

Dynamic Financial Contracting
Lecture Note # 2
The discrete time infinite horizon case

Bruno Biais

Toulouse School of Economics

November 2009

Preliminary & incomplete.

Comments welcome

1 Introduction

In this second lecture note we keep the same framework as in the first one. But we look at the infinite horizon case. We consider a dynamic moral hazard model in which high effort is always requested. The analysis is directly in line with that presented in Biais, Mariotti, Plantin and Rochet (Review of Economic Studies, 2007 — hereafter BMPR.) In BMPR the contracting friction arises because of unobservable cash—diversion. As we have seen in the first lecture note, when high effort is requested, there is an isomorphism between the moral hazard and the cash diversion model. Like BMPR, the analysis presented here can be understood as the infinite horizon counterpart to the finite horizon model of DeMarzo and Fishman (Review of Financial Studies, 2007). Section 2 of this lecture note presents the optimal contract, Section 3 presents its implementation, Section 4 concludes.

2 The optimal contract

Timing: At period n the timing is the following:

- The continuation utility of the agent W_n and the value function of the principal F_n are evaluated given H_n , the information available at the beginning of period n .
- The new size of operation of the firm X_n is determined.
- The agent privately decides whether to exert effort or not.
- The cash flow (μ or $\mu - C$) is realized.
- The agent receives the transfer $T_n = X_n t_n$.

Contract: For each time n , the contract specifies, as a function of the past history H_n , the actions to be taken: X_n , t_n and e_n , thus pinning down W_n . We focus on the case where the contract requests high effort to be always exerted. Also we assume the parties can fully commit to the contract, so there is only one source of friction: the unobservability of effort.

Continuation utility: The continuation expected utility of the agent at period n is:

$$W_n = E\left[\sum_{k=0}^{\infty} \frac{X_{n+k} t_{n+k}}{(1+\rho)^k} \middle| H_n\right].$$

This can be expressed recursively as

$$W_n = E\left[X_n t_n + \frac{W_{n+1}}{1+\rho} \middle| H_n\right]. \quad (1)$$

To reason in size adjusted terms we set

$$W_n = X_{n+1} w_n. \quad (2)$$

Substituting (2) into (1)

$$X_{n-1} w_n = E\left[X_n t_n + \frac{X_n w_{n+1}}{1+\rho} \middle| H_n\right]. \quad (3)$$

Dividing both sides by X_{n-1} ,

$$w_n = x_n E[t_n + \frac{w_{n+1}}{1 + \rho} | H_n], \quad (4)$$

where $x_n = \frac{X_n}{X_{n-1}}$ can be interpreted as the downscaling factor (it can be taken out of the expectation since, at the beginning of period n it is known, in contrast with t_n and w_{n+1} which are known only after \tilde{C}_n is realized.)

The state variable: In principle the continuation utility of the agent, as well as the value function of the principal, depend on all the information available in H_n . But, since the horizon is infinite the optimal contract and the utilities of the player are not time dependent. Therefore, to reduce the burden of notation, we will omit the time indices whenever it is not confusing, and correspondingly we will denote the continuation utility of the agent by w . Furthermore, it can be shown that the only state variable one needs to remember to characterize the size-adjusted utilities of the players is w . A formal proof of that property is given in the verification theorem in BMRV.

The dynamics of w : w is a conditional expectation. Hence its changes are driven by the arrival of innovations in the information set of the players. In the present case these innovations reflect the cash flow realizations. Hence the continuation utility of the agent will evolve as a function of the cash flow realizations. It will move from w to w^+ after a success and from w to w^- after a failure. Similarly, the size-adjusted transfer received by the agent just after a success is t^+ , while its counterpart just after a failure is t^- . Hence the equation specifying the dynamics of w (4) can be rewritten as

$$w = x[\{(1 - \lambda)t^+ + \lambda t^-\} + \frac{1}{1 + \rho}\{(1 - \lambda)w^+ + \lambda w^-\}], \quad (5)$$

which is the counterpart of equation (6) in BMRV.

Value function: The value function of the principal at period n is:

$$F_n = E[\sum_{k=0}^{\infty} \frac{X_{n+k}(\mu - \tilde{C}_{n+k} - t_{n+k})}{(1 + r)^k} | H_n].$$

This can be expressed recursively as

$$F_n = E[X_n(\mu - \tilde{C}_n - t_n) + \frac{1}{1+r}F_{n+1}]. \quad (6)$$

In size-adjusted terms we can define f_n by

$$X_{n-1}f_n = F_n.$$

Thus we can rewrite (6) as

$$X_{n-1}f_n = E[X_n(\mu - \tilde{C}_n - t_n) + \frac{1}{1+r}X_n f_{n+1}]. \quad (7)$$

Dividing both sides by X_n we have

$$f_n = x_n E[(\mu - \tilde{C}_n - t_n) + \frac{1}{1+r}f_{n+1}]. \quad (8)$$

Since the size-adjusted value function is a function of the state variable w , it will move from $f(w)$ to $f(w^+)$ after a success and from $f(w)$ to $f(w^-)$ after a failure. Hence the equation specifying the dynamics of f can be rewritten as

$$f(w) = x[(\mu - \lambda C) - \{(1 - \lambda)t^+ + \lambda t^-\} + \frac{1}{1+r}\{(1 - \lambda)f(w^+) + \lambda f(w^-)\}]. \quad (9)$$

which is the counterpart of equation (5) in BMPR.

Incentive compatibility: The agent prefers to exert effort iff

$$\begin{aligned} & \{(1 - \lambda)t^+ + \lambda t^-\} + \frac{1}{1+\rho}\{(1 - \lambda)w^+ + \lambda w^-\} \\ \geq & B + \{(1 - (\lambda + \Delta\lambda))t^+ + (\lambda + \Delta\lambda)t^-\} + \frac{1}{1+\rho}\{(1 - (\lambda + \Delta\lambda))w^+ + (\lambda + \Delta\lambda)w^-\}. \end{aligned}$$

Thus the incentive compatibility condition is as given in the next lemma:

Lemma 1: *The incentive compatibility condition is*

$$(t^+ + \frac{w^+}{1+\rho}) - (t^- + \frac{w^-}{1+\rho}) \geq \frac{B}{\Delta\lambda}. \quad (10)$$

Note that (10) is the counterpart of equation (7) in BMPR. The interpretation is the following: The agent is incentivized to exert effort if his reward (wage plus continuation utility) is sufficiently larger after success than after failure. The wedge between the two is the ratio reflecting the magnitude of the moral hazard problem: $B/\Delta\lambda$. This is in line with the results obtained in the two period case (see equation (13) in Lecture Note 1.) Correspondingly the continuation utility of the agent increases after success and decreases after failure. This is in line with result obtained in the two period case, and stated in the first lecture note as Lemma 1.

Social surplus: It is convenient to consider the size-adjusted social surplus generated by the project, defined as

$$v(w) = f(w) + w.$$

Adding equations (5) and (9), we obtain a recursive formulation for v :

$$v(w) = x[(\mu - \lambda C) + \frac{(1 - \lambda)f(w^+) + \lambda f(w^-)}{1 + r} + \frac{(1 - \lambda)w^+ + \lambda w^-}{1 + \rho}].$$

Note that current transfers simplify out of the expression for utilitarian welfare. This rewrites as:

$$v(w) = x[(\mu - \lambda C) + \frac{(1 - \lambda)(f(w^+) + w^+) + \lambda(f(w^-) + w^-)}{1 + r} - \frac{(1 - \lambda)w^+ + \lambda w^-}{1 + r} + \frac{(1 - \lambda)w^+ + \lambda w^-}{1 + \rho}].$$

Or:

$$v(w) = x[(\mu - \lambda C) + \frac{(1 - \lambda)v(w^+) + \lambda v(w^-)}{1 + r} + \frac{(1 + r)[(1 - \lambda)w^+ + \lambda w^-] - (1 + \rho)[(1 - \lambda)w^+ + \lambda w^-]}{(1 + r)(1 + \rho)}].$$

This simplifies to:

$$v(w) = x[(\mu - \lambda C) + \frac{(1 - \lambda)v(w^+) + \lambda v(w^-)}{1 + r} - (\rho - r) \frac{[(1 - \lambda)w^+ + \lambda w^-]}{(1 + r)(1 + \rho)}], \quad (11)$$

which is the counterpart of equation (8) in BMPR.

Rewriting the constraints: Recall that, by shifting attention from the value function of the value function of the principal to the social surplus we have simplified away the current transfers. After some manipulations, one can also eliminate the current transfers from the constraints and obtain the following result:

Lemma 2: *The set of constraints (5), (11) and $t \geq 0$ is equivalent to*

$$w \geq x \left[\frac{w^-}{1 + \rho} + (1 - \lambda) \frac{B}{\Delta \lambda} \right], \quad (12)$$

$$w \geq x \left[\frac{(1 - \lambda)w^+ + \lambda w^-}{1 + \rho} \right], \quad (13)$$

and $t \geq 0$.

Proof of necessity:

For simplicity, we hereafter only establish that (5) and (11) imply (12) and (13) (sufficiency is proved in BMPR.)

First note that (5) and the non-negativity of transfers imply (13). Second note that given a triple (x, w^+, w^-) one can find a pair of non negative transfers such that (5) and (11) only if the set of positive real numbers (t^+, t^-) such that

$$\frac{w}{x} - \frac{(1 - \lambda)w^+ + \lambda w^-}{1 + \rho} \geq (1 - \lambda)t^+ + \lambda t^-, \quad (14)$$

and

$$t^+ - t^- \geq \frac{B}{\Delta \lambda} - \frac{w^+ - w^-}{1 + \rho},$$

is not empty. Setting t^- to 0 relaxes both constraints. So the pair of inequalities become

$$\frac{w}{x(1 - \lambda)} - \frac{(1 - \lambda)w^+ + \lambda w^-}{(1 + \rho)(1 - \lambda)} \geq t^+, \quad (15)$$

and

$$t^+ \geq \frac{B}{\Delta \lambda} - \frac{w^+ - w^-}{1 + \rho}.$$

This is feasible only if

$$\frac{w}{x} - \frac{(1-\lambda)w^+ + \lambda w^-}{1+\rho} \geq (1-\lambda)\left[\frac{B}{\Delta\lambda} - \frac{w^+ - w^-}{1+\rho}\right]. \quad (16)$$

That is

$$\frac{w}{x} - (1-\lambda)\frac{B}{\Delta\lambda} \geq \frac{w^-}{1+\rho}, \quad (17)$$

which implies (12).

QED

Downsizing: (12) and the non—negativity of continuation utilities imply that, when $w < (1 - \lambda)B/\Delta\lambda$, then we must have $x < 1$. When the continuation utility of the agent is small, he does not have a lot of skin in the game. Therefore it is difficult to incentivize him. Hence, it is necessary to downsize the project (i.e., to set $x < 1$) to reduce the private benefits from shirking, to make effort incentive compatible. Since the project has positive net present value, partial liquidation is inefficient in the first best. Thus, downsizing is a distortion implied by the moral hazard problem. This distortion will arise when, after a series of failures, w had to be reduced a lot. This is similar to the result, presented in Proposition 2 of the first lecture note, that downsizing maybe necessary after failure.

The Bellman equation: Building on the above remarks we can state the next lemma:

Lemma 3: *The optimal contract solves:*

$$V(w) = \max_{x, w^+, w^-} x[(\mu - \lambda C) + \frac{(1-\lambda)V(w^+) + \lambda V(w^-)}{1+r} - (\rho - r)\frac{[(1-\lambda)w^+ + \lambda w^-]}{(1+r)(1+\rho)}] \quad (18)$$

subject to

$$w \geq x\left[\frac{w^-}{1+\rho} + (1-\lambda)\frac{B}{\Delta\lambda}\right] \quad (19)$$

and

$$w \geq x\left[\frac{(1-\lambda)w + \lambda w^-}{1+\rho}\right]. \quad (20)$$

Existence and properties of the solution to the Bellman equation: The following proposition is proved in BMPR:

Proposition 1: *There exists a continuous and bounded solution V to the Bellman equation (18). It is non decreasing, concave and equal to 0 at 0. There are two thresholds*

$$w^m \geq w^l \geq (1 - \lambda)B/\Delta\lambda$$

such that:

- If $w \in [0, w^l)$ then $x = w/w^l < 1$ while if $w \geq w^l$ then $x = 1$.
- V is linear over $[0, w^l)$, increasing over $[w^l, w^m)$, and constant above w^m .

The intuition is the following:

- As mentioned above, in the discussion of Lemma 2, when when w is low (in particular when it is lower than $(1 - \lambda)B/\Delta\lambda$) then downsizing is necessary for the incentive compatibility and promise keeping constraints to hold. On the other hand, downsizing is costly, since it amounts to destroying useful production capacity. Hence, in the optimal contract downsizing arises only when w is low (lower than w^l).
- The solution is non decreasing because w increases after success, which reduces the risk of inefficient liquidation and therefore increases social surplus. Also, the solution is concave because the risk of inefficient liquidation when w is low creates an endogenous aversion to variability in w .

- The solution vanishes at 0, because when the continuation utility of the agent reaches 0, he can't be incentivized to work any longer, so the project has to be liquidated.
- At the opposite extreme, when w is very high, liquidation risk becomes very low. At this point, since the agent is more impatient than the principal, it is optimal to let her consume rather than letting w grow via promises of future consumption.

The value function V cannot be greater than its first-best counterpart: $\frac{\mu - \lambda C}{r}$. Hence when w goes to infinity, $V(w) - (\rho - r)\frac{w}{1+\rho}$ goes to minus infinity. Consequently, there is a value of w at which $V(w) - (\rho - r)\frac{w}{1+\rho}$ is maximized. Denote it w_r .

The dynamics of w and payments in the optimal contract: As shown in BMPR, the dynamics of transfers and payments depend on the level of w . Consistent with the above discussion, the lower w , the more binding the constraints. Thus, we have to consider four different regions:

- When w is very large, none of the constraints bind. As shown in BMPR, this is the case when $w \geq w^m$. In that case, it is optimal to move w to the value which maximizes (18) (without constraints.) Now, the value function V cannot be greater than its first-best counterpart: $\frac{\mu - \lambda C}{r}$. Hence when w goes to infinity, $V(w) - (\rho - r)\frac{w}{1+\rho}$ goes to minus infinity. Consequently, there is a value of w at which $V(w) - (\rho - r)\frac{w}{1+\rho}$ is maximized. Denote it w_r . This is the value to which it is optimal to move w when it starts above w^m .
- For all values of w lower than or equal to w^m , at least one of the constraints binds. At (19) w^m new IC is just binding and all the other constraints are slack. Hence:

$$w^m = \frac{w^r}{1 + \rho} + (1 - \lambda)\frac{B}{\Delta\lambda}.$$

- What of $w \in [w^l, w^m]$? In that interval, we must insure that (19) holds. As discussed above, it is costly to reduce w , since the value function is concave, reflecting inefficient liquidation when

w is very low. Hence, for $w \in [w^l, w^m)$, (19) binds. Thus:

$$w^- = (1 + \rho)[w - (1 - \lambda)\frac{B}{\Delta\lambda}].$$

As shown in BMPR, if $w \geq w^m - (1 - \lambda)\frac{B}{\Delta\lambda}$, then (20) is slack, so that we can set $w^+ = w^r$.

But, if $w < w^m - (1 - \lambda)\frac{B}{\Delta\lambda}$, then (20) binds. In that case

$$w^+ = (1 + \rho)[w + \lambda\frac{B}{\Delta\lambda}].$$

- Finally, when $w < w^l$ there is downsizing.

Define $\tilde{\varepsilon}_n$, as the random variable taking the value λC when there is a success and $-(1 - \lambda)C$ when there is a failure. Its expectation is

$$E(\tilde{\varepsilon}_n | H_n) = (1 - \lambda)\lambda C - \lambda(1 - \lambda)C = 0,$$

hence it is a martingale difference. Note that $\tilde{\varepsilon}_n = E(\tilde{C}_n) - \tilde{C}_n = \lambda C - \tilde{C}_n$. Thus, $\tilde{\varepsilon}_n$ reflects the intrinsic risk in the project. It is not linked to the moral hazard problem, and would arise also in the first best case where effort is observable and contractible.

Define

$$k = \frac{B}{\Delta\lambda C}.$$

The numerator is the private benefit obtained via shirking, while the denominator is the loss in expected cash flows due to lack of effort. Because we assume effort is efficient in the first best, we have $k \leq 1$. k is a normalized measure of the magnitude of the moral hazard problem.

Note that w^l and w^m are reflecting boundaries for w : when hits the upper bound w is reflected downward due due to transfers, while when it hits the lower bound it is reflected upward via downsizing. Between these two reflecting boundaries, the dynamics of w is

$$w_{n+1} = (1 + \rho)(w_n + k\varepsilon_n),$$

or

$$\frac{w_{n+1}}{1 + \rho} - w_n = k\varepsilon_n. \quad (21)$$

Thus, w_n is a discounted martingale. Its volatility is the product of the intrinsic volatility of the cash flow process (that of ε_n) multiplied by k , the magnitude of the agency problem. In other words, incentive compatibility requests that the sensitivity of the continuation utility of the agent to the performance of the firm be equal to the magnitude of the moral hazard problem k .

3 Implementation

The abstract optimal contract presented above can be implemented with realistic financial instruments. The project is operated as a firm, with a CEO (the agent) and outside investors (the principals). The latter hold securities, which are limited liability claims on the cash flows of the firm. These securities can be held diffusely. Since diffuse claims are difficult to renegotiate, this lends support to our assumption that the parties can commit to the contract. The activity of the firm is reflected in its balance sheet. On the asset side are the project, and the cash reserves of the firm, denoted by M_n , or, in size adjusted terms: $X_n m_n$. On the liability side are debt and equity. The former is owned by the outside financiers, the latter is owned in proportion k by the manager (inside equity) and in proportion $1-k$ by the outside investors (outside equity).

Thus, when the moral hazard problem is more severe, the manager owns a greater fraction of the equity. This suggests that empirical studies interpreting large degree of equity ownership (give references... Shleifer ?) may be misguided. Our analysis implies that the fraction of equity ownership, along with other financial structure variables is endogenous, and reflects deep parameters, such as k .

The coupon on the debt is $D_n = X_n d_n$. The dividends on equity are denoted by $E_n = X_n e_n$. The dynamics of the cash reserves of the firm are

$$M_{n+1} = (1 + r)(M_n + (\mu - \tilde{C}_n)X_n - D_n - E_n),$$

or, in size adjusted terms:

$$m_{n+1} = (1+r)(m_n + (\mu - \tilde{C}_n) - d_n - e_n).$$

The right-hand-side is the amount of cash held by the firm at the end of period n , equal to initial cash, plus net cash-flows from operation, minus payments to investors. The left-hand-side is that amount of cash, capitalized from period n to period $n+1$ at the interest rate serve on cash reserves: r .

To simplify the analysis, focus on the case where the discount rate of the principal and that of the agent are equal $\rho = r$. In that case, the coupon paid each period to the bondholders is equal to the expected cash flow from operations:

$$d_n = \mu - \lambda C.$$

Thus, in the region where there are no dividend and no downsizing, the dynamics of cash reserves is

$$m_{n+1} = (1+r)(m_n + (\mu - \tilde{C}_n) - (\mu - \lambda C)) = (1+r)(m_n + (\lambda C - \tilde{C}_n)) = (1+r)(m_n + \tilde{\varepsilon}_n),$$

or

$$\frac{m_{n+1}}{1+r} - m_n = \tilde{\varepsilon}_n. \quad (22)$$

Thus, in the region where there is no downsizing nor dividend, m_n is a discounted martingale. Comparing (21) and (22), one can see that the relation between continuation utility and cash reserves is simply

$$m_n = \frac{w_n}{k}.$$

To finalize the description of the implementation, note what happens at the upper and lower bound of the region in which m_n is a discounted martingale:

- When m_n hits $m^l = \frac{w^l}{k}$ the firm must be downsized. Note that m_n , which is the ratio of the liquid assets held by the firm M_n to the size of its physical assets X_n , can be interpreted as a liquidity ratio. Hence, in our implementation, the firm must hold a minimum liquidity ratio. When its actual liquidity ratio falls below that, the firm must be downsize, to the point that the size of its assets is low enough the the liquidity constraint holds.

- When m_n hits $m^h = \frac{w^h}{k}$ dividends are paid. Note that m_n which results from the accumulation of cash flow from operations and interests on liquid assets minus payments to investors can be interpreted as retained earnings. Hence, in our implementation, the firm pays dividends only when retained earnings reach milestones. This is consistent with the findings of Kaplan and Stromberg.