Motivating with Simple Contracts*

Juan F. Escobar
University of Chile†

Carlos Pulgar‡
University of Chile

May 2016

Abstract

In practice, incentive schemes are rarely tailored to the specific characteristics of contracting parties. However, according to economic theory, optimal contracts should be highly dependent on individual conditions. We reconcile these observations in the context of a principal-agent model with both moral hazard and adverse selection. Motivating an agent could be increasingly costly to the principal because a more productive agent could also be more able to manipulate the terms of the contract. As a result, the principal may optimally pool some types by offering a contract with constant transfer and bonus. We also explore parameterizations where the optimal contract is fully separating but simple contracts attain a significant portion of the optimal welfare.

JEL classification numbers: D86, L51, L22

Keywords: Moral hazard, adverse selection, regulation, simple contracts

*This is a revised version of the master thesis by Carlos Pulgar at the University of Chile (2011). Escobar acknowledges financial support from Conicyt (Fondecyt No. 1130435) and from the Institute for Research in Market Imperfections and Public Policy, MIPP, ICM IS130002, Ministerio de Economía, Fomento y Turismo.

†Center for Applied Economics, Department of Industrial Engineering, University of Chile, Beauchef 851, Santiago, Chile, E-mail: `jescobar@dii.uchile.cl`

‡Center for Applied Economics, Department of Industrial Engineering, University of Chile, Beauchef 851, Santiago, Chile, E-mail: `pulgroso@gmail.com`

1 Introduction

In many contractual relationships, moral hazard and adverse selection frictions coexist. Consider an authority that must regulate a monopoly. The monopoly has private information about its technology, but can also exert an unobservable effort (physical or managerial investments) that could increase consumers' welfare. The total welfare is therefore randomly determined by both the firm's private information and its effort choice. Another example is that of a manager who has private information about his own skills to run a firm, but can also exert efforts to increases the firm's profits.

We study a relation in which a principal designs a contract to an agent who exerts an unverifiable effort and has varying degrees of productivity. As the contract theory literature has long recognized, even when productivity need not be commonly observed, it is beneficial for the principal to make a variety of offers, each of them targeting a particular type of agent (Myerson 1981, Laffont and Tirole 1986). However, such menus are rarely used in practice (McAfee and Schwartz 1994, Lafontaine and Slade 2001, Crew and Kleindorfer 2002, Chiappori and Salanié 2003, Vogelsang 2006, Armstrong and Sappington 2007).[1] The main contribution of this work is to provide a simple economic mechanism that bridges the gap between contract theory and some of its practical applications. In a nutshell, we show how the interaction of moral hazard and adverse selection opens up gaming opportunities for the agent that are absent in simpler models. As a result, it may be optimal for the principal to resort to simple contracts –that consist of a fixed transfer and a fixed performance bonus.

Our environment is similar to the moral hazard model in Grossman and Hart (1983), with the added complication that the distribution over signals is a function of the agent's private type (and the agent is risk-neutral). The agent exerts a privately observed effort $e \geq 0$ which, together with its private efficiency type $\theta \in [0, 1]$, determines a distribution over signals $y \in \{L, H\}$. Our assumption is that the larger the pair effort-type (e, θ), the more likely the high signal H. More subtly, effort and type are strict complements in that the marginal probability of a high signal with respect to effort (strictly) increases in the efficiency type. For example, a highly skilled manager can make a firm significantly more successful by marginally increasing his effort, whereas a barely skilled manager cannot. In the regulation example, an inefficient monopoly may attempt

[1] For example, Lafontaine and Slade (2001) note that, according to economic theory, optimal franchising contracts "should differ by outlet within a chain as well as across chains. Contracts that are observed in practice, in contrast, are remarkably insensitive to variations in individual... conditions. Indeed, most firms use a standard business-format franchise contract –a single combination of royalty rate and franchise fee– for all franchised operations joining the chain at a point in time." Vogelsang (2006) remarks that the US Federal Communications Commission attempted price regulation with menus but "this practice was abandoned after a few years. I am not aware of any serious discussion of the menu approach in practice."

to economize on its marginal production costs by redesigning its managerial structure, but such managerial changes will be much more beneficial when the firm's logistics and operations departments (about which the firm is much more informed than the regulator) are already efficient. While there are applications where our complementarity assumption does not hold, we think it is a good approximation for many real-world contractual situations.

In our framework, the principal can offer transfers and rewards (or bonuses) targeting different types of agents. These transfers and rewards will be optimally selected by the agent and determine the incentives the agent has to exert effort. Our first observation is that a high type agent must make strictly more effort than a low type. Intuitively, since the marginal probability of success increases in type, a high type agent has stronger incentives to choose high effort than a low type. These incentives are reinforced by the fact that to ensure self-selection, a high-type agent must be given a more generous performance bonus. An implementable effort scheme must have a slope bounded away from 0.

The problem of optimal contract design must balance two forces. One the one hand, for any implementable effort scheme, the information rent that a given type receives is determined by the room that he has to pretend to be a low productivity type, exert less effort, and save on effort costs. When the marginal product of types is (sufficiently) increasing, this sort of deviation will be attractive for the agent and the principal will need to give up substantial information rents to motivate the agent. Motivating the agent is therefore expensive and to economize on information rents the principal would rather implement a relatively flat effort scheme. On the other hand, to implement a relatively flat effort scheme the principal would need to offer a decreasing bonus and violate the slope constraint described in the paragraph above. It is therefore optimal for the principal to offer a simple, bunching contract at least to some subset of types.

We stress that under some parameterizations, the solution to the optimal contracting problem may be a fully separating menu. However, we show that simple contracts still do remarkably well. The intuition is that even when the shape of information rents may favor the design of separating bonuses and transfers, simple contracts can get close to that solution by properly distorting the size of the bonus. Using a Cobb-Douglas parameterization, we show that in the most pessimistic scenario, a principal using a simple contract gets at least 56% of the total welfare using separating contracts. This implies that the unmodeled benefits of simplicity (such as low administrative costs, reduction in renegotiation risks once types have been revealed, avoidance of choice overload effects) can be moderate for simple contracts to emerge as the unique optimal design.

In our model, a candidate contract for the principal is to sell the assets to the agent – who is assumed to be risk-neutral and wealth-unconstrained – and negotiate the selling price. After all, once assets are transferred, the agent has all the incentives to choose an efficient effort level (Grossman and Hart 1983). The principal would not know the value of the assets to the agent but, as in other adverse selection problems (Myerson 1981), the principal could optimally design a take-it-or-leave-it offer. Our characterization shows that negotiating a transfer of assets with the agent is not optimal for the principal. On the contrary, the principal optimally keeps a fraction of the production and distorts the effort in a less radical manner.[2] This observation provides an explanation for why governments keep participation in privatized firms, even when fully transferring the assets would make the privatized firm more efficient (Lewis and Sappington 2000).

We consider an alternative formulation where the benefits accruing to the principal are just a function of the agent's effort (and not the agent's type). In this alternative model, the agent's type determines the monitoring function but not the production function. This is the case, for example, when the signals are the result of an unproductive auditing process over which the agent is privately informed (for example, because the agent is more familiar with the auditing industry). Under this specification, the forces for pooling are even stronger than in our baseline model. Indeed, we show conditions under which all types are pooled, and effort is distorted both at the bottom and at the top. When signals are unproductive, simplicity becomes a more prominent feature of optimal contracts.

Several authors have studied mixed models of adverse selection and moral hazard (Picard 1987, Caillaud, Guesnerie, and Rey 1992, Jullien, Salanie, and Salanie 2007, Lewis and Sappington 2000, Balmaceda 2013, Gottlieb and Moreira 2013). The model in Lewis and Sappington (2000) is the most closely related to ours, but they add a limited liability constraint. Lewis and Sappington (2000) study a multi-agent model, but they do note that with one agent the optimal contract is pooling.[3] The force in that model is totally different from ours because the binding limited liability constraints push the fixed transfers to 0 and, as a result, there is no room to set a type dependent bonus. Ollier and Thomas (2013) also study a model similar to ours but they restrict attention to contracts that satisfy ex-post participation so the model becomes one with type-dependent outside options. Compared to all these studies, the simplicity of our model allows us to focus on the basic tensions that arise because the agent could deviate in two different dimensions –reporting and effort–, abstracting away from other important considerations, and clarifying the role that each informational friction has in the optimal contract.

Our characterization of incentive compatible menus is simple, but new in mixed models. Nailing down transfers and bonuses as a function of the effort scheme allows us to formulate the problem of optimal contract design as a simple optimal control problem. This problem

[2] When the principal attempts to sell the firm, the effort ends up being 0 if the firm is not sold.

[3] Lewis and Sappington (2000) also show that with two or more agents, there is an additional instrument which is the probability of not assigning the good.

is similar to others appearing in pure adverse selection problems (Myerson 1981, Lewis and Sappington 1989, Maggi and Rodriguez-Clare 1995, Jullien 2000, Ollier and Thomas 2013). In those models, some bunching can be optimal due to the the existence of countervailing incentives. We add to this literature by providing a new model and a new condition –namely, a restriction on the shape of the monitoring technology– that determines whether some pooling is optimal.

We also contribute to the literature on the performance of simple incentive contracts. Rogerson (2003) shows that a simple regulatory scheme can secure at least three-quarters of the surplus attained by the optimal (and complicated) regulatory policy. See also Chu and Sappington (2007), Chu and Sappington (2009), Bose, Pal, and Sappington (2011) and Garrett (2014). We provide conditions under which a simple contract is actually optimal and show that when those conditions are not met a simple contract can attain a significant portion of welfare achieved by the optimal contract.[4]

The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 characterizes incentive compatible schemes. Section 4 characterizes optimal contracts. Section 5 discusses the model with unproductive signals. Section 6 concludes.

2 Model

Our model has one principal and one agent. The agent makes an unobservable effort $e \in \mathbb{R}_+$ conditional on his privately known type $\theta \in [0, 1]$. The effort and his type jointly determine the realization of a random variable y with support $\{L, H\}$. Larger (finite) space of signals can be accommodated. In particular, we write

$$\mathbb{P}[y = H \mid e, \theta] = p(e, \theta)$$

with $p(e, \theta) \in [0, 1]$. The cost of effort e is given by $c(e) = e$. Given a monetary transfer w, the agent's payoff is $u(w) - e$, where u is strictly increasing. We will assume that $u(w) = w$ keeping in mind that our results about implementable schemes remain valid even when u is not linear as one can redefine the problem to work with the utility transfer $w' = u(w)$.

We will maintain the following restrictions throughout the paper.

Condition 1. *p is twice continuously differentiable and $p_e, p_\theta > 0$, $p_{ee} < 0$, and $p_{e\theta} > 0$.*

These restrictions simplify the analysis as they allow us to take first order conditions to characterize the solutions to the agent's maximization problem. We also impose that both e and

[4] The issue of simplicity in optimal contracts has a well-established tradition. Holmstrom and Milgrom (1987) show a set-up where optimal contracts are linear in total output. See Carroll (2015) for a recent treatment.

θ make high signals more likely. The restriction $p_{e\theta} > 0$ is a complementarity condition that says that a marginal increase in effort e is more productive the higher the type θ.

The principal can offer transfers w_L and w_H conditional on the realized output $y \in \{H, L\}$. The payoff to the agent is then given by

$$p(e,\theta)w_H + (1 - p(e,\theta))w_L - e.$$

Throughout the paper, we will relabel transfers by working with the fixed transfer $w \equiv w_L \in \mathbb{R}$ and a bonus $B = w_H - w_L$. We therefore define

$$U(e,\theta,w,B) = w + Bp(e,\theta) - e.$$

We restrict bonuses to be nonnegative, $B \geq 0$.[5] As our problem has adverse selection, the principal will offer menus of transfers, bonuses, and efforts $(w, B, e)\colon [0,1] \to \mathbb{R}^3$. The agent will report a type θ' and that report will determine the transfer $w(\theta')$ and the bonus $B(\theta')$ (that the agent receives only if $y = H$). The report θ' also determines an effort $e(\theta')$ that the principal *suggests* to the agent.[6] A menu $(w, B, e)\colon [0,1] \to \mathbb{R}^3$ is *incentive compatible* if for all $\theta \in [0,1]$,

$$(\theta, e(\theta)) \in \arg\max \Big\{ U(e', \theta, w(\theta'), B(\theta')) \mid \theta' \in [0,1], e' \in \mathbb{R} \Big\}. \tag{2.1}$$

That is, incentive compatibility imposes that it is in the agent's interest to report his true type and to follow the effort suggestion. Observe that nothing deters the agent from deviating by lying about his true type and/or making any effort. The value of the optimization problem (2.1) will be denoted $U(\theta \mid w, B, e)$ where we highlight the dependance of this quantity on the incentive compatible menus being offered.

We say that a profile of efforts (or effort scheme) $e\colon [0,1] \to \mathbb{R}$ is *implementable* if there exists $(w, B)\colon [0,1] \to \mathbb{R} \times \mathbb{R}_+$ such that (w, B, e) is incentive compatible. If (w, B) is such that for some interval $[\theta_1, \theta_2] \subseteq [0,1]$, $(w(\theta, B(\theta)) = (w(\theta_1, B(\theta_1)))$ for all $\theta \in [\theta_1, \theta_2]$, we say that (w, B) is *simple* or *bunching* at $[\theta_1, \theta_2]$ and e is implemented by a *simple* or *bunching contract* in $[\theta_1, \theta_2]$. If $\theta_1 = 0$ and $\theta_2 = 0$, we will say that (w, B) is simple or bunching and e is implemented by a simple or bunching contract.

[5] It will become clear that this is without loss. When $B < 0$ the agent does not make any effort. Therefore, we can set the bonus to 0 and keep incentives by redefining the fixed transfer w.

[6] Due to the generalized revelation principle, the restriction to this family of menus is without loss; see Myerson (1982) and Laffont and Martimort (2009).

3 Characterizing Incentive Compatible Contracts

It will be useful to define

$$G(e,\theta) = \frac{p_{e\theta}(e,\theta)}{-p_{ee}(e,\theta)}$$

for each $e \geq 0$ and $\theta \in [0,1]$. G captures the complementarity of effort and types in the probability of a high signal. Observe that G is strictly positive under our working assumptions. We also write $\gamma(\theta)|_0^\theta = \gamma(\theta) - \gamma(0)$, where $\gamma\colon [0,1] \to \mathbb{R}$.

We will now present our first result characterizing incentive compatibility.

Proposition 1. *Let $(w, B, e)\colon [0,1] \to \mathbb{R} \times \mathbb{R}_+^2$ with $e(\theta) > 0$ for all θ. If (w, B, e) is incentive compatible, the equilibrium value of type θ satisfies*

$$U(\theta \mid w, B, e) = U(0 \mid w, B, e) + \int_0^\theta \frac{p_\theta(e(s), s)}{p_e(e(s), s)} ds. \tag{3.1}$$

Moreover, the following are equivalent:

a. (w, B, e) is incentive compatible.

b. e is almost everywhere differentiable and

(i) For all $\theta \in [0,1]$, $w(\theta) = w(0) + e(\theta)|_0^\theta - \frac{p(e(\theta),\theta)}{p_e(e(\theta),\theta)}|_0^\theta + \int_0^\theta \frac{p_\theta(e(s),s)}{p_e(e(s)),s)} ds$;

(ii) For all $\theta \in [0,1]$, $B(\theta) = \frac{1}{p_e(e(\theta),\theta)}$;

and

(iii) For almost all $\theta \in [0,1]$, $\frac{de}{d\theta}(\theta) \geq G(e(\theta), \theta)$.

This proposition fully characterizes incentive compatible menus. The first part nails down the equilibrium utility of an agent with type θ by showing that the marginal utility of type θ is given by the slope of the iso-probability curve $\{(e', \theta) \mid p(e', \theta) = p(e(\theta), \theta)\}$. To see this result, note that a type $\theta' = \theta + \epsilon$, with $\epsilon > 0$ small, can always claim to be of type θ and keep the probability of success equal to $p(e(\theta), \theta)$ by exerting effort $e' = e(\theta) - \frac{p_\theta(e(\theta),\theta)}{p_e(e(\theta),\theta)}\epsilon$. This means that the utility of type $\theta + \epsilon$ must be at least

$$U(\theta + \epsilon \mid w, B, e) \geq p(e(\theta), \theta)B(\theta) + w(\theta) - e(\theta) + \frac{p_\theta(e(\theta), \theta)}{p_e(e(\theta), \theta)}\epsilon = U(\theta \mid w, B, e) - \frac{p_\theta(e(\theta), \theta)}{p_e(e(\theta), \theta)}\epsilon.$$

Therefore, $\frac{U(\theta+\epsilon|w,B,e) - U(\theta|w,B,e)}{\epsilon} \geq \frac{p_\theta(e(\theta),\theta)}{p_e(e(\theta),\theta)}$. We can also consider the incentives of type θ to exert more effort and deduce that his utility $U(\theta \mid w, B, e)$ must exceed $U(\theta + \epsilon \mid w, B, e) - \epsilon \frac{p_\theta(e(\theta),\theta)}{p_e(e(\theta),\theta)}$.

These two inequalities combine to yield our envelope formula (3.1). Informational rents are therefore determined by the slopes of the iso-probability curves along the effort scheme $e(\theta)$ and hazard rates. The second part of Proposition 1 characterizes incentive compatibility. Conditions (i) and (ii) are familiar from Myerson (1981) in that they narrow down monetary transfers as a function of the allocation $e\colon [0,1] \to \mathbb{R}_+$ being implemented. The novelty in our framework comes from the observation that the slope $\frac{de}{d\theta}(\theta)$ is bounded below by a strictly positive amount, $G(e(\theta),\theta)$. In particular, and in contrast to the more standard set up, implementable effort schemes cannot pool types.

To see (iii), note that $e(\theta) > 0$ must solve $\max_{e'\geq 0} B(\theta)p(e',\theta) - e'$. This implies that $B(\theta)p_e(e(\theta),\theta) = 1$. Since $B(\theta)$ is increasing –otherwise, some type of agent would find attractive to misreport his true type– $p_e(e(\theta),\theta)$ must be decreasing. This implies that $p_{ee}(e(\theta),\theta)\frac{de}{d\theta} + p_{e\theta}(e(\theta),\theta) \leq 0$ and therefore (iii) holds. In other words, condition (iii) arises due to the fact that the agent must have incentives to reveal truthfully (so that B increases) and to exert effort (so that $B(\theta)p_e(e(\theta),\theta) = 1$). The combination of adverse selection and moral hazard results in the differential inequality $\frac{de}{d\theta} \geq G(e(\theta),\theta)$.

As suggested by the above discussion, the bonus is an important determinant of the slope of e.

Corollary 1. *Let (w, B, e) be incentive compatible. Then, for all $\theta \in [0,1]$ in which e is differentiable and $e(\theta) > 0$,*

$$\dot{e}(\theta) = G(e(\theta),\theta) \quad \textit{iff} \quad \dot{B}(\theta) = 0.$$

This corollary provides necessary and sufficient conditions for the emergence of constant bonuses in a neighborhood of a type θ. This corollary will allow us to deduce the simplicity of a contract just by checking whether the differential inequality binds. Also note that incentive compatibility implies that $w(\theta) = w(\theta')$ whenever $B(\theta) = B(\theta')$.

4 Optimal Contracts

4.1 Formulation

In this section we consider the design of optimal contracts and show how the fact that the contract is signed under uncertain monitoring (e.g. there is both adverse selection and moral hazard) impacts the efficiency and rents of optimal arrangements.

The gross benefit given an effort $e \geq 0$ and a type $\theta \in [0,1]$ is $\Pi(e,\theta)$. Define $U^P = \Pi(e,\theta) - T$, and $U^A = T - e$, where $T = w + p(e,\theta)B$. We will assume that the principal's goal is to maximize

$U^P + \lambda U^A$, with $\lambda \in [0,1[$. In a standard principal-agent framework (Grossman and Hart 1983), the principal's goal equals U^P and therefore $\lambda = 0$. However, we consider a slightly more general formulation to accommodate applications to regulation and procurement (Baron and Myerson 1982). In regulation models, U^P represents the consumers' surplus while U^A represents the firm's surplus and the principal (or regulator) maximizes weighted total surplus.

We focus on contracts that maximize the expectation of $U^P + \lambda U^A$, over all menus $(w, B, e) \colon [0,1] \to \mathbb{R}^3$. The restrictions in the problem are the incentive compatibility constraints and the participation constraints of the agent:

$$U(\theta \mid w, B, e) \geq 0, \quad \forall \theta \in [0,1].$$

The participation constraints capture the idea that the agent, knowing the type θ, has an outside option that guarantees a payoff normalized to 0.

Before solving for the optimal contract, we note that if there is either pure moral hazard (types are observable) or pure adverse selection (actions are observable), the principal can always attain the first best (Section 6.3 in Bolton and Dewatripont 2005). All distortions in our model will come from the interaction between adverse selection and moral hazard.

To solve the optimal contract with adverse selection and moral hazard, we restrict attention to contracts in which $e(0) > 0$ and use Proposition 1 and integration by parts to deduce a formula for the expected transfer:

$$\begin{aligned}
&\int \{w(\theta) + p(e(\theta),\theta)B(\theta)\} f(\theta)d\theta \\
&\quad = w(0) + p(e(0),0)\frac{1}{p_e(e(0),0)} - e(0) + \int \left(e(\theta) + \int_0^\theta \frac{p_\theta(e(s),s)}{p_e(e(s),s)} ds \right) f(\theta)d\theta \\
&\quad = U(0 \mid w, B, e) + \int \left(e(\theta) + \frac{1-F(\theta)}{f(\theta)} \frac{p_\theta(e(\theta),\theta)}{p_e(e(\theta),\theta)} \right) f(\theta)d\theta
\end{aligned}$$

where $U(\theta \mid w, B, e)$ is the utility level that type $\theta \in [0,1]$ attains. The objective function becomes

$$\int \Big\{ \Pi(e(\theta),\theta) - e(\theta) - (1-\lambda)Q(e(\theta),\theta) \Big\} f(\theta)d\theta - (1-\lambda)U(0 \mid w, B, e), \tag{4.1}$$

where $Q(e,\theta) = \frac{1-F(\theta)}{f(\theta)} \frac{p_\theta(e,\theta)}{p_e(e,\theta)}$ represents the informational rents that the agent receives. The function $Q(e,\theta)$ increases in e and therefore it is effectively the cost of providing incentives to an agent of type θ.

Since $U(\theta \mid w, B, e)$ increases in θ, ensuring participation of all types boils down to

$$U(0 \mid w, B, e) \geq 0. \tag{4.2}$$

From the previous section, the incentive constraints reduce to

$$\frac{de}{d\theta} \geq G(e(\theta), \theta) \tag{4.3}$$

The problem of maximizing weighted payoffs subject to participation and incentive constraints can be written as

$$\max_{(e,w,B)\colon [0,1]\to\mathbb{R}^3} \int \Big(\Pi(e(\theta),\theta) - e(\theta) - (1-\lambda)Q(e(\theta),\theta)\Big) f(\theta)d\theta - (1-\lambda)U(0 \mid w, B, e) \tag{4.4}$$

subject to (4.2)-(4.3). This maximization problem (4.5) has some features that are familiar from standard mechanism design problems (Myerson 1981). The objective function equals expected total welfare, $\Pi(e,\theta) - e$, minus (weighted) informational rents, $(1-\lambda)Q(e,\theta)$. The incentive constraint takes a rather special form as it restricts $e\colon [0,1] \to \mathbb{R}_+$ to have a slope bounded away from 0. One alternative to solve this problem is to take

$$\hat{e}(\theta) \in \arg\max_{e\geq 0} \Big(\Pi(e,\theta) - e - (1-\lambda)Q(e,\theta)\Big) \tag{4.5}$$

and $w(0)$ such that $U(0 \mid w, B, e) = 0$. This results in a natural candidate to solve (4.4). Indeed, when $\hat{e}$ satisfies (4.3), $\hat{e}$ is a indeed solution to (4.4). In particular, when (4.3) is slack for all θ, the solution is implemented by a non-bunching, type-dependent, menu $(w, B)\colon [0,1] \to \mathbb{R}$ as in standard adverse selection problems. In this case, the solution has the familiar features from screening models: (i) there are no distortions at the top; (ii) the effort scheme increases in types.

However, even under our working assumptions and natural restrictions on the distribution F, $\hat{e}$ need not satisfy restriction (4.3) everywhere. This happens as the principal's objective function is total surplus minus information rents, whereas the incentive constraints impose bounds on the slope of implementable effort schemes that need not be consistent with the principal's goal.

The problem of finding an optimal effort scheme reduces to the following optimization problem:

$$\max_{e\colon [0,1]\to\mathbb{R}_+} \int \Big(\Pi(e(\theta),\theta) - e(\theta) - (1-\lambda)Q(e(\theta),\theta)\Big) f(\theta)d\theta \tag{4.6}$$

subject to

$$\frac{de(\theta)}{d\theta} \geq G(e(\theta), \theta). \tag{4.7}$$

Denote $H(\theta) = \frac{1-F(\theta)}{f(\theta)}$ and $\Phi(e,\theta) = \frac{p_\theta(e,\theta)}{p_e(e,\theta)}$.[7] We will impose the following restriction in the rest of the paper.

Condition 2. *(a) Π and Q are twice continuously differentiable and $-\Pi_{ee}, \Phi_{ee} > 0$; (b) G is twice continuously differentiable and $G_{ee} \leq 0$; (c) $\hat{e} > 0$.*

These are convexity restrictions on the problem that ensure the validity of necessary optimality conditions. Note that under these restrictions, $\hat{e}(\theta)$ turns out to be the unique maximizer of (4.5). Condition (c) simplifies the analysis as, without it, one would need to consider cases in which the agent does not produce and is not rewarded.

We can exploit optimality conditions for optimal control problems,[8] to find solutions to (4.6). Those optimality conditions can be used to prove the following useful characterization.

Lemma 1. *Let $e^*\colon [0,1] \to \mathbb{R}_+$ be a solution to (4.6). If $\frac{de^*}{d\theta} > G(e^*(\theta),\theta)$ in some interval $[\underline{\theta}, \bar{\theta}]$, then $e^*(\theta) = \hat{e}(\theta)$ in $[\underline{\theta}, \bar{\theta}]$.*

This lemma says that whenever $e^*(\theta)$ satisfies (4.7) with slack, it must be a point wise solution to the objective function. This result is similar to those appearing in the analysis of mechanism design problems where the monotonicity constraint binds (Fudenberg and Tirole 1991, Jullien 2000).

4.2 Productive Signals

We consider the case in which the signal $y \in \{H, L\}$ equals production. We therefore write the production function as $\Pi(e,\theta) = \Delta p(e,\theta)$, where $\Delta > 0$ is the value of a high over a low signal.

The following result shows a necessary and sufficient condition for the effort scheme $\hat{e}$ to violate the slope condition (4.7).

Lemma 2. *For all $\theta \in [0,1]$*

$$\frac{d\hat{e}}{d\theta} \leq G(\hat{e}(\theta),\theta) \text{ iff } \quad \Big(H_\theta(\theta)\Phi_e((\hat{e}(\theta),\theta)) + H(\theta)\Phi_{e\theta}(\hat{e}(\theta),\theta)\Big) + \frac{p_{e\theta}(\hat{e}(\theta),\theta)}{-p_{ee}(\hat{e}(\theta),\theta)} H(\theta)\Phi_{ee}(\hat{e}(\theta),\theta) \geq 0.$$

[7] $\frac{1}{H(\theta)} = \frac{f(\theta)}{1-F(\theta)}$ is the hazard rate of the distribution F.

[8] The optimization problem (4.6) can be formulated as an optimal control problem by replacing the differential inequality $\frac{de}{d\theta} \geq G(e(\theta),\theta)$ by

$$\frac{de}{d\theta} = G(e(\theta),\theta) + u(\theta)$$

with $u(\theta) \geq 0$. We formulate the problem in the space of absolutely continuous functions $e\colon [0,1] \to \mathbb{R}_+$ and impose the differential inequality everywhere. We conjecture this formulation is without loss.

To understand this result, note that

$$\hat{e}(\theta) = \arg\max_{e\geq 0}\{\Delta p(e,\theta) - e - (1-\lambda)H(\theta)\Phi(e,\theta)\}$$

whereas the first best solution $e^{FB}(\theta) = \arg\max_{e\geq 0}\{\Delta p(e,\theta) - e\}$. The slope of the first best solution equals $G(e^{FB}(\theta),\theta)$ by construction. Whether the slope of $\hat{e}$ is greater than or equal to $G(\hat{e}(\theta),\theta)$ will depend on the sign of the cross derivative of $H(\theta)\Phi(e,\theta)$. When the cross derivative is greater than or equal to 0 , then the slope of $\hat{e}$ will be less than or equal to $G(\hat{e}(\theta),\theta)$. Intuitively, when $Q_{e\theta} \geq 0$, the marginal information cost increases in θ and therefore this is a force towards a flatter $\hat{e}$. This is precisely the content of the first term on the right side. The second term is a correction accounting for the fact that a very convex Q is also a force towards a flatter $\hat{e}$.

The condition on the right side in Lemma 2 does not hold at $\theta = 1$. When H is increasing and $\Phi_{e\theta} < 0$, $\hat{e}$ is the optimal effort scheme and the optimal contract is such that all types are separated. This is presented in the following result.

Proposition 2. *Suppose that H is decreasing, $\Phi_{e\theta} < 0$, and $\Phi_{ee} = 0$. Then, the optimal contract is such that $B^*\colon [0,1] \to \mathbb{R}_+$ is strictly increasing and $w^*\colon [0,1] \to \mathbb{R}$ is strictly decreasing.*

Proposition 2 applies when $p(e,\theta) = (e\theta)^{1/2}$ and $F(\theta) = \theta$. Yet, no natural economic restrictions can be given to sustain those conditions. For example, when $F(\theta) = \theta^\gamma$, with $\gamma > 0$, and we want to model a principal who is pessimistic about the parameter θ, we would restrict $\gamma < 1$.[9] In this case, H is increasing for low values of θ and therefore Proposition 2 does not apply. As shown in the following proposition, the shape of the contract can be very different in those cases.

Proposition 3. *Assume that $p(e,\theta) = e^\alpha h(\theta)$, with $0 < \alpha < 1$ and $h' > 0$. Assume that there exists $\bar{\theta} \in]0,1[$ such that*

$$\frac{d}{d\theta}\Big(\ln(\frac{h'(\theta)}{h(\theta)})\Big) \geq \frac{d}{d\theta}\Big(\ln(\frac{1}{H(\theta)})\Big) \quad \textit{iff} \quad \theta \leq \bar{\theta}.$$

Then, there exists $\theta^* \geq \bar{\theta}$ such that the following hold.

a. $\frac{de^*}{d\theta} = G(e^*(\theta),\theta)$ *for all* $\theta \leq \theta^*$*. For* $\theta > \theta^*$, $e^*(\theta) = \hat{e}(\theta)$.

b. The optimal payment scheme satisfies $(B^*(\theta), w^*(\theta)) = (B^*(\theta^*), w^*(\theta^*))$ *for all* $\theta \leq \theta^*$. B^* *is strictly increasing in* $[\theta^*,1]$ *and* w^* *is strictly decreasing in* $[\theta^*,1]$*. For all* $\theta \in [0,1[$, $B^*(\theta) < \Delta$.

[9] When $\gamma < 1$, $F(\theta) = \theta^\gamma$ is dominated (in the first-order stochastic dominance sense) by the uniform distribution in $[0,1]$.

c. e^* and $\hat{e}$ cross once over the interval $[0, \bar{\theta}]$, and $e^*(0) < \hat{e}(0)$.

Moreover, when $\theta^* \in]\bar{\theta}, 1[$, it satisfies the equation

$$\frac{H(\theta^*)h'(\theta^*)}{h(\theta^*)} \int_0^{\theta^*} h(\theta)^{\frac{1}{1-\alpha}} f(\theta)d\theta = \int_0^{\theta^*} h(\theta)^{\frac{\alpha}{1-\alpha}} H(\theta)h'(\theta)f(\theta)d\theta.$$

This proposition shows that when the rate at which $\frac{h'}{h}$ grows is above the rate at which $\frac{1}{H(\theta)}$ grows in some interval $[0, \bar{\theta}]$, then all types $[0, \bar{\theta}]$ are offered the same bonus and the same transfer. When $\frac{h'}{h}$ is sufficiently increasing, marginal informational rents increase with θ. In this case, to economize on informational rents, the effort scheme $\hat{e}$ increases slowly. But the principal faces an incentive compatibility restriction and such effort scheme is not feasible. As a result, the optimal effort scheme e^*, (4.7) binds on $[0, \bar{\theta}]$. Corollary 1 implies that e^* is implemented by a bunching contract in $[0, \bar{\theta}]$.

The optimal solution distorts the effort scheme of all types in $[0, \bar{\theta}]$. Observe that even when the principal is not constrained to pool types in $[\bar{\theta}, \theta^*]$ it is optimal to do it. Intuitively, by choosing $\theta^* > \bar{\theta}$, the principal distorts the effort scheme of types in $[\bar{\theta}, \theta^*]$ but reduces the distortions to low types. Figure 1 illustrates the optimal effort scheme e^*.

The following examples show parameterizations where Proposition 3 can be applied.

Example 1. *Suppose that $p(e, \theta) = (\frac{e}{2-\theta})^{1/2}$ and $F(\theta) = \theta^{1/2}$. The sufficient condition for bunching holds by taking $\bar{\theta} = 6 - 4\sqrt{2}(\approx .34)$. All types below $\theta^* \approx .67$ choose the same contract.*

Example 2. *Suppose that $p(e, \theta) = \frac{1+\theta^2}{2} e^{1/2}$ and $F(\theta) = \theta$. The sufficient condition for bunching holds by taking $\bar{\theta} = -1 + (2)^{1/2}(\approx .41)$. All types below $\theta^* \approx .64$ choose the same contract.*

Example 3. *Suppose that $p(e, \theta) = \frac{1}{3^n} \exp(\theta^n) e^{1/2}$ with $n \geq 1$ and $F(\theta) = \theta$. The sufficient condition for bunching holds by taking $\bar{\theta} = \frac{n-1}{n}$. The cutoff type $\theta^* = \theta^*(n)$ converges to 1 as n goes to infinity.*

4.3 Bounds on Simple Contracts

The results in the previous sections show conditions under which the optimal effort scheme is implemented by a bunching contract. Yet, it is still possible that at the optimal effort scheme the slope condition (4.7) is slack everywhere and the optimal menu $(B(\theta), w(\theta))$ fully separates types. This is the case, for example, when $p(e, \theta) = e^\alpha \theta^{1-\alpha}$, $\Pi(e, \theta) = \Delta p(e, \theta)$, $F(\theta) = \theta$. In this section we ask how important is the principal's payoff loss when the contract is restricted to be simple.

Define $W(\lambda, \alpha)$ as the value of the objective function (4.6) subject to the constraint that the contract is simple. As shown in Corollary 1, the restriction to simple contracts is captured by the differential equation

$$\frac{de}{d\theta}(\theta) = G(e(\theta), \theta).$$

On the other, let $\hat{W}(\lambda, \alpha)$ be the value of the objective (4.6) given $\hat{e}$. Given our parametric restrictions, the real number $\hat{W}$ is the value of the optimal contracting problem (4.6) subject to (4.7). Defining

$$R(\lambda, \alpha) = \frac{W(\lambda, \alpha)}{\hat{W}(\lambda, \alpha)}$$

we get the total value that simple contracts can achieve as a percentage of the optimal value (when incentive and participation constraints are taken into account).

Proposition 4. *Assume $p(e, \theta) = e^{\alpha}\theta^{1-\alpha}$, $\Pi(e, \theta) = \Delta p(e, \theta)$, $F(\theta) = \theta$, with $\lambda < 1$ and $\alpha \in]0, 1[$. Then,*

$$R(\lambda, \alpha) = \frac{1}{2} \frac{1}{\left(1 + (1-\lambda)\frac{1-\alpha}{\alpha}\right)^{\frac{\alpha}{1-\alpha}} \int_0^1 \frac{\theta}{\left[1+(1-\lambda)\frac{1-\alpha}{\alpha}\left(\frac{1-\theta}{\theta}\right)\right]^{\frac{\alpha}{1-\alpha}}} d\theta}.$$

In particular, for all λ and all α,

$$R(\lambda, \alpha) \geq \frac{1}{1/4 + 2\int_{\frac{1}{2}}^{1} \theta e^{\left(\frac{2\theta-1}{\theta}\right)(1-\lambda)} d\theta},$$

$R(1, \alpha) = 1$ and $R(\lambda, \alpha) \geq 0.56$.

This result shows that even when simple contracts may not be optimal, they can perform remarkably well. In particular, simple contracts capture at least 56% of the welfare that unrestricted optimal contracts achieve. The (unmodeled) benefits of simplicity need not be substantive to favor the use of an optimal simple contract.[10]

To find the optimal simple contract, we find a value for the initial conditions such that the curve $\frac{de}{d\theta} = G(e(\theta), \theta)$ gets a total welfare the closest to that of $\hat{e}(\theta)$. When $\lambda = 1$, this is simple as $\hat{e}$ is actually implemented by a simple contract. But this is not feasible in general and the simple contract will give up some rents. See Figure 2.

[10] When $\alpha = 1/2$ and $\lambda = 0$, $R(\lambda, \alpha = 0.75)$.

5 Discussion: Certain Production Functions

We now study a variation of our main model and assume that the production function is certain.

Definition 1. *The production function* $\Pi(e,\theta)$ *is* certain *if* $\Pi(e,\theta)$ *depends on* e *but not on* θ. *In this case, we write* $\Pi(e,\theta) = \pi(e)$.

When $\Pi(e,\theta)$ is a non-trivial function of θ, the principal is not only uncertain about the monitoring technology but also about the production technology. While in standard moral hazard problems, production and monitoring coincide, in many cases of interest they differ. As in Holmstrom and Milgrom (1991), the principal may get signals $y \in \{H, L\}$ about the agent's effort e that need not be productive by itself. For example, the principal may randomly audit the agent's effort performance e, but the outcome of such audit (or signals) $y \in \{L, H\}$ is not productive. Moreover, the agent may have superior, private information about how likely different auditing outcomes y are determined by his effort, possibly because the agent has already experienced other audits or is more familiar with the auditing industry.

When the production function is certain, the first best effort is type-independent and determined by $\pi_e(e^{FB}) = 1$. When θ is commonly known, the principal can implement e^{FB} by offering a bonus $B^{FB}(\theta) = \frac{1}{p_e(e^{FB},\theta)}$ and extracting all the agent's rents by properly setting $w_L(\theta)$. The bonus B^{FB} decreases in θ.[11]

Proposition 5. *Suppose that the production function is certain and that*

$$\frac{(1-\lambda)H_\theta(\theta)\Phi_e(e,\theta)}{\pi_{ee}(e)-(1-\lambda)H(\theta)\Phi_{ee}(e)} + \frac{(1-\lambda)H(\theta)\Phi_{e\theta}(e,\theta)}{\pi_{ee}(e)-(1-\lambda)H(\theta)\Phi_{ee}(e)} \leq \frac{p_{e\theta}(e,\theta)}{-p_{ee}(e,\theta)}. \tag{5.1}$$

for all $e = \hat{e}(\theta)$ *and all* $\theta \in [0,1]$. *Then,*

1. *At the optimal effort scheme* e^*, $\frac{de^*}{d\theta}(\theta) = G(e^*(\theta),\theta)$ *for all* $\theta \in [0,1]$.
2. e^* *is implemented by a bunching contract* $(w^*, B^*) \in \mathbb{R} \times \mathbb{R}_+$.
3. e^* *and* $\hat{e}$ *cross once at some* $\theta' \in]0,1[$ *and* $e^*(\theta) \geq \hat{e}(\theta)$ *iff* $\theta \geq \theta'$.

Note that the optimal effort scheme is distorted everywhere. More specifically, there exists a cutoff type θ^* such that $e^*(\theta) < e^{FB}$ for $\theta < \theta^*$ and $e^*(\theta) > e^{FB}$ for $\theta > \theta^*$. Since a constant

[11] When θ is not known, this contract is clearly not incentive compatible. One could think that the principal could just sell the firm to the agent and implement the first-best effort $e^{FB} \in \arg\max \{\Pi(e) - e\}$. This contract, however, is not available in our setting as the only contractible variable is $y \in \{H, L\}$. The value of the firm $\pi(e)$ cannot be transferred to the agent.

effort scheme is not feasible, in the optimal contract higher types exaggerate their efforts whereas lower types exert lower effort levels. See Figure 3.

When signals $y \in \{H, L\}$ are also productive, production is uncertain. Suppose that the production function takes the form

$$\Pi(e,\theta) = \pi(e) + \Delta p(e,\theta)$$

where $\Delta > 0$ is the added value of a high over a low signal. When (5.1) holds with slackness for $e = \hat{e}^0(\theta)$ and

$$\hat{e}^0(\theta) \in \arg\max_{e \geq 0} \Big(\pi(e) - c(e) - (1-\lambda)Q(e,\theta)\Big),$$

it is easy to see that when Δ is small enough, the effort scheme e^* solving (4.5) satisfies all the conclusions of Proposition 5.

6 Conclusions

In their survey of the optimal regulation literature, Armstrong and Sappington (2007) note that policy makers "propose relatively simple regulatory policies that appear to have some desirable properties, even if they are not optimal in any precise sense." The purpose of this paper is to provide a rationale for the use of simple contracts in a rather rich model of adverse selection and moral hazard. Motivating an agent to exert effort could be increasingly expensive as he becomes more efficient and therefore it is optimal for the principal to offer a simple contract in which the effort scheme increases slowly. While our model is highly specialized, we think that the tradeoff identified in this work will appear in more complicated models of moral hazard and adverse selections (for example, when the agent is risk averse, or when the principal needs to motivate two or more privately informed agents).

It is also possible to use our model to explore issues such as specific human capital accumulation[12] or the conditions under which the assets should be sold to the agent.[13] These extensions are left for future research.

[12]Pulgar (2011) studies a model in which specific human capital θ is privately acquired by the agent at a private cost $c(\theta)$ before the principal-agent relationship described in Section 2 takes place. The equilibrium distribution of types (which critically depends on the monitoring technology) is derived and several comparative statics results are obtained.

[13]Selling the firm to the agent is not optimal in our baseline model in Section 2, but it could be optimal in the model in Section 5 when the incentive contract is not flexible enough.

Appendix

Proofs for Section 3

Proof of Proposition 1. Define the utility function

$$U(w, B, \theta) = w + F(B, \theta)$$

where $F(B, \theta) = \max_{e \geq 0} \Big(Bp(e, \theta) - w \Big)$. Then, we are in the standard mechanism design setting, where B is the quantity and w is the transfer (Myerson 1981, Milgrom 2004). The added restriction is that the function $e(\theta)$ is optimal for the agent. We can therefore use well-known results on incentive compability to deduce that (w, B, e) is incentive compatible if

$$B \text{ is nondecreasing and } w(\theta) = U(0 \mid w, B, e) - F(B(\theta), \theta) + \int_0^\theta B(s) p_\theta(e(s), s) ds$$

and $e(\theta) \in \arg\max_{e'}\{B(\theta)p(e, \theta) - e\}$. Note that the conditions (i) B is nondecreasing, and (ii) $e(\theta) \in \arg\max_{e'}\{B(\theta)p(e, \theta) - e\}$ can be equivalently written as

$$B(\theta) p_e(e(\theta), \theta) = 1$$

and the function $\theta \mapsto p_e(e(\theta), \theta)$ is decreasing. Imposing $\frac{d}{d\theta}\big(p_e(e(\theta), \theta)\big) \leq 0$ for almost all θ, the Proposition follows. □

Proofs for Section 4

We now write the Hamiltonian of the problem of optimal control:

$$H(e, u, \theta, \beta) = -\Big(\Pi(e, \theta) - e - (1 - \lambda)Q(e, \theta)\Big) + \beta(G(e, \theta) + u).$$

The optimality conditions are

$$\frac{d\beta}{d\theta}(\theta) = H_e(e(\theta), u(\theta), \beta(\theta), \theta)$$

$$u(\theta) \in \arg\min_{u' \geq 0}\{H(e(\theta), u', \theta, \beta(\theta))\},$$

and

$$\frac{de}{d\theta} = G(e(\theta), \theta) + u(\theta)$$

with $\beta(1) = \beta(0) = 0$. These optimality conditions are necessary and sufficient since Condition 2 implies that the problem is convex (Bertsekas 1995). Since $u(\theta)$ minimizes the Hamiltonian, $\beta(\theta) \geq 0$ for all θ.

Proof of Lemma 1. To see this lemma, just note that if $\frac{de^*}{d\theta} > G(e^*, \theta)$, then $u^*(\theta) > 0$. It follows that $\beta(\theta) = 0$ (otherwise $u^*(\theta) > 0$ would not minimize the Hamiltonian). It follows that $\frac{d\beta}{d\theta} = 0$ and therefore $e^*(\theta)$ satisfies $\Pi_e(e^*(\theta), \theta) = 1 + (1-\lambda)Q_e(e^*(\theta), \theta)$. Since $\Pi(e, \theta) - e - (1-\lambda)Q(e, \theta)$ is strictly concave in e, we deduce that $e^*(\theta) = \hat{e}(\theta)$. $\square$

Proof of Lemma 2. For all $\theta \in [0, 1]$, the solution $\hat{e}(\theta)$ satisfies the first order condition

$$\Delta p_e(e, \theta) = 1 + H(\theta)\Phi_e(e, \theta).$$

Taking derivatives with respect to θ, we obtain

$$\frac{d\hat{e}}{d\theta}\Big(\Delta p_{ee}(\hat{e}(\theta), \theta) - H(\theta)\Phi_{ee}(\hat{e}(\theta), \theta)\Big) = H(\theta)\Phi_{e\theta}(\hat{e}(\theta), \theta) - \Delta p_{e\theta}(\hat{e}(\theta), \theta) + H_\theta \Phi_e(\hat{e}(\theta), \theta)$$

It follows that $\frac{d\hat{e}}{d\theta} \leq G(\hat{e}(\theta), \theta)$ iff

$$\Big(H_\theta(\theta)\Phi_e((\hat{e}(\theta), \theta)) + H(\theta)\Phi_{e\theta}(\hat{e}(\theta), \theta)\Big)(-p_{ee}(\hat{e}(\theta), \theta)) + p_{e\theta}H(\theta)\Phi_{ee}(\hat{e}(\theta), \theta) \geq 0.$$

$\square$

Proof of Proposition 3. First note that in $[0, \bar{\theta}]$, $\frac{de^*}{d\theta} = G(e^*(\theta), \theta)$. Otherwise, Lemma 1 would imply that $e^*(\theta) = \hat{e}(\theta)$. Since $\frac{d\hat{e}}{d\theta} < G(\hat{e}(\theta), \theta)$ for $\theta \in [0, \bar{\theta}]$, we would deduce that e^* is not feasible. Let $[0, \theta^*]$ be the largest interval such that $\frac{de^*(\theta)}{d\theta} = G(e^*(\theta), \theta)$ on $[0, \theta^*]$. We claim that for $\theta > \theta^*$, $e^*(\theta) = \hat{e}(\theta)$. To see this, note that for $\theta > \theta^*$ close to θ^*, $\frac{de^*(\theta)}{d\theta} > G(e^*(\theta), \theta)$ by construction of θ^* and therefore $e^*(\theta) = \hat{e}(\theta)$ for all $\theta \geq \theta^*$ close to θ^*. Since $\hat{e}(\theta)$ satisfies the differential inequality in $[\theta^*, 1]$, it is clear that the objective cannot be improved on the interval $[\theta^*, 1]$ by picking an effort scheme different from $\hat{e}$. Finally to see the formula for θ^* just note that θ^* must be a solution to the maximization problem

$$\max_{\theta^* \in [\bar{\theta}, 1]} \int_0^{\theta^*} \Big(\Pi(e^*(\theta), \theta) - e^*(\theta) - (1-\lambda)Q(e^*(\theta), \theta)\Big) d\theta + \int_{\theta^*}^1 \Big(\Pi(\hat{e}(\theta), \theta) - \hat{e}(\theta) - (1-\lambda)Q(\hat{e}(\theta), \theta)\Big) d\theta$$

subject to

$$e^*(\theta) = G(e^*(\theta), \theta), \forall \theta \in [0, \theta^*], \text{ and } e^*(\theta^*) = \hat{e}(\theta^*).$$

Denoting

$$A(\theta^*) = \Big(\frac{\Delta\alpha}{1 + \frac{H(\theta^*)h'(\theta^*)(1-\lambda)}{\alpha h(\theta^*)}}\Big)^{1/(1-\alpha)}$$

and solving the differential equation, we can rewrite the constraints in the optimization as

$$e^*(\theta) = A(\theta^*)h(\theta)^{1/(1-\alpha)}.$$

We therefore reduce the problem to

$$\max_{\theta^*\in[\bar{\theta},1]} \int_0^{\theta^*} \Big(\Delta A(\theta^*)^\alpha h^{1/(1-\alpha)} - A(\theta^*)h^{1/(1-\alpha)} - A(\theta^*)h^{1/(1-\alpha)}(((1-\lambda)H(\theta)h(\theta)')/\alpha h(\theta)\Big) f(\theta)d\theta +$$
$$\int_{\theta^*}^1 \Big(\Delta\hat{e}(\theta)^\alpha h(\theta) - \hat{e}(\theta) - \hat{e}(\theta)(((1-\lambda)H(\theta)h'(\theta))/\alpha h(\theta))\Big) f(\theta)d\theta$$

Taking first order conditions and rearranging terms, we obtain the condition for θ^* in the statement of the result. □

Proof of Proposition 4. Let us first find the formula for R. Solving the unrestricted problem we get

$$\hat{W} = \int_0^1 \theta \frac{\Delta^{\frac{1}{1-\alpha}}\left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right)}{\left(1 + (1-\lambda)\left(\frac{1-\theta}{\theta}\right)\left(\frac{1-\alpha}{\alpha}\right)\right)^{\frac{\alpha}{1-\alpha}}} d\theta$$

To find the value of the optimal bunching contract, note that for any such contract it must be that $e(\theta) = A\theta$ for some A. Plugging into the welfare function and optimizing the value of A we obtain

$$A^{1-\alpha} = \frac{\Delta\alpha}{1 + (1-\lambda)\left(\frac{1-\alpha}{\alpha}\right)}.$$

The welfare results in

$$W = \frac{\Delta^{\frac{1}{1-\alpha}}}{2} \frac{\left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right)}{\left(1 + (1-\lambda)\left(\frac{1-\alpha}{\alpha}\right)\right)^{\frac{\alpha}{1-\alpha}}}$$

We deduce that

$$R(\lambda,\alpha) = \frac{\frac{1}{2}}{\left(1 + (1-\lambda)\left(\frac{1-\alpha}{\alpha}\right)\right)^{\frac{\alpha}{1-\alpha}} \int_0^1 \frac{\theta}{\left(1+(1-\lambda)\left(\frac{1-\theta}{\theta}\right)\left(\frac{1-\alpha}{\alpha}\right)\right)^{\frac{\alpha}{1-\alpha}}} d\theta}.$$

Now,let $u(\theta) = \frac{1-\theta}{\theta} > 0$ and $r = \frac{\alpha}{1-\alpha} > 0$. We can write

$$R(\lambda,\alpha)^{-1} = 2\int_0^1 \theta \left[\frac{r/(1-\lambda)+1}{r/(1-\lambda)+u(\theta)}\right]^r d\theta \leq 2\Big(\int_0^{1/2} \theta d\theta + \int_{1/2}^1 \theta \left[\frac{r/(1-\lambda)+1}{r/(1-\lambda)+u(\theta)}\right]^r d\theta\Big).$$

For $\theta > 1/2$, the function $r \mapsto \left[\frac{r/(1-\lambda)+1}{r/(1-\lambda)+u(\theta)}\right]^r$ is increasing (just note that the derivative is strictly positive). Therefore, for all $\theta > 1/2$ and all $r > 0$

$$\theta\left[\frac{r/(1-\lambda)+1}{r/(1-\lambda)+u(\theta)}\right]^r < \lim_{r\to\infty}\theta\left[\frac{r/(1-\lambda)+1}{r/(1-\lambda)+u(\theta)}\right]^r = \theta\exp\big((1-u(\theta))(1-\lambda)\big).$$

It follows that

$$R(\lambda,\alpha)^{-1} \leq \frac{1}{4} + 2\int_{1/2}^{1}\theta\exp\big(\frac{2\theta-1}{\theta}(1-\lambda)\big)d\theta$$

and therefore

$$R(\lambda,\alpha) \geq \frac{1}{1/4 + 2\int_{\frac{1}{2}}^{1}\theta e^{\left(\frac{2\theta-1}{\theta}\right)(1-\lambda)}\,d\theta}$$

In particular $R(1,\alpha) = 1$ for all α and $R(\lambda,\alpha) \geq 0.56$. □

References

Armstrong, M., and D. Sappington (2007): "Recent Developments in the Theory of Regulation," in *Handbook of Industrial Organization*. Elsevier.

Balmaceda, F. (2013): "On the Optimality of One Size Fits All Contracts: The Limited Liability Case," Discussion paper.

Baron, D. P., and R. B. Myerson (1982): "Regulating a Monopolist with Private Information," *Econometrica*, 50(4), 911–930.

Bertsekas, D. (1995): *Dynamic Programming and Optimal Control*, vol. 1.

Bolton, P., and M. Dewatripont (2005): *Contract Theory*. MIT press.

Bose, A., D. Pal, and D. Sappington (2011): "On the Performance of Linear Contracts," *Journal of Economics and Management Strategy*, 20(1), 159–193.

Caillaud, B., R. Guesnerie, and P. Rey (1992): "Noisy Observation in Adverse Selection Models," *The Review of Economic Studies*, 59(3), 595–615.

Carroll, G. (2015): "Robustness and Linear Contracts," *The American Economic Review*, 105(2), 536–563.

Chiappori, P., and B. Salanié (2003): "Testing Contract Theory: A Survey of Some Recent Work," .

Chu, L., and D. Sappington (2007): "Simple Cost-Sharing Contracts," *American Economic Review*, 97(1), 419–428.

Chu, L., and D. E. Sappington (2009): "Procurement Contracts: Theory vs. Practice," *International Journal of Industrial Organization*, 27(1), 51–59.

Crew, M. A., and P. Kleindorfer (2002): "Regulatory Economics: Twenty Years of Progress?," *Journal of Regulatory Economics*, 21(1), 5–22.

Fudenberg, D., and J. Tirole (1991): *Game Theory*. MIT Press.

Garrett, D. (2014): "Robustness of Simple Menus of Contracts in Cost-Based Procurement," *Games and Economic Behavior*, 87, 631–641.

Gottlieb, D., and H. Moreira (2013): "Simultaneous Adverse Selection and Moral Hazard," Discussion paper.

Grossman, S., and O. Hart (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, pp. 7–45.

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

——— (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, pp. 24–52.

Jullien, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93(1), 1–47.

Jullien, B., B. Salanie, and F. Salanie (2007): "Screening Risk-Averse Agents under Moral Hazard: Single-Crossing and the CARA Case," *Economic Theory*, 30(1), 151–169.

Laffont, J.-J., and D. Martimort (2009): *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

Laffont, J.-J., and J. Tirole (1986): "Using Cost Observation to Regulate Firms," *The Journal of Political Economy*, pp. 614–641.

Lafontaine, F., and M. Slade (2001): "Incentive Contracting and the Franchise Decision," in *Game Theory and Business Applications*.

Lewis, T., and D. Sappington (1989): "Inflexible Rules in Incentive Problems," *The American Economic Review*, pp. 69–84.

——— (2000): “Motivating Wealth-Constrained Actors,” *American Economic Review*, pp. 944–960.

Maggi, G., and A. Rodriguez-Clare (1995): “On Countervailing Incentives,” *Journal of Economic Theory*, 66(1), 238–263.

McAfee, R., and M. Schwartz (1994): “Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity,” *The American Economic Review*, pp. 210–230.

Milgrom, P. (2004): *Putting Auction Theory to Work*. Cambridge University Press.

Myerson, R. (1982): “Optimal Coordination Mechanisms in Generalized Principal-Agent Problems,” *Journal of Mathematical Economics*, 10(1), 67–81.

Myerson, R. B. (1981): “Optimal Auction Design,” *Mathematics of Operation Research*, 6(1), 58–73.

Ollier, S., and L. Thomas (2013): “Ex Post Participation Constraint in a Principal-Agent Model with Adverse Selection and Moral Hazard,” *Journal of Economic Theory*, 148(6), 2383–2403.

Picard, P. (1987): “On the Design of Incentive Schemes under Moral Hazard and Adverse Selection,” *Journal of Public Economics*, 33(3), 305–331.

Pulgar, C. (2011): “Riesgo Moral y seleccion adversa: Informacion privada sobre la productividad y el monitoreo,” Master’s thesis.

Rogerson, W. (2003): “Simple Menus of Contracts in Cost-Based Procurement and Regulation,” *The American Economic Eeview*, 93(3), 919–926.

Vogelsang, I. (2006): “Electricity Transmission Pricing and Performance-Based Regulation,” *The Energy Journal*, pp. 97–126.

List of Figures

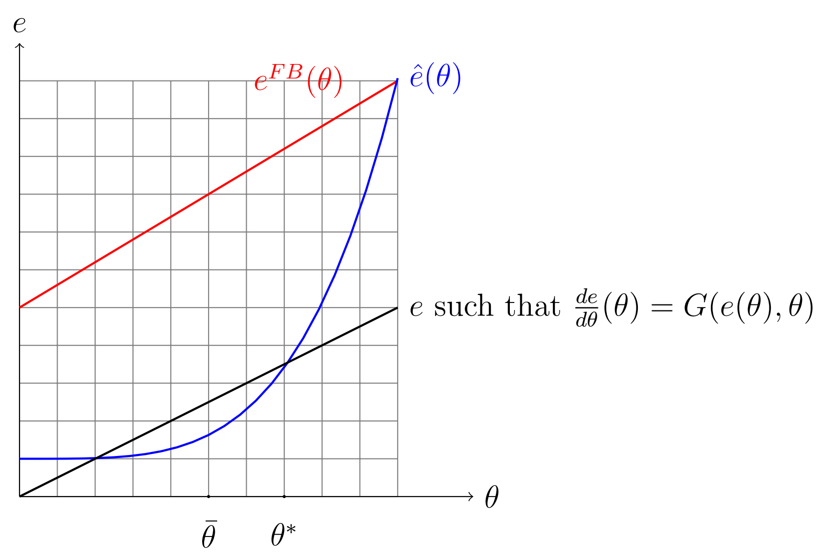

Figure 1: The solution $e^*(\theta)$ equals $e(\theta)$ for $\theta \leq \theta^*$ and equals $\hat{e}(\theta)$ for $\theta > \theta^*$.

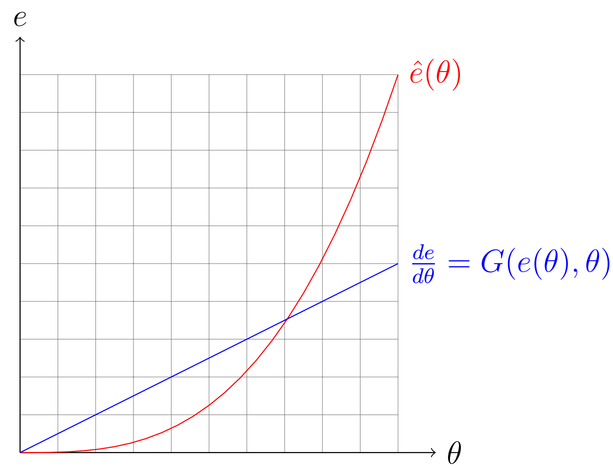

Figure 2: The optimal simple contract is found by optimally choosing initial conditions to approximate the unrestricted solution $\hat{e}$.

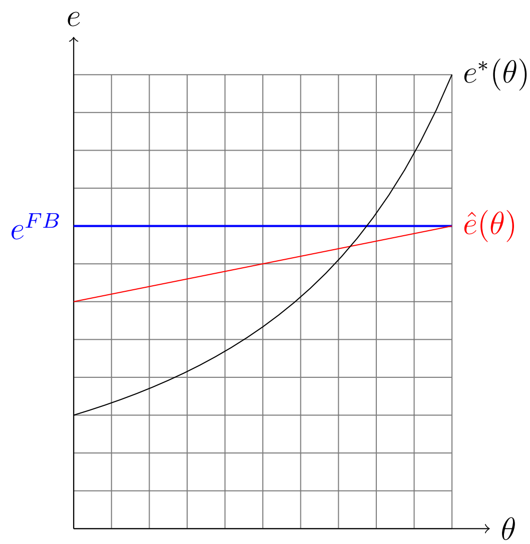

Figure 3: $\hat{e}$ does not satisfy the slope restriction. The optimal effort scheme is e^*.