Procurement with Adverse Selection and Dynamic Moral Hazard

Terence R. Johnson*
University of Notre Dame

First draft: 3/2012
This draft: 12/2012

Abstract
Consider a principal seeking to contract with one of many firms to undertake a risky project in an environment that suffers from both adverse selection and dynamic moral hazard. The likelihood of project success at each moment depends on the amount of work already completed, which is observed only by the contracted firm. Firms can siphon a portion of the funds intended for the project and use the rest to create an illusion of productivity. Consequently, inefficient firms can bid for contracts simply to siphon, efficient firms who win may shirk, and firms who succeed can siphon funds rather than report success to the principal. I show that under standard assumptions, dynamic contracts can be constructed that mitigate all inefficiency arising from dynamic moral hazard. These contracts can be implemented through a generalization of a procurement auction that features endogenous penalties for failure and time-varying awards for success.

JEL Classification Codes: D44, C73, D86
Keywords: Auction Theory, Dynamic Mechanism Design, Dynamic Moral Hazard, Costly State Falsification

1 Introduction

In many economic settings, risk evolves over time in ways that dynamically alter the incentives of the participants. Even in relatively mundane procurement settings like highway construction, there are jobs that become unexpectedly complicated, costly, and sometimes end in default. These failures can not only be costly to the principal but also to the firms involved, harming their reputations and financial solvency.

A prominent example of the drawbacks of traditional approaches to procurement is the Big Dig, a massive highway construction project in Boston, Mass. The project was planned to be completed in 1998 at a cost of $6 billion. It was ultimately finished in 2007 at a cost of $14.6 billion, and a number of lawsuits are currently pending against contractors involved in the project. The process by which the project was procured is generally called the design-bid-build method, where

---

*I am grateful to Thomas Gresik and participants at the Spring 2012 Midwest Microeconomic Theory Conference for helpful comments. 917 Flanner Hall, University of Notre Dame. tjohns20@nd.edu.
the principal specifies the design details and issues a request for proposals, firms bid competitively, and a contract is awarded to the lowest responsive bidder. Indeed, this was the method used for the Big Dig: “When the preliminary engineering on the Big Dig started in the 1980s, the Massachusetts Highway Department had strict bidding requirements governed by Massachusetts law,” says Carl Gottschall, project administrator at FHWA’s Massachusetts Division Office [2]. “Design-bid-build was our only option for project delivery.” Gottschall goes on to say, “Building such infrastructure within a dense urban core ... would have made it almost impossible to pin down a price up front.” Attempting to negotiate a single price before work begins, however, is a limitation that contributes to the massive cost overruns and defaults that can accompany risky projects. This paper shows that by incorporating dynamic elements into the economic environment, time-varying contracts can be structured that mitigate misconduct on the side of the contractors and give rise to novel market designs.

Consider an economic environment in which a principal would like to hire one of a number of firms to undertake a project, but the firms’ costs per unit time are privately known only to them. Initially, a “market for contracts” occurs in which the project is allocated to at most one firm, after which the firm dynamically maximizes its payoff from the project. The likelihood of the project’s success is random and commonly known, and depends on the accumulated stock of effort exerted by a winning firm. The principal cannot observe whether the agent is exerting effort or whether the project is complete, but can verify ex post whether success has occurred. Despite having capital for the purposes of collateral, the firms are not sufficiently liquid to fund work on their own without “cutting corners”, requiring the principal to finance the project. This is modelled by imposing a non-pecuniary efficiency penalty on firms who exert effort when the funding is less than their true marginal cost, and can be interpreted as, for example, the expected cost of an accident when insufficient safety measures are employed in the work, or the added difficulties imposed by using equipment or techniques that are ill-suited to the work. The principal can set a terminal date at which the relationship is terminated, pay the firm a potentially time-varying flow payment, punish the firm for failing to succeed by the terminal date through forfeit of a bond, and pay a time-varying award to the firm if success occurs.

Working against the principal, however, is the agent’s ability to siphon funds: Rather than expend effort on work that will lead towards success, the agent can divert funds to private con-
sumption and use the rest to give the illusion of productivity. This monitoring structure creates three separate challenges. First, firms may attempt to win the contract even though they have no intention of exerting any effort, and are merely planning to siphon all the funds. Second, a firm that has worked and succeeded may then begin siphoning funds, waiting to exercise the option of revealing success at a later, more lucrative date. Finally, late in its contract, a firm might cease exerting effort and begin siphoning funds, since the likelihood of success fails to justify further effort. For example, a construction firm might succeed at the crux of a large scale project, then delay completion of less demanding tasks over time to stretch out the payments from the principal. Once the firm comes forward to claim success, however, bridge inspectors and civil engineers can judge whether the final result is complete. Or, the firm might work for a time, realize the project is impossible, and siphon the remaining stream of flow payments rather than admit the bad news to the principal and default immediately.

Under the standard assumption of a log concave distribution of types, Section 3 shows that an optimal dynamic direct revelation mechanism mitigates all these sources of inefficiency, allowing the principal to implement the same effort by the firm he would select if he could directly observe the project’s progress. In equilibrium, firms who overstate their true costs exert effort at all dates, while firms who understate their true costs stop work early and siphon the remaining funds. To deter such deviations, the optimal mechanism adopts a “cost-plus” approach, in which the agent is paid a fee to cover expenses plus a bonus upon success. Such contracts provide the best motivation for firms to keep working late in a contractual agreement by pushing all of the informational rents onto the award, thereby providing positive incentives for effort and immediate revelation of success. To discourage shirking, a bond is selected that punishes the firm for failure, thereby keeping it “on the hook” late in the contractual horizon. It is not enough, however, to set the bond at the maximum amount siphonable from the project, since this ignores incentive constraints across types. One contribution of the paper is to show how the bond plays a non-trivial role not only in preventing spurious bidding, but also in determining incentive compatibility and ensuring that the proper level of effort is exerted.

While the paper focuses on the simple cases of an increasing and decreasing hazard rate of success, the framework can be made significantly more general than it might appear at first by allowing the hazard rate to take more general forms. For example, a project may have a zero
probability of success for some time after which the hazard rate increases, and then decreases as the long delay reveals that the project is actually unlikely to succeed at all. Many settings for which a stochastic framework might sound unnatural — building a bridge in fairly straightforward ex ante conditions, for example — are actually very appropriate, given that unexpected delays might reveal that the difficulty of the project was initially underestimated.

An indirect implementation is proposed in Section 4 that has a number of interesting features. Firms bid in terms of their cost per unit time, the lowest bidder satisfying a reserve price requirement is awarded the project, and is paid its stated bid for each moment of time over the course of the contract. The award function is composed of three terms: A bonus for efficiency, a “side-bet” concerning the loss of the bond, and the portion of the expected cost that the firm could siphon. The bonus for efficiency depends on the second-highest bid, rewarding the winning firm for the cost savings it generates relative to the next-highest firm eligible to win, as well as the longer time horizon that it brings to the project. Upon taking expectations, this becomes the informational rent. The “side-bet” is the odds ratio of success multiplied by the bond, which rewards the firm for taking on the risk of the project and ensures effort late in the contractual horizon. The final term depends on the firm’s ability to siphon, and compensates it for revealing success immediately to the principal. The bonus for efficiency bears similarity to the Vickrey-Clarke-Groves mechanism, but the dynamic procurement auction developed here maximizes the principal’s payoff and is not constructed to internalize each agent’s influence on social welfare. In addition, the bond and award functions are determined endogenously by the bidding, which reduces the sensitivity of the mechanism to the principal’s beliefs about the distribution of private information as well as provides a useful interpretation of the optimal direct revelation mechanism.

The possibility of a firm failing to succeed and thereby forfeiting a bond is related to the literature on firm defaults at auctions, as in Board [4], Calveras et al. [6], Zheng [35], and Waehrer [34]. These papers assume that not all of the information is available at an initial date when contracting takes place, and the resolution of some residual uncertainty makes the initial agreement untenable. The current paper departs from such a framework by imagining that even if the agent worked in good faith, the project might take an arbitrarily long amount of time to finish. In the presence of moral hazard, this means that the principal cannot commit to perpetual funding, or winning firms would have an incentive to siphon forever. The current paper contributes to the literature by providing an
environment in which success is fundamentally uncertain, and future re-contracting is undesirable because it would undermine the principal’s ability to deter spurious bidding and wasteful shirking. In this sense, defaults are not something that could be avoided through more careful contracting, but necessary evils that result from the informational features of the environment.

The closest literature considers how to design markets for agency contracts, including McAfee and McMillan [21] and [22], and Laffont and Tirole [16]. These papers show how the principal can trade off moral hazard against adverse selection to improve his payoff. In the current paper, however, it is shown that inefficiency arising from moral hazard can be entirely mitigated. This is possible because there is no limit to the amount of collateral or size of the bond that can be demanded of the firm, so that efficient effort provision can be made to maximize the expected payoff of the firm as well as the principal. In the literature that focuses on moral hazard alone, models like Sannikov [31] have a much richer dynamic principal-agent problem, but do not consider the market structure that gives rise to the contractual setting in which the principal and agent bargain. By incorporating the market for contracts, the current paper provides a tractable framework for investigating dynamic principal-agent problems with an adverse selection component.

One of the key features of the model in the current paper is that a winning firm can engage in dynamic costly state falsification, where it sacrifices a portion of the payments received from the principal to given an impression of effort. This captures a dynamic where completely shutting down operations would be too brazen, but the firm can “keep the lights on” at work to hide its shirking and siphoning. This is similar to Lacker and Weingberg [15], Crocker and Morgan [12], and other works that allow an agent to expend effort to keep the principal from learning that they have engaged in inefficient or wasteful behavior. While these papers adopt a static moral hazard framework in which the principal gets a noisy signal that is correlated with the agent’s effort, the signal of effort in the current paper is whether or not the agent has come forward to claim success or defaulted. Another difference is that on the equilibrium path, no siphoning or costly state falsification ever occurs, because the penalty for failure can be made sufficiently severe that the firm wants to work at all dates.

Within the growing field of dynamic mechanism design, the current paper is most similar to Board [5]. In that paper, the author considers the problem of auctioning an option contract to agents who have a privately known benefit of exercising, but a time-varying expected cost. Board
shows that the welfare-maximizing mechanism does not feature payments conditional on exercise, but the revenue-maximizing mechanism does, leading to delays in exercise. The current paper adds a dynamic moral hazard problem, so that agents are not only choosing a stopping time, but also how much effort to invest in the project’s success. Doepke and Townsend [14] consider a discrete-time framework with dynamic adverse selection and moral hazard, closest in spirit to the optimal taxation or optimal unemployment benefits literatures. They provide a revelation principle in which allocations are implemented by presenting the agents with a “promised utility” that represents the discounted benefit of each actions, and then show how optimal contracts can be computed efficiently. The current paper adds a market for contracts, and involves a single project that either succeeds or fails, rather than an infinitely-lived relationship. However, the firms in the current paper draw only a single type at the beginning of the game, so there is no dynamic adverse selection element, while Doepke and Townsend allow types to evolve according to very general rules. Finally, Pavan et al [?] study a very general dynamic mechanism design framework, exploiting what they call impulse response functions to study how an agent’s current behavior influences the payoffs of future types. The firms’ types in the current paper are static, but the simpler framework allows a study of indirect implementation and the kinds of markets for contracts that exploit competition among the agents to implement the profit-maximizing outcome. In contrast to Board [5], Doepke and Townsend [14], and Pavan et al [?], the current paper also exploits a continuous time framework, rather than discrete time. This allows an elegant characterization of an agent’s optimal dynamic behavior using the adjoint equation, rather than systems of inequalities. Implementing a particular allocation then hinges on studying the properties of a single function characterized by an ordinary differential equation and an economically meaningful transversality condition.

A large literature considers procurement auctions, both theoretically and empirically. Dasgupta and Spulber [13] study a procurement model similar to Myerson [27], and show that in the presence of strictly convex costs for the firms there can be dual- and multi-sourcing. The contribution of the current paper is to incorporate a dynamic moral hazard problem, although dual- or multi-sourcing are never optimal in equilibrium here, since once a firm fails, there is no benefit to transferring the project to a less efficient firm who would have received a shorter time horizon in the first place. Manelli and Vincent [17] show that in a procurement setting where sellers privately know the value of their good to the buyer, making a series of take-it-or-leave-it offers to potential sellers
can dominate auctions. This occurs because of the dependence of the buyer’s value for the good on the seller’s private information, which is intrinsically different from a situation in which the sellers all provide the same good but have different, privately known costs of providing it. Since the principal in the current paper receives the value $v$ whenever a fully funded firm completes the project and any winning firm is fully funded, this phenomenon does not appear. In a paper with theoretical and empirical components, Bajari and Lewis [3] look at recent changes in Caltrans’ procurement auctions that incorporate time-varying elements into the contracts, and find evidence that incentivizing firms along a time dimension leads to significant welfare gains. The current paper provides a useful example of explicitly incorporating the time horizon into the design of the mechanism, which might be adopted to other questions and sets of assumptions. In addition, the static projects considered in the literature are similar to the dynamic outcome in the current paper when the hazard rate of success is monotone increasing: In this case, all winning firms are contracted to work until they succeed, and the optimal contracts become substantially simpler. This nests the regular procurement auction as a special case of the current model in which the longer the firm works, the more likely it is to succeed.

Lastly, the current paper assumes that firms can only work sequentially, but a large literature considers optimal contests and tournaments. Particularly, Che and Gale [8] and Moldovanu and Sela [26] consider environments where prizes are offered for success, and the principal must decide on the optimal contest architecture. The results of the current paper would change if firms could work in parallel, since competition could be used to deter shirking behavior. However, in many environments firms cannot feasibly work in parallel; for example, only one construction site typically exists. Similarly, many markets feature the kind of contracting here, where a principal issues a request-for-proposals but does not award the project to multiple firms and force them to compete for a prize.

1.1 Outline

Section 2 formally describes the model. Section 3 shows how a firm’s payoff can be computed in this continuous time, dynamic environment, which allows incentive compatibility and individual rationality constraints to be clearly defined. Section 4 then constructs a profit-maximizing mechanism. This process is complicated by the fact that incentive compatibility and individual rationality
constraints must be considered not only for honest reports and the subsequent optimal effort provision schedule, but also for out-of-equilibrium deviations. Section 5 then proposes the dynamic procurement auction and shows that it implements the same outcome as the principal-optimal direct revelation mechanism. Section 6 concludes.

2 Model

A principal would like to hire a firm to undertake a risky project, but the market suffers from both adverse selection and dynamic moral hazard. There are \( i = 1, 2, \ldots, n \) firms, who each have a privately known marginal cost per unit time, \( c_i \). Let \( c = (c_i, c_{-i}) \) be the vector of firm marginal costs, with \( c_{-i} = (c_1, c_2, \ldots, c_{i-1}, c_{i+1}, \ldots, c_n) \). Each \( c_i \) is drawn independently from a commonly known probability distribution function \( F(c_i) \) with a strictly positive, continuous density \( f(c_i) \) on \( [\underline{c}, \bar{c}] \), where \( \underline{c} > 0 \) and \( \bar{c} \) is finite. A completed project yields value \( v \) to the principal, and here is an observable fixed cost of attempting the project, \( K \), which is borne by the principal without loss of generality\(^1\). The principal and firms are perfectly patient, maximize their expected utility, and are risk neutral.

Time is the union of an allocation stage, during which the market-for-contracts meets, and a construction stage, \( [0, \bar{T}] \), during which a winning firm solves its project management problem. At each moment during the construction stage, a hired firm either chooses to work and produce a unit of effort \( \epsilon_{it} = 1 \), or shirk and produce no progress, \( \epsilon_{it} = 0 \). The stock of accumulated effort is given by

\[
E_{it} = \int_0^t \epsilon_{iz} dz
\]

For any time-subscripted variable \( x_t, \dot{x}_t \) denotes the partial derivative of \( x_t \) with respect to \( t \), so that

\[
\dot{E}_{it} = \epsilon_{it}
\]

The probability that a firm succeeds given accumulated effort \( E_{it} \) is \( G(E_{it}) \), a differentiable distribution function with density \( g(E_{it}) \). The hazard rate of success is given by

\[
h(E_{it}) = \frac{g(E_{it})}{1 - G(E_{it})}
\]

\(^1\)See the discussion following Proposition 4.1
In the main analysis, the hazard rate is assumed to be decreasing and have strictly positive support on \([0, \bar{T}]\), but the less interesting case in which the hazard rate is strictly increasing can be solved using the same methods, and the differences are mentioned in the text.

During the allocation stage, the principal announces a contract, giving the probability that each firm is selected to attempt the project, a terminal date \(T_i\) at which time a selected firm enters into default and forfeits a bond \(B_i\), a flow payment \(r_{it}\) to the firm at time \(t\), and an award function \(w_{it}\) specifying a payment to the firm for revealing success at time \(t\) to the principal. The firms can each accept the terms of the contract and participate, or reject and take a payoff of zero.

The firms face unlimited liability, so that the bond can be set to any amount. However, firms do not have funds to cover the variable costs of the project themselves, and working without sufficient funding imposes additional, non-pecuniary costs. In particular, if a firm receives a flow payment \(r_{it} < c_i\), it suffers an efficiency penalty: In addition to a pecuniary cost of \(c_i\), it forfeits non-pecuniary effort costs equal to \(\gamma(r_{it}, c_i)\), which is strictly positive and decreasing when \(r_{it} < c_i\), zero when \(r_{it} \geq c_i\), convex, and satisfies
\[
\lim_{r_{it} \to c_i^-} \frac{\partial \gamma(r_{it}, c_i)}{\partial r_{it}} = 0
\]
For example, the function \(\gamma(r_{it}, c_i) = (\min\{r_{it} - c_i, 0\})^2\) has these properties.

The principal cannot observe if the firm is working or not, so the firm can divert funds away from the project by engaging in costly behavior that gives the appearance of productive effort, but does not lead to a higher likelihood of completion. In particular, a firm can siphon funds at a rate \(0 \leq \lambda \leq 1\), keeping that fraction of any payments received in cash for itself and setting \(\epsilon_{it} = 0\).

From the principal’s perspective, activity appears to be taking place, but this is just to hide the agent’s efforts to extract money from the contract.

2.1 Direct revelation mechanisms

For the firms to decide whether reporting honestly at the allocation stage is profit-maximizing, they must assess the dynamically evolving value of the project. Since they have private information not only about their initial type, but also about the accumulated level of effort and whether or not the project has succeeded, the incentive constraints are not limited to honesty in initial reporting, but also obedience in the undertaking of the project at every point in time.
Formally, a direct mechanism is a set of functions

\[ m(c) = \{ P_i(c), r_{it}(c), w_{it}(c), B_i(c), T_i(c), s_{it}(c), \epsilon_{it}(c) \}_{i,t} \]

whose domains are the type spaces of the agents, where

1. The assignment function \( P_i(c_i, c_{-i}) \), specifies the probability of assigning the project to firm \( i \)
2. The flow payment, \( r_{it}(c_i, c_{-i}) \), specifies the payment at each date \( t \)
3. The award, \( w_{it}(c_i, c_{-i}) \), specifies the payment to a firm who reports success at time \( t \)
4. The (surety) bond, \( B_i(c_i, c_{-i}) \), specifies a penalty in case of failure
5. The terminal date, \( T_i(c_i, c_{-i}) \), specifies when flow payments stop and an award is no longer offered
6. The delay function, \( s_{it}(c'_i, c_i, c_{-i}) \), specifies the amount of time a firm should delay before revealing success to the principal, given that success occurred at time \( t \), and the agent’s initially reported type is \( c'_i \)
7. The effort function, \( \epsilon_{it}(c'_i, c_i, c_{-i}) \), specifies the effort level the firm for each date \( t \) when the agent’s initial report is \( c'_i \)

The functions \( w_{it}(c), r_{it}(c), T_i(c), \) and \( B_i(c) \) are all assumed to be piecewise continuously differentiable.

This specification of a mechanism is without loss of generality because these functions cover all possible scenarios the principal might witness. In particular, each firm provides at most two pieces of information: The initial report of its type and the date at which it succeeds (and no report at all if it fails). Consequently, the set of functions given above spans all possible decisions and payments contingent on observable events.

Definition 2.1 A direct mechanism is incentive compatible if (i) each firm finds it optimal to report its marginal cost per unit time honestly at the allocation stage, (ii) the proposed effort functions \( \epsilon_{it}(c'_i, c_i, c_{-i}) \) are optimal for all dates \( t \) for a firm with true type \( c_i \) reported its type as \( c'_i \), and (iii) the proposed delay \( s_{it}(c'_i, c_i, c_{-i}) \) is optimal whenever success occurs at date \( t \) and a firm with
true type $c_i$ reported its type as $c'_i$. A direct mechanism in which agents have incentives to report truthfully and behave obediently is a direct revelation mechanism. A direct revelation mechanism is individually rational if a firm that participates honestly at the allocation stage and then adopts the recommended effort and reporting strategies at each subsequent date gets an expected payoff of at least zero, its outside option.

In short, a mechanism is incentive compatible if it is a Bayesian Nash equilibrium for the firms to be honest about their costs during the allocation stage, and obedient concerning the proposed effort and delay for all type-report pairs and success dates during the construction phase. Note that for a firm to decide whether or not to report its type honestly at the allocation stage, it must contemplate how it will behave once it has won the project. The constraints in (ii) and (iii) account for this by requiring the mechanism to provide payoff-maximizing effort and delay plans for all type-report pairs $(c_i, c'_i)$, even those that will never occur on the equilibrium path\(^2\). By an extension of the revelation principle due to Doepke and Townsend [14] that accommodates hidden information and hidden actions, the principal can restrict attention to direct revelation mechanisms that induce honesty at the allocation stage and obedience at the construction phase without loss of generality.

### 3 Project Value and Incentive Compatibility

A contractual arrangement in many procurement design models is typically static, or involves a small number of periods across time. This section shows how to dynamically value a contract where the likelihood of success changes across the project horizon, thereby characterizing the agent’s payoffs in the construction phase. A small amount of discounting of the form $e^{-\rho t}$ is assumed, and the payoffs are then derived as the time preference parameter $\rho$ goes to zero\(^3\).

After winning the project, a firm’s behavior is characterized by two functions: The effort that the firm exerts at each date, $\epsilon_{it}(c'_i, c_i, c_{-i})$, and the delay between success and reporting this to the principal, $s_{it}(c'_i, c_i, c_{-i})$, where $c'_i$ is the agent’s report, $c_i$ is the agent’s true type, and $c_{-i}$ are the reports of the other firms. Since the winning firm receives a flow of payments over time, it may

\(^2\)While the accumulated effort $E_{it}(c'_i, c)$ also becomes private information as a consequence of dynamic hidden actions once the contract is assigned, by the principle of optimality, it is never optimal for a firm of type $c_i$ reporting $c'_i$ to deviate from the plan proposed to it by the mechanism immediately following the market for contracts, so we need not include obedience constraints at all states $(c'_i, c_i, c_{-i}, E_{it})$.

\(^3\)All results in the paper can be understood as holding in a model with discounting for $\rho$ sufficiently close to zero. Allowing arbitrary discounting complicates the analysis since sufficiently impatient firms ($\rho \to \infty$) will always shirk.
be to his advantage to wait to exercise the option of revealing success to the principal. Define the 
**optimal delay given an initial report type** $c_i'$ and **success at time** $t$, $s_{it}(c_i', c_i, c_{-i})$, as the solution to

$$
\tilde{w}_{it}(c_i', c_{-i}) = \max_{s \leq T_i(c_i')} e^{-\rho(s-t)}w_{is}(c_i', c_{-i}) + \lambda \int_t^s e^{-\rho(z-t)}r_{iz}(c_i', c_{-i})dz
$$

Let $\tilde{w}_{it}(c_i', c_{-i})$ be the **effective award**. Note that the true $c_i$ appears nowhere in the maximization problem, so the optimal delay depends only on the firm’s reported type and the terms of the contract, not on the firm’s true costs. A sufficient condition for the solution to be well-defined is that the functions $r_{it}(c_i', c_{-i})$ and $w_{it}(c_i', c_{-i})$ be continuous, which is assumed. Since the objective function is continuous in the delay, by the theorem of the maximum the value function is continuous, so the effective award $\tilde{w}_{it}(c_i', c_{-i})$ is a continuous function.

Given the effective award, the value of a given effort plan can be characterized. The expected value to the agent of exerting effort for a moment of time $\Delta$ is given by

$$
J_i(t, E_{it}(c_i', c), c_i', c) = h(E_{it}(c_i', c))\Delta e^{-\rho\Delta}\tilde{w}_{it}(c_i', c_{-i}) + \Delta(r_{it}(c_i', c_{-i}) - c_i - \gamma(r_{it}(c_i', c_{-i}), c_i))
$$

$$
+ (1 - h(E_{it}(c_i', c))\Delta) e^{-\rho\Delta}J_i(t + \Delta, E_{it}(c_i', c) + \Delta, c_i', c)
$$

The first term is the current probability of success times the effective award, the second term is the net flow profits, and the third term is the current probability of failure times the discounted value of reaching time $t + \Delta$ with a stock of accumulated effort of $E_{it}(c_i', c) + \Delta$. This expresses the value of exerting effort at time $t$ as a discounted expected value. Similarly, the payoff to exerting no effort is

$$
J_i(t, E_{it}(c_i', c), c_i', c) = e^{-\rho\Delta}J_i(t + \Delta, E_{it}(c_i', c), c_i', c) + \Delta \lambda r_{it}(c_i', c_{-i})
$$

so that the firm siphons the current flow payment at rate $\lambda$ and enters the next moment $t + \Delta$ without accumulating any additional effort.

By standard limiting arguments (see the proof of Proposition 3.1), these can be converted into differential equations that describe the evolution of the value of the project:

$$
\frac{dJ_i(t, E_{it}(c_i', c), c_i', c)}{dt} = (\rho + h(E_{it}(c_i', c)))J_i(t, E_{it}(c_i', c), c_i', c)
$$

$$
- (h(E_{it}(c_i', c))\tilde{w}_{it}(c_i', c_{-i}) + r_{it}(c_i', c_{-i}) - c_i - \gamma(r_{it}(c_i', c_{-i}), c_i))
$$
and
\[
\frac{dJ_i(t, E_{it}(c'_i, c), c'_i, c))}{dt} = \rho J(t, E_{it}(c'_i, c), c'_i, c) - \lambda r_{it}(c'_i, c_{-i})
\]

In the differential equation corresponding to exerting effort, the change in the current value of the project is equal to the discounted expected value, less a term that reflects the lost opportunity for success at time \(t\). In the differential equation corresponding to shirking, the firm has not accumulated any additional effort, so the “opportunity to succeed” at \(E_{it}(c'_i, c)\) is still available, but some time has been lost and some value siphoned out of the contract.

Since the differential equations characterize the evolution of the value of the contract, a value function for any candidate policy \(\epsilon_{it}(c'_i, c)\) can be constructed by working backwards from the terminal value, \(-(1 - G(E_{iT_i(c'_i, c_{-i})}(c'_i, c)))B_i(c'_i, c_{-i})\).

**Proposition 3.1** Any winning firm solves the linear optimal control problem

\[
\max_{\epsilon_{it}(c'_i, c)} \mathcal{J}_i(c'_i, c, \epsilon_{it}(c'_i, c)) = \max_{\epsilon_{it}(c'_i, c)} \int_0^{T_i(c'_i, c_{-i})} \left\{ g(E_{it}(c'_i, c))\tilde{w}_{it}(c'_i, c_{-i}) + (1 - G(E_{it}(c'_i, c))(r_{it}(c'_i, c_{-i}) - c_i - \gamma(r_{it}(c'_i, c_{-i}), c_i)) \right\}
\]
\[
+ (1 - \epsilon_{it}(c'_i, c)) (1 - G(E_{it}(c'_i, c))) \left\{ \lambda r_{it}(c'_i, c_{-i}) \right\} \right\} dt - (1 - G(E_{iT_i(c'_i, c_{-i})}(c'_i, c_{-i})))B_i(c'_i, c_{-i})
\]

subject to \(\epsilon_{it}(c'_i, c) \in \{0, 1\}\) and \(\dot{E}_{it}(c'_i, c) = \epsilon_{it}(c'_i, c)\). The functions

\[
y_{it}^1(c'_i, c, E_{it}(c'_i, c)) = g(E_{it}(c'_i, c))\tilde{w}_{it}(c'_i, c) + (1 - G(E_{it}(c'_i, c))(r_{it}(c'_i, c_{-i}) - c_i - \gamma(r_{it}(c'_i, c_{-i}), c_i)
\]

and

\[
y_{it}^0(c'_i, c, E_{it}(c'_i, c)) = \lambda r_{it}(c'_i, c_{-i})
\]

give the current flow value of exerting effort and siphoning, respectively. The payoff-maximizing effort plan \(\epsilon_{it}(c'_i, c)\) necessarily satisfies

\[
\epsilon_{it}(c'_i, c) = \begin{cases} 
1 & \text{if } y_{it}^1(c'_i, c, E_{it}(c'_i, c)) - y_{it}^0(c'_i, c, E_{it}(c'_i, c)) + \mu_{it}(c'_i, c) \geq 0 \\
0 & \text{otherwise} 
\end{cases}
\]
where

\[
\mu_{it}(c'_i, c) = \epsilon_{it}(c'_i, c) \left\{ -g'(E_{it}(c'_i, c))\tilde{w}_{it}(c'_i, c_{-i}) + g(E_{it})(r_{it}(c'_i, c_{-i}) - c_i - \gamma(r_{it}(c'_i, c_{-i}), c_i)) \right\} \\
+ (1 - \epsilon_{it}(c'_i, c))g(E_{it}(c'_i, c))\lambda r_{it}(c'_i, c_{-i})
\]

with transversality condition \(\mu_{iT_i}(c'_i, c_{-i})(c'_i, c) = g(E_{iT_i}(c'_i, c_{-i}))B_i(c'_i, c_{-i})\).

This converts the problem of choosing the effort policy that maximizes the discounted expected value of the contract into a linear optimal control program, where \(E_{it}(c'_i, c)\) is the state variable, and \(\epsilon_{it}(c'_i, c) \in \{0, 1\}\) is the control, which switches back and forth between exerting effort and siphoning. The co-state variable \(\mu_{it}(c'_i, c)\) denotes the discounted value to the winning firm of increasing the stock of accumulated effort at date \(t\), given its report \(c'_i\) and the true types \(c\). The transversality condition reflects the fact that at the terminal date, the marginal benefit of a higher stock of accumulated effort is the expected gain of avoiding default. The key condition is that the firm exerts effort \(\epsilon_{it}(c'_i, c) = 1\) only if

\[
y^1_{it}(c'_i, c, E_{it}(c'_i, c)) - y^0_{it}(c'_i, c, E_{it}(c'_i, c)) + \mu_{it}(c'_i, c) \geq 0 \tag{1}
\]

This characterization of effort provision is the main advantage of using a continuous time framework rather than discrete time. Incentive compatibility can now be characterized by focusing on the properties of a single function \(\mu_{it}(c'_i, c)\) rather than a larger number of truth-telling and obedience inequalities.

Define \(\epsilon_i(c'_i, c) = \{\epsilon_{it}(c'_i, c)\}_{t \in [0, T_i(c'_i, c_{-i})]}\) as the optimal plan given the initial report \(c'_i\) and true types \(c\). Let \(J_i(c'_i, c) = J_i(c'_i, c, \epsilon_i(c'_i, c))\) denote the optimized value of the project for a firm of type \(c_i\) reporting \(c'_i\). A firm’s payoff from submitting report \(c'_i\) with true type \(c_i\) is given by the direct utility function

\[
U_i(c'_i, c_i) = \mathbb{E}_{c_{-i}} \left[ P_i(c'_i, c_{-i}) J_i(c'_i, c) \right].
\]

A mechanism is incentive compatible for type \(c_i\) if, for all \(c'_i \neq c_i\),

\[
U_i(c_i, c_i) \geq U_i(c'_i, c_i),
\]
and an incentive compatible mechanism is \textit{individually rational for type} \( c_i \) if

\[ U_i(c_i, c_i) \geq 0. \]

The \textit{indirect utility function} is given by

\[ U_i(c_i) = \max_{c_i'} \mathbb{E}_{c_{-i}} \left[ P_i(c_i', c_{-i}) J_i(c_i', c) \right] \]

Since the optimized value of the project is determined through \( s_{it}(c_i', c_{-i}) \) and \( \epsilon_{it}(c_i', c) \), it is of central interest whether various derivatives of \( J_i(c_i', c, \epsilon_{i}(c_i', c)) \) with respect to \( c_i \) and \( c_i' \) can be computed, and what consequences this has for incentive compatibility of a mechanism.

**Proposition 3.2** Where differentiable, the value function \( J_i(c_i', c) \) satisfies

\[ \frac{\partial J_i(c_i', c_i, c_{-i})}{\partial c_i} = \int_0^{T_i(c_i', c_{-i})} -\epsilon_{it}(c_i', c)(1 - G(E_{it}(c_i', c))) \left( 1 + \frac{\partial \gamma(r_{it}(c_i', c_{-i}), c_i)}{\partial c_i} \right) dt \]

At a point of non-differentiability, the above expression is well-defined as a sub-gradient in \( c_i \).

\( J_i(c_i', c_i, c_{-i}) \) is upper semi-continuous in \( c_i' \).

Let \( c^*_i \) be the worst-off type that participates in the mechanism. Then in any mechanism in which \((c^*_i(c_i', c), s^*_{it}(c_i', c_{-i}))\) is optimal, the mechanism is incentive compatible at the allocation stage iff for all types \( c_i \),

\[ U_i(c_i) = u^*_i + \int_{c_{-i}}^{c_i} \mathbb{E}_{c_{-i}} \left[ P_i(y, c_{-i}) \frac{\partial J_i(y, y, c_{-i})}{\partial c_i} \right] dy, \tag{2} \]

for all types \( c_i \) and deviations \( c_i' \),

\[ \int_{c_i'}^{c_i} \mathbb{E}_{c_{-i}} \left[ P_i(y, c_{-i}) \frac{\partial J_i(y, y, c_{-i})}{\partial c_i} - P_i(c_i', c_{-i}) \frac{\partial J_i(c_i', y, c_{-i})}{\partial c_i} \right] dy \leq 0, \tag{3} \]

and the worst-off type for firm \( i \), \( c^*_i \) receives a payoff \( u^*_i \geq 0 \).

This proposition characterizes incentive compatibility, and how the the value of the project varies in each firm’s cost and report. In particular, since \( J_i(c_i', c_i, c_{-i}) \) is upper semi-continuous in \( c_i' \), an optimal report will always exist for the firms. While differentiability of \( J_i(c_i', c_i, c_{-i}) \) in \( c_i \) cannot be guaranteed, this will not ultimately affect calculations of the payoff to deviating, so the above characterizations of \( J_i(c_i', c_i, c_{-i}) \) and \( U_i(c_i', c_i) \) are sufficient for what follows.
4 Optimal Contracting

By the Revelation Principle, the principal can restrict attention to direct revelation mechanisms that (i) induce honesty at the allocation stage and (ii) obedience at the construction phase. Necessary and sufficient conditions for honest reporting are given by Proposition 3.2, and are standard. The obedience constraints require the mechanism to propose an effort and delay strategy that is optimal for all \((c'_i, c_i, c_{-i})\), not just honest reporting, \((c_i, c_i, c_{-i})\). By imposing these extra constraints, the mechanism is forced to consider how firms will act following a false report, and firms can contemplate the consequences of deviating at the allocation stage. Consequently, the principal’s full problem can be stated as

\[
\max_{m(c)} \mathbb{E}[c] \left[ \sum_i P_i(c) \left\{ \int_{t=0}^{T_i(c)} (1 - \epsilon_{it}(c_i, c))(1 - G(E_{it}(c_i, c)) \right. \\
+ \epsilon_{it}(c_i, c)(1 - G(E_{it}(c_i, c))) \left. \right\} \right]
\]

subject to the envelope payoff representation (2), monotonicity condition (3), and for all \((c'_i, c)\),

\[
\dot{E}_{it}(c'_i, c) = \epsilon_{it}(c'_i, c), \\
\dot{\mu}_{it}(c'_i, c) = (1 - \epsilon_{it}(c'_i, c))g(E_{it}(c'_i, c))\lambda r_{it}(c'_i, c_{-i}) \\
+ \epsilon_{it}(c'_i, c) \{-g'(E_{it}(c'_i, c)) \bar{w}_{it}(c'_i, c_{-i}) + g(E_{it}(c'_i, c_{-i}) - c_i - \gamma(r_{it}(c'_i, c_{-i}, c_i))\}, \\
\epsilon_{it}(c'_i, c) = \begin{cases} 
1 & \text{if } y_{it}(c'_i, c, E_{it}(c'_i, c)) - y_{it}^0(c'_i, c, E_{it}(c'_i, c)) + \mu_{it}(c'_i, c) \geq 0 \\
0 & \text{otherwise},
\end{cases}
\]

and

\[
s_{it}(c'_i, c_{-i}) \in \text{Argmax}_{s} w_{is}(c'_i, c) + \lambda \int_{0}^{s} r_{iz}(c'_i, c)dz
\]

where (4) is the law of motion for the state variable \(E_{it}(c'_i, c)\), (5) is the law of motion for the co-state variable \(\mu_{it}(c'_i, c)\), (6) is the obedience constraint on the optimal effort plan, and (7) is the obedience constraint associated with the delay strategy.
By identifying common terms between the principal’s objective and the agent’s direct utility function, the envelope payoff representation can be used to simplify the problem. The problem is further simplified by dropping not only the monotonicity condition — which is standard — but all obedience constraints that correspond to dishonest reports and those that correspond to the delay strategies $s_{it}(c^*_it, c_{-i})$, yielding a relaxed program

$$\max_{m(c)} \mathbb{E}_c \left[ \sum_i P_i(c) \left\{ \int_{t=0}^{T_i(c)} (1 - \epsilon_{it}(c_i, c))(1 - G(E_{it}(c_i, c))(-1 - \lambda) r_{it}(c) \right. \right.$$  

$$\left. \epsilon_{it}(c_i, c)(1 - G(E_{it}(c_i, c)) \left( h(E_{it}(c_i, c))v - \psi(r_{it}(c), c_i) - (1 - \lambda) \int_0^{s_{it}(c)} r_{iz}(c)dz \right) dt - K \right\} \right]$$

subject to

$$\dot{E}_{it}(c_i, c) = \epsilon_{it}(c_i, c),$$

$$\dot{\mu}_{it}(c_i, c) = (1 - \epsilon_{it}(c_i, c))g(E_{it}(c_i, c))\lambda r_{it}(c, c_{-i}) + \epsilon_{it}(c_i, c) \left\{ -g'(E_{it}(c_i, c))\tilde{w}_{it}(c_i, c_{-i}) + g(E_{it}(r_{it}(c_i, c_{-i}) - c_i - \gamma(r_{it}(c^*_it, c_{-i}), c_i)) \right\},$$

$$\epsilon_{it}(c_i, c) = \begin{cases} 1 & \text{if } y^1_{it}(c_i, c, E_{it}(c_i, c)) - y^0_{it}(c_i, c, E_{it}(c_i, c)) + \mu_{it}(c_i, c) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where

$$\psi_i(r_{it}(c), c_i) = c_i + \gamma(r_{it}(c), c_i) + \frac{F(c_i)}{f(c_i)} \left( 1 + \frac{\partial \gamma(r_{it}(c), c_i)}{\partial c_i} \right)$$

is the firm’s virtual cost per unit time.

While it is standard in mechanism design to drop a number of constraints at this stage of the analysis and later verify they are satisfied at the solution to the relaxed problem, ignoring the omitted obedience constraints potentially renders the solution useless. However, this approach turns out to be successful under standard assumptions on the virtual cost and hazard rate, as will be shown in the remainder of Section 4.

**Proposition 4.1** Suppose $F(c_i)$ is log-concave and the hazard rate is decreasing for all $t$. In the solution to the relaxed program, $T^*_i(c_i)$, $r_{it}(c_i)$, $\epsilon^*_it(c_i)$, $s^*_it(c^*_it)$, and $c^*_i$ do not depend on $c_{-i}$. Let

$$\phi_i(c_i) = \int_0^{T^*_i(c_i)} (1 - G(t)) \left( h(t)v - \psi_i(r_{it}(c_i), c_i) \right) dt - K$$
The solution to the relaxed program is \( r^*_{it}(c_i) = c_i, s^*_{it}(c_i) = 0, \)

\[
P^*_i(c_i, c_{-i}) = \begin{cases} 
1 & \text{if } c_i = \min_k c_k \text{ and } \phi_i(c_i) \geq K \\
0 & \text{otherwise},
\end{cases}
\]

\[
T^*_i(c_i) = \begin{cases} 
(h)^{-1}\left(\frac{\psi(c_i)}{v}\right) & \text{if } \phi(c_i) \geq K \\
0 & \text{otherwise},
\end{cases}
\]

\[
\epsilon^*_it(c_i) = \begin{cases} 
1 & \text{if } 0 \leq t \leq T_i(c_i) \\
0 & \text{otherwise}
\end{cases}
\]

and the worst-off type \( c^*_i \) is defined by \( \phi_i(c^*_i) = K \).

Since the principal is the residual claimant on social surplus, he suffers any inefficiencies arising from breaks in effort provision, delays, or inefficient funding. Consequently, in the solution to the relaxed problem, the project is awarded to the firm with the lowest marginal cost that justifies incurring the fixed cost \( K \), the winning firm works continuously and reports success immediately, and a winning firm’s flow payment is equal to its true marginal cost.

In the case of a strictly increasing hazard rate \( h(t) \), the solution is slightly different: Any firm contracted to work is asked to work until \( \bar{T} \). This is because an increasing hazard rate implies that the principal’s relaxed maximization problem in \( T_i(c_i, c_{-i}) \) is convex, and the solution is at a corner of the interval \([0, \bar{T}]\). Since this case removes one of the margins of the contract, the analysis largely simplifies to a regular procurement model, nesting a standard private-values procurement problem in this framework. In the case of an arbitrary, differentiable hazard rate \( h(t) \), there will be intervals on which the hazard rate is increasing, and some types receive the time horizon allocated at the upper endpoint while the rest receive the time horizon allocated at the lower endpoint, while wherever the hazard rate is decreasing, the types will receive different contracts. This does not lead to pooling in the classic sense, because the types who receive the same time horizon can still be screened other margins, such as the probability of winning and the compensation received.
4.1 Construction of an optimal DRM

If the solution to the relaxed problem can be implemented by a direct revelation mechanism that satisfies the constraints of the original problem, then such a mechanism is optimal. However, this requires constructing award and bond functions that satisfy the monotonicity constraint as well as all of the dropped obedience constraints. Such a mechanism is now constructed by working backwards through the firm’s problem, finding necessary and sufficient conditions on the mechanism to implement the relaxed solution along the way. Section 4.1.1 shows that to implement no delay in revelation of success, the functional form of the award function is uniquely defined up to an additive constant that depends on the bond function. Using this functional form, Section 4.1.2 derives the optimal plans for all hypothetical type-report pairs, allowing calculation of off-equilibrium path payoffs. Section 4.2.2 then uses these characterizations to construct a bond function which implements the relaxed solution and satisfies all the constraints of the full problem, thereby solving the optimal contracting problem.

4.1.1 Strategic delay

In the final stage of the game, the firm privately knows whether or not success has occurred, and the principal would like to implement a particular schedule of delays in the revelation of this information. Since the firm cannot fool the principal by reporting success early, it can only delay revelation until the most lucrative date. This is equivalent to a secondary mechanism design problem, in which the success date \( t \) is privately known by the firm, and the principal would like to select the delay schedule \( s_{it}(c_i) \) that maximizes his payoff:

\[
\max_{s_{it}(c_i)} \int_0^{T_i(c_i)} \left( 1 - G(t) \right) \left\{ h(t) (v - w_{is_{it}}(c_i)) - s_{it}(c_i)c_i - c_i \right\} dt + (1 - G(T_i(c_i))) B_i(c_i)
\]

subject to incentive compatibility for every \( t \),

\[
s_{it}(c_i) \in \text{Argmax}_{t \leq s \leq T_i(c_i)} w_{is}(c_i) + \lambda (s - t)c_i
\]
and ex ante individual rationality for each $c_i$ in $[c, c^*_i]$,

\[
p_i(c_i) \left\{ \int_0^{T_i(c_i)} g(t) \left( w_{it}(c_i) + \lambda c_i s_{it}(c_i) \right) - (1 - G(T_i(c_i))) B_i(c_i) \right\}
\]

\[= \int_{c_i}^{c^*_i} p_i(x) \int_0^{T_i(x)} (1 - G(z)) dz dx \]

where $\Lambda_i(c_i)$ is the agent’s informational rent and $p_i(c_i) = \mathbb{E}_{c_{-i}} [P_i(c_i, c_{-i})]$.

This problem is equivalent to asking the firm to reveal the date at which success occurs, $t$, and then suggesting a time to delay before reporting this to the principal. The obedience constraints require that the suggestion is actually optimal, so that a winning firm behaves as intended. The ex ante individual rationality constraint appears because at the allocation stage, the agent must receive an expected payoff of $\Lambda_i(c_i)$, before finding out whether he will succeed or not. The success time $t$ can then be treated as a piece of private information that is distributed $g(t)$, turning this into an essentially static mechanism design problem subject to an ex ante participation constraint that must match up with the expected payoffs at the end of the allocation stage, before $t$ is revealed to the firm.

**Proposition 4.2** Let

\[w_{it}(c_i) = w_{i0}(c_i) - \lambda c_i t\] (8)

where the base award is given by

\[w_{i0}(c_i) = \frac{\Lambda_i(c_i)/p_i(c_i) + (1 - G(T_i(c_i))) B_i(c_i) + \lambda c_i \int_0^{T_i(c_i)} g(z) dz}{g(T_i(c_i))}\] (9)

1. Suppose the award function $w_{it}(c_i)$ is positive at all dates. If the bond satisfies

\[B_i(c_i) \geq G(T_i(c_i)) \left( \lambda c_i T_i(c_i) + \frac{\lambda c_i}{h(T_i(c_i))} \right) - \frac{\Lambda_i(c_i)}{p_i(c_i)} - \lambda c_i \int_0^{T_i(c_i)} zg(z) dz,\] (10)

then continuous effort and immediate revelation are incentive compatible in the secondary problem.
2. Suppose the award function \( w_{it}(c_i) \) is negative for some dates near \( T_i(c_i) \). If the bond satisfies

\[
B_i(c_i) \left( 1 - \frac{1 - G(T_i(c_i))}{G(T_i(c_i))} \right) \geq \frac{\lambda c_i}{h(T_i(c_i))} + \frac{\Lambda_i(c_i)}{p_i(c_i)} + \lambda c_i \int_0^{T_i(c_i)} g(z) zdz,
\]

then continuous effort and immediate revelation are incentive compatible in the secondary problem.

This proposition characterizes the award and bond functions that achieve immediate revelation in the secondary problem and well as continuous effort. As time goes on, the prize for completion erodes at the same rate at which funds are siphoned, making firms indifferent between delaying and reporting success immediately. To discourage shirking, the bond is chosen in tandem to keep the firm on the hook late in the contractual horizon. The key is to find the date at which the firm is most likely to give up, and ensure that the award and bond functions satisfy (1). Doing this provides two additional constraints on the bond functions that must be satisfied to ensure the unconstrained solution is a constrained optimum.

This proposition shows that “late penalties” can be an endogenous phenomenon that arises when the award function becomes negative. This shows how the contracts used by Caltrans in Bajari and Lewis [3] could be close to optimal, and why many other markets employ late fees or other punishments when firms fail to reach deadlines: Despite receiving a negative award for success, firms continue to work in fear of the prospect of an even more significant financial loss from the bond. However, if

\[
1 - \frac{1 - G(T_i(c_i))}{G(T_i(c_i))} < 0
\]

then \( G(T_i(c_i)) < 1/2 \), and the left-hand side of (11) is negative while the right-hand side is positive, leading to a contradiction. Then if the award function becomes negative and a firm is more likely to fail than succeed at time zero, it is impossible to simultaneously implement continuous effort and immediate reporting introducing additional complications into the problem. Consequently, the remainder of the paper focuses on the case when the bond is positive at all dates.
4.1.2 Optimal plans after dishonest reports

To ensure incentive compatibility at the allocation stage, the optimal effort strategies for dishonest reports must be characterized as well. In particular, there are two qualitatively different types of harmful effort plans: Those in which the firm has no intention of exerting any effort at any date, and those in which the deviating firm intends to exert some effort at some date.

**Proposition 4.3** Suppose the award function satisfies (8), (9) and (10). Then at the relaxed solution,

1. If any types are excluded from participation, so that \( c_i^* < \bar{c} \), the bond function must satisfy

\[
B_i(c_i) \geq \lambda c_i T_i^*(c_i) \tag{12}
\]

2. A firm reporting \( c_i' > c_i \) exerts effort for all dates \( t \in [0, T_i^*(c_i')] \)

3. A firm reporting \( c_i' < c_i \) exerts effort for all dates \( 0 < t \leq \sigma_i^*(c_i', c_i) \leq T_i^*(c_i') \) and exerts no effort between \( \sigma_i^*(c_i', c_i) < t \leq T_i^*(c_i') \), where the stopping time \( \sigma_i^*(c_i', c_i) \) is the unique solution to the equation

\[
(1 - G^*(\sigma_i^*(c_i', c_i))(c_i' - c_i - \gamma(c_i', c_i) - \lambda c_i') + g^*(\sigma_i^*(c_i', c_i))(w_{i\sigma_i^*(c_i')}(c_i') + B_i(c_i')) - g^*(\sigma_i^*(c_i', c_i))(\lambda c_i'(T_i^*(c_i') - \sigma_i^*(c_i', c_i))) = 0 \tag{13}
\]

The stopping time \( \sigma_i^*(c_i', c_i) \) is decreasing in \( c_i \).

This proposition shows that firms who over-state their marginal cost will, like honest firms, work for the entire contractual horizon, while firms who under-state their marginal cost will “give up” early and siphon funds for some interval of time near the terminal date. To deter spurious bidding in which firms intend never to exert any effort, it must be unprofitable to bid, win, and siphon for all dates. If there are some inefficient types who are excluded ex ante this is certainly a concern, but from a broader perspective, violations of (12) invites socially inefficient entry and arbitrage by insincere agents.
Figure 1 illustrates the optimal stopping time $\sigma_i^*(c_i', c_i)$ as a function of the report $c_i'$. When a firm reports honestly, it works continuously from time zero to time $T_i^*(c_i)$ by construction. Deviations to a higher report imply that (1) is still satisfied at all dates, so for $c_i' > c_i$, $\sigma_i^*(c_i', c_i) = T_i^*(c_i')$. However, when deviating to a lower report, the payoff of continued effort near the terminal date drops below the value of siphoning, so that the opportunity cost of exerting effort is greater than the possibility of achieving success and avoiding the bond payment, so consequently $\sigma_i^*(c_i', c_i) < T_i^*(c_i')$.

4.1.3 Incentive compatible bonds

The only remaining degree of freedom in designing the mechanism is the bond function, $B_i(c_i)$. To exploit the results of the previous two sections, it must be chosen so that the award function is weakly positive for all true reports at all dates, the conditions in (8), (9) and (10) are satisfied, and the monotonicity condition holds.

Definition 4.1 Define $m^*(c) = \{P_i^*(c), r_{it}^*(c_i), w_{it}^*(c_i), B_i^*(c_i), T_i^*(c_i), s_{it}^*(c_i), \epsilon_{it}^*(c_i)\}_i$ as the direct mechanism where $P_i^*(c), r_{it}^*(c_i), T_i^*(c_i)$ and $s_{it}^*(c_i)$ are given by Proposition 4.1, the award function $w_{it}^*(c_i)$ satisfies (8) and (9), the effort schedule is given by

$$
\epsilon_{it}^*(c_i', c_i) = \begin{cases} 
1 & \text{if } t \leq \sigma_i^*(c_i', c_i) \lor T_i^*(c_i') \\
0 & \text{otherwise}
\end{cases}
$$
where \( \sigma_i^*(c_i', c_i) \) is the unique solution to (13), and the bond is given by

\[
B_i^*(c_i) = G(T_i^*(c_i)) \left( \lambda c_i T_i^*(c_i) + (1 - \lambda) \log \left( \frac{h(T_i^*(c_i))}{T_i^*(c_i)} \right) \right) + \\
\max \left\{ \frac{\lambda c_i^*}{h(T_i^*(c_i))}, \frac{\lambda c_i^*}{1 - G(T_i^*(c_i))}, \frac{1}{G(T_i^*(c_i))} \int_{0}^{T_i^*(c_i)} (1 - G(z))dz(c_i^* - (1 - \lambda)c_i) \right\} \\
- \frac{\Lambda_i(c_i)}{p_i^*(c_i)} - \lambda c_i \int_{0}^{T_i^*(c_i)} g(z)dz
\] (14)

This mechanism combines the solution to the principal’s relaxed problem in Proposition 4.1, the award function developed in Section 4.1.1, and the optimal behavior after dishonest reports characterized in Section 4.1.2 with a particular bond function, \( B_i^*(c_i) \), which is constructed to ensure that the award function is positive at all dates and (2), (10), and (12) are satisfied.

The main result of the paper is:

**Theorem 4.1** Suppose \( F(c_i)/f(c_i) \) is increasing in \( c_i \) and \( h(t) \) is decreasing in \( t \). Then \( m^*(c) \) implements the solution to the relaxed problem and satisfies all the constraints of the full problem.

This shows that the solution to the relaxed problem can be achieved with a mechanism that satisfies all of the incentive constraints faced by the principal. Furthermore, this is the same allocation that would be chosen if the principal could directly observe the progress of the project. Thus, the firm’s ability to extract rents is determined entirely by its private information, and not the potential for moral hazard. The proof of the theorem reduces incentive compatibility and obedience to a set of conditions that the bond condition must satisfy, and then constructs a feasible candidate. The indirect implementation in the next section shows how a number of terms in the bond and award functions above can be interpreted as the expectation of the total cost differences between the lowest-cost firm and the next-lowest bidder, providing additional economic intuition.

The main practical lesson of Theorem 4.5 is that it is not enough to set the bond at the amount siphonable from the project, \( \lambda c_i^* T_i^*(c_i) \), since this ignores the incentive constraints across types. The terms of the contract need to vary with the horizon so that inefficient firms never find it optimal to report a low type, receive a much longer horizon, and siphon large amounts of funds once the project becomes unprofitable. In short, the bond plays a non-trivial role not just preventing spurious bidding, but also in maintaining incentive compatibility in reporting and ensuring that the proper level of effort is exerted.
5 Dynamic Procurement Auctions

This section shows that the direct revelation mechanism developed in the previous section can be implemented through a bidding game that has elements of both a first- and second-price auction.

Let \( b = (b_1, ..., b_I) \) be the vector of firm bids, with \( b_{(n)} \) the lowest bid and \( b_{(n-1)} \) the second-lowest, and consider the dynamic procurement auction:

1. The principal announces a schedule \( T(b_i) \) relating the bid of firm \( i, b_i \), to a schedule of terminal dates, satisfying

\[
T(b_i) = h^{-1} \left( \frac{\psi_i(b_i, b_i)}{v} \right)
\]

and the reserve bid, \( b^*_i \), above which the principal does not award the project, given by

\[
G(T(b^*_i))v - \psi(b^*_i, b^*_i) \int_0^{T(b^*_i)} t g(t) dt = K
\]

2. Each participating firm \( i \) simultaneously submits a sealed bid \( b_i \).

3. The lowest-bidding firm satisfying \( b_{(n)} \leq b^*_i \) wins, and is paid its bid, \( r_{it}(b_{(n)}) = b_{(n)} \), each moment of the entire contract period \( T(b_{(n)}) \).

4. If a winning firm succeeds, it receives an award

\[
w_{i0}(b_{(n)}, b_{(n-1)}) = \frac{a(b_{(n)}, b_{(n-1)} \lor b^*_i) + (1 - G(T_i(b_{(n)})))B_i(b_{(n)}) + \lambda b_{(n)} \int_0^{T(b_{(n)})} g(z)dz}{G(T(b_{(n)}))} - \lambda b_{(n)} t
\]

\[
w_{it}(b_{(n)}, b_{(n-1)}) = w_{i0}(b_{(n)}, b_{(n-1)} \lor b^*_i) - \lambda b_{(n)} t
\]

where

\[
a(b_{(n)}, b_{(n-1)}) = (b_{(n-1)} \lor b^*_i) \int_0^{T(b_{(n-1)} \lor b^*_i)} (1 - G(z))dz - b_{(n)} \int_0^{T(b_{(n)})} (1 - G(z))dz
\]

\[
\text{Cost Savings} + \int_{b_{(n)}}^{b_{(n-1)} \lor b^*_i} x(-T'(x))(1 - G(x))dx \quad \text{(15)}
\]

\[
\text{Efficiency Bonus}
\]
If a winning firm fails to complete the project by $T(b)$, it forfeits a bond

$$B_i^*(b_n, b_{(n-1)}) = G(T_i^*(b_n)) \left( \lambda b_n T_i^*(b_n) + (1 - \lambda) \log \left( \frac{h(T_i^*(c))}{F_i^*(b_n)} \right) \right) + \max \left\{ \lambda b_i^* \frac{h(T_i^*(b_n))}{1 - G(T_i^*(b_n))}, \lambda b_i^* T_i^*(b_n) \int_0^{T_i^*(b_n)} \frac{1}{G(T_i^*(b_n))} (1 - G(z)) dz \right\}$$

$$- a(b_n, b_{(n-1)}) - \lambda b_n \int_0^{T_i^*(b_n)} g(z) dz$$

In this indirect mechanism, firms bid competitively, stating a cost-per-unit time. The lowest-bidding firm that is eligible to win receives its bid per unit of time spent working on the project, while its award and the bond are computed using the next highest-bid, similar to a second-price auction. The function $a(b_n, b_{(n-1)})$ is chosen so that the expectation conditional on $b_{(n-1)}$ being the $(n-1)$-st lowest draw from a sample of $n$ draws from the distribution $F(c_i)$, conditional on $b_n$ being the lowest, equals the informational rent term $\Lambda_i(b_n)$.

**Theorem 5.1** Suppose that $F(c_i)/f(c_i)$ is increasing in $c_i$ and the hazard rate of success $h(t)$ is decreasing in $t$. Then the dynamic procurement auction implements the same allocation as $m^*(c_i)$, so it is an optimal indirect implementation. It is a Bayesian Nash equilibrium to bid $b_i = c_i$.

This design combines a number of features common to well-known mechanisms, but is not equivalent to any of them. Similar to a first-price auction, the lowest eligible bidder wins and is paid its bid at each moment to fund the project. Unlike a first-price auction, these bids are honest in equilibrium, and exactly equal the firms’ true costs of operation. Similar to a second-price auction, the winner’s payoff depends on the next highest bid, since it appears in $a(b_n, b_{(n-1)})$. In particular, $a(b_n, b_{(n-1)})$ rewards the winner for a lower marginal cost in two ways through two terms. The cost savings in (14) is the reduction in total cost relative to the next-best alternative, given that the winner and the next-best firm work the same amount of time. The winner then also receives the efficiency bonus in (14) which corresponds to the longer terminal date awarded to the more efficient firm, similar to a Vickrey-Clarke-Groves mechanism.

A novel feature of the mechanism is that the bond is determined endogenously through the bidding and, indeed, must be determined endogenously. The construction of the bid occurs in the secondary problem along with the award function, and this places restrictions on what kinds of bid functions are compatible with a given award. If the award function is determined through the
function $a(b_n, b_{n-1})$, then the bond must satisfy the restrictions imposed by (9) and (10). The only way to guarantee this without literally adopting the optimal direct revelation mechanism is to make bonding endogenous. This emphasizes the fact that while bonding is often assumed only to deter spurious bidders from participation, it also plays a role in maintaining incentive compatibility, and arbitrary bond schedules can undermine incentives more generally.

### 6 Conclusion

This paper provides a number of useful results about the design of contracts in markets that suffer from both adverse selection and dynamic moral hazard. The ability of firms to engage in siphoning behavior creates a number of channels through which the principal can be harmed, requiring more sophisticated contracts that explicitly incorporate the dynamic nature of the market. This results in “cost-plus” payments, where the firm is compensated for its cost per unit time, and all informational rents are deferred until success is achieved. If the firm fails to succeed, however, it forfeits a bond chosen to satisfy incentive compatibility conditions as well as ensure the firm keeps working late in the contract when it might otherwise prefer to give up and siphon. Lastly, an indirect mechanism is proposed that determines the award and bond functions endogenously through the bidding, and shares features with both first-price auctions and Vickrey-Clarke-Groves mechanisms. This provides a useful benchmark for future research on the design of dynamic markets for contracts.

The most significant drawback of the present analysis is that it assumes that the amount which a firm can post as a bond is unlimited. This presumes that the project is “small” relative to the size of the firm’s capital, and that failure does not threaten the existence of the firm. In practice, firms will be constrained by the value of their real, pledgeable assets, and this limited liability will have to be considered as a constraint on the principal’s design of the mechanism. In situations where intellectual property such as human capital or patents make up the majority of the firm’s value, it would be difficult or impossible — both practically and legally — to construct agreements where high bond values are at stake. Since the bond essentially plays the role of a “side bet” between the principal and firm that allows them to gamble over the project’s value and the informational rents in a way that deters siphoning, there are a number of potential solutions. One option is for the principal to restrict attention to only those firms who can afford the risk of a high bond, thereby reducing competition. Another approach is for the principal to solve the problem in which
the limited liability constraint binds at the optimum, and perhaps use a negative bond or pooling in the secondary mechanism design problem. That analysis would present a number of additional challenges, but could provide useful and more general results.

The issue of negative award functions arises in Proposition 4.2, but the prevalence of this practice in the real world makes it worthy of further study in light of the fact that whenever high cost firms are given a task they are considered unlikely to complete at the allocation stage, negative award functions necessarily result in inefficiency through delays in revelation of success or inefficient effort provision. The main advantage of picking an award that penalizes the firm near the deadline is that it allows a lower default penalty, since some of the punishment for failure is shifted into the dynamic payments. If a high-quality firm optimally works such a long time that its limited liability constraint binds, switching from a high bond and strictly positive award at $T_i(c_i)$ to a lower bond and weakly negative award near the terminal date might relax these constraints. However, it is unclear which firms have the most expensive contracts, since $c_i T_i^*(c_i)$ is not necessarily a monotone or single-peaked function.

In risky industries in which projects end in default, the principal may take steps to “check in” with the firm before the deadline or better track the firm’s efforts over the course of the project. If the project is awarded to the most efficient firm, the principal will want to renegotiate ex post after a failure, and potentially forgive the bond. A model that incorporates noisy signals of a firm’s progress (or lack thereof) might provide a basis for contracts in which renegotiation or monitoring is endogenous. As in this paper, this potential for renegotiation will likely impact the optimal bond structure, and understanding the interplay of bonding and renegotiation may provide some guidance for procurement in large scale projects where failures or missed deadlines are common.

An even more comprehensive theory of industry management might be developed by allowing firms to keep their own “buffer-stocks” of funds, where the likelihood and survival of the firm depend on these hidden assets. As Doepke and Townsend [14] show, the problem of hidden actions and hidden savings can be formalized in a general framework, incorporating default issues similar to Calveras et al. [6]. While this would seem to be a straightforward extension, the incorporation of binding pure state constraints into optimal control problems presents some unexpected challenges. In particular, the costate variables can jump discontinuously as a state variable moves from binding to non-binding, making the derivation of necessary conditions for optimality somewhat complicated.
(see [32], p. 332–333). While this problem is easily solved for a particular, fixed mechanism, it becomes necessary to characterize the binding and non-binding periods of time in order to derive an envelope representation of a winning firm’s payoff, which is a complication beyond the scope of this paper. By further understanding how bidding becomes a signal of a firm’s health in a dynamic model and how firms exploit the ability to secretly save, however, could provide useful insights into how competition can become destructive.

Finally, this paper is an independent private values framework, in which each firm’s cost draw is independent of the others. In situations where each firm receives information that is useful to the others in deciding the value of the project, the optimal mechanisms will likely be different. Alternatively, uncertainty could be modelled about the hazard rate, \( h[E|s_1, ..., s_n] \), where \( s_i \) is the signal received by firm \( i \) about the likelihood of success. Better and more efficient markets might be designed where information is aggregated to incorporate winner’s curse phenomena.
References


7 Appendix: Proofs

Proof of Proposition 3.1

Proof: A firm $i$ who holds the rights to the project receives a value from exerting effort for a small period of time $\Delta$ of

$$J_i(t, E_{it}, c_i', c) = e^{-\rho\Delta} h(E_{it}) \Delta \tilde{w}_{it}(c_i', c_{-i}) + (1 - h(E_{it}) \Delta) J_i(t + \Delta, E_{it} + \Delta, c_i', c)$$

$$+ \frac{\Delta \tilde{r}_{it}(c_i', c_{-i}) - c_i - \gamma(r_{it}(c_i', c_{-i}), c_i))}{c_i}$$
Re-arranging and dividing by $\Delta$ yields

$$- \frac{J_i(t + \Delta, E_{it} + \Delta, c_{i}', c) - J_i(t, E_{it}, c_{i}', c)}{\Delta} + \left( \frac{1 - e^{-\rho \Delta}}{\Delta} + h(E_{it}) \right) J_i(t + \Delta, E_{it} + \Delta, c_{i}', c)$$

$$= h(E_{it})e^{-\rho \Delta} \tilde{w}_{it}(c_{i}', c_{-i}) + r_{it}(c_{i}', c_{-i}) - c_i - \gamma(r_{it}(c_{i}', c_{-i}), c_i)$$

and taking the limit as $\Delta \to 0$ gives

$$- \frac{dJ_i(t, E_{it}, c_{i}')}{dt} + (\rho + h(E_{it})) J_i(t, E_{it}, c_{i}', c) = h(E_{it})\tilde{w}_{it}(c_{i}', c_{-i}) + r_{it}(c_{i}', c_{-i}) - c_i - \gamma(r_{it}(c_{i}', c_{-i}), c_i)$$

Solving this first-order differential equation in $t$ yields an expression for the value function on an interval of time $[t_a, t_b]$ during which effort is exerted,

$$(1 - g(E_{it_a}(c_{i}', c))) e^{-\rho t_a} J_i(t_a, E_{it_a}(c_{i}', c), c_{i}', c) = \int_{t_a}^{t_b} (1 - g(E_{iz}(c_{i}', c))) e^{-\rho z} \left\{ h(E_{iz}(c_{i}', c)) \tilde{w}_{iz}(c_{i}', c_{-i}) + r_{iz}(c_{i}', c_{-i}) - c_i - \gamma(r_{iz}(c_{i}', c_{-i}), c_i) \right\} dz$$

$$+ (1 - g(E_{it_b}(c_{i}', c))) e^{-\rho t_b} J_i(t_b, E_{it_b}(c_{i}', c), c_{i}', c)$$

Then as $\rho \to 0$, the value of working for an interval $[t_a, t_b]$ equals

$$(1 - g(E_{it_a}(c_{i}', c))) J_i(t_a, E_{it_a}(c_{i}', c), c_{i}', c) = \int_{t_a}^{t_b} (1 - g(E_{iz}(c_{i}', c))) \left\{ h(E_{iz}(c_{i}', c)) \tilde{w}_{iz}(c_{i}', c_{-i}) + r_{iz}(c_{i}', c_{-i}) - c_i - \gamma(r_{iz}(c_{i}', c_{-i}), c_i) \right\} dz$$

$$+ (1 - g(E_{it_b}(c_{i}', c))) J_i(t_b, E_{it_b}(c_{i}', c), c_{i}', c)$$

and the value of siphoning for an interval $[t_a, t_b]$ equals, noting that $E_{it_a}(c_{i}', c) = E_{it_b}(c_{i}', c)$,

$$(1 - g(E_{it_a}(c_{i}', c))) J_i(t_a, E_{it_a}(c_{i}', c), c_{i}', c) = \int_{t_a}^{t_b} \lambda(1 - g(E_{it_a}(c_{i}', c))) r_{iz}(c_{i}', c_{-i}) dz + (1 - g(E_{it_a}(c_{i}', c))) J_i(t_b, E_{it_b}(c_{i}', c), c_{i}', c)$$

Using the boundary condition that failure gives a terminal payoff of $-(1 - g(E_{it}(c_{i}', c_{-i}))) B_t(c_{i}', c_{-i})$, we can work backwards to generate the linear optimal control problem,

$$\max_{c_{i}', c} J_i(c_{i}', c, e_{i}(c_{i}', c)) =$$

$$\max_{c_{i}', c} \int_{t_a}^{T_i(c_{i}', c_{-i})} \left\{ g(E_{it}(c_{i}', c)) \tilde{w}_{it}(c_{i}', c_{-i}) + (1 - g(E_{it}(c_{i}', c))) \left[ r_{it}(c_{i}', c_{-i}) - c_i - \gamma(r_{it}(c_{i}', c_{-i}), c_i) \right] \right\} dt$$

$$+ (1 - e_{it}(c_{i}', c)) \left( 1 - g(E_{it}(c_{i}', c)) \right) \left\{ \lambda r_{it}(c_{i}', c_{-i}) \right\} \left( 1 - g(E_{it}(c_{i}', c)) \right) B_t(c_{i}', c_{-i})$$

$$+ (1 - e_{it}(c_{i}', c)) \left( 1 - g(E_{it}(c_{i}', c)) \right) \{ \lambda r_{it}(c_{i}', c_{-i}) \} \left( 1 - g(E_{it}(c_{i}', c)) \right) B_t(c_{i}', c_{-i})$$
subject to $\epsilon_{it}(c'_i, c) \in \{0, 1\}$ and $\dot{E}_{it}(c'_i, c) = \epsilon_{it}(c'_i, c)$.

The Hamiltonian of the optimal control problem is

$$H(\epsilon_{it}, E_{it}; t, c'_i, c) = \epsilon_{it}y^1_{it}(c'_i, c, E_{it}) + (1 - \epsilon_{it})y^0_{it}(c'_i, c, E_{it}) + \mu_{it}(c'_i, c)\epsilon_{it}$$

Due to the linearity of the Hamiltonian, by Pontryagin’s necessary conditions, the optimal control takes a bang-bang form:

$$\epsilon_{it}(c'_i, c) = \begin{cases} 1 & \text{if } y^1_{it}(c'_i, c, E_{it}(c'_i, c)) - y^0_{it}(c'_i, c, E_{it}(c'_i, c)) + \mu_{it}(c'_i, c) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and the co-state variable evolves as

$$\frac{\partial H}{\partial E_{it}} = \dot{\mu}_{it}(c'_i, c) = \epsilon_{it}(c'_i, c) \left\{ -g'(E_{it}(c'_i, c))\dot{w}_{it}(c'_i, c_{-i}) + g(E_{it})(r_{it}(c'_i, c_{-i}) - c_i - \gamma(r_{it}(c'_i, c_{-i}), c_i)) \right\} + (1 - \epsilon_{it}(c'_i, c))g(E_{it}(c'_i, c))\lambda r_{it}(c'_i, c_{-i})$$

The transversality condition for $E_{it}(c'_i, c)$ is $\mu_{iT_i(c'_i, c_{-i})}(c'_i, c) = g(E_{iT_i(c'_i, c_{-i})}(c'_i, c))B_i(c'_i, c_{-i})$, since the marginal benefit of having exerted another instant of effort is the probability of avoiding the bond payment times the loss. □

**Proof of Proposition 3.2**

**Proof:** Properties of the derivative or subgradient of the value function and upper semi-continuity in parameters are proved in Seierstad and Sydsaeter [32] (p. 220) or Clarke et al. [11] (p.105). The characterization of incentive compatibility is proved in Milgrom and Segal, Proposition 4, [25]. The monotonicity condition is a consequence of rearranging the direct and indirect utility functions. □

**Proof of Proposition 4.1**

**Proof:** Since delay after success is always costly for the principal, the unconstrained solution is to set $s_{it}(c_i, c_{-i}) = 0$ for all $c_i$.

Whenever the coefficient of $\epsilon_{it}(c_i, c)$ is positive, it should be set to one, and zero otherwise. Whenever $\epsilon_{it}(c_i, c)$ is one, $r_{it}(c_i, c_{-i})$ should be set equal to $c_i$ to avoid incurring the efficiency penalty $\gamma(r_{it}(c'_i, c_{-i}), c_i))$. When $\epsilon_{it}(c_i, c)$ is set equal to zero and $r_{it}(c'_i, c_{-i}) \geq 0$, the principal incurs a loss through siphoning. Since the agents have no time preference, it can be arranged so that the firm works continuously, so that all productive dates occur from time zero to time $T'$, and then the principal sets $T_i(c'_i) = T'$. This implies that $G(E_{it}(c_i, c)) = G(t)$.

Then the principal’s payoff becomes

$$\max_{m(c)} \mathbb{E}_c \left[ \sum_i P_i(c) \left\{ \int_{t=0}^{T_i(c)} (1 - G(t))(h(t)v - \psi(c_i, c_i)) dt - K \right\} \right]$$

Maximizing over $T_i(c)$ for each firm $i$ yields the first-order necessary condition $h'(T_i(c_i, c_{-i}))(v - \psi(c_i, c_i)) = 0$ and second-order sufficient condition $h''(T_i(c_i, c_{-i}))(v < 0$. So that there is an interior
solution that depends only on the winner’s marginal cost, $c_i$, if the hazard rate is decreasing at the optimum. If $h'(t) > 0$ for all $t$, the objective is convex and any critical point is a local minimum, so that there is a corner solution where a winning firm is contracted to work until success occurs. Note that since $c_{-i}$ appears nowhere in $h(T_i(c_i, c_{-i}))(v - \psi_i(c_i, c_i))$, $T_i(c_i, c_{-i})$ is not a function of $c_{-i}$.

Lastly, the inner integral $\phi_i(c_i) = \int_0^{T_i(c_i)} (1 - G(t)) (h(t)v - \psi(c_i, c_i)) \, dt - K$ is decreasing in $c_i$ if $h(t)$ is a decreasing function, so that the optimal decision is

$$P_i(c_i, c_{-i}) = \begin{cases} 1 & \text{if } c_i = \min_k c_k \text{ and } \phi_i(c_i) \geq K \\ 0 & \text{otherwise} \end{cases}$$

□

Proof of Proposition 4.2

Proof: Since the firm’s payoff once it succeeds does not depend on its private information, the delay is independent of the firm’s private information and relies only on the report. Consider the maximization problem

$$\tilde{w}_{is}(c'_i) = \max_{0 \leq s \leq T_i(c'_i)} w_{is}(c'_i) + \lambda sc'_i$$

A necessary condition at any optimal $s^*_i(c'_i) < T_i(c'_i)$ is

$$\tilde{w}_{is}(c'_i) + \lambda c'_i = 0$$

Solving the equation for $w_{is}(c'_i)$ yields (8),

$$w_{is}(c'_i) = w_{i0}(c'_i) - \lambda t c'_i$$

where the arbitrary constant $w_{i0}(c_i)$ will be chosen to satisfy the ex ante individual rationality constraint. Substituting the proposed award function into the maximization problem yields

$$\max_{0 \leq s \leq T_i(c'_i)} w_{i0}(c'_i)$$

which is independent of $s$. Consequently, reporting immediately is a solution for all $c'_i$. If this indifference is a cause for concern, adding an arbitrarily small but positive and decreasing function will break the indifference in favor of immediate reporting.

We now ensure that types who report honestly receive the correct expected payoff. Substituting the award function into the ex ante individual rationality constraint yields

$$p_i(c_i) \left\{ \int_0^{T_i(c_i)} f(z) \{w_{i0}(c_i) - \lambda c_i z\} \, dz - (1 - G(T_i(c_i))B_i(c_i) \right\} = \int_{c_i}^{c'_i} p_i(x) \int_0^{T_i(x)} (1 - G(z)) \, dz \, dx$$

\[ \Lambda_i(c_i) \]
from which (9) is derived,

\[ w_{i0}(c_i) = \frac{\Lambda_i(c_i)/p_i(c_i) + (1 - G(T_i(c_i))B_i(c_i) + \lambda c_i \int_0^{T_i(c_i)} zg(z)dz}{G(T_i(c_i))} \]

Returning to (1) with \( E_{it}(c_i, c_i) = t \), effort will be implemented by the \( c \) type at all dates \( t \) if,

\[ y_{it}^1(c_i, c_i, t) - y_{it}^0(c_i, c_i, t) + \mu_{it}(c_i, c_i) \geq 0 \]

For \( E_{it}(c_i, c_i) = t \), this equals

\[ g(t)w_{it}(c_i) + \int_t^{T_i(c_i)} -g'(z)w_{iz}(c_i)dz + g(T_i(c_i))B_i(c_i) - (1 - G(t))\lambda c_i \geq 0 \]

To ensure that this inequality holds for all \( t \in [0, T_i(c_i)] \), we look for a minimum to the left-hand side in \( t \). The first-order necessary condition is

\[ 2g'(t^*)w_{it^*}(c_i) \leq 0 \]

Since \( w_{it}(c_i) \) is monotone decreasing in time and \( g'(t) \neq 0 \), the first-order necessary condition has at most one interior critical point,

\[ t^* = \frac{w_{i0}(c_i)}{\lambda c_i} \]

Depending on the sign of \( w_{iT_i(c_i)}(c_i) \), there are two candidate solutions: \( t^* \) and \( T_i(c_i) \). If \( w_{iT_i(c_i)}(c_i) > 0 \), the solution is \( T_i(c_i) \) because the first-order condition has no zeros and is monotone decreasing, while if \( w_{iT_i(c_i)}(c_i) < 0 \), then \( t^* \) is the solution because the first-order necessary condition has a unique zero and the second-order sufficient condition is satisfied. This generates two cases.

In the first case where \( w_{it}(c_i) \) is positive at all dates, evaluating (1) at \( T_i(c_i) \) and substituting in \( w_{it}(c_i) \), the inequality becomes

\[ h(T_i(c_i)) (w_{i0}(c_i) - \lambda c_i T_i(c_i) + B_i(c_i)) - \lambda c_i \geq 0 \]

because

\[ w_{i0}(c_i) + B_i(c_i) = \frac{\Lambda_i(c_i) + B_i(c_i) + \lambda c_i \int_0^{T_i(c_i)} zg(z)dz}{G(T_i(c_i))}. \]

Consequently, implying that exerting effort at the final date of the contract gives a better payoff than siphoning at \( T_i(c_i) \). Re-arranging the inequality to solve for \( B_i(c_i) \) provides the inequality in (10),

\[ B_i(c_i) \geq G(T_i(c_i)) \left( \lambda c_i T_i(c_i) + \frac{\lambda c_i}{h(T_i(c_i))} \right) - \frac{\Lambda_i(c_i)}{p_i(c_i)} - \lambda c_i \int_0^{T_i(c_i)} zg(z)dz \]

In the second case where \( w_{it}(c_i) \) becomes negative for some dates near \( T_i(c_i) \), evaluating (1) at
\[ t^* = \frac{w_{it}(c_i)}{(\lambda c_i)} \] yields
\[ \int_{t^*}^{T_i(c_i)} -g'(z)w_{iz}(c_i)dz + g(T_i(c_i))B_i(c_i) - (1 - (G(t^*))\lambda c_i \geq 0 \]

Integrating the first term by parts yields
\[ g(T_i(c_i))(B_i(c_i) - w_{iT_i(c_i)}(c_i)) + \lambda c_i(G(T_i(c_i)) - G(t^*)) - (1 - (G(t^*))\lambda c_i \geq 0 \]

and substituting in the award function and re-arranging yields the condition in (11).

**Proof of Proposition 4.3**

**Proof:** Consider deviations \( c'_i > c_i \). Evaluating the left-hand side of (1) at \( E_{it}(c'_i, c_i) = t \) yields
\[ g^1_{it}(c'_i, c_i, E_{it}(c'_i, c_i)) - g^0_{it}(c'_i, c_i, E_{it}(c'_i, c_i)) + \mu_{it}(c'_i, c_i) \]
which equals
\[ g(t)w_{it}(c'_i) = \lambda c'_i - g(t)w_{it}(c'_i) + (g(t) + (1 - G(t))(c'_i - c_i) \]
The first three terms equal the right-hand side of (1) when the \( c'_i \) type reports \( c'_i \) honestly, so by construction must be positive. The last term is positive because \( c'_i > c_i \), and \( g(t) \) and \( 1 - G(t) \) are positive. Therefore, (1) holds at \( E_{it}(c'_i, c_i) = t \) for all \( c'_i > c_i \), and it is a profit-maximizing policy for firm that overbid to exert effort for all moments \( t \in [0, T_i(c'_i)] \).

Lying downward implies that (1) becomes
\[ g^1_{it}(c'_i, c_i, E_{it}(c'_i, c_i)) + (c'_i - c_i) - \gamma(c'_i, c_i))(1 - G(E_{it}(c'_i, c_i))) - g^0_{it}(c'_i, c_i, E_{it}(c'_i, c_i)) + \mu_{it}(c'_i, c_i) \geq 0 \]
The \( (g(t) + (1 - G(t))(c'_i - c_i - \gamma(c'_i, c_i)) \) term that was positive at \( E_{it}(c'_i, c_i) = t \) for upward deviations is now negative, so that it is now no longer guaranteed that the deviator works the entire horizon.

Note that the Hamiltonian of the optimal control problem is strictly decreasing in \( t \), since \( w_{it}(c'_i) \) is strictly decreasing in \( t \), so that if the firm ever works, it works at the beginning of the contract.

Let \( \sigma^*_i(c'_i, c_i) \) be the stopping time at which the \( c'_i \) type gives up on the project given true type \( c_i \). Then the optimal stopping time \( \sigma(c'_i, c_i)^* \) with \( c'_i < c_i \) solves
\[ \max_{\sigma} \int_0^\sigma (1 - G(z)) \{ h(z) w_{iz}(c'_i) + c'_i - c_i - \gamma(c'_i, c_i) \} dz + \int_\sigma^{T_i(c'_i)} (1 - G(\sigma))\lambda c'_i dz - (1 - G(\sigma))B_i(c'_i) \]
whose critical points are characterized by (8),
\[ (1 - G(\sigma^*(c'_i, c_i))(c'_i - c_i) - \gamma(c'_i, c_i) - \lambda c'_i) + g(\sigma^*(c'_i, c_i))(w_{\sigma^*(c'_i)}(c'_i) + B_i(c'_i)) \]
\[ - g(\sigma^*(c'_i, c_i))(\lambda c'_i(T_i^*(c'_i)) - \sigma^*(c'_i, c_i)) = 0 \]
which is a version of (1). Note that the second derivative in \( \sigma \) is
\[ h'(\sigma) \{ w_{\sigma^*}(c'_i) + B_i(c'_i) \} - h(\sigma)\lambda c'_i < 0, \]

36
so the problem is concave in $\sigma$ and has a unique solution. Note finally that $\sigma^*_i(c'_i, c_i)$ is decreasing in $c_i$, since the partial derivative of the first-order condition with respect to $c_i$ is

$$-(1 - G(\sigma))(1 + \frac{\partial \gamma(c'_i, c_i)}{\partial c_i}) < 0$$

□

**Proof of Theorem 4.4**

**Proof:** For types $c'_i > c_i$, incentive compatibility is satisfied, because

$$\frac{\partial^2 U_i(c'_i, c_i)}{\partial c'_i \partial c_i} = -p'_i(c'_i) \int_0^{T_i(c'_i)} (1 - G(z))dz - p_i(c'_i)(1 - G(T_i(c'_i)))T'_i(c'_i) \geq 0$$

so that supermodularity holds. By standard mechanism design arguments, this implies that deviations above one’s true type are unprofitable.

For deviations $c'_i < c_i$, the argument is more complicated because the firm gives up at some date near $T_i(c'_i)$, but $T_i(c'_i)$ is greater than $T_i(c_i)$, so the firm might gain by deviating and giving up early, leading to a profitable deviation. To deter such deviations, consider bond functions of the form

$$B_i(c_i) = G(T_i(c_i)) \{ \lambda c_i T_i(c_i) + \xi_i(c_i) \} - \frac{A_i(c_i)}{p_i(c_i)} = \lambda c_i \int_0^{T_i(c_i)} zg(z)dz$$

where $\xi_i(c_i)$ is an as-yet undetermined function. By developing criteria for which incentive compatibility is ensured in terms of $\xi_i(c_i)$, any $\xi_i(c_i)$ that also satisfies (10) and (12) will then solve the optimal contracting problem.

Note that if (12) holds, any type who fails to work receives a weakly negative payoff regardless of its policy, which is worse than reporting honestly. Therefore, we need only consider deviations in which the firm finds it profitable to work for some strictly positive period of time before giving up. As a slight abuse of notation, define the payoff function in $(c'_i, c_i)$ for a $c'_i < c_i$ type-report pair in which $\sigma^*(c'_i, c_i) > 0$ as

$$U_i(c'_i, c_i) = p_i(c'_i)J^*_i(\sigma^*(c'_i, c_i), c'_i, c_i)$$

Note that if $U_i(c'_i, c_i)$ is supermodular, then by standard arguments incentive compatibility follows (see Mas-Colell et al [19]). This function is supermodular if

$$p'_i(c'_i) \frac{\partial J^*_i(\sigma^*_i(c'_i, c_i), c'_i, c_i)}{\partial c_i} + p_i(c'_i) \left\{ \frac{\partial^2 J^*_i(\sigma^*_i(c'_i, c_i), c'_i, c_i)}{\partial \sigma \partial c'_i} \frac{\partial \sigma^*_i(c'_i, c_i)}{\partial c_i} + \frac{\partial^2 J^*_i(\sigma^*_i(c'_i, c_i), c'_i, c_i)}{\partial c_i \partial c'_i} \right\} \geq 0$$

Since $p'_i(c'_i) < 0$ and

$$\frac{\partial J_i(\sigma^*_i(c'_i, c_i), c'_i, c_i)}{\partial c_i} = - \int_0^{\sigma^*_i(c'_i, c_i)} (1 - G(z)) \left(1 + \frac{\partial \gamma(c'_i, c_i)}{\partial c_i}\right)dz < 0,$$
the first term is positive by (2). The second term is positive as long as

\[
\frac{\partial^2 J_i^*(\sigma_i^*(c'_i, c_i), c'_i, c_i)}{\partial \sigma_i \partial c'_i} \frac{\partial \sigma_i^*(c'_i, c_i)}{\partial c_i} \geq 0
\]

Using (1), Proposition 4.4, and the necessary condition for maximization in \( \sigma_i \) from (9), \( \partial \sigma_i^*(c'_i, c_i)/\partial c_i \) is unambiguously negative. We now derive conditions to ensure that \( \partial^2 J_i^*(\sigma_i^*(c'_i, c_i), c'_i, c_i)/\partial \sigma_i \partial c'_i \) is negative. Since

\[
\frac{\partial J_i (\sigma_i, c'_i, c_i)}{\partial \sigma_i} = (1 - G(\sigma_i))(c'_i - c_i - \gamma(c'_i, c_i) - \lambda c'_i) + g(\sigma_i)(w_i(\sigma_i) + B_i(c'_i)) - g(\sigma_i)(\lambda c'_i(T_i(c'_i) - \sigma)),
\]

and

\[
w_i(\sigma_i) + B_i(c'_i) = \xi_i(c'_i) + \lambda c'_i(T_i(c'_i) - \sigma),
\]

we have

\[
\frac{\partial J_i (\sigma_i, c'_i, c_i)}{\partial \sigma_i} = (1 - G(\sigma_i))(c'_i - c_i - \gamma(c'_i, c_i) - \lambda c'_i) + g(\sigma_i) \xi_i(c'_i)
\]

Computing the partial derivative of the above equation with respect to \( c'_i \) yields a sufficient condition for \( U_i(c'_i, c_i) \) to exhibit supermodularity,

\[
h(\sigma_i^*(c'_i, c_i)) \xi_i(c'_i) + (1 - \lambda) \leq 0.
\]

Now, since \( \sigma_i^*(c'_i, c_i) \) is decreasing in \( c_i \) and \( c'_i < c_i \), it follows that

\[
T_i^*(c'_i) = \sigma_i^*(c'_i, c'_i) > \sigma_i^*(c'_i, c_i)
\]

Since the above inequality implies that \( \xi_i(c'_i) < 0 \) and \( h() \) is a decreasing function, the following inequality is sufficient to guarantee supermodularity:

\[
h(T_i^*(c'_i)) \xi_i(c'_i) + (1 - \lambda) \leq 0. \tag{16}
\]

Now, exploiting the functional form assumption on \( B_i(c_i) \) yields four conditions that \( \xi_i(c'_i) \) must satisfy: The award function must be positive,

\[
\xi_i(c_i) \geq \frac{\lambda c_i \int_0^{T_i(c_i)} G(z)dz - \Lambda_i(c_i)/p_i(c_i)}{1 - G(T_i(c_i))} \tag{17}
\]

the condition in (10),

\[
\xi_i(c_i) \geq \frac{\lambda c_i}{h(T_i(c_i))} \tag{18}
\]

the no arbitrage condition in (12),

\[
\xi_i(c_i) \geq \frac{\lambda c_i \int_0^{T_i(c_i)} G(z)dz + \Lambda_i(c_i)/p_i(c_i)}{G(T_i(c_i))} \tag{19}
\]
and incentive compatibility in (15),

$$\xi_i(c_i) \geq \frac{-(1-\lambda)}{h(T^*_i(c_i))}$$  \hspace{1cm} (20)

These conditions collectively require that $\xi_i(c_i)$ be “sufficiently large” and decrease “sufficiently quickly,” and any function $\xi_i(c_i)$ that satisfies all four inequalities will achieve the desired outcome. It is straightforward to verify that the proposed bond does so, with

$$\xi_i(c_i) = (1-\lambda) \log\left(\frac{h(T^*_i(c_i))}{T^*_i(c_i)}\right)$$
\[+ \max \left\{ \frac{\lambda c_i^*}{h(T^*_i(c_i))}, \frac{\lambda c_i^*}{1-G(T^*_i(c_i))}, \frac{1}{G(T^*_i(c_i))} \int_0^{T^*_i(c_i)} (1-G(z))dz\right\} \hspace{1cm} (21)

which is shown in the extended appendix. In particular, the terms inside the max operator ensure that each of (17), (18), and (19) are satisfied, and the first term ensures that (20) is satisfied, since all the terms in the inside the max operator are decreasing, so that $\xi_i(c_i)$ is almost everywhere differentiable. □

**Proof of Theorem 5.1**

**Proof:** Let $f_{(n-1)}(c)$ be the density function of the $n−1$-st order statistic drawn from $F(c)$, given that $c_{(n−1)} > b_n$. First, note that

$$E_{b_n\{b_n \mid b_n \mid (n-1) > b_n\}}[a(b_n, b_{(n-1)}) | b_n > b_n] = \frac{\int_{b_n}^{b^*} a(b_n, b_{(n-1)})f_{(n-1)}(x)dx + \int_{b^*}^\infty a(b_n, b^*)f_{(n-1)}(x)dx}{1 - F_{(n-1)}(b_n)}$$

and after an integration by parts

$$= \left[ a(b_n, x)(1 - F_{(n-1)}(x)) \right]_{b_n}^{b^*} + \int_{b_n}^{b^*} \frac{\partial a(b_n, x)}{\partial x} (1 - F_{(n-1)}(x))dx + a(b_n, b^*)(1 - F_{(n-1)}(b^*))$$

$$= \left[ a(b_n, x)(1 - F_{(n-1)}(x)) \right]_{b_n}^{b^*} + \int_{b_n}^{b^*} \left( a(b_n, b^*) (1 - F_{(n-1)}(x))dx + a(b_n, b^*)(1 - F_{(n-1)}(b^*)) \right)$$

$$= \int_{b_n}^{b^*} \int_0^{T(x)} (1 - G(z))dz (1 - F_{(n-1)}(x))dx$$

implying

$$E[a(b_n, b_{(n-1)}) | b_n > b_n] = \frac{\int_{b_n}^{b^*} p_i(x) \int_0^{T(x)} (1 - G(z))dz dx}{p_i(b_n)} = \frac{\Lambda(b_n)}{p_i(b_n)}$$

So the expectation of $a(b_n, b_{(n-1)})$ conditional on winning is equal to the informational rent conditional on winning.

Substituting this into the direct utility function where it appears in the bond and award function generates the same objective for a firm as in the direct revelation mechanism in Section 4. Since truth-telling is an equilibrium strategy there, it is here as well. Since the allocation of the project
and the payoffs are all equivalent in expectation, this is a profit-maximizing indirect implementation.
8 Extended Appendix

Verification of Conditions (17)-(20):

To see that (20) holds, note that

\[
\frac{\partial J_i}{\partial \sigma_i} = (1 - G(\sigma_i)(c'_i - c_i - \gamma(c'_i, c_i) - \lambda c'_i) + g(\sigma_i)(w_{i0}(c'_i) + B_i(c'_i)) - g(\sigma_i)(\lambda c'_i(T_i(c'_i) - \sigma))
\]

yields

\[
\frac{\partial J_i}{\partial \sigma_i} = g(\sigma_i)\xi_i(c'_i) + (1 - G(\sigma_i)(c'_i - \lambda c'_i - c_i - \gamma(c'_i, c_i)))
\]

and taking the partial derivative with respect to \(c'_i\) yields (20).

To see where the other three conditions come from, note that

\[
w_{i0}(c'_i) = \frac{\Lambda_i(c'_i)/p_i(c'_i) + (1 - G(T_i(c'_i))B_i(c'_i) + \lambda c'_i\int_0^{T_i(c'_i)} z g(z)dz}{G(T_i(c'_i))}
\]

and substituting in the bond

\[
B_i(c_i) = G(T_i(c_i)) \{ \lambda c_i T_i(c_i) + \xi_i(c_i) \} - \frac{\Lambda_i(c_i)}{p_i(c_i)} - \lambda c_i \int_0^{T_i(c_i)} z g(z)dz
\]

yields

\[
w_{i0}(c'_i) = (1 - G(T_i(c'_i)) \{ \lambda c_i T_i(c_i) + \xi_i(c_i) \} + \frac{\Lambda_i(c'_i)}{p_i(c'_i)} + \lambda c'_i\int_0^{T_i(c'_i)} z g(z)dz
\]

Substituting this into (10) and (12) and re-arranging yields (17) to (19):

\[
\xi_i(c_i, c_{-i}) \geq \frac{\lambda c_i \int_0^{T_i(c_i)} G(z)dz - \Lambda_i(c_i)/p_i(c_i)}{1 - G(T_i(c_i))}
\]

\[
\xi_i(c_i) \geq \frac{\lambda c_i}{h(T_i(c_i))}
\]

\[
\xi_i(c_i) \geq \frac{\lambda c_i \int_0^{T_i(c_i)} G(z)dz + \Lambda_i(c_i)/p_i(c_i)}{G(T_i(c_i))}
\]

Verification that (20) satisfies (16)-(19):

To verify that the proposed \(\xi_i(c_i)\) function satisfies the sufficient conditions, note first that all of the terms inside the max operator are decreasing, so that \(\xi_i(c_i)\) is almost everywhere differentiable, and where it is non-differentiable, any supergradient is negative. Since the terms inside the max
operator are all decreasing and differentiable, denote the derivative of the appropriate term at \( c_i \) as \( \eta'(c_i) \), yielding

\[
\xi_i'(c_i) = -\frac{(1 - \lambda)}{h(T_i^*(c_i))} h'(T_i^*(c_i)) T_i'(c_i) - \eta'(c_i) \leq -\frac{(1 - \lambda)}{h(T_i^*(c_i))} h'(T_i^*(c_i)) T_i'(c_i)
\]

Since the first-order necessary condition for \( T_i^*(c_i) \) is

\[
h'(T_i^*(c_i)) T_i'(c_i) = 1 + \frac{d}{dc} \left[ \frac{F(c_i)}{f(c_i)} \right] \geq 1
\]

it follows that

\[
\xi_i'(c_i) \leq -\frac{(1 - \lambda)}{h(T_i^*(c_i))} h'(T_i^*(c_i)) T_i'(c_i) \leq -\frac{(1 - \lambda)}{h(T_i^*(c_i))}
\]

and (20) is satisfied.

Verifying the other three conditions relies on the string of inequalities

\[
\Lambda_i(c_i') = \int_{c_i}^{c_i'} \frac{p_i(x)}{p_i(c_i)} \int_0^{T_i(x)} (1 - G(z))dzdx 
\]

\[
\leq \int_{c_i}^{c_i'} p_i(x)dx \int_0^{T_i(c_i)} (1 - G(z))dz 
\]

\[
\leq \int_{c_i}^{c_i'} dx p_i(c_i) \int_0^{T_i(c_i)} (1 - G(z))dz 
\]

\[
= (c_i' - c_i) \int_0^{T_i(c_i)} (1 - G(z))dz
\]

Using this to bound the informational rent, and recalling that

\[
\xi_i(c_i) = (1 - \lambda) \log \left( \frac{h(T_i^*(c_i))}{T_i'(c_i)} \right) 
\]

\[
+ \max \left\{ \frac{\lambda c_i^*}{h(T_i^*(c_i))}, \frac{\lambda c_i^* T_i^*(c_i)}{1 - G(T_i^*(c_i))}, \frac{1}{G(T_i^*(c_i))} \right\} \int_0^{T_i^*(c_i)} (1 - G(z))dz(c_i - (1 - \lambda)c_i) \right\}
\]

(23)

it is easily checked that the presence of the first term inside the max operator implies that (18) is satisfied, the second term implies (17) is satisfied, and the third term implies that (19) is satisfied.