

Relational Incentive Contracts

Jonathan Levin

May 2006

These notes consider Levin's (2003) paper on relational incentive contracts, which studies how self-enforcing contracts can provide incentives in agency settings. The principal applied motivation is that contractual relationships often function quite well without every detail of the relationship being codified in a contingent court-enforceable contract. There are many reasons for this: information may be available to the parties that is unverifiable in court; writing or enforcing contracts can be time-consuming and expensive; some contracts (such a vote-buying) may be illegal; the court system may be non-existent or poorly functioning. The paper tries to identify the extent to which a long-term relationship sustained by goodwill or reputation can substitute for the kind of perfect enforcement mechanisms typically assumed in incentive theory.

The analysis uses the repeated game methods of Abreu, Pearce and Stacchetti (1990) and Fudenberg, Levine and Maskin (1994). Relational contracts are a special class of repeated games, however, because the ability to make monetary transfers simplifies the structure of optimal repeated game equilibria. Because parties can "settle up" period by period, rather than moving to a different continuation equilibrium, the analysis becomes much more straightforward.

1 The Model

There are a principal and an agent, both risk-neutral, who share a common discount factor $\delta < 1$. At each time t , they interact as follows. The principal first offers the agent a salary w_t . The parties then simultaneously choose whether to transact or separate. If they separate, they receive period payoffs $\bar{\pi}$ and \bar{u} . If both choose to trade, the agent chooses an effort $e_t \geq 0$ at cost $c(e_t, \theta_t)$, where $c_e, c_{ee} > 0$. Effort is not directly observed, but generated an output y_t , drawn from a distribution with density $f(\bullet|e)$. The principal

receives the output, pays the salary w_t and may also make a discretionary payment b_t (we also allow b_t to be negative in which case the agent makes the discretionary payment). Let $W_t = w_t + b_t$ denote the total payment.

The realized payments at time t are then $y_t - W_t$ for the principal, $W_t - c(e_t)$ for the agent. The joint surplus is $y_t - c(e_t)$. The first best effort level is e^{FB} , the solution to $\max_e s(e) = E[y|e] - c(e)$. Average payoffs in the repeated game are given by:

$$\begin{aligned}\pi &= (1 - \delta)E \left\{ \sum_{t=0}^{\infty} \delta^t (y_t - W_t) \right\} \\ u &= (1 - \delta)E \left\{ \sum_{t=0}^{\infty} \delta^t (W_t - c(e_t)) \right\}\end{aligned}$$

It is useful to define $s = \pi + u$ to be the average surplus in the repeated game, and $\bar{s} = \bar{\pi} + \bar{u}$ to be the surplus realized if the agents do not transact. We'll assume for simplicity that $s(e^{FB}) > \bar{s} > s(0)$.

Observe that the unique nash (and subgame perfect) equilibrium in a one-shot game (or if $\delta = 0$) is for the parties not to trade. Why? If they do, the principal certainly will not make a discretionary payment, so $b = 0$. So the agent has no incentive to exert effort and will choose $e = 0$. But then a greater surplus would be realized by not trading, so no salary can make trade desirable for both parties.

More is possible in the repeated game. To define repeated game strategies, let $h^t = (y_0, w_0, b_0, \dots, y_{t-1}, w_{t-1}, b_{t-1})$ denote a time t history. A strategy for the principal specifies a salary $w_t(h^t)$, a decision about whether or not to trade, and a contingent bonus $b_t(h^t, y_t)$. A strategy for the agent specifies whether or not to trade and an effort level $e_t(h^t, w_t)$. A *relational contract* specifies for any history h^t , an effort e_t , a salary w_t and a contingent bonus $b_t(y)$. A relational contract is *self-enforcing* if it corresponds to some perfect public equilibrium of the repeated game.

A useful observation is that if there is some self-enforcing contract (or PPE) that achieves a joint surplus s , then there are self-enforcing contracts that achieve any individually rational split of this surplus.

Proposition 1 *Suppose there is some self-enforcing contract with expected surplus s . Then any payoff vector u, π with $u + \pi = s$, $u \geq \bar{u}$ and $\pi \geq \bar{\pi}$ can be achieved with a self-enforcing contract.*

Proof. Suppose the PPE that has expected surplus s gives expected payoffs $\hat{u}, \hat{\pi}$ and involves a salary \hat{w} in the first period. Let $u \geq \bar{u}, \pi =$

$s - u \geq \bar{\pi}$ be given. Construct a new contract as follows. In the first period, the principal offers a salary $w = \hat{w} + (u - \hat{u})/(1 - \delta)$, following which play exactly follows the PPE that gives payoffs $\hat{u}, \hat{\pi}$. If the principal deviates, the players do not transact at date 0 or any following date. *Q.E.D.*

2 Stationary Contracts

A relational contract is *optimal* if no other self-enforcing contract achieves a higher surplus. A key simplifying result is that in searching for optimal contracts, it suffices to consider contracts that are stationary: i.e. that involve the same salary, effort and contingent bonus plan in every period on the equilibrium path.

Definition 1 *A contract is stationary if in every period on the equilibrium path $e_t = e$, $w_t = w$, and $b_t = b(y_t)$ for some $(e, w, b(y))$.*

Notice that stationary contracts assign the same continuation payoffs (and same continuation play) after every history on the eqm path. This is similar to optimal equilibria in games with perfect monitoring, but in sharp contrast to, say, equilibria in the Green and Porter model.

Proposition 2 *If an optimal contract exists, there is a stationary contract that is optimal.*

Proof. Let s^* denote the surplus achieved by an optimal contract. Suppose there is some optimal contract that achieves payoffs u, π , with $u + \pi = s^*$, involves a salary w_0 , effort e_0 and bonus payments $b_0(y_0)$ at $t = 0$ and specifies continuation payoffs $u_1(h^1), \pi_1(h^1)$ starting at $t = 1$. Note that for histories off the equilibrium payoff, we can without loss generality specify that the players cease to transact forever, as this is the worst possible punishment.

The first claim is that any optimal contract must be sequentially optimal. That is $s(e_0) = s^*$ and moreover $u_1(h^1) + \pi_1(h^1) = s^*$ for any h^1 on the equilibrium path. To see why the latter must be so, notice that increasing $\pi_1(h^1)$ improves the principal's incentives to deliver on discretionary payments without changing the agent's incentives at all. Therefore if $u_1(h^1) + \pi_1(h^1) < s^*$ for some h^1 on the equilibrium path, it would be possible to increase $\pi_1(h^1)$ and have a new self-enforcing contract with higher initial surplus. Therefore starting at time $t = 1$, any optimal contract must achieve surplus s^* for any history h^1 on the equilibrium path. Clearly a

higher surplus is not possible starting at $t = 1$ as s^* is the highest possible equilibrium surplus. But then to achieve s^* on average from date $t = 0$, it must be the case that $s(e_0) = s^*$.

Having established sequential optimality, we now use the (possibly non-stationary) optimal contract to construct a stationary contract that achieves the same surplus. Let u, π be individually rational payoff vectors with $u \geq \bar{u}$, $\pi \geq \bar{\pi}$ and $u + \pi = s^*$. Let $e = e_0$, so that $s(e) = s^*$. Define payments $w, b(y)$ to satisfy:

$$\begin{aligned} u &= w + \mathbb{E}_y[b(y)|e] - c(e) \\ b(y) + \frac{\delta}{1-\delta}u &= b_0(y_0) + \frac{\delta}{1-\delta}u_1(w_0, e_0, y). \end{aligned}$$

Consider the agent's expected future payoff at the point in time he chooses his action. By construction it is the same under the stationary contract as it is in the first period of the optimal contract. As $e_0 = e$ was optimal in the first period of the optimal contract, the same is true in the stationary contract.

Next consider the parties' expected future payoff at the point in time they choose whether or not to make the discretionary payment. The agent's payoff is:

$$b(y) + \frac{\delta}{1-\delta}u = b_0(y_0) + \frac{\delta}{1-\delta}u_1(w_0, e_0, y),$$

i.e. identical to in the first period of the optimal contract. The principal's payoff is:

$$-b(y) + \frac{\delta}{1-\delta}\pi = -b_0(y_0) + \frac{\delta}{1-\delta}\pi_1(w_0, e_0, y),$$

i.e. identical to in the first period of the optimal contract (note we've used the fact that $\pi_1 + u_1 = \pi + u = s^*$). So both parties are willing to make the discretionary payment rather than walk away, and we have identified a stationary contract that is self-enforcing and generates the optimal surplus s^* (indeed with an arbitrary individually rational split). *Q.E.D.*

An implication of this result is that to characterize optimal contracts, one can consider only stationary contracts. The basic logic of the result is very simple. In the model, the parties have two instruments to provide incentives: contingent transfers made today and continuation payoffs. These instruments are perfect substitutes. If we start with an optimal contract where the principal provides incentives using variation in continuation payoffs, we can always replace this variation with variation in transfers payments today yielding a stationary contract that provides the same incentives.

Remark 1 *The optimal stationary contract constructed in the proof of Proposition 2 is not renegotiation proof because observable deviations (e.g. refusal to make specified payments) are punished with termination of the relationship. Levin (2003) argues that one can construct optimal contracts that are strongly renegotiation proof, however. The reason is that among the set of optimal (stationary) contracts are contracts that hold each of the two parties to their outside options, \bar{u} and $\bar{\pi}$ respectively.*

3 Optimal Contracts

The next step is to characterize optimal stationary contracts. A stationary contract consists of an effort level e , a salary w and a contingent payment plan $b(y)$. The next result explains exactly what stationary contracts are self-enforcing.

Proposition 3 *There exists a stationary contract that implements effort e if and only if there is some payment schedule $W(y)$ such that*

$$e \in \arg \max_{\hat{e}} \mathbb{E}_y[W(y)|\hat{e}] - c(\hat{e}) \quad (\text{IC})$$

and

$$\frac{\delta}{1-\delta}(s(e) - \bar{s}) \geq \max_y W(y) - \min_y W(y) \quad (\text{DE})$$

Proof. Note that to construct a self-enforcing contract it is natural to punish any departure from the contract with the worst possible continuation payoff, namely the separation payoffs $\bar{u}, \bar{\pi}$. Given this, a stationary contract $\{e, w, b(y)\}$ will be self-enforcing if and only if it (1) gives the principal a period expected utility $\pi \geq \bar{\pi}$ and the agent a period expected utility $u \geq \bar{u}$, (2) satisfies the incentive compatibility constraint

$$e \in \arg \max_{\hat{e}} \mathbb{E}_y[w + b(y)|\hat{e}] - c(\hat{e}),$$

and (3) satisfies two constraints on the discretionary transfer payment $b(y)$: for all y ,

$$\begin{aligned} b(y) &\leq \frac{\delta}{1-\delta}(\pi - \bar{\pi}) \\ -b(y) &\leq \frac{\delta}{1-\delta}(u - \bar{u}) \end{aligned}$$

To prove the result, we first show that (IC) and (DE) are necessary for there to be a self-enforcing contract that implements e . Suppose $\{e, w, b(y)\}$

is self-enforcing. Define $W(y) = w + b(y)$. Then $e, W(y)$ satisfies (IC) and (DE).

Conversely, suppose $e, W(y)$ satisfies (IC), (DE). Let u, π be given with $u + \pi = s(e)$, $u \geq \bar{u}$ and $\pi \geq \bar{\pi}$. Construct stationary payments $w, b(y)$ that satisfy::

$$\begin{aligned} u &= \mathbb{E}_y[w + b(y)|e] - c(e) \\ b(y) &= W(y) - \min_{\tilde{y}} W(\tilde{y}). \end{aligned}$$

By construction, the stationary contract $\{e, w, b(y)\}$ (1) generates individually rational payoffs $u \geq \bar{u}$, $\pi \geq \bar{\pi}$ with $u + \pi = s(e)$, (2) as a consequence of (IC), satisfies the incentive compatibility constraint above, and (3) as a consequence of (DE), satisfies the restrictions on discretionary transfers. *Q.E.D.*

The result says that stationary contracts must satisfy two natural constraints: a standard incentive compatibility constraint for the agent's effort choice and a dynamic enforcement constraint. The latter requires that discretionary payments are not too small (to prevent the agent from walking away), nor too large (to prevent the principal from walking away). This limited variation in payments is what distinguishes optimal self-enforced contract from optimal contracts that are court-enforced.

Given the above result, it's pretty straightforward to characterize optimal contracts. To do so, it's useful to impose two assumptions: namely that the distribution of output as a function of effort, $F(y|e)$, satisfies the monotone likelihood ratio property (MLRP) and is concave in effort (CDFC). These assumptions are strong, but fairly standard in incentive theory. They imply that the incentive constraint above can be replaced by a first-order condition for the agent's optimal effort choice.

The optimal contract $\{e, W(y) = w + b(y)\}$ is then the solution to the following problem:

$$\begin{aligned} \max_{e, W(y)} \quad & s = \mathbb{E}[y|e] - c(e) \\ \text{s.t.} \quad & \int_Y W(y) \frac{f_e}{f}(y|e) dF(y|e) - c'(e) = 0 \\ & \frac{\delta}{1 - \delta} (s - \bar{s}) \geq \max_y W(y) - \min_y W(y) \end{aligned}$$

The optimal contract take a very simple form: a fixed salary plus a bonus if output exceeds some threshold. The size of the salary can be varied to achieve different divisions of the joint surplus.

Proposition 4 *Under MLRP and CDFC, the optimal contracts are “one-step”, i.e. there is some \hat{y} with $W(y) = \underline{W}$ if $y \leq \hat{y}$ and $W(y) = \overline{W}$ if $y \geq \hat{y}$.*

Proof. The marginal benefit to raising $W(y)$ for some y is $\min W < W(y) < \max W$ is

$$\mu \cdot \frac{f_e}{f}(y|e),$$

where $\mu > 0$ is the Lagrange multiplier on the incentive compatibility constraint. The MLRP assumption means that $(f_e/f)(y|e)$ is increasing in y for a fixed e , so there will be some \hat{y} s.t. the marginal benefit is positive for all $y > \hat{y}$ and negative for all $y < \hat{y}$. The result follows immediately. *Q.E.D.*

4 Comments

1. The stationarity result is more general than is outlined here. Essentially it follows from two observations. The first is that the combination of risk-neutrality (quasi-linear utility) and monetary transfers allow the parties to replace variation in continuation payoffs with variation in present transfers, i.e. to “settle up” immediately. The second is that in a model where the principal’s actions are observable, optimal contracts will be sequentially optimal, so transfers can be balanced. More generally, for instance in some moral hazard in teams problems, optimal contracts might involve money-burning (a deliberate destruction of surplus).
2. As a result, the only difference between standard (static) incentive theory and relational incentive theory is the presence of the dynamic enforcement constraint. As a consequence, many applications are possible: hidden action as above, hidden information, multiple agents (Levin, 2002), the use of both verifiable and observable but non-verifiable information (in Baker, Gibbons and Murphy 1993), team production (Rayo ’01). It is also possible to incorporate explicit (payoff-relevant) state variables.

3. Self-enforcement has interesting implications for the use of hidden information screening contracts. To study hidden information we assume that the agent privately observes some iid cost shock θ_t drawn from a distribution $P(\cdot)$ and chooses output y_t at cost $c(y_t, \theta_t)$. Levin (2003) shows that in this setting optimal contracts either achieve the first-best or they will involve production distortions for *all* cost types. Moreover, second-best contracts always involve pooling. For low discount factors, optimal contracts require all types to pool on a single level of effort. For medium discount factors, optimal contracts separate high cost types but pool low-cost types. For high discount factors, the first best separating contract is possible.
4. The last section of Levin (2003) considers a variant of the hidden action model where output is privately observed by the principal, rather than commonly observed. The principal can then issue a report about the agent's performance (a subjective evaluation). This model is much more complicated because it involves *private monitoring*. Two issues arise. The first is that an optimal contract must provide incentives for the agent to exert effort and for the principal to monitor honestly. As a result, equilibrium contracts cannot be sequentially optimal; joint surplus must vary over time. The second question that arises is how the principal should release information over time. Discounting means that the principal cannot wait forever to make payments. But concealing information makes it easier to provide incentives for the agent (this is in insight of Abreu, Milgrom and Pearce, 1991). Levin (2003) restricts attention to "full-review" contracts (i.e. PPE) and shows that one-step termination contracts are optimal. MacLeod (2003) and Fuchs (2005) provide further analyses.
5. An important early paper on relational contracts by MacLeod and Malcolmson (1989) provides a full characterization of self-enforcing contracts under the assumption of perfect information. Not surprisingly, stationary contracts are optimal, the key enforcement condition being that $s(e) \geq \frac{\delta}{1-\delta}c(e)$. Their paper also goes a step further by nesting the agency model in a market equilibrium setting where principals and agents match to start relationships. This is a great paper that didn't get nearly the attention it deserved when it was published.

References

- [1] Abreu, D., D. Pearce and E. Stacchetti (1990) "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*.
- [2] Abreu, D., P. Milgrom and D. Pearce (1991) "Information and Timing in Repeated Partnerships," *Econometrica*.
- [3] Baker, G., R. Gibbons and K. Murphy, (1993) "Subjective Performance Measures in Optimal Incentive Contracts," *Quart. J. Econ.*
- [4] Fuchs, W. (2005) "Contracting with Repeated Moral Hazard and Subjective Evaluations," Working Paper.
- [5] Fudenberg, D., D. Levine and E. Maskin (1994) "The Folk Theorem with Imperfect Public Information," *Econometrica*.
- [6] Levin, J. (2002) "Multilateral Contracting and the Employment Relationship," *Quart. J. Econ.*
- [7] Levin, J. (2003) "Relational Incentive Contracts," *Amer. Econ. Rev.*
- [8] MacLeod, B. (2003) "Optimal Contracting with Subjective Evaluation," *Amer. Econ. Rev.*
- [9] MacLeod, B. and J. Malcomson (1989) "Implicit Contracts, Incentive Compatibility and Involuntary Unemployment," *Econometrica*.
- [10] Rayo, L. (2004) "Relational Team Incentives and Ownership," Working Paper.