The Moral Hazard Model with Multiple Activities

Stefan Hupfeld[1]

February 21, 2005

Seminar Paper for the
Doctoral Course on Contract Theory
by Prof. Dr. Fabel
at the University of Konstanz

Abstract: If a principal - agent relationship is such that an action of the agent affects both principal and agent, but the action itself is not perfectly observable, we speak of a moral hazard problem. Here, an extension is presented, namely the case of a variety of possible actions that affect both the principal and the agent, while the degree of observability may differ among the actions. The optimal contract derived under these conditions appears to be a function of the whole variance-covariance matrix of the disturbance terms and of the first and second derivatives of the principal's and the agent's utility function, respectively.

[1]Department of Economics, University of Konstanz, Box D136, 78457 Konstanz, Germany, Stefan.Hupfeld@uni-konstanz.de

1. Introduction

Think of an *agent* that takes some action, and this action has three main properties: (1) it takes effort to venture in this action, (2) it is hidden to the *principal*, and (3) the outcome of the action affects the utility of both the agent and the principal. The first is represented as disutility (or cost function) to the agent, the second is represented by noise that distorts the otherwise deterministic transformation of action into outcome or a signal[2], and third, the principal draws utility directly from the outcome, while the agent receives some payment conditioned on outcome or signal. The fact that the principal cannot infer on whether the agent has taken this action (or to what extent) creates *moral hazard*. Usually, the parties' attitudes to risk differ: often the agent is modelled as risk averse, while risk neutrality is presumed for the principal.

The principal now faces an optimization problem of the following kind: Letting the agent heavily enjoy the outcome would align the goals of both parties, but the agent would carry the risk of the noise; fully insuring the agent on the other hand will drive the agent to take the lowest effort possible, as his payment does not depend on outcome anymore.

Now suppose that the agent can venture in more than only one activity. Each single action[3] may affect utilities of the agent and the principal in different ways, to the limit in which an action taken by the agent is not even argument of the principal's utility. In addition, the actions may differ in their degree of observability and in their interaction that creates disutility to the agent. Under additional restrictions, the explicit formulation of the optimal contract can be derived.

In order to achieve consistency with the lecture "Contract Theory" held by Prof. Dr. Fabel at the University of Konstanz, I follow the notation of this lecture, wherever possible.

2. The Moral Hazard Model with a Single Activity

Let the following definitions and assumptions hold throughout this section. The agent behaves as if he were maximizing a utility function denoted by $U(q,t)$, the agent's type is θ, and q is the quantity supplied by the agent. The agent's type θ is distributed according to $F(\theta)$ on the support Θ. The principal receives the quantity $Q(q,\theta)$, has the utility $W(Q,t)$, where $t(Q)$ denotes the payment made to the agent. The functions are assumed to exhibit the following properties or are assumed to take the following forms, respectively:

$$
\begin{aligned}
U &= u(t) - q \quad \text{with} \quad \frac{\partial u}{\partial t} > 0 \quad \text{and} \quad \frac{\partial^2 u}{\partial t^2} < 0 && (1)\\
W &= Q - t && (2)\\
\frac{\partial Q}{\partial q} &> 0, \quad \frac{\partial^2 Q}{\partial q^2} < 0, \quad \frac{\partial Q}{\partial \theta} > 0 && (3)
\end{aligned}
$$

Note that $U(q,t)$ could be more general, as long as it satisfies $\partial U/\partial q < 0$, $\partial U/\partial t > 0$ and $\partial^2 U/\partial q^2 \partial^2 U/\partial t^2 - 2\partial^2 U/\partial q \partial t \geq 0$. In the case of utility being additively separable in payment and effort, the latter condition implies (weak) convexity of disutility. Depending on

[2] Salanié (1997) speaks of a *technology*.

[3] Holmstrom and Milgrom (1991) refer to it as a *task*.

the agent's type θ, the principal has to solve

$$\max_{q(\theta),t(\theta)} \int_{\theta\in\Theta} [Q(q,\theta) - t(\theta)]\, dF(\theta), \tag{4}$$

subject to the participation constraint and the incentive compatibility constraint:

$$\int_{\theta\in\Theta} U(\theta)dF(\theta) \geq \overline{U} \tag{5}$$

$$\theta = \arg\max_{\hat{\theta}\in\Theta} U(\hat{\theta},\theta) \tag{6}$$

The formulation in 4, 5, and 6 is called the *state-space* representation. Salanié (1997) uses the *output-space* formulation, which is feasible once the principal is confronted with one agent only:

$$\max_{q,t(Q)} \int_Q [Q - t(Q)]\, dH(Q) \tag{7}$$

$$\text{s.t.} \quad \int_Q U(q,t(Q))dH(Q) \geq \overline{U} \tag{8}$$

$$q = \arg\max_{\hat{q}} \int_Q U(\hat{q},t(Q))dH(Q), \tag{9}$$

where $H(Q)$ denotes the distribution of outcome Q, given q. In order to solve this problem, it is convenient to substitute the incentive compatibility constraint in equation 9 with the help of the first order approach: Under certain assumptions (namely, the *convexity of distribution function condition* and the *monotone likelihood ratio condition* as defined below), the conditions for a global maximum of the agent's utility coincide with the first order conditions for a local and interior extremum.

Definition 1 (Convexity of the Distribution Function Condition, CDFC)**.** *$H(Q|q)$ satisfies the convexity of distribution function condition if $\int \frac{\partial H(Q|q)}{\partial q} < 0$ and $\int \frac{\partial^2 H(Q|q)}{\partial q^2} > 0$.*

The first part of the definition is not crucial as at coincides with the concept of first order stochastic dominance: The distribution of $Q|q_i$ dominates the distribution of $Q|q_j$ if $q_i > q_j$, meaning that higher values of Q are more likely the higher q. Unfortunately, the second part of the definition implies a non-decreasing density (for the case in which the distribution function is differentiable), which is not satisfied by the normal, e.g.

Definition 2 (Monotone Likelihood Ratio Condition, MLRC)**.** *$H(Q|q)$ satisfies the monotone likelihood ratio condition if $\partial \frac{h_q(Q)}{h(Q)}/\partial Q \geq 0$.*

Given these assumptions, the optimal contract that solves the principal's optimization problem exhibits the following property:

Proposition 1. *The optimal contract $t^*(Q)$ is such that the payment to the agent is increasing in the observed output, hence $\frac{\partial t^*(Q)}{\partial Q} > 0$.*

The proof can be found in Laffont and Martimort (2002), chapter 5. The second best level of effort is not necessarily lower than the first best level, which can also be found in Laffont and Martimort (2002).

3. The Moral Hazard Model with Multiple Activities

3.1. The Setup and Assumptions

Now suppose that the number of activities that the agent can take effort in in is not restricted to one - it is now a vector of n distinct actions:

$$q = (q_1, ..., q_n) \tag{10}$$

Denote the *set* of possible actions by $\mathcal{Q}$. The agent's utility is such that each single activity creates disutility:

$$U(t, q) = U(t; q_1, ..., q_n) \tag{11}$$

The agent's utility function is assumed to be additively separable in payment and effort, such that equation 11 can be rewritten as

$$U(t, q) = u(t) - C(q). \tag{12}$$

The function C is strictly convex in all elements of q. The properties of $C(q)$ define the relation among different tasks. Positive cross derivatives constitute substitutes[4], which is a plausible assumption once the total effort taken in different activities can be interpreted as time, e.g. In order to transform output supplied q into output received Q, the dimension has to be reduced to one:

$$Q : q = (q_1, ..., q_n) \in \mathbb{R}^n \to \mathbb{R} \tag{13}$$

This is necessary as Q is directly part of the principal's utility, so it has to be a uni-dimensional index. The principal is again assumed to be risk neutral, so utility of output received is linear in all elements; one possibility is

$$Q = \sum_{i=1}^{k} w_i q_i, \tag{14}$$

where w_i is the weight the principal puts on the specific task. Multiplying by a constant $1/n$, output received can also be thought of as the (weighted) average of all actions. Note

[4] Holmstrom and Milgrom (1991) claim that positive cross derivatives of the cost function imply that the respective tasks are substitutes, which they are, regarding the uncompensated price elasticity; hence if α_i as the price of effort in task i is increased, effort in task j is decreased.

that Salanié (1997) implicitly uses $w_i = 1$ for all i. The number of elements that affect utility of the principal can simply be as great as the number of possible actions, such that $k = n$. In general, there may be more possible actions, so $k \leq n$. Denote the set of actions that create utility to the principal by $\mathcal{K}$.

Payments can be conditioned on observed outcome, but perfect observation is not possible. The technology that transforms n activities into n signals is

$$x = \mu(q) + \theta. \tag{15}$$

The distribution of the noise term θ follows a multinomial normal with zero expectations and a variance-covariance matrix Σ. For simplicity, assume that q and x are of the same dimension, which is not as restrictive as it may seem; although equal dimensions imply that the principal observes *something* about every element of q, the contract does not have to condition on every element as well. This is the case once an infinitely large variance of θ_i disturbs the observation of the respective q_i.

3.2. The Optimal Contract

3.2.1. Elements of the Optimal Contract

Imposing some more restrictions on the model, one can derive properties of the optimal contract. It is assumed that the optimal contract entails a strictly positive payment to the agent, hence that there is an interior solution. The set of possible contracts is restricted to the set of affine functions:

$$t(x) = \alpha' x(q, \theta) + \beta \tag{16}$$

The "incentive part" α of the contract is now a vector. The restriction to affine functions is less of an assumptions, but rather a result: Holmstrom and Milgrom (1987) develop a moral hazard model in continuous time. Under the assumption that the error term follows the Brownian motion (which is the continuous time equivalent to the assumption of normality), the optimal payment contract t is an affine function of the observed signals.

The optimal contract in the case of multiple actions does not only consist of the affine payment function in equation 16, but also of the subset of actions that the agent is allowed to take. Define this subset (the allowance) as $\mathcal{L}$ with l elements, while q is still of dimension n. We have $n \geq l$. The set of allowed actions may then be limited to those that benefit the principal (hence to dimension k), but it may also include actions that are *not* argument of the principal's utility. In the latter case the agent can be driven to concentrate on these specific actions by increasing the power of incentives. So in general, the dimension of possible actions is greater than the dimension of allowed actions, which again is greater than the dimension of actions that are argument of the principal's utility:

$$n \geq l \geq k \qquad \text{and} \tag{17}$$
$$\mathcal{K} \subseteq \mathcal{L} \subseteq \mathcal{Q}. \tag{18}$$

There is some anticipation of what Holmstrom and Milgrom (1994) call *supermodularity*[5] of functions: Think of the second best optimum as being described by a design or combination of parameters. Here, the optimal choice of $\mathcal{L}$ and α is not independent, since $\mathcal{L}$ is merely a function of α then a parameter of its own. In principle, the higher the possible α, depending on the variance of measurement, the greater can be the allowance, just as Holmstrom and Milgrom (1991) name it: "... responsibility and authority should go hand in hand."[6] So there are values of α and $\mathcal{L}$ that in general come together.

Assume further that utility of the agent is exponential in payment:

$$U = -e^{-rt} - C(q), \tag{19}$$

with r being the risk aversion parameter. The assumption of normality and equation 19 together allow to express the agent's utility in terms of the certainty equivalent, which is computed as

$$CE = \alpha' E(x) + \beta - C(q) - \frac{1}{2} r\alpha'\Sigma\alpha \tag{20}$$
$$= \alpha' \mu(q) + \beta - C(q) - \frac{1}{2} r\alpha'\Sigma\alpha \tag{21}$$

The last term of 21 represents a risk premium. The principal's expected utility $E(W)$ is

$$E(W) = E(Q(q)) - \alpha'\mu(q) - \beta. \tag{22}$$

The parameter β is not part of the joint surplus $CE + E(W)$; if all bargaining power is as usually assigned to the principal, β is set in order to fulfill the participation constraint of the agent with equality.[7]

3.2.2. Optimization and the First Order Approach

The size of α is chosen in order to maximize the joint surplus:

[5] In plain words, supermodularity is a property of functions that contains information on the cross derivatives. A function is supermodular, if the elements of every pair of parameters are complements, hence if the respective cross derivatives are greater or equal to zero.

[6] Or as Uncle Ben to Peter Parker in *Spiderman*: "With great power comes great responsibility."

[7] Holmstrom and Milgrom (1991) implicitly define the participation constraint such that the certainty equivalent of the reservation utility is zero. Since by not taking part in the joint project the agent will not bear any risk, this is equivalent to $\overline{U} = 0$.

$$\alpha^* = \arg\max Q - C(q) - \frac{1}{2} r\alpha' \Sigma \alpha \tag{23}$$
$$\text{s.t.} \quad q = \arg\max \alpha' \mu(q) - C(q), \tag{24}$$

the latter being the incentive constraint. Applying the first order approach, this can be restated as

$$\alpha = C', \quad C' = \begin{pmatrix} \partial C/\partial q_1 \\ \partial C/\partial q_2 \\ \dots \end{pmatrix}. \tag{25}$$

In general, the first order approach is feasible once utility is additively separable, and the distribution of output (in the uni-dimensional case) has to meet the conditions in definitions 1 and 2. Unfortunately, Holmstrom and Milgrom (1991) do not provide conditions under which the first order approach is valid in the multivariate case. Nevertheless, consider the standard conditions:

1. The agent's utility is **additively separable** in payment and costs.
2. The **monotone likelihood ratio condition** (as introduced in definition 2 for the uni-dimensional case) is satisfied: The technology that transforms q into the signal x satisfies

$$\frac{\partial x_i}{\partial q_i} > 0, \tag{26}$$

   so the probability of a higher signal rises with higher output supplied. Note that x_i replaces Q in the multi-dimensional case, as Q is by definition of dimension one and cannot serve as a signal for q_i. At the same time we *can* have

$$\frac{\partial Q}{\partial q_i} \geq 0, \tag{27}$$

   so even quantity received depends positively on output supplied. The latter is especially satisfied if Q takes the form as given in equation 14 with all $w_i > 0$.
3. The **convexity of distribution function condition** causes some trouble. By assumption, θ is distributed according to a multivariate normal; the signal, however, is an additive function of output supplied and the agent's type. So for any *given* output supplied, the distribution is a normal with a mean of q, and the distribution is not convex, as the density of the normal is not increasing over its whole support.

Holmstrom and Milgrom (1991) and Holmstrom and Milgrom (1994) impose the following conditions instead:

$$\left(\frac{\partial C}{\partial q}\right)' \Sigma \frac{\partial C}{\partial q} \tag{28}$$

has to be convex in q.[8] Furthermore, utility of the agent has to be exponential in payment, such that it can easily be expressed (in conjunction with normality) as its certainty equivalent. It remains open whether there are more general conditions in the sense of MLRC and CDFC for the validity of the first order approach in the multi-dimensional case.

3.2.3. The Solution

To simplify notation, we impose additional restrictions: First, the number of possible activities is limited to two, $\mu(q) = q$, and Σ is diagonal, hence all covariances are zero. Then, the first order conditions yield

$$\begin{pmatrix} \alpha_1^* \\ \alpha_2^* \end{pmatrix} = \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + r \begin{pmatrix} \partial^2 C/\partial q_1^2 & \partial^2 C/\partial q_1 \partial q_2 \\ \partial^2 C/\partial q_2 \partial q_1 & \partial^2 C/\partial q_2^2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right)^{-1} \begin{pmatrix} \partial Q/\partial q_1 \\ \partial Q/\partial q_2 \end{pmatrix}' . \tag{29}$$

The proof can be found in the appendix. Equation 29 is central to the following arguments. Recall the conjecture that the optimal contract does not necessarily have to condition on all observable quantities. This is possible for q_2 once $\sigma_2^2 \to \infty$, i.e. once the noise is too large to infer on q_2 at all. Then it follows that

$$\alpha_2^* = 0. \tag{30}$$

Again, in anticipation of supermodularity, α_i and σ_i^2 are in general not independent; in Holmstrom and Milgrom (1991), the variance-covariance matrix Σ is still exogenous, whereas Holmstrom and Milgrom (1994) endogenize it by introducing the possibility of monitoring, which decreases the variance in measurement. High powered incentives usually come together with high costs of reducing the variance (and therefore with a high variance), and vice versa. The size of incentives for the first task however depends on the size and sign of the cross derivatives of C. Assume $C(q) = c(q_1 + q_2)$, then it follows that $\partial C/\partial q_1 = \partial C/\partial q_2$. The optimal incentives for task one are then given by

$$\alpha_1^* = \frac{\partial Q/\partial q_1 - \partial Q/\partial q_2 \frac{\partial^2 C/\partial q_1 \partial q_2}{\partial^2 C/\partial q_2^2}}{1 + r\sigma_1^2 \left(\partial^2 C/\partial q_1^2 - \frac{(\partial^2 C/\partial q_1 \partial q_2)^2}{\partial^2 C/\partial q_2^2} \right)} \tag{31}$$

Obviously, the power of incentives for a specific action i (hence, the size of α_i) increases the more weight is put on that specific task, hence with $\partial Q/\partial q_i$. It decreases with $\partial Q/\partial q_j$, $j \neq i$, if the respective actions are substitutes in the cost function, and it increases with $\partial Q/\partial q_j$ if q_i and q_j are complements. If the two activities are complements in the agent's cost function (negative cross derivatives), the second activity can be induced by rewarding the first one. If

[8] See appendix for details.

the activities are substitutes (which is the case under $c(q_q+q_2)$, as given here), then incentives for the first task actually *reduce* the effort level in the second activity, since effort spent on it would decrease the possibility to get reward for effort spent on the first activity.

This can be proven by differentiating equation 25 with respect to q:

$$\begin{aligned}
\begin{pmatrix} \partial\alpha_1/\partial q_1 & \partial\alpha_1/\partial q_2 \\ \partial\alpha_2/\partial q_1 & \partial\alpha_2/\partial q_2 \end{pmatrix} &= \begin{pmatrix} \partial^2 C/\partial q_1^2 & \partial^2 C/\partial q_1\partial q_2 \\ \partial^2 C/\partial q_2\partial q_1 & \partial^2 C/\partial q_2^2 \end{pmatrix} && (32) \\
\Rightarrow \begin{pmatrix} \partial q_1/\partial\alpha_1 & \partial q_1/\partial\alpha_2 \\ \partial q_2/\partial\alpha_1 & \partial q_2/\partial\alpha_2 \end{pmatrix} &= \begin{pmatrix} \partial^2 C/\partial q_1^2 & \partial^2 C/\partial q_1\partial q_2 \\ \partial^2 C/\partial q_2\partial q_1 & \partial^2 C/\partial q_2^2 \end{pmatrix}^{-1} && (33) \\
&= \begin{pmatrix} \frac{C_{22}}{-C_{12}^2+C_{11}C_{22}} & -\frac{C_{12}}{-C_{12}^2+C_{11}C_{22}} \\ -\frac{C_{12}}{-C_{12}^2+C_{11}C_{22}} & \frac{C_{11}}{-C_{12}^2+C_{11}C_{22}} \end{pmatrix}, && (34)
\end{aligned}$$

where C_{ij} denote the respective second partial derivatives. The signs of all diagonal elements are unique by assumption of the convex cost function; assuming that tasks are substitutes, we find that all off-diagonal elements fulfill

$$\frac{\partial q_i}{\partial\alpha_j} < 0 \quad \forall i \neq j. \tag{35}$$

To sum up: in order to induce effort in a specific activity, the principal provides incentive payment for this activity, or, if this is not feasible due to infinitely large noise, he increases incentives for complementary tasks and reduces incentives for substitutes. In general, the power of incentive payments for a specific task depends positively on the weight the principal puts on this task, and negatively on the weight of other tasks. Moreover, α is determined by the degree of convexity of the cost function, the degree of complementarity of the specific task with all other tasks, it is negatively related to the variance of the type parameter regarding this specific task, and the covariance of the type with all other types. Formally, we have

$$\begin{aligned}
\frac{\partial\alpha_i}{\partial(\partial Q/\partial q_i)} &> 0 && (36) \\
\frac{\partial\alpha_i}{\partial\sigma_i^2} &\leq 0 && (37) \\
\frac{\partial\alpha_i}{\partial r} &\leq 0 && (38)
\end{aligned}$$

Equations 36 to 38 are only proven for $n = l = k = 2$. Since α depends on first and second derivatives of C and Q, respectively, the expression for the vector α become very complex, even for $n = 3$. Therefore, I can only conjecture that the main mechanisms are true for all higher dimensions as well. It might well be necessary to impose further restrictions upon the nature of the cost function and the variance-covariance matrix in order to maintain the results. The complexity can already be seen in the two-dimensional case, once all elements of Σ are non-zero and finite; the optimal solution is then given by

$$\begin{pmatrix} \alpha_1^* \\ \alpha_2^* \end{pmatrix} = \begin{pmatrix} \frac{Q_1(1+rC_{22}\sigma_2^2)-rC_{12}\sigma_{12}Q_2}{1+r(C_{11}\sigma_1^2(1+rC_{22}\sigma_2^2)-rC_{12}^2\sigma_{12}^2+C_{22}\sigma_2^2)} \\ \frac{Q_2(1+rC_{11}\sigma_1^2)-rC_{12}\sigma_{12}^2Q_1}{1+r(C_{11}\sigma_1^2(1+rC_{22}\sigma_2^2)-rC_{12}^2\sigma_{12}^2+C_{22}\sigma_2^2)} \end{pmatrix}. \tag{39}$$

The respective indices denote first and second partial derivatives. By assumption of a convex cost function (all C_{ii} positive), we have

$$\frac{\partial \alpha_1}{\partial \frac{\partial Q}{\partial q_1}} > 0 \tag{40}$$

$$\frac{\partial \alpha_1}{\partial \frac{\partial Q}{\partial q_1}} < 0 \quad \text{iff} \quad \text{sign}(C_{12}) = \text{sign}(\sigma_{12}) \tag{41}$$

$$\frac{\partial \alpha_1}{\partial \sigma_1^2} < 0 \tag{42}$$

$$\frac{\partial \alpha_1}{\partial \sigma_2^2} \gtrless 0. \tag{43}$$

The `sign` function gives 1 for weak positive values and 0 for negative values. We see that even here, the effect of the variance in the observation of the other activity is ambiguous; its sign depends on the convexity of the cost function and the degree of complementarity of actions in the cost function. The explicit solution for the optimal α_1 in the case of three possible actions is given in the appendix.

4. Conclusion

Allowing for more than one action, the optimal contract depends on more parameters than the one of the uni-dimensional case. In general, *all* first derivatives of the principal's utility function, all second derivatives of the cost function, as well as all elements of the variance-covariance matrix influence the power of incentives. The latter two are responsible for the fast increase of the complexity of the solution, as it includes square matrices, so that n possible actions yield incentives that depend on parameters in the order of magnitude n^2. Still, the model allows to find some explicit results even in the general case, independent of the number of activities.

Nevertheless, the assumptions under which the optimal contract is derived are quite narrow; one may find more general restrictions that still allow the application of the first order approach.

A. Convexity of Expression 28

We need

$$\left(\frac{\partial C}{\partial q}\right)' \Sigma \frac{\partial C}{\partial q} \tag{44}$$

to be convex in q. In the two-dimensional case, the above expression simplifies to

$$C_1^2\sigma_1^2 + 2C_1C_2\sigma_{12} + C_2^2\sigma_2^2. \tag{45}$$

The condition for expression 45 being convex in q is the positive semi-definiteness of its Hessian; under the assumption of all third derivatives of C being zero, the determinant of its Hessian is given by

$$4C_{11}(C_{12}^2 - C_{11}C_{22})(C_{12}\sigma_{12}^2 - C_{22}\sigma_1^2\sigma_2^2), \tag{46}$$

which is greater or equal than zero if and only if

$$C_{12}^2 \geq C_{11}C_{22} \quad \text{and} \quad \sigma_{12}^2 C_{12} \geq \sigma_1^2\sigma_2^2 C_{22} \tag{47}$$

$$\text{or} \quad C_{12} \leq C_{11}C_{22} \quad \text{and} \quad \sigma_{12}^2 C_{12} \leq \sigma_1^2\sigma_2^2 C_{22}. \tag{48}$$

B. Proof of Equation 29

Proof. The objective function is

$$Q - C(q) - \frac{1}{2} r\alpha'\Sigma\alpha, \tag{49}$$

which is constrained by incentive compatibility constraint, using the first order approach:

$$\alpha_i = \frac{\partial C}{\partial q_i} \tag{50}$$

Whether the participation constraint $U \geq \overline{U}$ is binding, remains to be checked. The first order condition of the optimization problem is:

$$\frac{\partial Q}{\partial q} - \frac{\partial C}{\partial q} - r\frac{\partial^2 C}{\partial q \partial q'}\Sigma\frac{\partial C}{\partial q} = 0 \tag{51}$$

Using the IC constraint again, we have

$$\frac{\partial Q}{\partial q} = \alpha\left(I + r\frac{\partial^2 C}{\partial q \partial q'}\Sigma\right) \tag{52}$$

$$\Leftrightarrow \quad \alpha^* = \left(I + r\frac{\partial^2 C}{\partial q \partial q'}\Sigma\right)^{-1}\frac{\partial Q}{\partial q} \tag{53}$$

□

C. Optimal Contract if $n = 3$

The advantage of the presented approach is the fact that under the given assumptions there exists a complex, but explicit solution.[9]

$$
\begin{aligned}
&(Q_3\ (-r\, C_{13}\, \sigma_3^2 - r^2\, C_{13}\, C_{22}\, \sigma_2^2\, \sigma_3^2 + r^2\, C_{12}\, C_{23}\, \sigma_2^2\, \sigma_3^2))\ /\\
&\quad (1 + r\, C_{11}\, \sigma_1^2 + r\, C_{22}\, \sigma_2^2 - r^2\, C_{12}^2\, \sigma_1^2\, \sigma_2^2 + r^2\, C_{11}\, C_{22}\, \sigma_1^2\, \sigma_2^2 + r\, C_{33}\, \sigma_3^2 - r^2\, C_{13}^2\, \sigma_1^2\, \sigma_3^2 +\\
&\qquad r^2\, C_{11}\, C_{33}\, \sigma_1^2\, \sigma_3^2 - r^2\, C_{23}^2\, \sigma_2^2\, \sigma_3^2 + r^2\, C_{22}\, C_{33}\, \sigma_2^2\, \sigma_3^2 - r^3\, C_{13}^2\, C_{22}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 +\\
&\qquad 2\, r^3\, C_{12}\, C_{13}\, C_{23}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 - r^3\, C_{11}\, C_{23}^2\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 - r^3\, C_{12}^2\, C_{33}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 + r^3\, C_{11}\, C_{22}\, C_{33}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2) +\\
&(Q_2\ (-r\, C_{12}\, \sigma_2^2 + r^2\, C_{13}\, C_{23}\, \sigma_2^2\, \sigma_3^2 - r^2\, C_{12}\, C_{33}\, \sigma_2^2\, \sigma_3^2))\ /\\
&\quad (1 + r\, C_{11}\, \sigma_1^2 + r\, C_{22}\, \sigma_2^2 - r^2\, C_{12}^2\, \sigma_1^2\, \sigma_2^2 + r^2\, C_{11}\, C_{22}\, \sigma_1^2\, \sigma_2^2 + r\, C_{33}\, \sigma_3^2 - r^2\, C_{13}^2\, \sigma_1^2\, \sigma_3^2 +\\
&\qquad r^2\, C_{11}\, C_{33}\, \sigma_1^2\, \sigma_3^2 - r^2\, C_{23}^2\, \sigma_2^2\, \sigma_3^2 + r^2\, C_{22}\, C_{33}\, \sigma_2^2\, \sigma_3^2 - r^3\, C_{13}^2\, C_{22}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 +\\
&\qquad 2\, r^3\, C_{12}\, C_{13}\, C_{23}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 - r^3\, C_{11}\, C_{23}^2\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 - r^3\, C_{12}^2\, C_{33}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 + r^3\, C_{11}\, C_{22}\, C_{33}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2) +\\
&(Q_1\ (1 + r\, C_{22}\, \sigma_2^2 + r\, C_{33}\, \sigma_3^2 - r^2\, C_{23}^2\, \sigma_2^2\, \sigma_3^2 + r^2\, C_{22}\, C_{33}\, \sigma_2^2\, \sigma_3^2))\ /\\
&\quad (1 + r\, C_{11}\, \sigma_1^2 + r\, C_{22}\, \sigma_2^2 - r^2\, C_{12}^2\, \sigma_1^2\, \sigma_2^2 + r^2\, C_{11}\, C_{22}\, \sigma_1^2\, \sigma_2^2 + r\, C_{33}\, \sigma_3^2 - r^2\, C_{13}^2\, \sigma_1^2\, \sigma_3^2 +\\
&\qquad r^2\, C_{11}\, C_{33}\, \sigma_1^2\, \sigma_3^2 - r^2\, C_{23}^2\, \sigma_2^2\, \sigma_3^2 + r^2\, C_{22}\, C_{33}\, \sigma_2^2\, \sigma_3^2 - r^3\, C_{13}^2\, C_{22}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 +\\
&\qquad 2\, r^3\, C_{12}\, C_{13}\, C_{23}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 - r^3\, C_{11}\, C_{23}^2\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 - r^3\, C_{12}^2\, C_{33}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2 + r^3\, C_{11}\, C_{22}\, C_{33}\, \sigma_1^2\, \sigma_2^2\, \sigma_3^2)
\end{aligned}
$$

Figure 1: α_1^* with zero covariances

$$
\begin{aligned}
&(Q_3\ (r^2\ (C_{12}\, \sigma_2^2 + C_{11}\, \sigma_{12} + C_{13}\, \sigma_{23})\ (C_{23}\, \sigma_3^2 + C_{12}\, \sigma_{13} + C_{22}\, \sigma_{23}) -\\
&\qquad r\ (C_{13}\, \sigma_3^2 + C_{11}\, \sigma_{13} + C_{12}\, \sigma_{23})\ (1 + r\ (C_{22}\, \sigma_2^2 + C_{12}\, \sigma_{12} + C_{23}\, \sigma_{23}))))\ /\\
&\quad (-r\ (C_{23}\, \sigma_2^2 + C_{13}\, \sigma_{12} + C_{33}\, \sigma_{23})\ (-r^2\ (C_{12}\, \sigma_1^2 + C_{22}\, \sigma_{12} + C_{23}\, \sigma_{13})\ (C_{13}\, \sigma_3^2 + C_{11}\, \sigma_{13} + C_{12}\, \sigma_{23}) +\\
&\qquad r\ (1 + r\ (C_{11}\, \sigma_1^2 + C_{12}\, \sigma_{12} + C_{13}\, \sigma_{13}))\ (C_{23}\, \sigma_3^2 + C_{12}\, \sigma_{13} + C_{22}\, \sigma_{23})) +\\
&\quad (1 + r\ (C_{33}\, \sigma_3^2 + C_{13}\, \sigma_{13} + C_{23}\, \sigma_{23}))\ (-r^2\ (C_{12}\, \sigma_1^2 + C_{22}\, \sigma_{12} + C_{23}\, \sigma_{13})\ (C_{12}\, \sigma_2^2 + C_{11}\, \sigma_{12} + C_{13}\, \sigma_{23}) +\\
&\qquad (1 + r\ (C_{11}\, \sigma_1^2 + C_{12}\, \sigma_{12} + C_{13}\, \sigma_{13}))\ (1 + r\ (C_{22}\, \sigma_2^2 + C_{12}\, \sigma_{12} + C_{23}\, \sigma_{23}))) +\\
&\quad r\ (C_{13}\, \sigma_1^2 + C_{23}\, \sigma_{12} + C_{33}\, \sigma_{13})\ (r^2\ (C_{12}\, \sigma_2^2 + C_{11}\, \sigma_{12} + C_{13}\, \sigma_{23})\ (C_{23}\, \sigma_3^2 + C_{12}\, \sigma_{13} + C_{22}\, \sigma_{23}) -\\
&\qquad r\ (C_{13}\, \sigma_3^2 + C_{11}\, \sigma_{13} + C_{12}\, \sigma_{23})\ (1 + r\ (C_{22}\, \sigma_2^2 + C_{12}\, \sigma_{12} + C_{23}\, \sigma_{23})))) +\\
&(Q_2\ (r^2\ (C_{13}\, \sigma_3^2 + C_{11}\, \sigma_{13} + C_{12}\, \sigma_{23})\ (C_{23}\, \sigma_2^2 + C_{13}\, \sigma_{12} + C_{33}\, \sigma_{23}) -\\
&\qquad r\ (C_{12}\, \sigma_2^2 + C_{11}\, \sigma_{12} + C_{13}\, \sigma_{23})\ (1 + r\ (C_{33}\, \sigma_3^2 + C_{13}\, \sigma_{13} + C_{23}\, \sigma_{23}))))\ /\\
&\quad (-r\ (C_{23}\, \sigma_2^2 + C_{13}\, \sigma_{12} + C_{33}\, \sigma_{23})\ (-r^2\ (C_{12}\, \sigma_1^2 + C_{22}\, \sigma_{12} + C_{23}\, \sigma_{13})\ (C_{13}\, \sigma_3^2 + C_{11}\, \sigma_{13} + C_{12}\, \sigma_{23}) +\\
&\qquad r\ (1 + r\ (C_{11}\, \sigma_1^2 + C_{12}\, \sigma_{12} + C_{13}\, \sigma_{13}))\ (C_{23}\, \sigma_3^2 + C_{12}\, \sigma_{13} + C_{22}\, \sigma_{23})) +\\
&\quad (1 + r\ (C_{33}\, \sigma_3^2 + C_{13}\, \sigma_{13} + C_{23}\, \sigma_{23}))\ (-r^2\ (C_{12}\, \sigma_1^2 + C_{22}\, \sigma_{12} + C_{23}\, \sigma_{13})\ (C_{12}\, \sigma_2^2 + C_{11}\, \sigma_{12} + C_{13}\, \sigma_{23}) +\\
&\qquad (1 + r\ (C_{11}\, \sigma_1^2 + C_{12}\, \sigma_{12} + C_{13}\, \sigma_{13}))\ (1 + r\ (C_{22}\, \sigma_2^2 + C_{12}\, \sigma_{12} + C_{23}\, \sigma_{23}))) +\\
&\quad r\ (C_{13}\, \sigma_1^2 + C_{23}\, \sigma_{12} + C_{33}\, \sigma_{13})\ (r^2\ (C_{12}\, \sigma_2^2 + C_{11}\, \sigma_{12} + C_{13}\, \sigma_{23})\ (C_{23}\, \sigma_3^2 + C_{12}\, \sigma_{13} + C_{22}\, \sigma_{23}) -\\
&\qquad r\ (C_{13}\, \sigma_3^2 + C_{11}\, \sigma_{13} + C_{12}\, \sigma_{23})\ (1 + r\ (C_{22}\, \sigma_2^2 + C_{12}\, \sigma_{12} + C_{23}\, \sigma_{23})))) +\\
&(Q_1\ (-r^2\ (C_{23}\, \sigma_3^2 + C_{12}\, \sigma_{13} + C_{22}\, \sigma_{23})\ (C_{23}\, \sigma_2^2 + C_{13}\, \sigma_{12} + C_{33}\, \sigma_{23}) +\\
&\qquad (1 + r\ (C_{22}\, \sigma_2^2 + C_{12}\, \sigma_{12} + C_{23}\, \sigma_{23}))\ (1 + r\ (C_{33}\, \sigma_3^2 + C_{13}\, \sigma_{13} + C_{23}\, \sigma_{23}))))\ /\\
&\quad (-r\ (C_{23}\, \sigma_2^2 + C_{13}\, \sigma_{12} + C_{33}\, \sigma_{23})\ (-r^2\ (C_{12}\, \sigma_1^2 + C_{22}\, \sigma_{12} + C_{23}\, \sigma_{13})\ (C_{13}\, \sigma_3^2 + C_{11}\, \sigma_{13} + C_{12}\, \sigma_{23}) +\\
&\qquad r\ (1 + r\ (C_{11}\, \sigma_1^2 + C_{12}\, \sigma_{12} + C_{13}\, \sigma_{13}))\ (C_{23}\, \sigma_3^2 + C_{12}\, \sigma_{13} + C_{22}\, \sigma_{23})) +\\
&\quad (1 + r\ (C_{33}\, \sigma_3^2 + C_{13}\, \sigma_{13} + C_{23}\, \sigma_{23}))\ (-r^2\ (C_{12}\, \sigma_1^2 + C_{22}\, \sigma_{12} + C_{23}\, \sigma_{13})\ (C_{12}\, \sigma_2^2 + C_{11}\, \sigma_{12} + C_{13}\, \sigma_{23}) +\\
&\qquad (1 + r\ (C_{11}\, \sigma_1^2 + C_{12}\, \sigma_{12} + C_{13}\, \sigma_{13}))\ (1 + r\ (C_{22}\, \sigma_2^2 + C_{12}\, \sigma_{12} + C_{23}\, \sigma_{23}))) +\\
&\quad r\ (C_{13}\, \sigma_1^2 + C_{23}\, \sigma_{12} + C_{33}\, \sigma_{13})\ (r^2\ (C_{12}\, \sigma_2^2 + C_{11}\, \sigma_{12} + C_{13}\, \sigma_{23})\ (C_{23}\, \sigma_3^2 + C_{12}\, \sigma_{13} + C_{22}\, \sigma_{23}) -\\
&\qquad r\ (C_{13}\, \sigma_3^2 + C_{11}\, \sigma_{13} + C_{12}\, \sigma_{23})\ (1 + r\ (C_{22}\, \sigma_2^2 + C_{12}\, \sigma_{12} + C_{23}\, \sigma_{23}))))
\end{aligned}
$$

Figure 2: α_1^* with non-zero and finite covariances

[9] The *Mathematica* notebook is available upon request.

References

Holmstrom, B., and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55(2), 303–328.

——— (1991): "Multitask principle-agent analyses: Incentive contracts, asset ownership and job design," *Journal of Law, Economics and Organization*, 7, 24–51.

——— (1994): "The Firm as an Incentive System," *American Economic Review*, 84(4), 972–991.

Laffont, A.-J., and D. Martimort (2002): *The Theory of Incentives.* Princeton University Press, Princeton.

Salanié, B. (1997): *The Economics of Contracts - A Primer.* MIT Press, Cambridge, Massachusetts.