncertainty in the environment.[16] On the other hand, when uncertainty is low or the relationship is not very durable, the first best can be achieved at lower cost by writing default rules and occasionally negotiating exceptions when low-probability events occur.[17] It is worth emphasizing that a contingent contract may be optimal even though the parties can avoid the cost of describing contingencies by writing spot contracts. The reason a contingent contract may be optimal is that it saves on the costs of describing behavior.

If c is higher, the optimal contracting plan may be incomplete. Contractual incompleteness can take the form of rigidity or discretion. Point (ii) states that high-cost tasks are left to the agent's discretion and intermediate-cost tasks are regulated by rigid rules. Low-cost tasks are regulated by contingent rules or by default rules cum exceptions. Since both of these rules achieve the first best outcome, the more efficient is the one that minimizes (the present expected value of) writing costs, hence their comparison is independent of δ_k. This is why this group of tasks is regulated entirely by contingent rules or entirely by default rules cum exceptions. Each of the three groups of tasks may be empty, depending on parameters.

If the contract is written once and for all at $t = 1$ (i.e. it is not modified over time), we interpret it as a long-term contract. Literally interpreted, this is a one-period contract that is renewed every period; however, an equivalent strategy would be to write a contract with no expiration date at $t = 1$, therefore we feel justified in viewing this as a long-term contract. Our model thus offers a simple (and potentially testable) prediction: a long-term contract is optimal if the relationship is relatively durable (d high) and uncertainty is relatively high (p low). Under these conditions, the only rules that can be optimal are contingent rules and rigid rules, both of which are written once and for all in the initial contract.[18] Note that the writing cost c has no role in determining whether or not a long-term contract is optimal; it only determines the

[16] Point (i) implies that a contingent contract is optimal if p is higher than the critical level $\hat{d}^{-1}(d)$, where $\hat{d}^{-1}(\cdot)$ is the inverse of $\hat{d}(\cdot)$.

[17] A real-world example of a default-cum-exceptions approach is perhaps given by trade agreements. The basic GATT/WTO articles on tariffs and quotas take the form of non-contingent default rules (typically, quotas are prohibited and tariffs are subject to a fixed upper bound). However, exceptions to these rules are negotiated from time to time, when circumstances are such that keeping low trade barriers is too costly (economically or politically) for a government.

[18] Note that the optimal contract is written once and for all also in the extreme case of $p = 1$ (no uncertainty), but this case is not very interesting.

extent of incompleteness.

Point (iii) and (iv) look more closely at how the optimal contract is affected by the parameters d and p. If the relationship is more durable, the optimal contract tends to be more contingent and to leave less discretion to the agent. If the relationship is sufficiently stable, the model predicts a complete contingent contract. If there is more uncertainty, noncontingent rules (rigid or default) are less attractive. This in turn makes the other two options, contingent rules and discretion, more attractive.

2.2. Persistent shocks

Thus far we have assumed that elementary events are serially uncorrelated. In this subsection we examine how results change when exogenous shocks are persistent. We will find that, when shocks are persistent, it may be optimal to use amendments, rather than exceptions, as a way to adapt the default rules to changing events.

We assume that, for every $t = 2; 3; \ldots$ and $s_t; s_{t-1} \in S$, the transition probabilities are described by:

$$\mu(s_t|s_{t-1}) = \left(\frac{1}{2} + \rho\right)^{\sum_k I_{[s_t(e_k)=s_{t-1}(e_k)]}(s_t)} \left(\frac{1}{2} - \rho\right)^{[N - \sum_k I_{[s_t(e_k)=s_{t-1}(e_k)]}(s_t)]}, \ \rho \in (0; \frac{1}{2}). \tag{2.6}$$

where $I_{[s(e_k)=s'(e_k)]}(s)$ is an indicator function that takes value one if $s(e_k) = s'(e_k)$ and zero otherwise. (We do not need to assume anything about $\mu(s_1)$.) In words, for each t and k, the probability that $s_t(e_k)$ is equal to $s_{t-1}(e_k)$ is $(\frac{1}{2} + \rho)$, and the probability that $s_t(e_k)$ is different from $s_{t-1}(e_k)$ is $(\frac{1}{2} - \rho)$, and there is "cross-sectional independence". The parameter $\rho \in (0; \frac{1}{2})$ captures the persistence of shocks in the external environment.

Note that, unlike in the previous section, we are assuming that the elementary events e_k and $\neg e_k$ are symmetric. This allows us to focus more sharply on the role of persistence.

It turns out that the qualitative results are very similar to those of section 2.1, with the following modification. A default rule cum exceptions can no longer be optimal, and there is a new candidate optimal rule: a default rule cum amendments. This is a default clause that is amended every time the realization of $s(e_k)$ changes. The fact that amendments are more efficient than exceptions, as a way to adapt default rules to the changing environment, is an intuitive consequence of persistence.[19] Note that, like contingent rules and default rules cum exceptions, default rules cum amendments implement the first best at all times.

The following proposition highlights the changes in results relative to the previous section.

Proposition 2. If $\mu(s_t|s_{t-1})$ is described by (2.6), Proposition 1 still holds, provided "exceptions" is replaced with "amendments" and p is replaced with ρ.

This proposition suggests that serial correlation in the states has similar implications as 'intrinsic' asymmetries among states (which we considered in the previous section), with the difference that amendments are now preferred to exceptions. The general insight is that low uncertainty about the future state makes contract modifications (amendments or exceptions) preferable to contingent clauses. When low uncertainty is due to persistence, amendments are preferred. When it is due to intrinsic asymmetries between likely and unlikely states, exceptions are preferred.

Note that here a sufficient condition for a long-term contract to be optimal is $d > \bar{d}(\rho)$, where $\bar{d}(\cdot)$ is an increasing function. Again, we can interpret this result as saying that a long-term contract tends to be optimal when the relationship is relatively durable and uncertainty is relatively high.

A more general model would allow for intrinsically asymmetric states and persistent shocks. This would be substantially more complicated to analyze, but we conjecture that the main qualitative insights would not change, except that the optimal contracting plan would probably involve the use of both exceptions and amendments.

2.3. Language matters

Our model yields distinct predictions on the role of writing costs for the structure of contracts in a long-term relationship. In this section we argue that these predictions depend on our

[19] Exceptions and amendments are equivalent in the knife-edge case $\rho = 0$.

language-based approach, and would differ radically if writing costs were modeled in a different way. To make this point, we consider a simple alternative model of writing costs, more similar in spirit to Dye's (1985a) model (see also section 3 in MacLeod (2000)).

Consider the following modification of our model. Suppose there are M states, $s \in \{s_1; \ldots; s_M\}$ and M behaviors, $b \in \{b_1; \ldots; b_M\}$. The first best-correspondence is one-to-one, and in particular it specifies that the agent should do b_j if and only if the state is s_j; for all $j \in M$. The cost of describing a state and the cost of describing a behavior are both equal to c. Suppose c is small, so that it is optimal to implement the first best outcome. Keep all other assumptions of our model unchanged.

In this model, if the number of states M is sufficiently large, a sequence of spot contracts is optimal, and in particular it dominates a contingent contract. To see this, note that the cost of a complete contingent contract is $2cM$, while the discounted cost of a complete sequence of spot contracts does not exceed $\frac{c}{1-d}$. The intuition is that spot contracting (i) avoids the costs of describing the states of nature, and (ii) requires describing a (weakly) smaller number of behaviors than a contingent contract.

Thus, this alternative specification of writing costs implies that spot contracting is optimal, in stark contrast with our model. This should clarify our statement that the nature of language matters greatly for the predictions of the theory.

2.4. The Maskin-Tirole argument

Maskin and Tirole (1999) have argued that the presence of unforeseen contingencies (or, by a straightforward extension of their argument, the costs of describing contingencies) does not imply inefficiencies in contracting, provided that parties can design an appropriate message-based mechanism to be played after contingencies are observed and before actions are taken. Since in our setting parties are allowed to contract in 'spot' fashion, i.e. after contingencies are observed and before actions are taken, mechanisms à la Maskin-Tirole are redundant, hence none of our conclusions changes if such mechanisms are available. In particular, it remains true that the costs of describing contingencies cause inefficiencies in contracting, which is in contrast with Maskin and Tirole's argument. The reason for this apparent divergence in conclusions is that we allow for a repetitive relationship and for costs of describing behavior, which Maskin and Tirole do not.

Recall that in our model, under some parameter values, parties choose to write a contingent contract and incur the corresponding costs of describing contingencies, even if they have the option of writing spot contracts. Suppose the cost of describing an elementary event is distinct from the cost of describing an elementary task, and consider increasing the former, keeping the latter constant. Can this increase inefficiency? The answer is yes. As we have seen, there is a parameter region in which it is optimal to write contingent clauses. This is *a fortiori* true in the extended parameter space where the cost of describing contingencies is distinct from the cost of describing behavior. Therefore, starting from this parameter region, an increase in the costs of describing contingencies decreases the net surplus.

3. Informal versus formal contracting

In reality, long-term relationships are often managed by informal (self-enforcing) contracts. It also happens frequently that a relationship is governed by a combination of informal and formal contracts. In this section we examine how the predictions of the model change when parties have the option of using informal as well as formal contracts.

Informal contracts have an advantage over formal contracts, namely that they can be based on informal communication (i.e. communication for the only purpose of reciprocal understanding), as opposed to formal communication (i.e. communication for the purpose of making the contract enforceable in court). Arguably, the cost of the latter is higher than the cost of the former, because for the contract to be enforced by courts it must be written according to the commonly accepted legal standards, which may be quite cumbersome to meet. In particular, it is not sufficient that the language used in the formal contract be common knowledge to the contracting parties, it has to be common knowledge to the parties and the courts, and this may require effort and skills (or lawyers).[20]

[20] See footnote 1 for a discussion of our terminology of formal vs. informal communication.

The shortcoming of informal contracts, on the other hand, is the absence of an external enforcement mechanism. Since an informal contract must satisfy self-enforcement constraints, it may have to be distorted away from the perfect-enforcement optimum. In what follows we will examine more closely this tradeoff between formal and informal contracting.

We will assume for simplicity that the cost of informal contracting is zero. We could allow for a positive cost of informal contracting, but this would change the main results in an obvious direction, tilting the balance in favor of formal contracting. Also, what matters most for the tradeoff between formal and informal contracting is the differential cost of formal versus informal contracting, and this is captured in our model by the parameter c.

We will first analyze the optimal informal contract, supposing that formal contracts are not available, then we will consider the possibility of using both formal and informal contracts.

3.1. The optimal informal contract

Consider the game of section 2.1, and modify it only by assuming that formal contracts are not enforceable, or that the writing cost c is infinite. Of course, in this case the equilibrium of the one-shot game entails no exchange between principal and agent. In the infinite-horizon game, however, there are subgame perfect equilibria that support some exchange, including possibly the efficient outcome. These are equilibria where decisions are conditioned on past behavior, so that players can be 'punished' for deviations. In these equilibria, players follow a set of norms that regulate equilibrium behavior as well as a set of norms that regulate off-equilibrium (punishment) behavior. We interpret such a system of norms as a 'self-enforcing' contract.

Since a self-enforcing contract is in essence an equilibrium of the infinite-horizon game, it should be thought of as a long-term contract, thus it is natural to suppose that such a contract is selected once and for all at the outset and is not modified over time. We will focus on contrained Pareto-efficient subgame perfect equilibria, that is, subgame perfect equilibria that are not Pareto-dominated by some other subgame perfect equilibria. We will often refer to these simply as "efficient" informal contracts.

To analyze an informal contract, it is convenient to think of it as composed of three elements: (a) A set of clauses that specify the agent's equilibrium behavior, $g^I = (\gamma^I_k)_{k \in N}$; $\gamma^I_k \in \{C^I_k; R^I_k; D\}$, where the meaning of C^I_k and R^I_k is the same as the corresponding formal clauses, except that they are communicated informally, hence they are not enforceable by courts.[21] (b) An equilibrium wage profile $(m_t)_{t=1}^{\infty}$, where $m_t : S \to \mathbb{R}$. (Note that we allow the wage in period t to depend on the current state s_t, thus m_t is a random variable.)[22] (c) A "punishment clause" that specifies what happens after a deviation. In this setting there is no loss of generality in assuming a simple trigger punishment whereby, after any deviation, the relationship breaks down forever.[23] This way of formalizing a self-enforcing contract is slightly different than the conventional game-theoretic representation of a subgame perfect equilibrium, but it is convenient because it parallels the notation we used for formal contracts, and will be useful when we combine formal and informal contracts.

We can examine informal contracts in a similar way as we did for formal contracts, except that now we have to impose incentive constraints for the players.

Formally, let $M_t = \sum_{\tau=t}^{\infty} d^{\tau - t} E(m_\tau)$ denote the expected present value of wages from period t onward (since states are iid, there is no need to distinguish between conditional and unconditional expectation of the wage in period t). Also, let K^I_C (resp. K^I_R) be the set of tasks regulated by a contingent (resp. rigid) clause. We look for a solution to the following problem (recall that we are assuming $\mu_1(1; \ldots; 1) = 1$, while μ_t is given by (2.4) for $t \geq 1$):

$$\max_{K^I_C; K^I_R; (m_t)_{t=1}^{\infty}} \sum_{k \in K^I_C \cup K^I_R} (1 - \delta_k) + \frac{d}{1-d} \left[\sum_{k \in K^I_C} (1 - \delta_k) + \sum_{k \in K^I_R} p(1 - \delta_k) \right] \qquad \text{(P)}$$

subject to

[21] It can be easily shown that there is no loss of generality from restricting to time-independent clauses and excluding the informal rigid clause $\overline{R}^I_k$.

[22] There is no need to consider wage processes with longer memory. Also, it can be shown that there is no gain from making the wage contingent on the agent's past actions on the equilibrium path.

[23] This is the right juncture to discuss our timing assumption. Recall that, in each period, the principal makes the payment and then the agent acts. First note that this timing is preferable to one in which the payment is made simultaneously to the agent's actions, because the principal has a weaker incentive to cheat under the sequential timing. On the other hand, assuming the opposite sequence (the principal makes the payment after the agent acts) would fail to capture an important feature of many agency relationships: that punishments and/or rewards for the agent's actions are delayed; in practice this may happen because of lags in the observation of the agent's actions, or because payments are more infrequent than the agent's actions.

$$\sum_{k\in K_C^I\cup K_R^I}\delta_k \le dM_{t+1} - \frac{d}{1-d}\left(\sum_{k\in K_C^I}\delta_k + \sum_{k\in K_R^I}p\delta_k\right) \text{ for all } t = 1, 2, \ldots, \qquad (IC_A^1)$$

$$M_t - \sum_{k\in K_C^I\cup K_R^I}\delta_k - \frac{d}{1-d}\left(\sum_{k\in K_C^I}\delta_k + \sum_{k\in K_R^I}p\delta_k\right) \ge 0 \text{ for all } t = 1, 2, \ldots. \qquad (IC_A^2)$$

$$\sum_{k\in K_C^I}1 + \sum_{k\in K_R^I}s(e_k) + \frac{d}{1-d}\left(\sum_{k\in K_C^I}1 + \sum_{k\in K_R^I}p\right) - m_t(s) - dM_{t+1} \ge 0 \text{ for all } t \text{ and } s \text{ with } \mu_t(s) > 0. \qquad (IC_P)$$

Inequality (IC_A^1) ensures that the agent has no incentive to shirk. To understand this constraint, note that the agent has the highest incentive to shirk when $s(e_k) = 1$ for all the tasks k regulated by a rigid clause. Inequality (IC_P) says that in every period and state the principal must have an incentive to pay the equilibrium wage. If the principal is to cheat on the wage, he might as well pay a zero wage, but then the agent will quit immediately, therefore the incentive constraint reduces to the condition that the value of the relationship is nonnegative for each given state s. If all transfers go from the principal to the agent ($m_t \ge 0$ for all t), incentive constraints (IC_A^1) and (IC_P) are sufficient for an equilibrium. But if some transfer goes from the agent to the principal ($m_t < 0$ for some t), we must ensure that the agent has incentive to make this payment. Inequality (IC_A^2) incorporates such condition. Note that, if $m_t \ge 0$ for all t, inequality (IC_A^2) is implied by (IC_A^1).

Note an important difference with respect to the analysis of formal contracts. In spite of our assumptions of separable payoffs and independent elementary events, here we cannot optimize separately task by task, as we did in the case of formal contracts. The reason is that, even though the objective function in problem (P) is separable in the N tasks, the constraints are not task-by-task constraints, hence the optimal choice of clause for a given task depends on the way the other tasks are regulated.

It turns out that the rigid clause R_k^I is dominated by the contingent clause C_k^I for all k, because a rigid clause entails a lower net surplus for the principal and does not relax the relevant constraints relative to a contingent clause. Thus, the only candidate clauses for each task k are C_k^I and the empty clause, D.

The next proposition characterizes the efficient informal contracts:

Proposition 3. At an efficient informal contract:

(i) There exists a critical level δ^* such that tasks with $\delta_k < \delta^*$ are regulated by informal contingent rules, and tasks with $\delta_k \geq \delta^*$ are left to the agent's discretion.

(ii) Wages satisfy
$$\begin{cases} m_t = \bar{m} \in [\frac{1}{d}\sum_{k \in K_C^I} \delta_k; \sum_{k \in K_C^I} 1] \text{ for } t = 2; 3; \ldots \\ m_1 \in [\frac{1}{1-d}\sum_{k \in K_C^I} \delta_k - \frac{d}{1-d}\bar{m}; \frac{1}{1-d}\sum_{k \in K_C^I} 1 - \frac{d}{1-d}\bar{m}] \end{cases}$$

Point (i) states that the optimal informal contract may be incomplete, and that the incompleteness always takes the form of discretion. It is interesting to contrast this result with the formal-contracting benchmark analyzed in the previous section. Contract incompleteness may arise from writing costs or from the presence of self-enforcement constraints. In the first case, incompleteness can take the form of rigidity and/or discretion, but in the second case it can only take the form of discretion. The reason is that discretion may relax the self-enforcement constraints, while rigidity cannot.

Low-cost tasks are included in the optimal informal contract, while high-cost tasks are left to the agent's discretion. Recall that it is first-best to perform all tasks (since $\delta_k < 1$ for all k), thus we can interpret this discretion as genuine incompleteness of the informal contract. The intuition is as follows. From the incentive constraints we can derive an aggregate "self-enforcement" constraint, which requires that the expected gross profit be at least as large as the (minimum) efficiency wage necessary to elicit the agent's effort. Including an additional task in the contract increases the efficiency wage and it increases the available surplus from the relationship. If the disutility from the task (δ_k) is low, introducing the task in the contract increases the efficiency wage by less than it increases the surplus, hence it relaxes the self-enforcement constraint. On the other hand, introducing a high-δ_k task tightens the self-enforcement constraint. If this constraint-tightening effect is mild, the task will still be introduced in the contract, but if it is strong the task will be left to the agent's discretion.

Point (ii) characterizes the set of efficient wage profiles. An efficient wage profile is composed of a first-period wage m_1 and a stationary wage $\bar{m}$ from $t = 2$ on. To understand the bounds on wages described in (ii), it is useful to consider the two extreme cases where all the surplus goes to the agent and where it all goes to the principal. We can think of these two cases as the two extreme distributions of bargaining power. If the principal has all the bargaining power, $\bar{m}$ will be set at its lower bound, and m_1 will be set at its lower bound given $\bar{m}$; we thus get $\bar{m} = \frac{1}{\delta}\sum_{k \in K_C^I} \delta_k$ and $m_1 = 0$. This can be interpreted as saying that the agent is offered an 'apprenticeship' in the first period, after which he is 'hired' (if he has not cheated) and paid an efficiency wage. The intuition is that an efficiency wage is needed only from $t = 2$ on, because the first-period wage does not affect the agent's incentive to cheat. The first-period wage can then be lowered to the point that the agent's payoff equals his maxmin level; this value turns out to be zero.

If, on the other hand, the agent has all the bargaining power, $\bar{m}$ will be set at its upper bound, and m_1 will be set at its upper bound given $\bar{m}$. This yields $\bar{m} = m_1 = \sum_{k \in K_C^I} 1$, i.e. the agent gets all the revenue in all periods. For intermediate distributions of bargaining powers, the first-period wage is typically lower than the wage in subsequent periods, or in other words, the wage profile typically has the 'apprenticeship' feature.[24]

3.2. Formal and informal contracting

In this section we consider the case in which both formal and informal contracts are available. The game is the same as in section 2.1. The only difference in the analysis with respect to that section is that now we seek to characterize the Pareto-efficient subgame perfect equilibria, rather than the Markov perfect equilibrium. These equilibria may involve formal norms and/or informal norms. We will sometimes refer to these simply as the "efficient contracts".

Our objective is to understand under what conditions it is efficient to combine formal and informal contracting, and if so, which tasks tend to be regulated by formal versus informal norms, and how the underlying parameters affect the relative importance of formal and informal contracting.

The set of efficient contracts can be derived in the following way. We look for a wage profile

[24] The claim is true without the qualifier "typically" if the self-enforcement constraint is binding. In this case, $\bar{m}$ is equal to the gross profit and we must have $m_1 \leq \bar{m}$ if the principal is to obtain a non-negative payoff. But, due to the discrete choice nature of the problem, there may be some slack in the constraint, allowing m_1 to be slightly above $\bar{m}$. In a model with a continuum of tasks and smooth payoff functions there would be no slack.

$(m_t)_{t=1}^{\infty}$ and, for each task k, one of six possible rules (informal contingent, informal rigid, formal contingent, formal rigid, formal default-cum-exceptions, or discretion) maximizing the net present value of the surplus subject to the constraint that players have no incentive to cheat on the informal rules. There is no loss of generality in supposing that in each period the wage is entirely specified in the formal contract, and that the principal pays just the formal wage (i.e. he pays no "bonus"). Also, it can be shown that there is no gain from specifying wages contingent on the agent's actions on the equilibrium path.[25]

To simplify the analysis, we assume the following parameter restrictions:

$$p = \frac{1}{2} \text{ and } d \geq \frac{2}{3} \tag{3.1}$$

This restriction ensures that there exists a credible punishment strategy that keeps the principal at his maxmin, as well as one that keeps the agent at his maxmin, as the following lemma states. This will considerably simplify the characterization of the efficient contracts. At the end of this section we will discuss how results are likely to change when condition 3.1 is not satisfied.

Lemma 1. Under assumption (3.1), the minimum subgame perfect equilibrium payoff for each player in each subgame is his maxmin (zero).

The punishment strategies that we construct to prove the lemma have roughly the following structure: after a deviation, the informal contract is abandoned and parties revert to the optimal formal contract, and all the surplus from this contract is given to the player that has not deviated. In Appendix we describe this punishment strategy in greater detail.

As in the previous section, informal rigid rules can be shown to be dominated. Note also that, under assumption 3.1, default-cum-exceptions rules are (weakly) dominated by contingent formal rules. Therefore, we have only four candidate rules for each k: (1) informal contingent; (2) formal rigid; (3) formal contingent, and (4) no rule (discretion). Therefore, taking Lemma 1 into account, the problem can be stated as

$$\max_{K_C; K_C^I; K_R; (m_t)_{t=1}^{\infty}} \sum_{k \in K_C^I} (1 - \delta_k) + \sum_{k \in K_C} [1 - \delta_k - 2c(1 - d)] + \sum_{k \in K_R} [(1 - d + dp)(1 - \delta_k) - c(1 - d)] \quad \text{(P')}$$

[25] There is no gain from making wages contingent on actions because this requires describing the relevant actions in a formal contract, and this is as costly as forcing those actions directly with a formal contract.

subject to

$$\sum_{k\in K_C^I} \delta_k \le dM_{t+1} - \frac{d}{1-d}\left(\sum_{k\in K_C^I\cup K_C} \delta_k + \sum_{k\in K_R} p\delta_k\right) \text{ for all } t = 1, 2, \ldots, \qquad (IC_A^{1'})$$

$$M_t - \sum_{k\in K_C^I\cup K_R\cup K_C} \delta_k - \frac{d}{1-d}\left(\sum_{k\in K_C^I\cup K_C} \delta_k + \sum_{k\in K_R} p\delta_k\right) \ge 0 \text{ for all } t = 1, 2, \ldots . \qquad (IC_A^{2'})$$

$$\sum_{k\in K_C^I\cup K_C} 1 + \sum_{k\in K_R} s(e_k) + \frac{d}{1-d}\left(\sum_{k\in K_C^I\cup K_C} 1 + \sum_{k\in K_R} p\right) - m_t(s) - dM_{t+1} \ge 0 \text{ for all } s \in S,\ t = 2, 3 \ldots . \qquad (IC'_P)$$

$$\sum_{k\in K_C^I} 1 + \sum_{k\in K_C} (1-2c) + \sum_{k\in K_R} (1-c) + \frac{d}{1-d}\left(\sum_{k\in K_C^I\cup K_C} 1 + \sum_{k\in K_R} p\right) - m_1 - dM_2 \ge 0 \qquad (PC_P)$$

where K_C, K_R and K_C^I are the sets of tasks regulated respectively by formal contingent, formal rigid and informal contingent rules, and (PC_P) is the "participation constraint" for the principal (the corresponding participation constraint for the informal contract problem was implicit in (IC_P), here we must take the writing costs into account).

This problem may admit multiple optimal choices of $(K_C; K_R; K_C^I)$, for the following reason. For example, suppose that, at a solution of the problem, task 1 is regulated with an informal rule and task 2 with a formal contingent rule. Now modify the contract by regulating task 1 with a formal contingent rule and task 2 with an informal rule. Since informal and formal contingent rules yield the same profit gross of (wage and) writing costs, this does not change the payoff and it may well be the case that the incentive constraints are still satisfied, in which case the modified contract is also optimal. Such indeterminacy prevents clean comparative statics. Therefore we use the following tie-breaking rule: if two contracts yield the same present value of net surplus, we assign a preference to the one that implies more slack in the aggregate self-enforcement constraint (this constraint was mentioned in the previous section and is defined more precisely in appendix).We call efficient a contract that solves problem (P') and satisfies our selection criterion.

Let F denote the ratio of formally-regulated tasks to informally-regulated tasks. This is a measure of the importance of formal contracting relative to informal contracting. The next proposition characterizes the efficient contracts and examines the impact of the key exogenous parameters.

Proposition 4. *It may be efficient to regulate some tasks by formal contract and others by informal contract. In particular, under assumption (3.1):*
(i) There exist three critical levels, $\delta' < \delta'' < \delta'''$, such that tasks with $\delta_k < \delta'$ are regulated by contigent informal rules, tasks with $\delta' < \delta_k < \delta''$ are regulated by contingent formal rules, tasks with $\delta'' < \delta_k < \delta'''$ are regulated by rigid formal rules, and tasks with $\delta_k > \delta'''$ are left to the agent's discretion.
(ii) As c decreases, F can either increase or decrease. In particular it is possible that, as c decreases, the number of formally-regulated tasks stays constant while the number of informally-regulated tasks increases. The same statements are true for an increase in d.

Note that, in principle, there can be three types of contractual incompleteness in the overall contract: discretion, rigidity of the formal contract, and rigidity of the informal contract. We have shown however that the third type of incompleteness cannot arise: rigidity can be a feature only of the formal contract. This is because rigidity in the informal contract does not help relax the incentive constraints, nor does it save on writing costs.

In the optimal contract, low-cost tasks are regulated informally, intermediate-cost tasks are regulated formally, and high-cost tasks are left to the agent's discretion. Within the group of tasks regulated formally, lower-δ_k tasks are handled by contingent rules and higher-δ_k tasks by rigid rules. Here we give a sketch of the argument. Let us ignore formal rigid rules for simplicity. As in the previous section, the incentive constraints can be reduced to an aggregate constraint, which requires that the expected gross profit be at least as large as the (minimum) wage necessary to convince the agent to carry out all the tasks specified in the contract. This wage is given by the total disutility incurred by the agent plus an efficiency-wage premium for each task regulated informally. The argument proceeds in two steps. First, consider two tasks characterized by different levels of δ_k, and suppose you have to regulate one task formally and the other informally; which one will you regulate formally? The increase in surplus is the same independently of which task is regulated formally, but the efficiency wage premium is higher for the higher-$\pm_k$ task, hence this task will be regulated formally. Second, consider a pair of tasks and suppose you have to regulate one task formally and leave the other to the agent's discretion; which one will you regulate formally? Intuitively, the lower-$\pm_k$ task will be regulated formally, because it entails a higher surplus. It is thus intuitive that formal contracting is preferred only for intermediate-$\pm_k$ tasks.

Not only are formal and informal contracting used jointly, but they are complementary, in the sense that they increase each other's value. This is because increasing the number of tasks regulated formally increases the surplus from the relationship, hence it helps relax the incentive constraints of both players, thus making it easier to introduce more tasks in the informal contract.[26]

Point (ii) highlights an interesting possibility: a decrease in the cost of formal contracting may result in a higher number of informal clauses with no change in the number of formal clauses. To explain this possibility, consider a small decrease in c. This may have the e¤ect of changing a rigid formal clause into a contingent formal clause. The resulting increase in the available surplus relaxes the incentive constraints, making it possible to introduce an additional task in the informal contract, in which case the total number of formal clauses does not change and the number of informal clauses increases.

Also noteworthy is the result that an increase in the discount factor d may decrease the relative importance of informal contracting. An increase in d has two e¤ects. First, since players discount the future less heavily, the incentive constraints are relaxed, hence it is easier to introduce additional tasks in the informal contract. Second, since writing costs (for contingent and rigid clauses) are paid only in the …rst period, the importance of these costs is reduced if d is higher, and this pushes in favor of introducing additional tasks in the formal contract. The net e¤ect can go either way.

Before concluding the section, we discuss the role of condition 3.1, which ensures that the minimum equilibrium payo¤ is zero for each player. We could not prove the lemma for a wider range of parameters, but we conjecture that it is valid for more general values of p. The

[26] This complementarity is of a similar nature as the one that arises in Baker et al. (1994). But see the next subsection for important di¤erences with respect to that paper.

assumption that d is relatively high, on the other hand, is clearly essential for the lemma. Consider the extreme case of d equal to zero. Then the only subgame perfect equilibrium is the one-shot Nash equilibrium, in which the principal o¤ers a formal contract and makes a positive pro…t. At any rate, even though this condition is needed for the lemma, we suspect it is not essential for our qualitative insights. If the condition is not satis…ed, it may not be possible to keep the principal at his maxmin payo¤ in the punishment phase, in which case the principal's incentive constraints will be more stringent, and this is likely to result in fewer tasks being included in the informal contract. However proposition 4(i) is still likely to hold, with the only amendment that default-cum-exceptions rules may be preferred to contingent rules. And proposition 4(ii) would obviously still hold, because it highlights a possibility rather than a general comparative-statics result

3.2.1. Relationship to Baker et al. (1994) and Pearce and Stacchetti (1998)

Here we discuss brie‡y the analogies and di¤erences between the model analyzed in the previous section and the two above-mentioned papers. In those papers, as in ours, a combination of formal and informal contracting may be optimal, and the two forms of contracting may be complementary, since the presence of a formal contract may relax the incentive constraints associated with the informal contract. However, this is where the analogy stops.

Baker et al. (1994) and Pearce and Stacchetti (1998) provide an explanation for the combined use of formal and informal payments (wages and bonuses), whereas our model explains why it may be e¢cient to regulate some tasks formally and some others informally (note that in our model there is no need for bonuses). Perhaps more importantly, the rationale for mixing formal and informal contracting is very di¤erent. In those models, formal and informal contracting are used together because some signals are veri…able and some are not. In our model, the combination of formal and informal contracting is not due to di¤erences in veri…ability or transaction costs across tasks; rather, it is due to the interaction between writing costs (which are symmetric across tasks) and self-enforcement constraints.

Moreover, a key element of our analysis is the distinction between rigidity and discretion, as the two forms of incompleteness that can arise in formal and informal contracting. Given the nature of the above mentioned models, they have little to say on this aspect of the problem.[27]

Our model also yields di¤erent predictions concerning the interplay between formal and informal contracting. One key result in Baker et al. (1994) is that the availbility of formal contracts may undermine informal contracts. In particular, if the veri…able signal is su¢ ciently precise – or in other words, if the imperfections in formal contracting are su¢ ciently small – an informal contract cannot be sustained. Therefore, a broad prediction of their model is that, if imperfections in formal contracting decrease over time, there comes a point at which informal contracting disappears. In our model, if formal contracting is close to perfect (i.e. if c is close to zero), the optimum typically involves both formal and informal contracting, possibly even a fully informal contract.[28] Thus, our analysis suggests that informal contracting need not disappear as the formal-contracting system becomes more e¢ cient.

The reason for this divergence in results lies in the punishment strategy. Baker et al. assume that, if a player cheats, parties revert to the optimal formal contract, with all the surplus from this contract accruing to the principal. This implies that, if formal contracting is close to perfect, it completely fails to deter the principal from cheating. However, we notice that in general this is not the most severe punishment, hence the equilibrium characterized by Baker et al. is in general not constrained Pareto e¢ cient. Our approach, on the other hand, is to characterize the constrained Pareto e¢ cient equilibria, which we are able to do under parameter restriction 3.1. In this parameter range, each player can be punished with his maxmin payo¤, hence changes in c do not a¤ect the severity of the punishment.

Recall also that in our model, as c decreases, the relative importance of formal contracting may decrease; this cannot happen in Baker et al. This e¤ect can arise in our model because of an interplay between the two forms of incompleteness (rigidity and discretion), which is absent in Baker et al.

Finally, our model yields di¤erent predictions on the e¤ect of changes in the discount factor d. In Baker et al., an increase in d always favors informal contracting. In our model, as we

[27] The discussion in the remainder of this section applies only to Baker et al. (1994).

[28] This is certainly true if d is su¢ ciently close to one: in this case, a complete informal contract satis…es the incentive constraints, and hence it is optimal regardless of c.

remarked earlier, the opposite may happen. This is due to the di¤erent nature of the contractual imperfection. In Baker et al., the contractual imperfection (i.e. the non-veri…ability of signals) is relevant in each period. In our model, on the other hand, writing costs are paid only in the …rst period (at least for contingent and rigid clauses), hence the dynamic implications of the two kinds of contractual imperfections are very di¤erent.

4. Conclusion

Here we mention two potentially interesting extensions of the model. We have assumed that coordinating on informal contracts is costless. It would probably be more realistic to assume that coordinating on an informal contract requires costly communication. Presumably, the costs of informal communication are similar to those of formal communication, though lower. Results would probably change in two ways if we introduced costs of informal communication. First, this would obviously tilt the balance in favor of formal contracting. Second, rigid informal rules might become optimal, as they save on communication costs relative to contingent informal rules. To the extent that informal communication is less costly than formal communication, however, it would still be true that informal contracts tend to be less rigid than formal contracts.

A second, potential extension of the model would allow for non-veri…ability of the agent's actions. This is a more 'traditional' source of contract incompleteness (see for example Baker et al., 1994, and Pearce and Stacchetti, 1998). We assumed full veri…ability of the agent's actions in order to isolate the implications of writing costs for the structure of dynamic contracts. As a topic for future research, however, it would be desirable to examine the interaction between these two sources of contract incompleteness, which are probably both empirically relevant.

5. Appendix

Proof of Proposition 1

Here we drop the assumption that $s(e_k) = 1$ for all k with certainty. Proposition 1 as stated in the text is still valid, provided we redefine two expressions:

By default rule cum exceptions, here we mean the following: The first period in which e_k occurs, say t', clause R_k is introduced in the default contract, and subsequently the exception $\tilde{R}_k$ is applied every : e_k occurs. If $s_1(e_k) = 0$, clause $\tilde{R}_k$ is introduced in the default contract, and replaced by R_k at t'.

By rigid rule, here we mean any plan that converges to a steady state where there is a default clause R_k or $\tilde{R}_k$ with no modifications. We will see that there are several plans that fall in this category and can be optimal under some parameters.

We can assume without loss of generality that the principal chooses at the beginning of each period t how he will react to the exogenous shock s_t (his optimally chosen reaction function yields ex post optimal decisions).

We can therefore analyze the resulting dynamic programming problem with restricted state space G, where the principal decides at the beginning of each period t which modifications of the default contract $\hat{\mathbf{e}}_{t-1}$ he will offer as a function of the external state s_t (yet to be observed). By the additive separability of payoffs we will obtain a value function $v : G \to \mathbb{R}$ where

$$v(\hat{\mathbf{e}}_{t-1}) = \sum_{k=1}^{N} v_k(\hat{\mathbf{e}}_{k;t-1}).$$

We now derive each component v_k ($k \in N$) of the value function. For each possible $\hat{\mathbf{e}}_{k;t-1}$; we can restrict our attention to the following candidate one-period decision rules:[29]

(i) Cont(k): include in the default contract a contingent clause independently of the realization s_t.

[29] Note that we can ignore the possibility of simply removing a clause, i.e. replacing it with $\emptyset$. This is due to our assumptions about the payoff structure: in this model having a non-empty clause cannot be worse than having an empty clause.

(ii) Amend(k): include in the default contract clause R_k if e_k occurs and $\overline{R}_k$ otherwise.

(iii) Amend($k,+$): include in the default contract clause R_k if e_k occurs and do nothing otherwise.

(iii') Amend($k,-$): include in the default contract clause $\overline{R}_k$ if $\neg e_k$ occurs and do nothing otherwise.

(iv) Except($k,-$): apply the exception $\overline{R}_k$ if $\neg e_k$ occurs and do nothing otherwise.

(iv') Except($k,+$): apply the exception R_k if e_k occurs and do nothing otherwise.

(v) Inaction(k): do not introduce any modification concerning aspect k, independently of the realization s_t.

All other decision rules can be shown to be suboptimal. Furthermore, it can be shown that $v_k(R_k) \geq v_k(\overline{R}_k)$. Intuitively, it is (weakly) better to have a rigid default clause prescribing the right action with probability $p > \frac{1}{2}$ rather than a rigid default clause prescribing the right action with probability $(1-p) < \frac{1}{2}$. This implies that Amend($k,-$) and Except($k,+$) cannot be (strictly) optimal and we can safely ignore them. Note also that, since Inaction(k) does not change the state variable, Inaction(k) is optimal in a given state if and only if it is optimal forever after. The same is true for Except($k,-$).

Now we consider all the possible values of coordinate k of the state variable g_{t-1}:

- $g_{k,t-1} = C_k$. Obviously in this case Inaction(k) is optimal in the current period and in any future period:

$$v_k(C_k) = 1 - \delta_k + dv_k(C_k).$$

Therefore

$$v_k(C_k) = \frac{1-\delta_k}{1-d}. \tag{5.1}$$

- $g_{k,t-1} = R_k$. In this case we can restrict our attention to three candidate decision rules:

(i) Cont(k), which yields $1 - \delta_k - 2c + dv_k(C_k) = \frac{1-\delta_k}{1-d} - 2c$ (by (5.1)),

(ii) Except($k,-$), which yields $1 - \delta_k - (1-p)c + dv_k(R_k)$;

(iii) Inaction(k), which yields $p(1-\delta_k)+dv_k(R_k)$:

It follows that

$$v_k(R_k) = \max\left\{\frac{1-\delta_k}{1-d}-2c;\ \frac{1-\delta_k-(1-p)c}{1-d};\ \frac{p(1-\delta_k)}{1-d}\right\}. \tag{5.2}$$

- $\hat{e}_{k;t-1}=\overline{R}_k$: In this case the candidate decision rules are:

  (i) Cont(k), which yields $\frac{1-\delta_k}{1-d}-2c$;

  (ii) Amend(k; +), which yields $1-\delta_k-pc+d\left[pv_k(R_k)+(1-p)v_k\left(\overline{R}_k\right)\right]$

  (iii) Inaction(k), which yields $(1-p)(1-\delta_k)+dv_k\left(\overline{R}_k\right)$:

It follows that

$$v_k\left(\overline{R}_k\right) = \max\left\{\frac{1-\delta_k}{1-d}-2c;\ \frac{1-\delta_k-pc+dpv_k(R_k)}{1-d(1-p)};\ \frac{(1-p)(1-\delta_k)}{1-d}\right\} \tag{5.3}$$

where $v_k(R_k)$ is given by (5.2).

- $\hat{e}_{k;t-1}=D$: In this case the candidate decision rules are:

  (i) Cont(k), which yields $\frac{1-\delta_k}{1-d}-2c$;

  (ii) Amend(k), which yields $1-\delta_k-c+d[pv_k(R_k)+(1-p)v_k\left(\overline{R}_k\right)]$;

  (iii) Amend(k; +), which yields $p(1-\delta_k-c)+d[pv_k(R_k)+(1-p)v_k(D)]$;

  (iv) Inaction(k), which yields $dv_k(D)$:

It follows that

$$v_k(D) = \max\left\{\frac{1-\delta_k}{1-d}-2c;\ 1-\delta_k-c+d[pv_k(R_k)+(1-p)v_k\left(\overline{R}_k\right)];\ \frac{p(1-\delta_k-c)+dpv_k(R_k)}{1-d(1-p)};\ 0\right\} \tag{5.4}$$

where $v_k(R_k)$ and $v_k\left(\overline{R}_k\right)$ are given respectively by (5.2) and (5.3). One can verify that the following intuitive inequalities hold:

$$v_k(D)\ \le\ v_k(\overline{R}_k)\ \le\ v_k(R_k)\ \le\ v_k(C_k): \tag{5.5}$$

Equations (5.1) to (5.4) fully characterize the component v_k ($k \in N$) of the value function for our problem. The optimal decision rule for each $\hat{e}_{k,t-1}$ can be derived using the same equations. For example, if $\hat{e}_{k,t-1} = R_k$, the optimal decision rule is Cont(k) or Except(k; −) or Inaction(k) depending on whether the maximum element of the set in the right hand side of (5.2) is the first, the second or the third one.

Using equations (5.1) to (5.4) and inequalities (5.5), one can derive that the only candidate optimal plans concerning aspect k are the following (a plan describes only the decisions at reachable states):

- C_k: Cont(k) at D and Inaction(k) at C_k. The value of this plan is: $\frac{1-\delta_k}{1-d} - 2c$;
- DE_k: Amend(k) at D, Amend(k; +) at $\overline{R}_k$, Exception(k; −) at R_k. The associated value is $\frac{1-\delta_k}{1-d} - \left(1 + \frac{2-d}{1-d} \cdot \frac{dp(1-p)}{1-d(1-p)}\right) c$;
- R_k^+: Amend(k) at D, Amend(k; +) at $\overline{R}_k$, Inaction(k) at R_k. The associated value is $\frac{\left[1-d(1-p^2)\right](1-\delta_k)}{(1-d)[1-d(1-p)]} - \left(1 + \frac{dp(1-p)}{1-d(1-p)}\right) c$;
- R_k^0: Amend(k) at D, Inaction(k) at R_k and $\overline{R}_k$. The associated value is $\frac{[1-2dp(1-p)](1-\delta_k)}{1-d} - c$;
- R_k^*: Amend(k; +) at D, Inaction(k) at R_k. The associated value is $\frac{p(1-\delta_k)}{1-d} - \frac{pc}{1-d(1-p)}$;
- D_k: Inaction(k) at D. The associated value is 0.

We can now prove parts (i), (ii) and (iii) of the proposition.

(i) If $1 - \delta_k \geq c$, then Inaction(k) is never optimal and Amend(k; +) is not optimal when there is no clause (i.e., at D). Therefore, if $c \leq 1 - \delta_N$, C_k and DE_k are better than any other plan, for all k. By comparing the respective associated values, one finds that a contingent contract is optimal if and only if

$$(1-d)[1-d(1-p)] < d(2-d)p(1-p)$$

It is direct to verify that this condition is satisfied if and only if $d > \overline{d}(p)$, where $\overline{d}(p)$ is an increasing function satisfying $\overline{d}(\frac{1}{2}) = \frac{2}{3}$ and $\overline{d}(1) = 1$. The claim for $c < c^*$ follows immediately.

(ii) As is apparent from the values listed above, C_k is preferred to DE_k if and only if d is higher than a threshold which depends only on p. To see how the ranking of different plans for aspect k depends on δ_k look at the coefficient of δ_k in their values. Since $\frac{1}{2} < p < 1$,

$$\frac{1}{1-d} > \frac{1-d(1-p^2)}{(1-d)\left[1-d(1-p)\right]} > \frac{1-2dp(1-p)}{1-d}, \frac{p}{1-d} > 0.$$

Taking the upper envelope of the plans values as (positive affine) functions of δ_k we obtain a decreasing convex function and two thresholds $\delta_* \leq \delta^*$ (functions of d, p and c); the envelope has slope $-\frac{1}{1-d}$ to the left of δ_* and is flat to the right of δ^*. This means that the D_k plan is optimal for $\delta_k > \delta^*$, one of the RR_k plans is optimal if $\delta_* < \delta_k < \delta^*$ and the C_k or DE_k plan is optimal for $\delta_k < \delta_*$.

For (iii)-(iv) compare the derivatives of the values of the different plans with respect to d; and p. ■

Proof of Proposition 2

Unlike in the proof of Proposition 1, due to the autoregressive nature of the shocks we cannot work with the restricted state space G, but we have to work with the unrestricted state space $G \times S$. The value function is still additively separable in the N dimensions:

$$v(\mathbf{e}_{t-1}; s_t) = \sum_{k=1}^{N} v_k(\mathbf{e}_{k,t-1}; s_t(e_k)).$$

We now derive each component v_k ($k \in N$) of the value function using a shortcut. We exploit the symmetries of the model to partition the state space for aspect k in four cells corresponding to the following situations:

D_k (Discretion): there is no clause regulating aspect k in the default contract.

M_k (Match): the default contract contains the rigid k-clause matching the current state of the environment.

NM_k (No Match): the default contract contains the rigid k-clause that does not match the current state of the environment.

C_k (Contingent rule): the default contract contains the efficient contingent k-clause.

For each possible situation, one can show that there are only three candidate one-period decision rules:

(i) Cont[k]: include in the default contract the contingent clause C_k.

(ii) Amend[k]: include in the default the rigid clause matching the current state of the environment (R_k if $s_k = 1$ and $\overline{R}_k$ if $s_k = 0$). Using this rule and independently of the current situation, the system makes a transition to situation M_k with probability $(\frac{1}{2} + \rho)$ and to situation NM_k with probability $(\frac{1}{2} - \rho)$.

(iii) Inaction[k]: do not introduce any modification concerning aspect k. Under this rule, if the situation is M_k (or NM_k) the system stays there with probability $(\frac{1}{2} + \rho)$ and makes a transition to NM_k (respectively M_k) with probability $(\frac{1}{2} - \rho)$.

Let us consider the value of each possible situation: with a slight abuse of notation we write $v_k(D_k)$, $v_k(M_k)$, $v_k(NM_k)$ and $v_k(C_k)$.

- C_k (Contingent rule): Inaction[k] is optimal in the current period and in any future period, thus we have

$$v_k(C_k) = \frac{1 - \delta_k}{1 - d}. \qquad (5.6)$$

- M_k (Match): In this case there are two candidate decision rules:

  (i) Cont[k], which yields $\frac{1-\delta_k}{1-d} - 2c$,

  (ii) Inaction[k], which yields $1 - \delta_k + d[(\frac{1}{2} + \rho)v_k(M_k) + (\frac{1}{2} - \rho)v_k(NM_k)]$;

  It follows that

$$v_k(M_k) = \max\left\{\frac{1 - \delta_k}{1 - d} - 2c; \frac{1 - \delta_k + (\frac{1}{2} - \rho)dv_k(NM_k)}{1 - d(\frac{1}{2} + \rho)}\right\} \qquad (5.7)$$

- NM_k (No Match): In this case there are three candidate decision rules:

  (i) Cont[k], which yields $\frac{1-\delta_k}{1-d} - 2c$,

  (ii) Inaction[k], which yields $d\left[(\frac{1}{2} - \rho)v_k(M_k) + (\frac{1}{2} + \rho)v_k(NM_k)\right]$;

  (iii) Amend[k], which yields $1 - \delta_k - c + d\left[(\frac{1}{2} - \rho)v_k(NM_k) + (\frac{1}{2} + \rho)v_k(M_k)\right]$

It follows that

$$v_k(NM_k) = \max\left\{\frac{1-\delta_k}{1-d} - 2c, \frac{d(\frac{1}{2}-\rho)v_k(NM_k)}{1-d(\frac{1}{2}+\rho)}, 1-\delta_k - c + d\frac{1-\delta_k - c + d(\frac{1}{2}+\rho)v_k(M_k)}{1-d(\frac{1}{2}-\rho)}\right\}. \quad (5.8)$$

- D_k (Discretion): In this case the candidate decision rules are:

  (i) Cont[k], which yields $\frac{1-\delta_k}{1-d} - 2c$;

  (ii) Amend[k], which yields $1-\delta_k - c + d[(\frac{1}{2}+\rho)v_k(M_k) + (\frac{1}{2}-\rho)v_k(NM_k)]$;

  (iii) Inaction[k], which yields $dv_k(D_k)$.

It follows that

$$v_k(D_k) = \max\left\{\frac{1-\delta_k}{1-d} - 2c, 1-\delta_k - c + d\left[(\frac{1}{2}+\rho)v_k(M_k) + (\frac{1}{2}-\rho)v_k(NM_k)\right], 0\right\} \quad (5.9)$$

Equations (5.6) to (5.9) fully characterize the component v_k ($k \in N$) of the value function for our problem. One can derive that the only candidate optimal plans concerning aspect k are the following

- C_k: Cont[k] in situation D_k and Inaction[k] in C_k. The value of this plan is: $\frac{1-\delta_k}{1-d} - 2c$.

- A_k: Amend[k] in situations D_k and NM_k, Inaction[k] in M_k. The associated value is $\frac{1-\delta_k}{1-d} - c\left[1 + \frac{d}{1-d}(\frac{1}{2}-\rho)\right]$.

- R_k: Amend[k] in situation D, Inaction[k] in M_k and NM_k. The associated value is $1-\delta_k - c + d(1-\delta_k)\left[\frac{1}{2(1-d)} + \frac{\rho}{1-2d\rho}\right]$.

- D_k: Inaction[k] at all states. The associated value is 0.

Parts (i)-(iv) of the proposition can then be shown with a similar logic as the one used in the proof of proposition 1. ■

Proof of Proposition 3

From inspection of the constraints, it is clear that we can focus on wage profiles where M_t is stationary from $t = 2$ on. In other words, a wage profile can be summarized by the pair $(M_1; M_2)$. Noting that constraint (IC_A^2) can be binding only for $t = 1$, we can write the set of constraints as

$$M_2 \geq \frac{1}{d} \sum_{k \in K_C^I \cup K_R^I} \delta_k + \frac{1}{1-d} \left(\sum_{k \in K_C^I} \delta_k + \sum_{k \in K_R^I} p\delta_k \right), \tag{5.10}$$

$$M_2 \leq \frac{1}{1-d} \left(\sum_{k \in K_C^I} 1 + \sum_{k \in K_R^I} p \right); \tag{5.11}$$

$$M_1 \geq \sum_{k \in K_C^I \cup K_R^I} \delta_k + \frac{d}{1-d} \left(\sum_{k \in K_C^I} \delta_k + \sum_{k \in K_R^I} p\delta_k \right), \tag{5.12}$$

$$M_1 \leq \sum_{k \in K_C^I \cup K_R^I} 1 + \frac{d}{1-d} \left(\sum_{k \in K_C^I} 1 + \sum_{k \in K_R^I} p \right). \tag{5.13}$$

Inequality (5.10) is derived from (IC_A^1) evaluated at $t = 1$. Inequality (5.11) is derived from (IC_P) evaluated at $t = 2$ by averaging over states and noting that, to make this constraint as loose as possible, we choose the state contingent wages so as to equalize the principal's payoff across states, while keeping the expected wage constant. Inequality (5.12) is derived from (IC_A^2) evaluated at $t = 1$. Inequality (5.13) is derived from (IC_P) evaluated at $t = 1$, making use of the assumption $s_1(e_k) = 1$ for all k.

Inequalities (5.10) and (5.11) together imply

$$\sum_{k \in K_C^I} (1 - \frac{\delta_k}{d}) + \sum_{k \in K_R^I} \left[p - (\frac{1}{d} - 1 + p)\delta_k \right] \geq 0; \tag{5.14}$$

while inequalities (5.12) and (5.13) imply that the NPV of the surplus must be non-negative:

$$\sum_{k \in K_C^I} (1 - \delta_k) + \sum_{k \in K_R^I} [(1 - d + dp)(1 - \delta_k)] \geq 0 \tag{5.15}$$

We claim that the optimal informal rules are given by the solution to the following auxiliary, finite problem:

$$\max_{K_C^I; K_R^I} \sum_{k \in K_C^I} (1 - \delta_k) + \sum_{k \in K_R^I} [(1 - d + dp)(1 - \delta_k)]$$

s.t. (5.14).

To see this, note that the objective function of the auxiliary problem is the same as in problem (P). Let $(K_C^I; K_R^I)$ be a solution. Since no-contract is a feasible choice that yields zero surplus, $(K_C^I; K_R^I)$ also satisfies (5.15). Since $(K_C^I; K_R^I)$ satisfies (5.14), there is a value M_2 such that $(K_C^I; K_R^I; M_2)$ satisfies (5.10) and (5.11). Furthermore, since $(K_C^I; K_R^I)$ satisfies (5.15), there is an initial wage $m_1 = M_1 - dM_2$ such that $(K_C^I; K_R^I; M_1)$ satisfies (5.12) and (5.13). Therefore the auxiliary problem is a reduced-form version of the original problem.

It is immediate to show that the rigid informal clause R_k^I is dominated by the contingent informal clause C_k^I for all k, because a rigid clause entails a lower net surplus for the principal and does not relax the principal's incentive constraint relative to a contingent clause. This latter statement follows from the fact that, since $\delta_k < 1$, we have $1 - \frac{\delta_k}{d} > p - (\frac{1}{d} - 1 + p)\delta_k$. Therefore, the only candidate clauses for each task k are C_k^I and the empty clause, $\emptyset$.

Given that there are only contingent clauses, constraint (5.14) yields

$$\sum_{k \in K_C^I} \left(1 - \frac{\delta_k}{d}\right) \geq 0.$$

Now we show that, for any pair of tasks $k; k'$, if $\delta_k > \delta_{k'}$ then it cannot be optimal to include task k in the contract and leave task k' out of the contract. Suppose this is the case. Then, we can improve the value of the objective without violating the constraint, by swapping the two tasks, so that task k' is included in the contract and task k is left out. Point (i) follows immediately.

To show point (ii), note that, since the optimal informal contract contains only contingent clauses, the agent's disutility is independent of the state s, hence we can focus on wage profiles that specify a first-period wage m_1 and a fixed wage $\bar{m}$ for $t = 2; 3; \ldots;$. We can derive the values of $(m_1; \bar{m})$ that satisfy the incentive constraints given the optimal informal clauses by plugging back $(m_1; \bar{m})$ in the four incentive constraints, yielding

$$\begin{cases} \frac{1}{d} \sum_{k \in K_C^I} \delta_k \leq \bar{m} \leq \sum_{k \in K_C^I} 1 \\ \sum_{k \in K_C^I} \delta_k \leq (1 - d)m_1 + d\bar{m} \leq \sum_{k \in K_C^I} 1 \end{cases}$$

Point (ii) follows immediately. ■

Proof of Lemma 1

Clearly, for each subgame the Markov perfect equilibrium gives all the net surplus to the principal and zero payoff to the agent. Thus the statement is true for the minimum equilibrium payoff of the agent.

We now exhibit a subgame perfect equilibrium keeping the principal at his maxmin. Consider the following strategies: for all $(\tilde{F}; s)$ (where $\tilde{F}$ is the default contract) the principal makes amendments and exceptions as in the Markov perfect equilibrium. Wages are determined according to a punishment phase. There are two punishment phases P_P and P_A. The system starts in phase P_P. When the system is in phase P_P the (offered) wage is the net profit generated by the offered contract. When the system is in phase P_A the (offered) wage is the disutility generated by the offered contract. As soon as player i deviates from his strategy the system switches immediately to phase P_i. If the system is in phase P_A the agent accepts the offered contract (and chooses the one-shot best response). Thus, in phase P_A the Markov perfect equilibrium is played. If the system is in phase P_P, the new default set of clauses is $\tilde{F}'$, the offered contract is $(F; m)$ and the state of nature is s, then the agent accepts if and only if

$$m - \delta(F; s) > dV(\tilde{F}'); \tag{5.16}$$

where $V(\tilde{F}')$ is the expected PDV of the flow of net surpluses generated by the Markov perfect equilibrium starting at default $\tilde{F}'$.

By construction, the agent has no incentive to deviate. In particular, suppose that the state of nature is s, the principal offers $(F; m)$ moving the default to $\tilde{F}'$ and, as a consequence the system enters (or stays in) phase P_P.

If $m - \delta(F; s) \leq dV(\tilde{F}')$, the agent is supposed to reject. The expected payoff if the agent conforms is $dV(\tilde{F}')$. The expected payoff of a one-shot deviation is $m - \delta(F; s)$, because after the deviation the system enters phase P_A where the agent gets his maxmin (zero). Therefore rejection is indeed a best response.

If $m - \delta(F; s) > dV(\tilde{F}')$ the agent is supposed to accept and this is obviously a best response.

Let us check that the principal has no incentive to deviate in phase P_P. Let the overall state be $(\tilde{F}; s)$, where $\tilde{F} = (\tilde{C}; \tilde{R}; \tilde{\overline{R}})$, $\tilde{C}$ is the set of contingent default clauses, $\tilde{R}$ is the set of rigid default clauses and $\tilde{\overline{R}}$ is the set of negative, rigid default clauses. Since the principal will be kept at his maxmin from the following period, the only way he can make a profitable one-shot deviation is to make amendments $F^A = (C^A; R^A; \overline{R}^A)$ and exceptions $F^A = (C^E; R^E; \overline{R}^E)$ and offer a wage m such that the resulting contract $F = f(\tilde{F}; F^A; F^E)$ and the resulting new default $\tilde{F}' = \tilde{f}(\tilde{F}; F^A)$ satisfy (5.16), making the agent accept, and furthermore $m < \pi(F; s) - Cost(F^A; F^E)$ (where $\pi(F; s)$ is the gross profit generated by F at s), so that he gets a positive payoff in the current period.

Clearly, a profitable deviation exists if and only if there exists $(F^A; F^E)$ and a state of nature s such that

$$\pi(F; s) - \delta(F; s) - Cost(F^A; F^E) > dV(\tilde{F}')$$

To show that this is not possible, we will provide a lower bound for the RHS of the above inequality and an upper bound for its LHS, and show that the former is bigger than the latter.

Note that

$$V(\tilde{F}') = V(\tilde{C}'; \tilde{R}'; \tilde{\overline{R}}') \geq \frac{1}{1-d}\left(\sum_{k \in \tilde{C}'} (1-\delta_k) + \sum_{k \in \tilde{R}' \cup \tilde{\overline{R}}'} \frac{1}{2}(1-\delta_k) + \sum_{k \in (C^E \cup R^E \cup \overline{R}^E) \setminus (\tilde{C}' \cup \tilde{R}' \cup \tilde{\overline{R}}')} (1-\delta_k - c)\right),$$

because it is possible to use the default clauses of $\tilde{F}$ in all future periods (each rigid clause k yields $1-\delta_k$ with probability $p = \frac{1}{2}$ in each period) and to make exceptions in every period with the clauses included in F^E but not included in $\tilde{F}$:

On the other hand,

$$\pi(F; s) - \delta(F; s) - Cost(F^A; F^E) \leq \sum_{k \in \tilde{C}'} (1-\delta_k) + \sum_{k \in \tilde{R}' \cup \tilde{\overline{R}}'} (1-\delta_k) + \sum_{k \in (C^E \cup R^E \cup \overline{R}^E) \setminus (\tilde{C}' \cup \tilde{R}' \cup \tilde{\overline{R}}')} (1-\delta_k - c).$$

But $d > \frac{2}{3}$ implies

$$\sum_{k \in \tilde{C}'} (1-\delta_k) + \sum_{k \in \tilde{R}' \cup \tilde{\overline{R}}'} (1-\delta_k) + \sum_{k \in (C^E \cup R^E \cup \overline{R}^E) \setminus (\tilde{C}' \cup \tilde{R}' \cup \tilde{\overline{R}}')} (1-\delta_k - c) <$$

$$< \frac{d}{1-d}\left(\sum_{k\in\tilde{C}'}(1-\delta_k)+\sum_{k\in\tilde{R}'\cup\bar{R}'}\frac{1}{2}(1-\delta_k)+\sum_{k\in(C^E\cup R^E\cup\bar{R}^E)\setminus(\tilde{C}'\cup\tilde{R}'\cup\bar{R}')}(1-\delta_k-c)\right).$$

thus a profitable deviation is impossible. ■

Proof of Proposition 4

We can apply a similar logic as for the proof of proposition 3. Again we can focus on wage profiles where M_t is stationary from $t = 2$ on. We can write the set of constraints as

$$M_2 \geq \frac{1}{d}\sum_{k\in K_C^I}\delta_k+\frac{1}{1-d}\left(\sum_{k\in K_C^I\cup K_C}\delta_k+\sum_{k\in K_R}\frac{1}{2}\delta_k\right), \tag{5.17}$$

$$M_2 \leq \frac{1}{1-d}\left(\sum_{k\in K_C^I\cup K_C}1+\sum_{k\in K_R}\frac{1}{2}\right); \tag{5.18}$$

$$M_1 \geq \sum_{k\in K_C^I\cup K_C\cup K_R}\delta_k+\frac{d}{1-d}\left(\sum_{k\in K_C^I\cup K_C}\delta_k+\sum_{k\in K_R}\frac{1}{2}\delta_k\right), \tag{5.19}$$

$$M_1 \leq \sum_{k\in K_C^I}1+\sum_{k\in K_C}(1-2c)+\sum_{k\in K_R}(1-c)+\frac{d}{1-d}\left(\sum_{k\in K_C^I\cup K_C}1+\sum_{k\in K_R}\surd\right). \tag{5.20}$$

Inequalities (5.17) and (5.18) yield

$$\sum_{k\in K_C^I}\left(1-\frac{\delta_k}{d}\right)+\sum_{k\in K_C}(1-\delta_k)+\sum_{k\in K_R}\frac{1}{2}(1-\delta_k)\geq 0; \tag{5.21}$$

while inequalities (5.19) and (5.20) imply that the NPV of the surplus net of writing costs must be non-negative:

$$\sum_{k\in K_C^I}(1-\delta_k)+\sum_{k\in K_C}[1-\delta_k-2c(1-d)]+\sum_{k\in K_R}\left[\left(1-\frac{d}{2}\right)(1-\delta_k)-c(1-d)\right]\geq 0. \tag{5.22}$$

The optimal clauses are given by the solution to the following auxiliary problem:

$$\max_{K_C;K_C^I;K_R}\sum_{k\in K_C^I}(1-\delta_k)+\sum_{k\in K_C}[1-\delta_k-2c(1-d)]+\sum_{k\in K_R}\left[\left(1-\frac{d}{2}\right)(1-\delta_k)-c(1-d)\right]$$

s.t. (5.21).

To see this note that the objective function of the auxiliary problem is the same as in problem (P'). Let $(K_C; K_C^I; K_R)$ be a solution. Since no-contract is a feasible choice that yields zero surplus, $(K_C; K_C^I; K_R)$ also satisfies (5.22). Since $(K_C; K_C^I; K_R)$ satisfies (5.21), there is a value M_2 such that $(K_C; K_C^I; K_R; M_2)$ satisfies (5.17) and (5.18). Furthermore, since $(K_C; K_C^I; K_R)$ satisfies (5.22), there is an initial wage $m_1 = M_1 - dM_2$ such that $(K_C; K_C^I; K_R; M_1)$ satisfies (5.19) and (5.20). Therefore the auxiliary problem is a reduced-form version of the original problem.

To prove point (i), we argue in three steps: (1) For any pair of tasks $k; k'$, if $\delta_k > \delta_{k'}$ then it cannot be optimal to regulate task k informally and task k' by formal contingent clause. Suppose this is the case, and consider swapping the two tasks, so that task k' is now regulated informally and task k by formal contingent clause. The value of the objective does not change, and the constraint gets relaxed; applying our selection criterion, this is preferable to the original contract. (2) For any pair of tasks $k; k'$, if $\delta_k > \delta_{k'}$ then it cannot be optimal to regulate task k by formal contingent clause and task k' by formal rigid clause. Suppose this is the case, and consider swapping the two tasks. This improves the value of the objective without violating the constraint. (3) For any pair of tasks $k; k'$, if $\delta_k > \delta_{k'}$ then it cannot be optimal to regulate task k by formal rigid clause and leave task k' out of the contract. Suppose this is the case, and consider swapping the two tasks. Again, this improves the value of the objective without violating the constraint. The claim follows right away.

To prove point (ii), we display a numerical example. Suppose there are four tasks, with $\delta_1 = 2/3$, $\delta_2 = 3/4$, $\delta_3 = 13/16$ and $\delta_4 = 29/32$, and suppose $d = 2/3$. In this case, we claim that there exists a critical level c^* such that, if $c < c^*$, the optimum is $\{C_1^I; C_2^I; C_3; C_4\}$, and if c is slightly higher than c^* the optimum is $\{C_1^I; C_2; R_3; D\}$: First note that for these parameter values, contracts $\{C_1^I; C_2^I; C_3; C_4\}$ and $\{C_1^I; C_2; R_3; D\}$ are implementable, i.e. satisfy (5.21), while contracts $\{C_1^I; C_2^I; C_3^I; C_4\}$ and $\{C_1^I; C_2^I; C_3; R_4\}$ are not. Also note that contract $\{C_1^I; C_2; R_3; D\}$ yields higher surplus than $\{C_1^I; R_2; R_3; R_4\}$. These facts imply that the best implementable contract among those that cost 4c is $\{C_1^I; C_2^I; C_3; C_4\}$, and the best implementable contract among those that cost 3c is $\{C_1^I; C_2; R_3; D\}$. This in turn implies that there is a critical level c^* such that for $c \in (0; c^*)$ the optimum is $\{C_1^I; C_2^I; C_3; C_4\}$, and for c in a right neighborhood of c^* the optimum is $\{C_1^I; C_2; R_3; D\}$: It follows that F may decrease if c decreases.

We leave it to the reader to construct an example in which a similar result obtains for an increase in d. ■

References

[1] ALLEN, F. (1985): "Repeated Principal-Agent Relationships with Lending and Borrowing," Economic Letters, 17, 27-31.

[2] ANDERLINI, L. and L. FELLI (1994): "Incomplete Written Contracts: Undescribable States of Nature," Quarterly Journal of Economics, 109, 1085-1124.

[3] ANDERLINI, L. and L. FELLI (1997): "Costly Coasian Contracts," mimeo, University of Cambridge.

[4] ANDERLINI, L. and L. FELLI (1999): "Incomplete Contracts and Complexity Costs," Theory and Decision, 46 (1), 23-50.

[5] BAKER, G., R. GIBBONS and K. MURPHY (1994): "Subjective Performance Measures in Optimal Incentive Contracts," Quarterly Journal of Economics, 109, 1125-1156.

[6] BATTIGALLI, P. and G. MAGGI (2000): "Imperfect Contracting," Working Paper S11, Department of Economics, Princeton University.

[7] BULL, C. (1987): "The Existence of Self-Enforcing Implicit Contracts," Quarterly Journal of Economics, 102, 147-159.

[8] CRAWFORD, V. (1988): "Long-Term Relationships Governed by Short-Term Contracts," American Economic Review, 78, 485-499.

[9] DYE, R. A. (1985a): "Costly Contract Contingencies," International Economic Review, 26, 233-50.

[10] DYE, R. A. (1985b): "Optimal Length of Labour Contracts," International Economic Review, 26, 251-270.

[11] FUDENBERG, D., B. HOLMSTROM and P. MILGROM (1990): "Short-Term Contracts and Long-Term Agency Relationships," Journal of Economic Theory, 51, 1-31.

[12] FUDENBERG, D. and J. TIROLE (1991): Game Theory. Cambridge MA: MIT Press.

[13] GALLAGER, R. (1996): Discrete Stochastic Processes. Dordrecht: Kluwer Academic Publishers.

[14] GRAY, J. (1976): "Wage Indexation: A Macroeconomic Approach," Journal of Monetary Economics, 2, 221-236.

[15] GRAY, J. (1978): "On Indexation and Contract Length," Journal of Political Economiy, 86, 1-18.

[16] HARRIS, M. and B. HOLMSTROM (1987): "On the Duration of Agreements," International Economic Review, 28, 389-406.

[17] HART, O. and B. HOLMSTROM (1987): "The Theory of Contracts," in T. Bewley (Ed.) Advances in Economic Theory, Fifth World Congress. Cambridge UK, C.U.P.

[18] KRASA, S. and S.R. WILLIAMS (1999): "Incompleteness as a Constraint in Contract Design," mimeo, University of Illinois.

[19] LAMBERT, R. (1983): "Long Term Contracting and Moral hazard," Bell Journal of Economics, 14, 441-452.

[20] LIPMAN, B (1997): "Limited Rationality and Endogenously Incomplete Contracts," mimeo, Department of Economics, University of Western Ontario.

[21] MacLEOD, W. B. (2000): "Complexity and Contract," Revue d'Economie Industrielle, 92, 149-178.

[22] MacLEOD, W. B. and J.M. MALCOMSON (1989): "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," Econometrica, 57, 447-480.

[23] MALCOMSON, J.M. and F. SPINNEWYN (1988): "The Multiperiod Principal-Agent Problem," Review of Economic Studies, 55, 391-407.

[24] MASKIN, E. and J. TIROLE (1999): "Unforeseen Contingencies and Incomplete Contracts," Review of Economic Studies, 66, 83-114.

[25] PEARCE, D. and E. STACCHETTI (1998): "The Interaction of Implicit and Explicit Contracts in Repeated Agency," Games and Economic Behavior, 23, 75-96.

[26] REY, P. and B. SALANIE (1990): "Long-Term, Short-Term and Renegotiation: On the Value of Commitment in Contracting," Econometrica, 58, 597-619.

[27] ROGERSON, W. (1985): "Repeated Moral Hazard," Econometrica, 53, 69-76.

[28] TIROLE, J. (1999): "Incomplete Contracts: Where Do We Stand?" Econometrica, 67, 741-781.

[29] TOWNSEND, R. (1982): "Optimal Multiperiod Contracts and the Gain From Enduring Relationships under Private Information," Journal of Political Economy, 90, 1166-1186.